國立臺灣大學生物資源暨農學院農藝學系

博士論文

Department of Agronomy
College of Bioresources and Agriculture
National Taiwan University
Doctoral Dissertation

開發最佳雙限制酶切位點標定法及其於甜瓜育種之應用

Developing optimal ddRADseq techniques and its application in melon breeding

王群山

Chun-San Wang

指導教授：黃永芬 博士

Advisor: Yung-Fen Huang, Ph. D.

中華民國 113 年 12 月

December 2024

doi:10.6342/NTU202500264

國立臺灣大學博士學位論文
口試委員會審定書

開發最佳雙限制酶切位點標定法及其於甜瓜育種之應用

Developing optimal ddRADseq techniques and the application in melon breeding

本論文係 王群山 君 (D99621105) 在國立臺灣大學農藝學系完成之博士學位論文，於民國 113 年 12 月 16 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

農業部桃園區農業改良場 場長
王毓華 研究員 王毓華

國立臺灣大學生態學與演化生物學研究所與植物科學研究所 教授
李承叡 博士 李承叡

國立臺灣大學園藝暨景觀學系 副教授
許富鈞 博士 許富鈞

國立臺灣大學農藝學系 副教授
陳凱儀 博士 陳凱儀

國立臺灣大學農藝學系 副教授
黃永芬 博士（指導教授） 黃永芬

誌謝

感謝口試委員王毓華場長、李承叡老師、許富鈞老師、陳凱儀老師與黃永芬老師在百忙之中對我的論文內容提出寶貴的建議，讓我精進論文寫作、報告技巧、強化思考與論述的能力。感謝農業部農業試驗所黃晉興博士與林思妤助理研究員、本實驗室成員劉地寬、張心怡、世界蔬菜中心林亞平博士與福萬福工作室魏甫錦博士，沒有各位的協助與支持，我也無法順利完成甜瓜白粉病遺傳與育種的研究及投稿。同時感謝我的指導老師胡凱康老師，在如此漫長的時間裡引導我，讓我學習如何成為自己想要成為的人。感謝我的指導老師黃永芬老師，讓我學會看見事實，讓我學習跳開困境的方法與動力。

感謝王裕文老師在入學考時給予我育種學的提點，讓我爭取到博士班的入場門票。感謝食科所江文章老師、鍾成沛學長、生機系李允中老師與何萬中先生在薏苡育種計畫的支持與協助，加深我實作的訓練。感謝博士班資格考的口試委員：黃懿秦老師、林順福老師、董致韡老師、陳凱儀老師與彭雲明老師讓我更深入了解遺傳與育種的基礎與拓展統計學的新知與應用。感謝黃文達主任的指導，讓我在農藝學找到起點。感謝農藝系廖振鐸老師、林彥蓉老師、劉力瑜老師與蔡欣甫老師協助論文寫作與指導寫作進度。感謝巴清雄老師與夏禾達利老師一家，讓我爭取延長學期的機會，才有產出這本論文的機會。讓我體驗不同文化的生活哲學，從艱難困境中保持「充滿力量的身體！」的樂觀態度，並找回自己純真的本質。感謝作物育種實驗室的學長姐與學弟妹們讓我獲得許多學習與成長的機會。感謝我的大學、碩士班與博士班同學陪伴我的校園生活。感謝所有我沒有感謝到的人，謝謝你們出現在我的生命裡，讓我豐富精彩！

最後，感謝我的家人和宜欣的家人，忍受我的固執、任性與漫長等待。終於，我要收書包回家了！謝謝你，宜欣，我的愛。

從來都沒有離終點這麼近。我感謝這趟旅程，由衷感謝

Everything that has a beginning has an end.

群山 敬上

2025.01.22

doi:10.6342/NTU202500264

摘要

分子標誌為遺傳研究與作物育種之重要工具。隨著次世代定序的發展，開發高通量分子標誌並進行基因型分型已成為作物遺傳育種研究的常用方法，但針對目標物種發展適合的最佳化流程仍為挑戰。為此，本研究奠基於次世代定序，優化雙限制酶切位點標定法 (double digest restricted associated DNA sequencing, ddRADseq) 之分子標誌開發與基因型分型，並以甜瓜抗白粉病之遺傳分析與育成抗病近同源系作為方法驗證。我們以電腦模擬限制酶於甜瓜基因體之切位並以小規模實驗評估限制酶組合。之後分別以 *Pst*I 和 *Xba*I 作為稀有切位限制酶與常見切位限制酶 *Taq*$^{\alpha}$I 組合並加入 *Sph*I 與 *Mse*I 排除序列，應用於單一 F_2 分離族群以建立高通量分子標誌基因型資料校正法及建立高密度連鎖圖譜之最佳化流程。基於此最佳化流程，以三個 F_2 族群（A6、B2 和 C4）定位甜瓜抗白粉病的數量性狀位基因座 (quantitative trait loci, QTL)。分別在 A6 族群中偵測到位於第 2 條染色體的 *qPM2*，在 B2 族群中偵測到位於第 5 條染色體的 *qPM5.B2* 與在 C4 族群中偵測到第 5 與第 12 條染色體上的 *qPM5.C4* 與 *qPM12*。其後，針對第 2、5 和 12 條染色體上的 QTL 開發一系列 TaqMan 分子標誌用於分子標誌輔助回交育種之前景選拔與重組選拔，並結合 ddRADseq 評估輪迴親基因體恢復率進行背景選拔。選用 A6 與 C4 族群的抗病親作為供給親，並分別與帶有綠色及橙色果肉的優良親本作為輪迴親進行分子標誌輔助回交增進優良甜瓜親本的白粉病抗病性。經過二到三個回交世代後，成功將單一抗病 QTL 導入優良輪迴親的遺傳背景中，最終在綠色和橙色果肉的輪迴親遺傳背景中分別獲得 6 個帶有白粉病抗病 QTL 的近同源系。從分子標誌開發完成到抗病近同源系的育成僅花費三年半的時間，顯示基於次世代定序的高通量基因分型技術可以有效與作物育種結合，提升作物分子育種的效率。

關鍵字： 雙限制酶切位標定法；基因型校正；分子標誌輔助回交；甜瓜 (*Cucumis melo* L.)；白粉病 (*Podosphaera xanthii*)；數量性狀基因座定位；TaqMan 分子標誌。

doi:10.6342/NTU202500264

Abstract

Molecular markers are important tools for genetic studies and crop breeding. With the development of next-generation sequencing, high-throughput genotyping has become a common tool for crop breeding. However, optimizing the genotyping process for each target crop is still challenging. Therefore, this study aimed to optimize the double digest restriction-site associated DNA sequencing (ddRADseq) genotyping platform and implement the optimized procedure in melon as a proof-of-concept. The framework consists of the identification of the genetic architecture of powdery mildew resistance in melon and the development of the powdery mildew resistance near-isogenic lines. We applied *in silico* digestion and empirical tests to evaluate candidate enzyme combinations in melon. Two candidate enzyme combinations were further applied to an F_2 population, which were *Pst*I-*Taq*αI-*Sph*I and *Xba*I-*Taq*αI-*Mse*I, for optimization of a high-throughput genotyping data error correction and linkage map construction. Based on the optimized procedure, quantitative trait locus (QTL) mapping of melon powdery mildew resistance was applied to three F_2 populations, A6, B2, and C4. QTL were identified on chromosomes 2, 5, and 12, which were *qPM2* in A6, *qPM5.B2* in B2 and *qPM5.C4*, and *qPM12* in C4. A series of TaqMan assays targeting QTL were developed and validated for foreground and recombinant selection, complemented with the ddRADseq genotyping system to evaluate the recurrent parent genome recovery. Three marker-assisted backcrossing (MABC) programs using resistant donor parents from A6 and C4 crossed with elite susceptible recurrent parents with green and orange fruit flesh were implemented. After two to three cycles of MABC, individual QTL was successfully introgressed into elite genetic backgrounds, giving six powdery mildew resistance near-isogenic lines in each green- and orange-flesh background. In three and a half years, we have achieved from marker development to the production of isogenic lines. This study demonstrated the power of high-throughput genotyping and its efficient implementation in molecular breeding.

doi:10.6342/NTU202500264

Keywords: double digest restriction-site associated DNA sequencing (ddRADseq); genotype correction; marker-assisted backcrossing (MABC); melon (*Cucumis melo* L.); powdery mildew (*Podosphaera xanthii*); quantitative trait loci (QTL); TaqMan.

doi:10.6342/NTU202500264

Table of Contents

doi:10.6342/NTU202500264

doi:10.6342/NTU202500264

doi:10.6342/NTU202500264

List of Tables

doi:10.6342/NTU202500264

doi:10.6342/NTU202500264

List of Figures

doi:10.6342/NTU202500264

doi:10.6342/NTU202500264

1. Introduction

1.1. Genetic markers in plant breeding

1.1.1. Conventional genetic markers

Genetic markers, inheritable biological features transmitted from parents to offspring across generations, play a critical role in understanding the genetic basis of traits. These markers can be broadly categorized into three types: morphological markers, biochemical markers, and molecular markers. Desirable characteristics for an ideal genetic marker include high-level of polymorphism, clear and distinct allele features, co-dominance, evenly distributed on the genome, easy detection, low-cost of marker development and genotyping, and high reproducibility (Xu 2010).

Morphological markers are visible phenotypic traits, such as plant height, fruit color, or seed shape. These markers are easy to evaluate without specialized equipment. However, the number of polymorphic traits is usually limited, and their expression can be easily influenced by environmental factors. Biochemical markers, such as isozymes, are features of biochemical processes at the protein level. These markers are less affected by the environment than morphological markers but remain constrained in their abundance and resolution (Lübberstedt et al. 2023). Molecular markers, such as restriction fragment length polymorphisms (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLP), simple sequence repeats (SSR), and single nucleotide polymorphisms (SNP), are identified through molecular biology techniques at the DNA level (Jiang 2015). The DNA markers require laboratory processes in controlled environments, focusing on the genetic material and provide high specificity, resolution, and genome-wide coverage, making them more reliable and closely to the criteria for ideal genetic markers.

RFLP were the first-generation molecular markers, where the polymorphisms arise from changes in restriction enzyme recognition sites causing detectable fragment size shifts. RAPD used random primers and polymerase chain reaction (PCR) to amplify genomic regions, with polymorphisms detected by the presence or absence of the amplified fragments. AFLP combined restriction enzyme digestion and PCR to amplify specific DNA fragments, where variations in cutting sites reveal polymorphisms by differences in DNA fragment patterns. The SSR markers amplify specific regions

doi:10.6342/NTU202500264

containing tandem repeat sequences, which heavily relies on a comprehensive sequence information to design primer pairs of the markers (Lübberstedt and Varshney 2013). These molecular markers generally assess DNA polymorphism through electrophoresis, which is commonly used to distinguish DNA fragment size. The polymorphisms of RFLP, AFLP, and RAPD are indirectly revealed through the differences in fragment size, but the polymorphism of SSR is directly shown through the fragment size with different numbers of tandem repeats.

SNP provides the ultimate and simplest form of molecular markers, defined by a single nucleotide base difference between genotypes, the smallest unit of genetic inheritance. The abundance of SNP and widespread distribution across the genome make them invaluable for comprehensive genetic analysis and breeding programs. SNP polymorphisms can be detected using various methods. CAPS (Cleaved Amplified Polymorphic Sequences) or allele-specific primers can generate PCR products with different fragment sizes, which can then be visualized through electrophoresis (Xu 2010; Jiang 2015). Alternatively, TaqMan or KASP (Kompetitive Allele-Specific PCR) marker systems utilize fluorescence-based probes to detect SNP polymorphisms with high precision and accuracy (Semagn et al. 2014). Similar to SSR markers, SNP markers require comprehensive sequence information for their development and application. However, with the advent of next-generation sequencing, it has become more efficient to discover, sequence, and genotype SNP markers across the genome of interest. This technological advancement has significantly enhanced the resolution and scalability of SNP-based genetic studies and breeding programs (Davey et al. 2011).

1.1.2. NGS-based high-throughput molecular markers

Next-generation sequencing (NGS) is defined as high-throughput sequencing methods that combine parallel sequencing processes to produce millions of sequences at once (Metzker 2010). Several NGS technologies have been developed, and after a few years of competition between the sequencing platforms, the most popular sequencing platforms are sequencing-by-synthesis (SBS) (Lübberstedt et al. 2023).

SBS relies on solid-phase amplification, which is composed of initial priming and extension of the single-stranded, single-molecule templates. Through bridge amplification, immobilized templates immediately adjacent primers to form clusters, with each cluster originating from a single DNA molecule on the glass slide (flow cell). This process generates clonally amplified templates, ensuring strong fluorescence signals that

doi:10.6342/NTU202500264

can be accurately detected by the imaging system using two lasers. The flow cell can produce 100 - 200 million spatially separated template clusters, providing free ends to which a universal sequencing primer can be hybridized to initiate the NGS reaction. SBS uses a cyclic reversible terminator comprised of nucleotide incorporation, fluorescence imaging, and cleavage. Four-color 3′-O-azidomethyl-deoxyribonucleoside triphosphates (dNTPs) blocking group to halt DNA strand extension after each nucleotide addition. A fluorescent dye labels each nucleotide for base identification. After imaging, the blocking group and dye are chemically removed, typically using tris (2-carboxyethyl) phosphine (TCEP) or similar agents, restoring the 3′-OH group for the next synthesis cycle. These chemical modifications ensure controlled, sequential nucleotide incorporation, making SBS highly accurate and suitable for high-throughput sequencing. The efficiency of the addition process with incomplete extension or addition of multiple nucleotides for the template ensemble occurs in a given cycle, resulting in lagging or leading-strand dephasing within the cluster. The dephasing signal within the cluster increases fluorescence noise, reduces base-calling accuracy, and limits read length (Metzker 2010).

The NGS sequencing library requires a specific sequence structure to conduct the sequencing process. At the core of the SBS sequencing library lies the inserted genomic DNA fragment, which serves as the target for sequencing. Surrounding insert DNA are binding sites for sequencing primer 1 and sequencing primer 2, positioned to facilitate the initiation of the SBS process and generate the read 1 and read 2 sequences. Further outward from these primer binding sites are the i5 and i7 index sequences, which are optionally included to enable sample multiplexing during sequencing runs. At the outermost edges of the fragments are the P5 and P7 adapters, located at the 5′ and 3′ ends, respectively. These adapters allow the fragments to anneal to complementary probes on the flow cell for solid-phase amplification and generate clusters. The total size of the sequencing library, including the DNA insert and all additional components, typically exceeds the insert size, generally around 140 bp or more. This size should be carefully considered during library preparation, particularly if the procedure includes library fragment size selection, to ensure the desired insert size range is achieved for sequencing.

With the introduction of NGS, the discovery, sequencing, and genotyping of molecular markers across the genome have become significantly more accessible and efficient, greatly expanding their application in genetic studies. Although whole-genome sequencing provides comprehensive coverage, its high cost and computational demands often make it impractical for large-scale studies. To address these challenges, genome

doi:10.6342/NTU202500264

complexity reduction methods have been developed to efficiently target informative regions of the genome, enabling cost-effective genotyping and marker discovery.

Building on traditional molecular techniques, many genome complexity reduction methods utilize restriction enzymes to selectively reduce genome complexity. These enzymes, or combinations of enzymes, play a critical role in targeting specific genomic regions for sequencing (Davey et al. 2011). These methods can be grouped as (1) restriction site-associated DNA (RAD) sequencing (Baird et al. 2008), (2) low coverage genotyping, including genotyping by sequencing (GBS) (Elshire et al. 2011), and (3) reduced-representation sequencing, such as two-enzyme GBS (Poland et al. 2012), and double digest restriction site-associated DNA (ddRAD) sequencing (Peterson et al. 2012). Beyond restriction enzyme-based approaches, other methods have been developed to enrich specific genomic regions for sequencing. For instance, sequence capture (Hoffberg et al. 2016) uses hybridization probes to isolate target regions, while PCR-based techniques (Sinn et al. 2022; Nishimura et al. 2024) employ random primers to amplify specific fragments. These methods should also be grouped as reduced-representation sequencing, which provides researchers with a variety of options to tailor their experiments based on genome complexity, budget, and the scope of their study.

The RADseq, GBS, and reduced-representation sequencing are started with restriction enzyme or enzyme combination treatment of genomic DNA. In RADseq, the digested genomic DNA fragments with sequence overhanding ends of restriction enzyme cutting site ligated with P1 adapter. This adapter contains an annealing site of sequencing primer 1 and the inline barcode nucleotide (4 or 5 bp) for sample identification. The adapter-ligated fragments from multiple samples were pooled, randomly sheared, and size-selected. The processed library then ligated a second P2 adapter in divergent ends of a Y-shape, which ensures the following PCR would only amplify the P1 adapter-ligated RAD tags in RADseq library (Baird et al. 2008). In GBS, genomic DNA is digested with the restriction enzyme *Ape*KI, generating fragments with sticky ends that are ligated to a barcode adapter for sample identification and a common adapter with an *Ape*KI-compatible sticky end. The adapters are designed with annealing sites for PCR primers, which include P5 and P7 sequences that enable the sequencing library to attach to the flow cell for cluster generation during sequencing. Unlike RAD-seq, GBS does not involve random shearing or size selection and uses non-phosphorylated adapters, eliminating the need for Y adapters. Following PCR amplification, the library undergoes

doi:10.6342/NTU202500264

quality evaluation to minimize adapter dimers and ensure high-quality preparation for sequencing (Elshire et al. 2011).

In both two-enzyme GBS and ddRAD-seq, genomic DNA is digested with an enzyme combination of a rare cutter and a common cutter, and the forward adapter and the reverse Y adapter, respectively, with overhangs were ligated to the digested genomic DNA. Furthermore, the application of Y-adapters was introduced to enhance the genome complexity reduction by generating an "uniform" library containing fragments bordered by different enzyme cutting sites (Poland et al. 2012). The key difference between the two methods lies in size selection: ddRAD-seq includes a size selection step to retain fragments within a specific size range, reducing library complexity and improving uniformity of coverage, whereas two-enzyme GBS typically omits this step, resulting in a more comprehensive but less uniform representation of the genome (Poland et al. 2012; Peterson et al. 2012). While the genome complexity is reduced to 1 to 5%, the read depth is increased sufficiently for a more reliable genotype calling. At the same time, more samples can be included within a single sequencing run.

In addition, an *in silico* digestion of the reference genome sequence can also provide guidelines for the choice of appropriate restriction enzyme combinations prior to library preparation (Lepais and Weir 2014; Wang et al. 2016a; Mora-Márquez et al. 2017; Chafin et al. 2018). If the *in silico* prediction can be followed by a small-scale empirical test, it can estimate the allocation of the sequence reads for optimized efficiency utilizing the sequence reads of the restriction enzyme combinations (Fu et al. 2016; Pértille et al. 2016; Li et al. 2020; Christiansen et al. 2021). Some studies conducted the empirical test on enzyme combinations, such as in tomato (Shirasawa et al. 2016), peach (Aballay et al. 2021), maize (Heffelfinger et al. 2014), rice (Fu et al. 2016), and chicken (Pértille et al. 2016). These studies provided different aspects to evaluate the enzyme combinations. The allocation of the sequencing resources and fragment polymorphisms under different enzyme combinations were not clearly recorded.

1.1.3. Applications of molecular markers to plant genetic analysis and breeding

Genetic factors can be classified into major genes or minor genes. A major gene provides a large effect to a trait which is less subject to environmental influences. The minor genes carry smaller effects and are easily influenced by the environment. Such minor genes, or quantitative trait loci (QTL), contribute to the variation of quantitative traits. Genetic markers provide the ability to locate QTL, to estimate the effect of these

doi:10.6342/NTU202500264

QTL. The common method to identify QTL involves controlled cross designs, where inbred parental lines are crossed to generate a segregating population, such as F_2, backcross (BC), or recombinant inbred lines (RILs) with segregated phenotypes and genotypes. Linkage disequilibrium is then used to detect QTL between phenotypic variations and marker genotypes within the segregating population. The development of many genome-wide markers, such as SNP, has greatly enhanced the resolution and accuracy of linkage mapping.

NGS-based methods enable low-cost per-data-point marker discovery and high-throughput genotyping of SNPs. Such availability enhances downstream genetic analysis, such as the construction of high-density linkage maps for QTL mapping and supports marker-assisted selection. The introduction of NGS provided a high-throughput and cost-effective solution for marker discovery. This has, in turn, accelerated the identification of QTL, thereby contributing to a better understanding of the genetic architecture of quantitative traits and facilitating the application of marker-assisted selection (MAS) in plant breeding programs.

Among MAS, Marker-assisted backcrossing (MABC) utilizes the association between target phenotype and marker genotype for indirect selection (Frisch 2004). MABC aims to introgress favorable alleles from the donor parent into the recurrent parent's genetic background. It involves three stages: foreground, recombinant, and background selection. Foreground and recombinant selections target the QTL-carrier chromosome with the objective of introducing an allele of interest with minimal linkage drag, while background selection aims to recover the recurrent parent genome (RPG) outside the target locus as much as possible (Collard and Mackill 2008). The effectiveness of MABC relies on adequate population size, the presence of polymorphic markers close to the target locus (Frisch et al. 1999a), and a broad distribution of markers within the genetic background (Frisch et al. 1999b). MABC has enhanced desirable characteristics in various crops, including submergence tolerance and salinity tolerance in rice (Neeraja et al. 2007; Marè et al. 2023), nutrient enrichment in maize (Singh et al. 2021; Chandrasekharan et al. 2022), and heat tolerance and fusarium head blight resistance in wheat (Zhang et al. 2021; Bellundagi et al. 2022).

doi:10.6342/NTU202500264

1.2. Melon (*Cucumis melo* L.)

1.2.1. Overview of melon

Melon (*Cucumis melo* L., 2n = 2x = 24) is an economically valuable crop in the Cucurbitaceae family, cultivated widely across the world from temperate to tropical regions due to its sweet taste, flavor, and fleshy fruit. The global production of melons was around 28.56 million tonnes, making it the third most-produced cucurbit crop after watermelon and cucumber (FAOSTAT 2022). Melon is a cross-pollinated species and is an important model plant for studying sex determination. Initially, melon flowers are bisexual but can undergo abortion of either the carpel or stamen during sex determination, leading to the development of andromonoecious (producing male and bisexual flowers) or gynomonoecious (producing female and bisexual flowers) plants (Aamir et al. 2021). Typically, commercially grown melons are monoecious or andromonoecious, with male flowers on the main stem and female or hermaphrodite flowers on the proximal nodes of lateral branches. The sex determination of melon flowers is position-dependent and environmental factors such as temperature, photoperiod, hormones, and developmental cues (Pitrat 2016; Aamir et al. 2021; Xu et al. 2022). Consequently, melon breeding programs must initially assess self-pollinated breeding lines before proceeding to evaluate the hybrids.

The first draft genome was published by Garcia-Mas et al. (2012), who sequenced the melon genome using the DHL92 line, which was a double-haploid homozygous line derived from a cross between PI 161375 (Songwhan Charmi, spp. *agrestis*) and the 'Piel de Sapo' T111 line (ssp. *inodorus*). The melon reference genome was sequenced using NGS (454 pyrosequencing) and assembled into 12 chromosomes, covering 83.3% of the genome in 375 Mb. Argyris et al. (2015) improved genome assembly as v3.5.1 using targeted SNP anchoring the melon scaffold.

1.2.2. Powdery mildew, a major disease for melon production

Powdery mildew significantly impacts melon production, causing economic losses across all growing areas. This fungal disease can affect seedlings, stems, leaves, and fruit, which typically display a powdery appearance due to the abundance of conidia. It ultimately results in fruit quality degradation and yield losses due to stunting and premature plant death, limiting melon cultivation in greenhouses and fields (Egel et al. 2022).

doi:10.6342/NTU202500264

Powdery mildew is caused by *Podosphaera xanthii* (Castagne) U. Braun & Shishkoff and *Golovinomyces cichoracearum* (DC.) V.P. Heluta. *P. xanthii* occurs more frequently in regions with high humidity and temperatures, such as subtropical and tropical areas. More than 20 races of *P. xanthii* have been identified based on the reactions to the differential set (McCreight 2006). Races 1, 2, and 5 are dominant races in Southern Europe; races 1 and 5 are prevalent in Japan; races 1, 2, and 3 in America; and races 1 and 2F are common in China (Zhang et al. 2013; Haonan et al. 2020). Races 1 and 5 were detected in Taiwan, with the former being predominant from 2001 to 2005, and the latter identified in 2005. From 2008 to 2010, both races were identified in Tainan, the primary region for melon production in Taiwan (Huang and Wang 2007; Wang 2016). The distribution of predominant physiological races is influenced by spatiotemporal situations and the specific melon cultivars grown in these regions (López-Martín et al. 2022). Powdery mildew management in melon involves cultural practices, biological controls, fungicides, and host resistance (Egel et al. 2022). Although fungicides are commonly used for melon powdery mildew control, they are not a sustainable solution since they can cause plant resistance breakdown and harm the environment. Therefore, resistant varieties remain a more sustainable and efficient option for powdery mildew control (Branham et al. 2021; Egel et al. 2022).

Identifying powdery mildew resistance QTL and the germplasm that carries them are essential for resistance breeding. Once the resistant materials and loci are identified, resistance loci can be effectively introduced into elite backgrounds using marker-assisted selection (Collard and Mackill 2008). Several studies have identified the genetic resources and QTL for powdery mildew resistance in melon. powdery mildew resistance QTL in melon have been identified on chromosomes 2, 4, 5, 9, 10, and 12 (Perchepied et al. 2005; Teixeira et al. 2008; Fukino et al. 2008; Yuste-Lisbona et al. 2010; Yuste-Lisbona et al. 2011a, b; Wang et al. 2011, 2016b; Beraldo-Hoischen et al. 2012; Zhang et al. 2013, 2023; Fazza et al. 2013; Ning et al. 2014; Kim et al. 2016; Li et al. 2017; Haonan et al. 2020; Cao et al. 2021; Branham et al. 2021; Cui et al. 2022; López-Martín et al. 2022). Regarding powdery mildew resistance materials, PI 414723, K7-1, and TARI-08874 were reported to carry a single QTL on chromosome 2 (Zhang et al. 2013; Fazza et al. 2013; Wang et al. 2016b). The melon variety Ano2 carried a QTL on chromosome 5 (Wang et al. 2011), and the accession AF125$^{Pm\text{-}1}$ carried resistance QTL on chromosome 9 (Teixeira et al. 2008), while the accessions wm-6, PI 124112, and MR-1 carried a single resistance QTL on chromosome 12 (Li et al. 2017; Cao et al. 2021; Zhang et al. 2023).

doi:10.6342/NTU202500264

Some accessions carried more than one resistance QTL. For instance, Edisto47 carried resistance QTL on chromosomes 2 and 5 (Ning et al. 2014); AR 5, PMR 5, and PMR 6 carried QTL on chromosomes 2 and 12 (Fukino et al. 2008; Kim et al. 2016; Haonan et al. 2020); and PI 124112 and TGR-1551 carried QTL on chromosomes 5 and 12 (Perchepied et al. 2005; Yuste-Lisbona et al. 2010; Yuste-Lisbona et al. 2011b, a; Beraldo-Hoischen et al. 2012; López-Martín et al. 2022). Based on the abovementioned studies, the polymorphism of the powdery mildew markers are not able to correspond to the resistance consistently while applying marker-assisted selection, suggesting the investigation of high-density markers are required.

1.3. Objective of the study

While the NGS-based high-throughput marker system was introduced in melon genetic studies and marker-assisted breeding in this study, the following factors from Davey et al. (2011) were considered to approach the optimal conditions for ddRADseq in melon breeding. (1) Study goals: in this study, we focused on analyzing the early segregating populations derived from inbred line cross design, such as F_2 and backcross population, with higher heterozygosity of the individuals. The expected number of markers was thousands of markers, but genotypes were required to have a higher accuracy. (2) Availability of a reference genome: the reference genome of melon (*Cucumis melo* L.) was available, and we especially involved the chloroplast and mitochondria sequences to increase the ability to identify unique loci from ddRADseq. (3) Expected degree of polymorphism: given the study goals in this study, the degree of polymorphism was determined by the parental inbred lines. Therefore, we expected the degree of polymorphism to range from 0.1 to 0.5, which could translate the number of fragments for the ddRAD library into the number of markers with the read depth to ensure the confidence of the genotypes. (4) Choice of restriction enzyme: this factor is one of the important issues in this study. We conducted *in silico* digestion of the reference genome and selected the restriction enzyme combinations for the empirical test. The metrics for evaluating restriction enzyme combinations are also important for this issue. (5) DNA sample preparation: We used DNA extraction kits to ensure the quality of the DNA in the samples remained the same. (6) Adapter design: the factor considered by Davey et al. (2011) included inline barcode adapters for sample multiplexing and avoiding the limitation of low diversity sequencing library in Illumina sequencing platform. Here, we applied dual-index barcoded systems for sample multiplexing in the Illumina platform

doi:10.6342/NTU202500264

(Scheben et al. 2017). The low diversity library issue was combined with spick in PhiX and dark sequencing cycles, which performed the sequencing process without capturing the fluorescent images. (7) PCR amplification: the dual-index system applied in this study. Therefore, the PCR step of the sequencing library was performed by each sample, and the bias of the PCR product occurred within each sample. The PCR product containing the adaptor dimers was removed through the process of fragment size selection. (8) Sequencing: the sequencing resources in this study expected that the F_2 population consisted of 96 to 192 individuals analyzed through Illumina HiSeq 2500 platform of Rapid mode with 300 M paired-end reads. Given the higher confidence of the genotype calls, we would expect a read depth of 30×. Each sample should consist of around 1M paired-end reads. This criterion suggested the sequence resources for the empirical test of restriction enzyme combinations.

In this study, we aimed to develop an optimization procedure to implement ddRADseq for genetic analysis and marker-assisted selection in plants. The powdery mildew resistance improvement in melon was a proof-of-concept case study. Therefore, the objectives of this study were (1) to develop *in silico* analysis tools for ddRADseq enzyme combination selection, (2) to verify the *in silico* digestion results by a small-scale empirical test, (3) to examine the two candidate enzyme combinations on a single F_2 population for choosing an optimal enzyme combination and develop genotype error correction tools for high-throughput SNPs, (4) to develop high-throughput markers in an optimized ddRADseq framework for melon, (5) to identify powdery mildew resistance QTL in three melon F_2 populations germplasms and to provide powdery mildew resistance-associated markers, and (6) to introgress resistant alleles into elite melon cultivars using MABC.

doi:10.6342/NTU202500264

2. Optimization of ddRADseq in melon

2.1. *In silico* analysis and empirical small-scale experiment for enzyme combination identification

2.1.1. Materials and Methods

2.1.1.1. *In silico* digestion

The melon reference genome DHL92 v3.5.1 (Argyris et al. 2015), including organelle sequences of the chloroplast (NC_015983) and mitochondria (JF412792 – JF412800) length of 156 Kb and 2.74 Mb, respectively, (Rodríguez-Moreno et al. 2011) were downloaded from the NCBI Nucleotide Database (https://www.ncbi.nlm.nih.gov/nuccore). *In silico* digestion was conducted using R (version 4.3.1, R Core Team 2023). The melon reference genome was converted to a BSgenome object (Pagès 2024) denoted as BSgenome.Cmelo.CRAG.3.5.1 library using BSgenomeForge package (Pagès and Kakopo 2024). The Biostrings package (Pagès et al. 2024) was utilized to identify the cutting sites of each restriction enzyme within the melon reference genome sequence.

The fragments produced by single-enzyme *in silico* digestion were sorted by genome position within each chromosome. The first fragment spanned from the start of the chromosome to the first cutting site, while subsequent fragments were found between each pair of cutting sites. The last fragment of each chromosome extended from the final cutting site to the end of the chromosome (Figure 1).

In two-enzyme combinations, the enzyme with fewer cutting sites was typically categorized as a rare cutter, and the other enzyme with a greater number of cutting sites was classified as a common cutter. The *in silico* digestion was performed as mentioned above, and the cleaved fragments were categorized based on the types of cutting sites that flank each fragment. If two rare cutters bordered the fragments, they were classified as AA-type fragments. If two common cutters bordered the fragments, they were classified as BB-type fragments. If the fragments were composed of a rare cutter and a common cutter, they were classified as AB-type fragments. These fragments produced by *in silico* digestion were classified as predicted fragments and were selected by size ranging from 160 to 460 bp. The predicted fragments were named by combining the chromosome name,

doi:10.6342/NTU202500264

the physical position of the flanking cutting sites, the type of fragment, and fragment length, which were further used for sequence verification.

We included a third enzyme C to exclude specific fragments from the sequencing library. Initially classified as AB-type fragments, those that contained at least one cutting site for enzyme C were cleaved into AC-, BC-, or CC-type fragments, only AB-type was included in ddRAD library (Figure 1).

2.1.1.2. DNA extraction and ddRAD library preparation

Genomic DNA was extracted from freeze-dried young melon leaves using a modified 1.25% SDS method (Jobes et al. 1995). Extracted DNA was purified using QIAquick 96 PCR Purification Kit (Qiagen, Hilden, Germany), quantified using Quant-iT™ PicoGreen™ dsDNA Assay Kits (Life Technologies, Oregon, USA), and adjusted to 10–15 ng/μL for library preparation.

The ddRAD library was prepared according to Peterson et al. (2012) with minor modifications as follows. The restriction enzymes, T4 DNA ligase, CutSmart Buffer, and rATP used for library construction were from New England Biolabs (Ipswich, MA, USA). For each sample, 300 ng DNA was double-digested using one of the rare cutters, *Hind*III-HF, *Eco*RI-HF, *Xba*I, *Sac*I-HF, *Pst*I-HF, and *Kpn*I-HF, followed with the common cutter, *Taq*$^{\alpha}$I, at 37°C and 65°C, respectively, for 30 minutes. Ligation was then proceeded overnight at 16°C, incorporating a 5-fold excess of P1 adapter of target rare cutter recognition sequences and Y (*Taq*$^{\alpha}$I) adapters, T4 DNA ligase, and rATP in 1X CutSmart Buffer. The calculated formulas for the input amounts of P1 and Y adapters for each restriction enzyme combination are provided in Supplementary Table 1. At the end of ligation, the ligase was heat-inactivated, and the ligation product was purified using 0.8X AMPure XP beads (Beckman Coulter, Brea, CA, USA) to eliminate short fragments and adapter dimers. PCR amplification was conducted using dual-indexed primers, including Nextera XT DNA Indexes v2 and Phusion polymerase (Thermo Fisher Scientific Baltics UAB, Vilnius, Lithuania) in a 20-cycle two-step PCR protocol, which included only the denaturation and extension steps. The final PCR products were purified using 0.8X AMPure XP beads, quantified with Quant-iT™ PicoGreen™ dsDNA Assay Kits, PCR products of all the samples were pooled in equal amounts for size selection (300 – 600 bp) using Blue Pippin (Sage Science, Beverly, MA, USA) in a 2% electrophoresis gel cassette. The size-selected ddRAD libraries of each enzyme combination were then sent

doi:10.6342/NTU202500264

to the Technology Commons, College of Life Science, National Taiwan University for sequencing libraries fragment length check using the Agilent 2100 Bioanalyzer with DNA High Sensitivity Chip (Agilent Technologies, Santa Clara, CA, USA). The size-checked ddRAD libraries were sent to the Seeing Bioscience Co., Ltd (New Taipei City, Taiwan) for paired-end (PE) 250 sequencing with extra dark cycles and a spike-in of 10% PhiX, utilizing the MiSeq v3 kit (Illumina, San Diego, CA, USA).

2.1.1.3. Bioinformatic analysis

The bioinformatic analysis was conducted for both *in silico* digestion fragments and empirical data. The *in silico* digestion predicted fragments with sizes ranging from 160 to 460 bp. Sequences up to 250 bp were extracted from each end of the size-selected fragments to simulate sequencing reads produced by the HiSeq2500 Rapid mode, resulting in Read 1 and Read 2 FASTA files. For fragments shorter than 250 bp, the entire sequence was extracted without exceeding the fragment length. For the empirical sequencing data, the FASTQ files were processed with AdapterRemoval v2.1.7 (Schubert et al. 2016) to truncate the Illumina adapter sequences, and the PE reads less than 100 bp were removed from further analysis.

Both *in silico* digested fragments and empirical sequencing reads were aligned to the DHL92 v3.5.1 reference genome using Bowtie2 v2.2.9 (Langmead and Salzberg 2012) with the parameters "--very-sensitive -X 800 --no-mixed --no-discordant". The aligned reads were sorted, compressed, and indexed into BAM files with SAMtools v1.3 (Li et al. 2009). Rsamtools v2.21.2 (Morgan et al. 2024) was used to extract the alignment results. The alignment results included total reads, PE read locations, insert size, mapping quality (MAPQ) score, and mapping categories: unmapped reads, PE reads that aligned to the reference genome exactly once, and PE reads aligned to more than one location. For mapped reads, the chromosome, physical position, and the estimated insert size of each read were verified. The MAPQ score represented the alignment score and related to the uniqueness of the PE read's location. An aligned read was unique if it had a much higher alignment score than all the other possible alignments. The bigger the gap between the best alignment's score and the second-best alignment's score, the more unique the best alignment, and the higher its mapping quality should be.

The aligned PE reads were extracted using the GenomicAlignments package (Lawrence et al. 2013), following a "Fragment analysis" approach shown in Figure 2. First, the findOverlaps function was applied, with the aligned PE reads as the query and

doi:10.6342/NTU202500264

the *in silico*-predicted fragments (AB type) as the subject, including fragment size ranging from 100 to 1000 bp in length. The PE reads were mapped to *in silico* fragments, allowing for an overlap up to 50 bp shorter than the full length of the predicted fragment. PE reads that did not align with any *in silico* digestion product were used to build "Empirical fragments". Given the properties of ddRADseq, all PE reads were expected to start at the rare cutter (R) and to end with the common cutter (C). To categorize these reads, the closest cutting site to each PE read was identified (Figure 3). For PE reads fall into the RC type, i.e., starting at the rare cutter and ending at the common cutter, those exceeded the length of dark cycle threshold might not align with the expected R or C sites. These reads were categorized as R-, C-, or -- types of empirical fragments (Heffelfinger et al. 2014).

The empirical PE reads were sorted by MAPQ score and the insert size in decreasing order (Figure 2). PE reads with MAPQ scores below 20 were removed, allowing a more accurate assessment of fragment attributes. The PE read with the highest MAPQ score and the largest insert size was selected as the initial empirical fragment. The initial empirical fragment was assigned as the subject, and the remaining PE reads were assigned as the query. Using findOverlaps to find the query PE reads located within the boundaries of the initial empirical fragment. Query PE reads with minimum overlap lengths differing by less than five base pairs from the empirical fragment were grouped as the same empirical fragment. Once the query PE reads were overlapped within the empirical fragment, these PE reads were removed from the query PE reads. The rest of the query PE reads were sorted, and the process was reiterated until no PE reads could be grouped into existing fragments. The rest of query PE reads were denoted as missed classified. At this point, the empirical fragment-building process was complete.

2.1.2. Results

2.1.2.1. *In silico* digestion

A total of 137 restriction enzymes were used in single-enzyme *in silico* digestion of the melon genome, including 87 six-cutter enzymes, 16 four-cutter enzymes, 15 five-cutter enzymes, 10 eight-cutter enzymes, and 9 seven-cutter enzymes. Among the 137 enzymes, we focused on nine commonly used enzymes in previous studies, which were *Ape*KI, *Eco*RI, *Hind*III, *Kpn*I, *Msp*I, *Pst*I, *Sac*I, *Taq*$^{\alpha}$I, and *Xba*I listed in Table 1 (Hamblin and Rabbi 2014; Fu et al. 2016; Shirasawa et al. 2016). Of the four-cutter enzymes listed

doi:10.6342/NTU202500264

in Table 1, *Taq*$^{\alpha}$I had the most frequent cutting sites in melon and was about five-fold more frequent than the other four-cutter enzyme, *Msp*I. The occurrence of cutting sites for *Ape*KI, five-cutter enzyme commonly used in single enzyme GBS, was even more frequent than *Msp*I. The six-cutter enzymes provided fewer cutting sites than the four- and five-cutter enzymes. *Hind*III and *Eco*RI had over 100K (where K = 1,000) cutting sites and could be categorized as high-frequency enzymes, while *Sac*I, *Pst*I, and *Kpn*I, with fewer than 50K sites each, were classified as low-frequency enzymes. *Xba*I, with approximately 85K cutting sites, fell between these two groups. The ratio for the number of cutting sites between rare cutters and the single common cutter ranged from 6 to 35-fold (Table 1). The highest cutting site ratio between rare and common cutter combinations was *Kpn*I and *Taq*$^{\alpha}$I, and the total fragments were almost two-fold (1.92) of the *Kpn*I cutting sites. This meant that the *Taq*$^{\alpha}$I cutting sites appear so frequently that both sides of *Kpn*I can be sampled as target fragments (AB type). The other rare cutters of the total fragment to the number of the cutting sites ranged from 1.6 to 1.8.

Based on the single-enzyme analysis, we selected *Taq*$^{\alpha}$I as common cutter and *Hind*III, *Eco*RI, *Xba*I, *Sac*I, *Pst*I, and *Kpn*I as rare cutters for two-enzyme combinations. We did not include *Msp*I as common cutter for comparison because its predicted number of cutting sites was too low for a successful preparation of the ddRAD library (Table 1). The total fragment counts, coverage, and selected fragments for the two-enzyme combinations are shown in Table 2, and the fragment length distribution is shown in Figure 4. The abundance of fragments produced by two-enzyme combinations followed the same trend as the number of fragments produced by single-enzyme digestion for rare cutter enzymes (Table 1 and Table 2). For two-enzyme *in silico* digestion, the rare cutters *Hind*III and *Eco*RI provided a higher number of fragments, 235K and 178K, respectively. *Xba*I produced 147K fragments, and *Sac*I, *Pst*I, and *Kpn*I produced the lowest number of fragments, 73K, 62K, and 47K, respectively (Table 2 and Figure 4). The extent of reduction varied between the number of fragments and the coverage of the genome. The reference genome coverage ranged from 6.5% for *Kpn*I-*Taq*$^{\alpha}$I to 25.4% for *Hind*III-*Taq*$^{\alpha}$I. The number of selected fragments (160 to 460 bp) was around one-third of the total number of fragments, ranging from 28.3% for *Sac*I-*Taq*$^{\alpha}$I to 32.8% for *Kpn*I-*Taq*$^{\alpha}$I. The coverage of selected fragments was one-fifth for total fragments, ranging from 17.0% for *Pst*I-*Taq*$^{\alpha}$I to 21.1% for *Hind*III-*Taq*$^{\alpha}$I, and the coverage of selected fragments relative to the whole genome was reduced to 1.1% for *Kpn*I-*Taq*$^{\alpha}$I to 5.3% for *Hind*III-*Taq*$^{\alpha}$I across the whole genome. Compared two-enzyme systems with single-enzyme digestion, the

doi:10.6342/NTU202500264

number of selected fragments increases by 4- to 20-fold, along with a similar increase in the ratio over total fragments (Table 1 and Table 2).

The *in silico* PE reads alignment results are shown in Table 3. The proportion of reads mapped exactly one time ranged from 68.0% for *Kpn*I-*Taq*$^{\alpha}$I to 82.4% for *Hind*III-*Taq*$^{\alpha}$I, with most combinations around 80%, except for *Kpn*I-*Taq*$^{\alpha}$I. On the other hand, reads mapped to the reference genome more than once showed an opposite trend. The unmapped reads ratios were similar across enzyme combinations, ranging from 0.6% for *Xba*I-*Taq*$^{\alpha}$I to 1.6% for *Sac*I-*Taq*$^{\alpha}$I. The *Hind*III-*Taq*$^{\alpha}$I, *Xba*I-*Taq*$^{\alpha}$I, *Pst*I-*Taq*$^{\alpha}$I, and *Kpn*I-*Taq*$^{\alpha}$I combinations formed the lower group, while the *Eco*RI-*Taq*$^{\alpha}$I and *Sac*I-*Taq*$^{\alpha}$I combinations were in the higher group of unmapped reads. The ratios of reads with MAPQ below 30 were 0.4 and 0.9% for *Xba*I-*Taq*$^{\alpha}$I and *Sac*I-*Taq*$^{\alpha}$I, respectively.

2.1.2.2. Empirical tests based on *in silico* recommendation

The PE read results of the empirical small-scale test in three samples are provided in Table 4. The average PE reads of each enzyme combination were between 0.94 M and 1.23 M PE reads for *Kpn*I-*Taq*$^{\alpha}$I and *Xba*I-*Taq*$^{\alpha}$I, respectively. The maximum and minimum PE reads ratios within each enzyme combination were between 1.03 and 1.30 for *Hind*III-*Taq*$^{\alpha}$I and *Pst*I-*Taq*$^{\alpha}$I, respectively, indicating consistent sequencing output across samples. The highest ratios of unique PE reads were for *Xba*I-*Taq*$^{\alpha}$I and *Pst*I-*Taq*$^{\alpha}$I, approximately 60%. *Eco*RI-*Taq*$^{\alpha}$I, *Kpn*I-*Taq*$^{\alpha}$I, and *Hind*III-*Taq*$^{\alpha}$I had ratios of 36.4, 30.4, and 25.9%, respectively. *Sac*I-*Taq*$^{\alpha}$I had the lowest ratio, at 18.2%. Most unique reads were originated from the nucleus while non-unique reads were mostly from organelle. Except for the *Xba*I-*Taq*$^{\alpha}$I, the proportions of PE reads mapped to organelle or nuclear genomes, relative to the total PE reads, were 11.7 and 15.8%, and the *Kpn*I-*Taq*$^{\alpha}$I was 26.4 and 35.0%. The proportion of unmapped PE reads, relative to the total PE reads, ranged from 6.9 to 13.3% for *Sac*I-*Taq*$^{\alpha}$I and *Xba*I-*Taq*$^{\alpha}$I. Among the enzyme combinations, the highest proportion of PE reads with MAPQ scores above 30 were 62.8% for *Xba*I-*Taq*$^{\alpha}$I and 63.8% for *Pst*I-*Taq*$^{\alpha}$I.

On the other hand, *Eco*RI-*Taq*$^{\alpha}$I and *Kpn*I-*Taq*$^{\alpha}$I displayed an approximately equal proportion of PE reads between MAPQ scores above 30 and those within the 0 to 9 range, with values of 32.6% and 46.5% for *Eco*RI-*Taq*$^{\alpha}$I, and 37.5% and 36.2% for *Kpn*I-*Taq*$^{\alpha}$I, respectively. *Hind*III-*Taq*$^{\alpha}$I and *Sac*I-*Taq*$^{\alpha}$I had the largest proportion of PE reads with low MAPQ scores (0 to 9), corresponding to 63.2% and 76.0%, respectively.

doi:10.6342/NTU202500264

The ddRADseq libraries were examined by electrophoresis through Bioanalyzer before sequencing (Figure 5). The electrophoresis results indicated that the *Hind*III-*Taq*$^{\alpha}$I and *Sac*I-*Taq*$^{\alpha}$I libraries exhibited a high-frequency peak in 290 bp (migration time in 69.2 sec) in 1,352 Fluorescent unit (FU) for *Hind*III-*Taq*$^{\alpha}$I and 367 bp (migration time in 75.5 sec) in 1,803 FU for *Sac*I-*Taq*$^{\alpha}$I. The high-frequency peak took 42 and 26% of the electrophoresis area. The *Kpn*I-*Taq*$^{\alpha}$I showed two high-frequency peaks in 514 and 590 bp (migration time in 84.3 and 88.0 sec), the *Xba*I-*Taq*$^{\alpha}$I contained a peak in 385 bp (migration time in 75.8 sec) in 970.8 FU and took 23.8% of the library. *Eco*RI-*Taq*$^{\alpha}$I and *Pst*I-*Taq*$^{\alpha}$I showed smooth profile with minor high-frequency peaks (Supplementary Table 2). Therefore, we assessed the insert DNA size distribution of the sequencing data to inspect the relationship between the electrophoresis results and the sequencing outcomes (Figure 5 and Figure 6). The insert size distribution of PE reads reflected the same trend of electrophoresis, especially for *Hind*III-*Taq*$^{\alpha}$I and *Sac*I-*Taq*$^{\alpha}$I exhibited a single high-frequency peak in 124 bp for *Hind*III-*Taq*$^{\alpha}$I and 199 bp for *Sac*I-*Taq*$^{\alpha}$I, accounting for 36.7% and 33.9% of total PE reads, respectively. *Xba*I-*Taq*$^{\alpha}$I exhibited a single high-frequency peak in 225 bp with approximately 8% of PE reads. The *Eco*RI-*Taq*$^{\alpha}$I, *Kpn*I-*Taq*$^{\alpha}$I, and *Pst*I-*Taq*$^{\alpha}$I showed multiple lower frequency peaks accounting for less than 10% of total PE reads (Supplementary Table 3). While the fragment analysis processed the sequencing data and identified the high-frequency peaks, the high-frequency peak often composed with PE reads with low MAPQ scores. Therefore, we applied the MAPQ score filter with the identified high-frequency peaks for clearly observe the insert size distribution of the sequencing library (Figure 6). The insert size distribution of *Xba*I-*Taq*$^{\alpha}$I, *Pst*I-*Taq*$^{\alpha}$I, and *Eco*RI-*Taq*$^{\alpha}$I exhibited higher counts relative to the *Kpn*I-*Taq*$^{\alpha}$I, *Hind*III-*Taq*$^{\alpha}$I, and *Sac*I-*Taq*$^{\alpha}$I showed lower counts of the distribution. The higher counts of the distribution of *Xba*I-*Taq*$^{\alpha}$I, *Pst*I-*Taq*$^{\alpha}$I, and *Eco*RI-*Taq*$^{\alpha}$I indicated the larger number of the remaining PE reads for the downstream analysis as better PE reads allocations.

The high-frequency fragments with similar sizes can originate from multiple genomic origins or can be an over-representation of specific regions. The distribution of PE reads across the genome, and fragment analysis were performed to identify high-frequency fragments, to determine their origins, and to reduce such fragments (Figure 7 – Figure 9). The major high-frequency fragments originated from chromosomes 0, 3, 4, 10, and the chloroplast (Supplementary Table 3). Chromosome 0 was concatenated with not being confidently assigned scaffolds and contigs. These high-frequency peaks can

doi:10.6342/NTU202500264

obscure the accurate representation of the insert size, as well as the aligned read position distribution. To reduce such interference, PE reads with MAPQ less than 20 were removed, which effectively reduced the noise (Figure 8). However, some regions still exhibited high-frequency PE reads at specific positions (Figure 9). Read filtering included MAPQ and the high-frequency fragments, shown in Table 4 and Table 5, the final remaining proportions of the PE reads in *Xba*I-*Taq*$^{\alpha}$I and *Pst*I-*Taq*$^{\alpha}$I were the highest, representing 68.4 and 61.1%, respectively. The remaining proportions, *Eco*RI-*Taq*$^{\alpha}$I, *Kpn*I-*Taq*$^{\alpha}$I, and *Hind*III-*Taq*$^{\alpha}$I, were moderate, 46.7, 41.6, and 32.5%, respectively. The remaining proportion for *Sac*I-*Taq*$^{\alpha}$I was the lowest, 19.1% of the total PE reads.

Results of fragment analysis for both *in silico* and empirical reads are listed in Table 5. The PE reads of *Hind*III-*Taq*$^{\alpha}$I and *Sac*I-*Taq*$^{\alpha}$I were represented mainly by empirical fragments, which were 58.1 and 66.8%. The *Xba*I-*Taq*$^{\alpha}$I, *Eco*RI-*Taq*$^{\alpha}$I, *Kpn*I-*Taq*$^{\alpha}$I, and *Pst*I-*Taq*$^{\alpha}$I were primarily represented by *in silico* fragments, ranging from 73.2 to 89.5% of the PE reads. The PE reads, which could not be classified into *in silico* or empirical fragments with MAPQ less than ten, accounted for between 1.5% and 4.7% of the PE reads across different enzyme combinations. By remaining PE reads with a MAPQ score below 20, we excluded most high-frequency fragments. This process allowed us to identify reliable fragments (Table 5, Figure 6 and Figure 9). Indeed, PE reads showing MAPQ>20 were mostly *in silico* fragments and accounted for 81.0% to 94.9% of the total available fragments. This demonstrates that the reliable fragments, whether *in silico* or empirical, were associated with PE reads with MAPQ scores above 20.

Among the two-enzyme combinations, *Hind*III-*Taq*$^{\alpha}$I, *Eco*RI-*Taq*$^{\alpha}$I, and *Xba*I-*Taq*$^{\alpha}$I produced a higher number of fragments (Table 6), from 57K for *Eco*RI-*Taq*$^{\alpha}$I to 77K for *Hind*III-*Taq*$^{\alpha}$I. At the same time, *Sac*I-*Taq*$^{\alpha}$I, *Pst*I-*Taq*$^{\alpha}$I, and *Kpn*I-*Taq*$^{\alpha}$I produced fewer fragments, from 20K for *Pst*I-*Taq*$^{\alpha}$I to 22K for *Sac*I-*Taq*$^{\alpha}$I of total fragments. The *in silico* predicted fragments accounted for most of the fragments, ranging from 40K (60.9.5%) for *Xba*I-*Taq*$^{\alpha}$I to 62K (80.3%) for *Hind*III-*Taq*$^{\alpha}$I. The empirical fragments accounted for a smaller proportion, making up only 3.4K (4.4%) for the C- type in *Hind*III-*Taq*$^{\alpha}$I to 11.7K (17.6%) for the R- type in *Xba*I-*Taq*$^{\alpha}$I.

The fragment depth distribution of each enzyme combination is shown in Figure 10, and summary statistics are listed in Table 6. The first quantile of fragment depth ranged from 1 to 2, while the median ranged from 3 to 14.7. Notably, the medians for *Pst*I-*Taq*$^{\alpha}$I and *Kpn*I-*Taq*$^{\alpha}$I were highest, at 14.7 and 11.0, respectively. The number of fragments was reduced by applying a fragment depth threshold, resulting in the fragment count

doi:10.6342/NTU202500264

being halved at its medians. The depth would correspond to the confidence of the genotypes, and we would evaluate the enzyme combinations by calculating the number of fragments remaining under a certain fragment depth threshold. Therefore, we calculated the numbers of fragments given the depth thresholds, as shown in Figure 11.

The number of fragments in *Hind*III-*Taq*$^{\alpha}$I, *Eco*RI-*Taq*$^{\alpha}$I, and *Xba*I-*Taq*$^{\alpha}$I was initially higher. With a depth threshold of 10 reads for *Eco*RI-*Taq*$^{\alpha}$I and *Hind*III-*Taq*$^{\alpha}$I, the number of fragments reduced to 57K to 17.5K in *Eco*RI-*Taq*$^{\alpha}$I and 77K to 7.5K in *Hind*III-*Taq*$^{\alpha}$I. Conversely, the number of fragments associated with *Xba*I-*Taq*$^{\alpha}$I demonstrated a moderate decrease, from 66K to 25K under depth above 10 reads. On the other hand, *Sac*I-*Taq*$^{\alpha}$I, *Pst*I-*Taq*$^{\alpha}$I, and *Kpn*I-*Taq*$^{\alpha}$I yielded fewer fragments. The number of fragments produced by *Sac*I-*Taq*$^{\alpha}$I decreased more rapidly than those produced by *Pst*I-*Taq*$^{\alpha}$I and *Kpn*I-*Taq*$^{\alpha}$I, and the *Pst*I-*Taq*$^{\alpha}$I displayed the most gradual decline. The number of fragments of *Sac*I-*Taq*$^{\alpha}$I was from 22K to 7.5K under a depth threshold of 10 reads, and *Pst*I-*Taq*$^{\alpha}$I and *Kpn*I-*Taq*$^{\alpha}$I were from 20K to larger than 10K in the depth of 10 reads.

2.1.3. Discussion

The single-enzyme and two-enzyme combination data clearly showed that the common cutter companion with the rare cutter could increase the genome coverage. According to the tested restriction enzymes, the number of the cutting sites of common cutters could exceed rare cutters by 10- to 30-fold (Table 1 and Table 2). However, the alignment of the *in silico* predicted fragment seems to have little difference between enzyme combinations, except the combination of *Kpn*I and *Taq*$^{\alpha}$I showed a larger difference between other enzyme combinations (Table 3). This indicated that *in silico* digestion provided potential combinations of rare cutters and common cutters suitable for the limited sequencing resources, but sequence alignment for the *in silico* predicted fragments to the reference genome did not provide many differences to differentiate the combinations. The *in silico* digestion of the reference genome considered only the nucleus and organelle sequences and could not represent the real biological conditions. For example, the composition of a single plant cell includes one nucleus along with varying numbers of chloroplasts and mitochondria. Therefore, empirical tests with small-scale samples were important in collecting true situations and clarified the allocation of the sequence reads of each enzyme combination.

During the ddRADseq library preparation process, the DNA concentration and library size were the only two major indicators to monitor the library status and its quality

doi:10.6342/NTU202500264

during library preparation. Through electrophoresis of the library, we could visually evaluate the quality of the library. High-frequency peaks might affect the downstream analysis by importing large PE read data between samples at a specific genomic region for variant calling, and it would cost memories of the server to identify variants.

The results of the empirical test of enzyme combinations showed that *Xba*I-*Taq*$^{\alpha}$I and *Pst*I-*Taq*$^{\alpha}$I provided a high proportion of unique PE reads and high MAPQ scores (Table 4 and Table 5, Figure 6). In addition, the number of fragments at different depth thresholds for these two combinations the number of fragments in *Xba*I-*Taq*$^{\alpha}$I and *Pst*I-*Taq*$^{\alpha}$I remained higher, even at a depth threshold of 30, which could give relatively high confidence for genotype calls. Meanwhile, these two enzyme combinations contained high-frequency fragments on chromosome 4 for *Xba*I-*Taq*$^{\alpha}$I and chromosome 10 for *Pst*I-*Taq*$^{\alpha}$I (Supplementary Table 3). The high-frequency fragments remain after filtering out the PE reads with MAPQ<20. Fortunately, the fragment analysis we performed could identify specific high-frequency fragments to remove through bioinformatic analysis. The *in silico* digestion we developed provided the optional feature for sequence exclusion with a third enzyme. The candidate restriction enzymes could be evaluated through in silico digestion to retain the original number of fragments in the enzyme combination or further reduce the number of fragments for exchanging for fragment depth of reliable genotype calls. Therefore, the three-enzyme combination, which incorporated a third restriction enzyme to exclude high-frequency fragments, was applied based on *Xba*I-*Taq*$^{\alpha}$I and *Pst*I-*Taq*$^{\alpha}$I to enhance the usability of PE reads. This approach aimed to allocate sequencing resources more efficiently by generating a higher proportion of unique reads for variant calling.

2.2. Compare two enzyme combinations in an F_2 population and improve the genotype quality for linkage map construction

2.2.1. Materials and Methods

2.2.1.1. The empirical test of the third restriction enzyme in three-enzyme ddRADseq

The empirical test of three-enzyme combinations, incorporating the third enzymes, *Cvi*AII, *Mse*I, and *Sph*I, was evaluated based on *Xba*I-*Taq*$^{\alpha}$I. These combinations, denoted as *Xba*I-*Taq*$^{\alpha}$I-*Cvi*AII, *Xba*I-*Taq*$^{\alpha}$I-*Mse*I, and *Xba*I-*Taq*$^{\alpha}$I-*Sph*I, were evaluated using three random samples from commercial varieties.

doi:10.6342/NTU202500264

2.2.1.2. Plant materials

Based on the results of the three-enzyme combination evaluation, the enzyme combinations *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I were analyzed in a 94-individual F_2 population. This population was derived from a cross between TARI-08874-7 (maternal parent) and TARI-08003-4 (paternal parent), with both parental lines having been self-pollinated for at least five generations.

2.2.1.3. DNA extraction and modified ddRAD library preparation

Genomic DNA was extracted following the procedure in 2.1.1.2 and was adjusted to 10–15 ng/μL for library preparation. The ddRAD library was prepared according to Peterson et al. (2012) with minor modifications as described in 2.1.1.2. The genomic DNA was double-digested using two-enzyme combinations, *Pst*I-*Taq*$^{\alpha}$I and *Xba*I-*Taq*$^{\alpha}$I, a third enzyme was added into the library after heat inactivation of the ligation product: Ten units (U) of *Cvi*AII was added to allow digestion at 25°C for 30 minutes, 10U of *Mse*I at 37°C for 120 minutes, and 5U of *Sph*I-HF at 37°C for 30 minutes.

Ligation purification, PCR, quantification of the PCR product, sample pooling, and size selection through Blue Pippin were as described in 2.1.1.2. The library for testing the third enzyme were sequenced using PE 250 sequencing (MiSeq, Illumina, San Diego, CA, USA) by Seeing Bioscience Co., Ltd (New Taipei City, Taiwan). The other two ddRAD libraries of the F_2 population with *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I were sent to the Core Facility of the Cancer Progression Research Center at National Yang Ming Chiao Tung University for PE250 sequencing HiSeq2500 Rapid mode (Illumina, San Diego, CA, USA). Both sequencing conditions were performed with extra cycles and spick-in PhiX.

2.2.1.4. The bioinformatics analysis

The FASTQ files were processed into BAM files according to the procedure described in 2.1.1.3. The optical duplicate reads in BAM files were removed using PICARD tools (http://broadinstitute.github.io/picard). Following the Best Practices recommendations of GATK v3.8 (McKenna et al. 2010), we used HaplotypeCaller for variant calling based on parental BAM files and generated the VCF file as the dbSNP to build the recalibration model for base quality score recalibration. The BAM files of the F_2 population were recalibrated based on parental VCF files. The non-unique PE reads in

doi:10.6342/NTU202500264

the recalibrated BAM files were removed using FilterSamReads. The variant calling was performed by converting the BAM files to gVCFs using HaplotypeCaller, aggregating the population-wised gVCF files, and the variants were called using GenotypeGVCFs to generate the VCF (DePristo et al. 2011; Van der Auwera et al. 2013). VCF files required filtering to eliminate low-quality variants using parameters such as QUAL, quality by depth (QD), and genotype quality (GQ). QUAL indicated the confidence of a variant based on sequencing depth, mapping quality, and base call accuracy, with higher values suggesting stronger evidence. QD normalized QUAL by alternative allele depth, correcting for inflated scores in high-coverage regions. GQ indicated confidence in the assigned genotype, reflecting the likelihood of alternative genotypes. The fragment analysis followed the same procedure as the mentioned in 2.1.1.3. The *in silico* predicted fragments produced by *Pst*I-*Taq*$^{\alpha}$I-*Sph*I, *Xba*I-*Taq*$^{\alpha}$I-*Cvi*AII, *Xba*I-*Taq*$^{\alpha}$I-*Mse*I, and *Xba*I-*Taq*$^{\alpha}$I-*Sph*I.

For linkage map construction, we used the R/qtl package (Broman et al. 2003) implemented in R software (version 4.3.1, R Core Team 2023). To prepare the input genotype data, the raw VCF files were subjected to sequential filtering with the following criteria to remove the undesired variants: (1) variants with more than two alleles, (2) the missing rate per variant above 0.5, (3) variants with a QUAL score less than 200 and a QD less than 10, and (4) a minor allele frequency (MAF) less than 0.01. The VCF files were later converted into R/qtl input format.

Segregation distortion of each marker was tested using the χ^2 goodness-of-fit (A : H : B = 1 : 2 : 1, where A and B indicate the two parental homozygous genotypes and H indicates the heterozygous genotype), with a threshold adjusted by Bonferroni correction (α = 0.05/number of markers). Markers were ordered based on physical positions, the marker phase was verified, and recombination fractions were estimated. Upon visual inspection of the recombination fraction patterns along the chromosomes, markers that were physically close to each other but showed no linkage were removed. Additionally, the marker order was manually refined based on the recombination fraction patterns or physical positions, as necessary. The genetic distances between markers were then re-estimated using the Kosambi mapping function. The number of crossovers was evaluated using countXO function implemented in R/qtl once the linkage map was constructed.

doi:10.6342/NTU202500264

2.2.1.5. Improve the genotype quality through genotype filtering or correction

High-throughput genotype data were often subjected to genotyping error. Therefore, we developed methods to improve genotype quality. The first method denoted as the Overall GQ filter, integrated GQ values extracted from the VCF file with genotype data and filtered out genotypes with GQ values below the threshold, marking them as missing. The Overall GQ filter threshold started from 0 to 80, by a step of 10, denoted as GQ0 to GQ80. Marking genotypes as missing increased the marker missing rate. Even as the GQ threshold was raised, the marker missing rate escalated significantly. Markers with missing rates exceeding the threshold were dropped, which further reduced the total number of markers. A second method, XOquality, was designed to identify the flanking GQ values of the crossover events. We applied the locateXO function in R/qtl to identify the crossover event with flanking endpoints, which provided the marker indices to include the GQ values from the VCF files and masked the genotypes with GQ less than the threshold as missing. After masking the genotype, the crossover events were re-identified iteratively for the next cycle. This process continued until both flanking genotypes associated with each crossover event had GQ values above the threshold, ensuring the crossover events were reliable. The GQ threshold for XOquality method was set as 90. The Genotype-Corrector (Miao et al. 2018) was commonly applied to identify genotype errors and to impute missing genotypes to prevent linkage map expansions frequently encountered in high-density maps. The Genotype Corrector used a sliding-window algorithm to correct the genotype within the windows across the chromosome. Thus, we sequentially test the window sizes from 5 to 21 markers with an increment of 2 markers. For convenience, we focused our analysis on chromosome 1.

2.2.2. Results

2.2.2.1. Evaluation of the third restriction enzymes for sequence exclusion

The tested three-enzyme combination, *Xba*I-*Taq*$^{\alpha}$I-*Cvi*AII, *Xba*I-*Taq*$^{\alpha}$I-*Mse*I, and *Xba*I-*Taq*$^{\alpha}$I-*Sph*I, contained 0.90 M to 1.11 M PE reads (Table 7). The unmapped reads were from 10.4% for *Xba*I-*Taq*$^{\alpha}$I-*Cvi*AII to 22.6% for *Xba*I-*Taq*$^{\alpha}$I-*Sph*I, which is higher than the two-enzyme test in terms of ratio and absolute read counts (Table 4). The proportion of unique reads ranged from 42.1 to 49.0% for *Xba*I-*Taq*$^{\alpha}$I-*Cvi*AII and *Xba*I-*Taq*$^{\alpha}$I-*Sph*I, respectively, with the major components originating from the nucleus. In

doi:10.6342/NTU202500264

contrast, the non-unique reads in *Xba*I-*Taq*$^{\alpha}$I-*Cvi*AII occupied 47.5% of total reads, relatively higher than *Xba*I-*Taq*$^{\alpha}$I-*Sph*I. For MAPQ, the PE reads with MAPQ above 30 were the major component for all three combinations from 52.0 to 60.7%. In addition, *Xba*I-*Taq*$^{\alpha}$I-*Sph*I efficiently removed most of the high-frequency fragments. However, *Xba*I-*Taq*$^{\alpha}$I-*Cvi*AII and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I still contained some proportion of high-frequency fragments on chromosome 4 and the chloroplast (Supplementary Table 3).

The PE reads located within the different types of fragments are shown in Table 6. The compositions of fragment types in *Xba*I-*Taq*$^{\alpha}$I-*Sph*I were similar to *Xba*I-*Taq*$^{\alpha}$I, but the *Xba*I-*Taq*$^{\alpha}$I-*Cvi*AII and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I contained higher RC types of the fragments, especially in *Xba*I-*Taq*$^{\alpha}$I-*Mse*I. The average fragment depths in *Xba*I-*Taq*$^{\alpha}$I-*Cvi*AII and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I were higher than in *Xba*I-*Taq*$^{\alpha}$I-*Sph*I, which indicated that the third restriction enzymes of the four-cutter enzymes, *Mse*I and *Cvi*AII, excluded more sequences than the six-cutter enzyme, *Sph*I.

The number of fragments under different fragment depth thresholds is shown in Figure 11. The number of fragments decreased rapidly in *Xba*I-*Taq*$^{\alpha}$I-*Cvi*AII while the threshold of fragment depth increased. When the fragment depth was above 12 to 15, the remaining fragments in *Xba*I-*Taq*$^{\alpha}$I-*Cvi*AII were higher than *Xba*I-*Taq*$^{\alpha}$I-*Mse*I and *Xba*I-*Taq*$^{\alpha}$I-*Sph*I, approximately 10K to 12K of the fragments. The number of fragments in *Xba*I-*Taq*$^{\alpha}$I-*Cvi*AII were higher than *Xba*I-*Taq*$^{\alpha}$I-*Mse*I and *Xba*I-*Taq*$^{\alpha}$I-*Sph*I at a higher depth, but the cost of *Cvi*AII and *Sph*I was higher than *Mse*I, and *Cvi*AII was less common than *Mse*I. Therefore, we decided to apply *Xba*I and *Pst*I as rare cutters, *Taq*$^{\alpha}$I as common cutter, and *Mse*I and *Sph*I as sequence-excluding enzymes to one F_2 population for the construction of a high-density linkage map.

2.2.2.2. Test of *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I in an F_2 population

For the test of *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I in the F_2 population, the average PE reads were 0.70 M in *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and 0.65 M in *Xba*I-*Taq*$^{\alpha}$I-*Mse*I, and the PE reads ratio between maximum and minimum samples was 2.06 in *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and 2.64 in *Xba*I-*Taq*$^{\alpha}$I-*Mse*I (Table 8). The PE reads composition in *Pst*I-*Taq*$^{\alpha}$I-*Sph*I was similar to the empirical test of *Pst*I-*Taq*$^{\alpha}$I, which was 58.9% and 27.1% of unique and non-unique reads (Table 4). The PE reads composition in *Xba*I-*Taq*$^{\alpha}$I-*Mse*I was relatively close to the empirical test of *Xba*I-*Taq*$^{\alpha}$I (Table 4), which was 53.8% and 30.5% in unique and non-unique reads (Table 7), but the unique reads were the major proportion of the PE reads composition.

doi:10.6342/NTU202500264

The common fragments, which were present in at least 50% of the F_2 population for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I, were categorized into AB, RC, R-, C-, and -- (Table 9). The selected variants with QUAL>200 and QD>10, MAF>0.01, and missing rate<0.5 were 7,252 for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and 31,3969 for *Xba*I-*Taq*$^{\alpha}$I-*Mse*I. *Pst*I-*Taq*$^{\alpha}$I-*Sph*I provided 14,509 fragments with the major type of AB fragments in 12,445, and the 4,049 fragments carried 6457 variants. *Xba*I-*Taq*$^{\alpha}$I-*Mse*I provided 43,571 fragments with the major type of RC fragments in 31,133, and the 11,862 fragments carried variants. The overall fragment polymorphism was 31.1% in *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and 35.3% in *Xba*I-*Taq*$^{\alpha}$I-*Mse*I.

2.2.2.3. Construction of high-density linkage maps

The total variants were 86,941 for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and 184,442 for *Xba*I-*Taq*$^{\alpha}$I-*Mse*I and were categorized into SNP, Insertion or Deletion (InDel), and multiple alternate alleles variants (Table 10). The major variants were SNP, which was 71,668 (82.4%) for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and 156,191 (84.7%) for *Xba*I-*Taq*$^{\alpha}$I-*Mse*I. The InDel was relatively few, with a proportion of 16.6% (14,418) and 14.2% (26,221) for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I, respectively. Based on the biological feature of a bi-parental F_2 segregating population, the maximum allele numbers were restricted to two. Therefore, variants with multiple alternate alleles were removed for further analysis. Combined the criteria of marker missing rate above 0.5, variants with quality (QUAL<200 and QD<10), and MAF less than 0.01, the remaining markers were 7,252 (56.9%) in *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and 3,1969 (58.7%) in *Xba*I-*Taq*$^{\alpha}$I-*Mse*I relative to the first filter. The filtered variants with chromosome assignment were 7,141 in *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and 31,350 in *Xba*I-*Taq*$^{\alpha}$I-*Mse*I and were input in R/qtl for linkage map construction.

To further improve marker quality, only markers showing missing rate less than 0.25 were retained for map construction, which were 6,066 for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and 17,713 for *Xba*I-*Taq*$^{\alpha}$I-*Mse*I. 885 and 3,330 markers in *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I were removed due to segregation distortion, and 25 and 275 markers in *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I were dropped due to no linkage with the flanking markers (Table 11). The integrated genotype data with GQ value masked the genotype with GQ less than zero as missing. This gave a final number of high-quality markers of 5,146 for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and 13,832 for *Xba*I-*Taq*$^{\alpha}$I-*Mse*I.

The linkage maps built with *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I variants are shown in Figure 12, and the map summary is in Table 12. The total linkage map length

doi:10.6342/NTU202500264

was 6,939.9 cM for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and 60,647.6 cM for *Xba*I-*Taq*$^{\alpha}$I-*Mse*I, and the large intervals between markers were 29.4 cM on chromosome 4 in *Pst*I-*Taq*$^{\alpha}$I-*Sph*I, and 460.5 cM on both chromosomes 2 and 3 in *Xba*I-*Taq*$^{\alpha}$I-*Mse*I. The length difference between the two maps was large, which reflected the great number of crossovers in *Xba*I-*Taq*$^{\alpha}$I-*Mse*I than in *Pst*I-*Taq*$^{\alpha}$I-*Sph*I (Figure 13 and Supplementary Figures). The origin of such a difference in the linkage map is commonly considered as the genotype error. Thus, we applied an overall genotype filter to retain markers of better quality.

The number of crossovers between two-enzyme combinations initially showed a large difference (Figure 13). The number of crossovers was reduced with a more stringent GQ threshold, and the difference between the two datasets was reduced. However, with the overall GQ filter, the number of markers decreases, especially for *Xba*I-*Taq*$^{\alpha}$I-*Mse*I, 47.5% (6,571) of markers were removed from GQ0 to GQ10 (Table 14). The number of crossovers was efficiently reduced toward 1 to 3 per chromosome, especially in the *Xba*I-*Taq*$^{\alpha}$I-*Mse*I genotype (Figure 13). Based on the result, we temporarily applied the GQ threshold of 50 to improve the genotype data and re-estimate the linkage map. The improved linkage map summary for both *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I is shown in Table 13, and the linkage maps are shown in Figure 14.

The relationship between genetic distance and physical distance can reveal the recombinant pattern across the genome. Assuming that both enzyme combinations produced a sufficient number of markers across each chromosome to capture nearly all crossover events, the recombination pattern for each chromosome should remain consistent. This consistency is based on the biological constraints of a single F_2 population. For the first linkage map, the *Xba*I-*Taq*$^{\alpha}$I-*Mse*I showed the nonstop increasing genetic distance across the chromosome (Figure 15). On the contrary, when genotypes with GQ less than 50 were removed, the relationship between genetic distance and physical distance became comparable, except for the slight difference on chromosome 10 (Figure 16).

2.2.2.4. Genotype correction efficiently improved marker quality

The genotype quality improvement methods included in this study, Overall GQ filter, XOquality, and Genotype-Corrector, were started from the basis of the Overall GQ filter, denoted as Overall in Table 15. The map lengths of chromosome 1 in the Overall GQ filter were reduced from 940.6 to 139.7 cM in *Pst*I-*Taq*$^{\alpha}$I-*Sph*I, and 6197.4 to 155.8 cM in *Xba*I-*Taq*$^{\alpha}$I-*Mse*I. While the Overall GQ filter was set at 80, the map lengths were 134.8

doi:10.6342/NTU202500264

cM and 140.6 cM for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I, but the number of markers was 300 and 118 for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I. These indicated the map length of chromosome 1 should be approximately between 135 and 140 cM, and the larger number of markers did not necessarily inflation the linkage map. The map length range of 135 to 140 cM was considered the convergence map length for chromosome 1. Regarding the second method, XOquality, based on the number of crossovers, it can perform well under different GQ-filtered genotypes in *Pst*I-*Taq*$^{\alpha}$I-*Sph*I, but the map length was larger in *Xba*I-*Taq*$^{\alpha}$I-*Mse*I. The linkage map length reduced with the increasing window size for Genotype-Corrector, especially when windows were larger than 7 markers (Table 15). The map lengths were 133.1 cM for the window size of 13 markers in GQ0 and 131.8 cM for the window size of 15 markers in GQ50, which were close to the Overall GQ filter result in *Pst*I-*Taq*$^{\alpha}$I-*Sph*I. When the window sizes were larger than 15 markers in different Overall GQ-filtered genotypes, the minimum map length shrank to 115.3 cM, which might indicate an over-correction in *Pst*I-*Taq*$^{\alpha}$I-*Sph*I (Figure 17). In the *Xba*I-*Taq*$^{\alpha}$I-*Mse*I genotype dataset, the map length was 290.1 cM for the window size of 21 markers in GQ0 and 169.1 cM for the window size of 19 markers in GQ10, which was the minimum map length under GQ0 and GQ10 genotype data after Genotype-Corrector corrected the genotype data.

While the GQ-filtered genotype data with GQ threshold larger than 20, the map length approximated the convergence map length that appeared from the windows size 17 markers in GQ20 genotype data in 134.5 cM, and window size 11 markers in GQ50 genotype data in 137.5 cM. The map length smaller than the convergence length would occur under window size larger than 17 and GQ threshold larger than 20 (Table 15). The results indicated that the overall genotype quality in *Pst*I-*Taq*$^{\alpha}$I-*Sph*I was better than *Xba*I-*Taq*$^{\alpha}$I-*Mse*I based on the map length in the Overall GQ filter result. However, the Genotype-Corrector corrected the genotype data well, at least under the Overall GQ 20 filtered genotype data. These provided guidance for genotype data filtering for linkage map construction.

2.2.3. Discussion

When including a third enzyme in the ddRAD library preparation, some extended questions needed to be answered, such as the suitable moment to introduce the third enzyme during the library preparation. In this study, we applied the third enzyme after ligation (Figure 1). Our results showed that a third enzyme could efficiently reduce the

doi:10.6342/NTU202500264

genome complexity. The *Sph*I seemed to perform as expected in removing the high-frequency fragments and rescuing the 5% of the PE reads for downstream analysis. *Cvi*AII and *Sph*I were costly, and the pricing issue became relevant for large-scale experiments. *Mse*I does not seem to be effective, given the results of *Xba*I-*Taq*$^{\alpha}$I-*Mse*I fragment analysis (Table 6). The composition of the PE reads with a third enzyme followed the trend of a two-enzyme empirical test (Table 5 and Table 7). Therefore, exploring additional rare cutters will be an opportunity for a better allocation of the composition of the PE reads. The rare cutters we had selected and tested, i.e. *Hind*III, *Eco*RI, *Xba*I, *Sac*I, *Pst*I, and *Kpn*I, can be considered as a standard testing set to screen and test for marker discovery and genotyping while dealing with a new species.

The *Xba*I-*Taq*$^{\alpha}$I-*Mse*I and *Pst*I-*Taq*$^{\alpha}$I-*Sph*I yielded different numbers of target fragments under similar sequencing resources. The *Pst*I-*Taq*$^{\alpha}$I-*Sph*I with a lower number of fragments and a higher genotype quality could be applied to highly heterozygotic segregating populations, such as F_2 and BC populations. The *Xba*I-*Taq*$^{\alpha}$I-*Mse*I or *Xba*I-*Taq*$^{\alpha}$I-*Sph*I yielded a higher number of fragments that should be applied to the highly homozygotic population, such as RILs for more markers across the genome and to identify the chromosome segments derived from the parental lines. In this study, we analyzed an F_2 population with two enzyme combinations. The linkage maps of the two datasets were expected to be similar since the data were collected on the same samples. The first version of the linkage maps showed a big difference between the two datasets. The genotype error was the major effect that caused the difference between the two linkage maps.

Linkage map construction is highly sensitive to data quality, as erroneous genotypes can disrupt marker ordering and inflate the genetic map, especially with increasing marker density. Although a reference genome can provide a framework for basic marker ordering, precise estimation of genetic distances depends on high-quality genotype data. Effective strategies for genotype correction and imputation in constructing high-density linkage maps include marker binning to consolidate redundant markers (Sim et al. 2012; Gonda et al. 2019), sliding windows to address local inconsistencies (Miao et al. 2018; Gao et al. 2019), and hidden Markov models for probabilistic error correction and genotype imputation (Fragoso et al. 2016; Lorieux et al. 2019; Furuta et al. 2023; Campos-Martin et al. 2023).

The genotype error rate is often unknown, making conservative filtering necessary, though it may result in excessive loss of genotype data. It is crucial to consider the starting

doi:10.6342/NTU202500264

quality of genotype data carefully. Our findings indicate that applying a GQ filter to exclude genotypes with a GQ score below 20 improves the data quality, making it more suitable for further genotype correction (Table 15 and Figure 17). However, this study also reveals that Genotype-Corrector may struggle to accurately identify errors without proper parameter tuning, potentially lead to under- or over-correction. To address this, employing a conservative GQ filter as a baseline before applying Genotype-Corrector with optimized parameters is recommended. Combining multiple tools and approaches to enhance genotype data quality can ultimately increase the reliability of the resulting linkage map.

3. Application of optimized ddRADseq in melon

3.1. Identify Powdery mildew resistance QTL in three F_2 populations

3.1.1. Materials and Methods

3.1.1.1. Three F_2 populations for QTL mapping

In this study, we used three F_2 populations, A6, B2, and C4, to map powdery mildew resistant QTL (Table 16). The A6 population, composed of 165 F_2 individuals, was derived from a cross between TARI-18-437 and TARI-18-494. The B2 population, composed of 179 F_2 individuals, was derived from a cross between TARI-18-410 and TARI-18-491, while the C4 population, with a population size of 179 F_2 individuals, was derived from a cross between TARI-18-491 and TARI-18-449. B2 and C4 shared TARI-18-491 as a common parent, but it served as the male parent for B2 and the female parent for C4.

The parental inbred lines, TARI-18-437 and TARI-18-410, were characterized by netted-rind, green-fleshed fruit, and resistance to *P. xanthii* race 1. TARI-18-437 was derived from a hybrid variety, and TARI-18-410 was generated from another accession after 16 and 14 generations of selfing, respectively. TARI-18-449 was a powdery mildew resistance inbred line developed through five generations of selfing of a hybrid variety with PI 124111 and PI 124112 in its pedigree. The powdery mildew resistance of the parental lines for the three F_2 populations is shown in Table 17.

doi:10.6342/NTU202500264

3.1.1.2. powdery mildew evaluation

The powdery mildew reaction was assessed using leaf disc inoculation, as described by Wang et al. (2016b). The experimental unit consisted of two leaf discs of 15-mm diameter per genotype, sampled from the second leaf when the melon plant was at the third-leaf stage. The leaf discs were placed on M-solution (10,000 ppm mannitol, 30 ppm benzimidazole, and 50 ppm tetracycline) in 60-mm Petri dishes. In each Petri dish, IRANH, a susceptible variety, and PMR45, a resistant variety, were used as experimental controls. For each genotype, the disease reaction was evaluated using two replications. A conidial suspension was uniformly sprayed over the leaf discs to a density of 50 – 100 spores per cm^2. Leaf discs were incubated at 24°C/18°C (day/night) with a 12-hour photoperiod for 12 days. The disease index (DI) was scored for each disc on a scale of 0 to 9, where 0 = no lesions; 1 = lesions covering 10% of the leaf area; 3 = lesions covering 50% of the leaf area; 5 = lesions covering 80% of the leaf area; 7 = lesions covering 100% of the leaf area, with thin spores on the leaf; and 9 = lesions covering 100% of the leaf area, with a thick brown disc of sporangia on the leaf discs. Plants with a mean DI<3.0 were considered resistant, while those with a mean DI≥3.0 were susceptible (Huang et al. 2002). For each batch of experiments, a non-inoculated set was set aside as a negative control. Disease rating was conducted in the fungal disease laboratory of the Plant Pathology division, TARI, Taichung, Taiwan, in spring 2017 and autumn 2018. Based on the leaf disc powdery mildew reaction evaluation on the melon differential set (McCreight 2006), the physiological race inoculated in this study is *P. xanthii* race 1.

3.1.1.3. ddRAD library preparation for genotyping in F_2 populations

Prior to library construction, we used *in silico* analysis on the melon reference genome DHL92 v3.5.1 (Argyris et al. 2015) and organelle genomes (Rodríguez-Moreno et al. 2011) to estimate adapter amounts and predict digested fragment sizes. Based on this analysis, the appropriate restriction enzyme sets for this study were *Pst*I, *Taq*$^{\alpha}$I, and *Sph*I. *Pst*I and *Taq*$^{\alpha}$I, serving as rare and common cutters, have average recognition site distances of 11,914 bp and 482 bp, respectively, which determined the adapter input amounts. *Sph*I was used to exclude target fragments containing the recognition site associated with rDNA sequences based on results from prior experiments.

Genomic DNA was extracted from freeze-dried young leaves using a modified 1.25% SDS method (Jobes et al. 1995). Extracted DNA was purified using the QIAquick

doi:10.6342/NTU202500264

96 PCR Purification Kit (Qiagen, Hilden, Germany). DNA concentration was quantified using Quant-iT™ PicoGreen™ dsDNA Assay Kits (Life Technologies, Oregon, USA) and adjusted to 10–15 ng/μL for library preparation. The ddRAD library was prepared according to Peterson et al. (2012) with minor modifications as follows. The restriction enzymes, T4 DNA ligase, CutSmart Buffer, and rATP used for library construction were from New England Biolabs, Ipswich, MA, USA. Initially, 300 ng of genomic DNA from each sample was double-digested using *Pst*I-HF and *Taq*αI at 37°C and 65°C, respectively, for 30 minutes. Ligation was then proceeded overnight at 16°C, incorporating a 5-fold excess of P1 (*Pst*I-HF) and Y (*Taq*αI) adapters, T4 DNA ligase, and rATP in 1X CutSmart Buffer. After inactivating the ligase, the library was treated with *Sph*I-HF at 37°C for 30 minutes to exclude sequences and then purified using 0.8X AMPure XP beads (Beckman Coulter, Brea, CA, USA) to eliminate short fragments and adapter dimers. PCR amplification was conducted using dual-indexed primers, including Nextera XT DNA Indexes v2 and Phusion polymerase (Thermo Fisher Scientific Baltics UAB, Vilnius, Lithuania) in a 20-cycle two-step PCR protocol, which included only the denature and extension steps. The final PCR products were purified using 0.8X AMPure XP beads, quantified with Quant-iT™ PicoGreen™ dsDNA Assay Kits, and pooled in equal amounts. The ddRAD library target fragments were size-selected, ranging from 300 to 600 bp, using Blue Pippin (Sage Science, Beverly, MA, USA) in a 2% electrophoresis gel cassette. The size-selected ddRAD libraries were then sent to the Core Facility of the Cancer Progression Research Center at National Yang Ming Chiao Tung University for PE250 sequencing with extra dark cycles and a spike-in of 10% PhiX, utilizing the HiSeq2500 Rapid mode (Illumina, San Diego, CA, USA).

3.1.1.4. Bioinformatic analysis workflow for SNP calling

Paired-end FASTQ files were processed with AdapterRemoval v2.1.7 (Schubert et al. 2016) and aligned to the DHL92 v3.5.1 reference genome using Bowtie2 v2.2.9 (Langmead and Salzberg 2012). Alignments were sorted, compressed, and indexed into BAM files with SAMtools v1.3 (Li et al. 2009), and optical duplicate reads were removed using PICARD tools (http://broadinstitute.github.io/picard). Following GATK v3.8 (McKenna et al. 2010) Best Practices recommendations, base quality score recalibration refined sequence quality scores through a recalibration model. This model, typically relying on a known variants database such as dbSNP to distinguish true genetic variants

doi:10.6342/NTU202500264

from sequencing errors, utilized VCF files from HaplotypeCaller of parental BAM files as a dbSNP substitute, improving variant calling accuracy. F_2 population BAM files were recalibrated, filtered with FilterSamReads, and then converted to gVCFs using HaplotypeCaller. Finally, aggregating all gVCFs through GenotypeGVCFs produced the population's VCF file (DePristo et al. 2011; Van der Auwera et al. 2013). The raw VCF files were subjected to filtering with specified criteria: (1) variants with more than two alleles were removed; (2) variants with a QUAL score less than 200, a Quality by Depth (QD) less than 10, and a minor allele frequency under 0.01 were excluded; and (3) genotypes with a Genotype Quality (GQ) from the VCF file's FORMAT fields below 20 were marked as missing, increasing the missing rate threshold to 0.25.

3.1.1.5. Linkage map construction and QTL mapping

For linkage map construction and QTL mapping, we used the R/qtl package (Broman et al. 2003) implemented in R software (version 4.3.1, R Core Team 2023). Within each F_2 population, the segregation distortion of each SNP was tested using the $\chi 2$ goodness-of-fit with a threshold adjusted by Bonferroni correction (α = 0.05/number of SNPs). SNPs were ordered based on physical positions, marker phasing was verified, and recombination fractions were estimated. Upon visual inspection of the recombination fraction patterns along the chromosomes, SNPs that were physically close to each other but showed no linkage were removed. Additionally, the SNP order was manually refined based on the recombination fraction patterns or physical positions, as necessary.

To prevent linkage map expansion frequently encountered in high-density maps, the Genotype Corrector (Miao et al. 2018) was applied to identify genotype errors and impute missing genotypes. Genotype Corrector used a sliding-window algorithm. The window size was set to 13 after empirical tests with the data from this study. The linkage map was then re-estimated using the Kosambi mapping function based on the corrected and imputed genotype data.

Both single QTL analysis and multiple interval mapping (Manichaikul et al. 2009) based on Haley and Knott regression (Haley and Knott 1992) were used for QTL mapping. For single QTL analysis, the empirical threshold was determined for each population based on 1,000 permutations (Churchill and Doerge 1994) at α = 0.05. For multiple interval mapping, 1,000 permutations for pairwise QTL mapping were first performed for each population. Thresholds for individual QTL and pairwise epistasis were derived from the permutation results at α = 0.01 with heavy interaction penalties only.

doi:10.6342/NTU202500264

3.1.2. Results

3.1.2.1. The DI of powdery mildew in three F_2 populations

The DI of powdery mildew in melon for the A6, B2, and C4 populations, their parental inbred lines, and F_1 individuals are shown in Figure 18. Disease reactions of most of the F_2 individuals were between those of the two parents for the three populations. The F_1 individuals showed a resistant phenotype, indicating that powdery mildew resistance was a dominant trait in the three populations. The phenotypic distribution of each F_2 population is largely skewed toward resistance, suggesting that powdery mildew resistance is controlled by a few major QTL in the populations of interest.

3.1.2.2. powdery mildew resistance QTL

For A6, B2, and C4, 127.14, 135.97, and 138.65 million paired-end reads were generated, respectively. On average, 0.76 million paired-end reads per sample were available for the three populations. The number of variants was 454,561 for A6, 444,854 for B2, and 465,971 for C4. After quality filtering, segregation distortion tests, and correction for genotyping errors, the final SNP data sets consisted of 3,466 SNPs for A6, 5,449 SNPs for B2, and 6,408 SNPs for C4. Across the full SNP data, 833 SNPs were shared by the three F_2 populations, while B2 and C4 shared up to 73% or 62% common SNPs, given that the two populations had a common parent (Figure 19). The total map lengths were 1,659.8 cM for A6, 1,586.0 cM for B2, and 1,609.2 cM for C4. SNPs were evenly distributed along the genome, except for some gaps on chromosomes 2 (B2, C4), 3 (A6), 5 (B2), 8 (A6), 9 (A6), and 10 (B2) (Figure 20 and Table 18).

The QTL identified in this study are summarized in Table 19 and Figure 20. One major QTL, *qPM2*, was identified in A6. It was located on chromosome 2 and explained up to 93% of the total phenotypic variance. In B2, one major QTL, *qPM5*, was identified on chromosome 5, explaining 80% of the phenotypic variance. Two QTL, *qPM5* and *qPM12*, were identified in C4, explaining up to 70% of the total phenotypic variance. All identified QTL exhibited dominant behavior. The confidence intervals for *qPM5*, extending to the closest flanking markers in the physical distance for the B2 and C4 populations on chromosome 5, were 25,267,104 – 25,725,099 bp and 25,203,821 – 25,678,875 bp, respectively. These overlapping intervals suggest that *qPM5* in the B2 and C4 populations can be considered the same QTL.

doi:10.6342/NTU202500264

3.1.3. Discussion

In this study, we have identified three major melon powdery mildew resistance QTL located on chromosomes 2, 5, and 12 in three F_2 populations and introduced these resistant QTL to the powdery mildew susceptible elite lines. According to the results of powdery mildew resistance QTL in Table 19, both the A6 and B2 populations carried single powdery mildew resistance QTL, *qPM2* and *qPM5*, respectively, and the C4 population carried two, *qPM5* and *qPM12*. The phenotypic variance explained (R^2) and the additive effect (a) of *qPM2* were larger than *qPM5*. While *qPM5* was accompanied by *qPM12* in the C4 population, the R^2 and the LOD scores of *qPM12* were larger than *qPM5*. Although the additive effect of *qPM12* was slightly lower than *qPM5*, the average DIs of PM in the near-isogenic lines carried *qPM5* were less than those carried *qPM12* under the green-fleshed genetic background, suggesting *qPM2* and *qPM12* contribute more powdery mildew resistance than *qPM5*.

To compare the resistance QTL across different studies, we blasted the reference genome DHL92 v4 (Castanera et al. 2020) using QTL flanking marker sequences and identified candidate genes within the interval between foreground markers using the DHL92 v4 annotation (Ruggieri et al. 2018). Under the interval between foreground markers, *qPM2*, *qPM5*, and *qPM12* covered 128, 98, and 116 genes, respectively (Supplementary Table 5 – Supplementary Table 7). Among this large number of genes, we will focus our discussion on candidate genes showing transcriptional or functional evidence from previous studies. Several studies have identified powdery mildew resistance QTL overlapping *qPM2* identified in the present study, including *Pm-pxA.II* and *Pm-pxB.II* (Fukino et al. 2008), *Pm-Edisto47-1* (Ning et al. 2014), *qPM2* (Wang et al. 2016b) *Pm-2F* (Zhang et al. 2013), *Pm2.1* (Haonan et al. 2020), and *Pm-II* (Kim et al. 2016). Among these QTL, *Pm-2F* and *Pm2.1* were identified through powdery mildew race 2F and *Pm-II* through race N5, while the other powdery mildew resistance QTL on chromosome 2 were identified through the inoculation of race 1. Therefore, there may be three different QTL co-located at this region. *MELO3C015353* and *MELO3C015354* may be the candidate genes because they were annotated as disease-resistance proteins with leucine-rich repeat (LRR) domains (Haonan et al. 2020). *qPM5* identified in this study was co-localized with several powdery mildew resistance QTL identified previously, including *Pm-AN* (Wang et al. 2011), *qPx1-5* (Branham et al. 2021), *Pm-R1-2*, and *PM-R5* (Yuste-Lisbona et al. 2011a, b). López-Martín et al. (2022) further narrowed down the

doi:10.6342/NTU202500264

region and identified the candidate genes, *MELO3C004297* and *MELO3C004311*, which encoded a branched-chain amino-acid aminotransferase-like protein and a tomato mosaic virus resistance protein N-like gene, respectively. *qPM12* identified in this study co-localized with QTL identified using different races: *Pm-pxA.XII* and *Pm-pxB.XII* (Fukino et al. 2008), *BPm12.1* (Li et al. 2017), *CmPMRl* (Cui et al. 2022), and *qPx1-12* (Branham et al. 2021), conferring resistance to race 1; *Pm-R1-2*-5 (Beraldo-Hoischen et al. 2012) to races 1, 2, and 5; *Pm-XII* to race N5 (Kim et al. 2016); *pm12.1* (Haonan et al. 2020) and *Cmpmr2F* to race 2F (Zhang et al. 2023); and *qCmPMR-12* (Cao et al. 2021) to non-specified races. López-Martín et al. (2022) further fine-mapped the QTL and identified one of the candidate genes, *MELO3C002504*, characterized as a cysteine-rich receptor-like protein kinase. Zhang et al. (2023) identified *MELO3C002403* as a candidate gene for *Cmpmr2F*. The transcription level of *MELO3C002403* in the resistant parent PI 124112 increased during powdery mildew inoculation. It encoded an allantoate amidohydrolase protein, primarily residing in the cytoplasm and cell membrane. *qPx1-12* harbored *MELO3C002392* and *MELO3C002393*, both were LRR receptor-like kinases (Branham et al. 2021). *MELO3C002434*, *MELO3C002438*, *MELO3C002439*, *MELO3C002440*, and *MELO3C002441* were Ankyrin repeat family proteins located within at least one of the following QTL: *BPm12.1*, *pm12.1*, *qCmPMR-12*, and *CmPMRl*. The parental lines for mapping *BPm12.1* and *qCmPMR-12* carried non-synonymous mutations at *MELO3C002434* and the expression level of *MELO3C002434* was significantly higher in the resistant parent after powdery mildew inoculation (Li et al. 2017; Cao et al. 2021). Therefore, it is a potential causal gene for *qPM12*. The genes *MELO3C002441* to *MELO3C002449* were located within *CmPMRl*. Among these genes, six carried non-synonymous SNPs between resistant and susceptible genotypes. In addition, *MELO3C002441*, *MELO3C002444*, and *MELO3C002448* were significantly upregulated after powdery mildew inoculation of resistant melon genotypes. Conversely, *MELO3C002446*, *MELO3C002447*, and *MELO3C002449* were downregulated during infection. *MELO3C002449* was associated with glycolytic enzyme activity but the functions of *MELO3C002446* and *MELO3C002437* remain unknown. Integrating results from different powdery mildew resistance studies, it seems that the powdery mildew resistance QTL are predominantly clustered on chromosomes 2, 5, and 12. QTL detected through the use of different races co-localized at these three regions. Therefore, these regions may harbor multiple resistance loci and are promising targets for melon powdery mildew resistance MABC.

doi:10.6342/NTU202500264

3.2. MABC for powdery mildew resistance of elite inbred lines

3.2.1. Materials and Methods

3.2.1.1. The recurrent parents for MABC

Marketing preferences for new melon varieties primarily depend on fruit quality traits, particularly flesh color. Thus, we selected TARI-18-431, with orange flesh and a globular shape, and TARI-18-432, with green flesh and a rounded shape, as the elite recurrent parents (Table 16). Both parents possess desirable horticultural traits for fruit rind netting and taste quality but without powdery mildew resistance (Table 17). Therefore, the MABC breeding programs in this study aimed to improve the powdery mildew resistance of these two elite inbred lines by using them as recurrent parents.

3.2.1.2. Conversion between physical and genetic distances

To integrate the relationship between genetic distance and physical distance from the three F_2 populations, the locally estimated scatterplot smoothing (LOESS) regression was employed. The regression model for each chromosome used 10% local neighborhood data points and incorporated a quadratic term (degree of 2) in the fitting predictors (Siberchicot et al. 2017). The LOESS regression model on the carrier chromosome was used to select candidate SNPs for foreground and recombinant selection.

3.2.1.3. TaqMan assay development for MABC

Prior to designing TaqMan assays, MABC parental lines were genotyped using ddRAD sequencing. Polymorphic SNPs on carrier chromosomes with high quality (QUAL>1000 and QD>10) were selected. Each selected SNP was at least 50 bp away from any other variants and positioned at least 30 bp from the edge of the target fragments. Sequences of target fragments carrying the selected SNPs were then designed using the Custom TaqMan® Assay Design Tool (Thermo Fisher Scientific 2024). The TaqMan assays for foreground selection were chosen based on the confidence intervals extending to the closest flanking markers. For recombinant selection, designed SNP markers were selected considering an average seed yield of approximately 200 – 500 per cross and the necessity of having at least one individual per cross for each MABC generation to manage risks effectively. The assays were chosen to cover at least 20 cM on both sides of the target region. TaqMan assays were validated using parental DNA samples. Recombinant

doi:10.6342/NTU202500264

and foreground selection were made using the TaqMan® SNP Genotyping Assay Mix (Life Technologies, Marsiling, Singapore) and TaqMan™ Universal PCR Master Mix II, no UNG (Thermo Fisher Scientific Baltics UAB, Vilnius, Lithuania) according to the manufacturer's instructions.

3.2.1.4. MABC for powdery mildew resistance introgression

The MABC aimed to enhance powdery mildew resistance in two elite recurrent parents. The process involved crossing donor parents with recurrent parents and selecting offspring according to target regions designated as *qPM2*, *qPM5*, and *qPM12* (Figure 21 – Figure 23).

At the BC_1F_1 generation, we aimed to first reduce the linkage drag around target loci via recombinant selection. Only individuals carrying homozygous for the recurrent parent at one recombinant marker and heterozygous genotypes at the other marker were passed to foreground selection. Those carrying resistance donor alleles showing heterozygous genotypes in two bracketed foreground markers were passed to background selection. Background selection was performed using ddRAD sequencing to well cover the whole genome, and the individuals showing high RPG recovery were then crossed with the recurrent parent to produce BC_2F_1 generation. The RPG recovery of the selected individuals was estimated using ddRAD genotype data. Heterozygous genotypes were assigned a value of 0.5, and homozygous genotypes of the recurrent parent were assigned a value of 1. These values were then weighted by the genetic distance between flanking markers and divided by the total genetic distance covered by the ddRAD markers on the non-carrier chromosomes.

At the BC_2F_1 generation, recombinant selection was first applied to screen for homozygous individuals at the second recombinant marker. Individuals showing homozygous genotypes at both recombinant markers were kept for foreground selection. Heterozygous individuals at foreground markers were selected for background selection using ddRAD sequencing. Individuals with the highest RPG recovery were backcrossed to the recurrent parents to generate BC_3F_1 individuals.

At the BC_3F_1 generation, only foreground selection was conducted. Subsequently, individuals with the highest RPG, as determined through ddRAD sequencing, were selected and self-pollinated to produce BC_3F_2 progeny. In the final BC_3F_2 generation, foreground selection was conducted to identify candidate individuals carrying homozygous powdery mildew resistance alleles. These individuals were then self-

doi:10.6342/NTU202500264

pollinated to establish the near-isogenic lines as the final outcome of the MABC in this study. The MABC schemes, as generally outlined above, were subject to minor adjustments depending on specific circumstances.

3.2.2. Results

3.2.2.1. TaqMan assays for MABC

To design markers for foreground selection, the confidence interval of each QTL was extended to the closest flanking markers (Table 19 and Supplementary Table 4). The *qPM2* region, located at 8.34 – 10.43 cM on chromosome 2 in the A6 population, had a corresponding physical distance from 778,333 to 1,142,653 bp. The *qPM5* region, located at 91.63 – 94.67 cM on chromosome 5 in the B2 population and at 86.45 – 90.58 cM on the same chromosome in the C4 population, had overlapping physical distances of 25,267,104 – 25,725,099 bp for B2 and 25,203,821 – 25,678,875 bp for C4, with a common fragment at 25,203,821 – 25,725,099 bp. The *qPM12* region, identified at 75.15 – 78.44 cM on chromosome 12 in the C4 population, had a corresponding physical distance from 22,493,040 to 22,879,440 bp.

Accounting for polymorphic SNPs between donor and recurrent parents, and the requirements for the design tool, the designed and validated TaqMan assays are shown in Table 20. The genetic distances between bracketed foreground selection markers ranged from 0.3 to 9.6 cM for *qPM2* in the orange-fleshed parent and green-fleshed parent, respectively. The genetic distances between markers for recombinant selection and foreground selection ranged from 2.3 to 15.3 cM. The positions of markers for foreground and recombinant selection are shown in Figure 24. Between BC_1F_1 and BC_2F_1, markers for foreground selection for *qPM2* in the orange-fleshed background and those for *qPM5* in orange- and -green-fleshed backgrounds were changed from F1.1 to F1.2 (Table 20 and Figure 24).

3.2.2.2. MABC process for developing powdery mildew resistance NILs carrying *qPM2*

For the recombinant selection at the BC_1F_1 generation, 41 out of 166 (24.7%) green-fleshed individuals and 28 out of 171 (16.4%) orange-fleshed individuals showed one marker as homozygous for the recurrent parent allele and the other as heterozygous (Table 20, Figure 21 and Figure 24). Among the selected individuals, 9 green-fleshed and 15

doi:10.6342/NTU202500264

orange-fleshed ones carried heterozygous genotypes for foreground markers F1, F1.1, and F2. The average RPG recovery for green- and orange-fleshed individuals was 68.0% and 66.1%, respectively. Individuals with the highest RPG recovery for green-fleshed background (#001) and orange-fleshed background (#093) were selected and crossed with the respective recurrent parent to form the BC_2F_1 generation. At the BC_2F_1 generation, a two-step selection was applied because the QTL was close to one end of chromosome 2 (Figure 20). Among the 174 individuals in the green-fleshed BC_2F_1 family, 64 carried the resistance allele at the heterozygous state, while 84 out of 182 individuals in the orange-fleshed family carried heterozygous genotypes. For background selection via ddRAD sequencing, the average RPG recovery reached 86.4% and 83.0% for green- and orange-fleshed backgrounds, respectively. Individuals showing the highest RPG recovery rates, i.e., 93.6% for green-fleshed background and 89.1% for orange-fleshed background, were self-pollinated to produce the BC_2F_2 generation. From this step, two families were maintained: #020 and #143 for green-fleshed background and #088 and #132 for orange-fleshed background. At the BC_2F_2 generation, only foreground markers were used to screen 20 individuals in each family. Individuals carrying homozygous resistance genotypes were selected and self-pollinated to generate BC_2F_3 individuals. For each near-isogenic line, 20 BC_2F_3 individuals were used to evaluate the powdery mildew reaction. All the near-isogenic lines showed a DI of zero, indicating the successful introgression of PM resistance to the elite backgrounds.

3.2.2.3. MABC process for developing powdery mildew resistance NILs carrying *qPM5*

For the recombinant selection of *qPM5* at the BC_1F_1 generation, 20 out of 180 (11.1%) green-fleshed individuals and 12 out of 138 (8.7%) orange-fleshed individuals showed one marker as homozygous for the recurrent parents allele and the other as heterozygous (Table 20, Figure 22 and Figure 24). Among the selected individuals, nine green-fleshed and five orange-fleshed ones carried heterozygous genotypes for foreground markers F1.1, and F2. The average RPG recovery for green- and orange-fleshed individuals was 67.7% and 67.4%, respectively. Individuals with the highest RPG recovery for green-fleshed background (#126) and orange-fleshed background (#093) were selected and crossed with the respective recurrent parent to form the BC_2F_1 generation. At the second recombinant selection performed at the BC_2F_1 generation, 88 out of 181 individuals from the green-fleshed family and 113 out of 184 individuals from

doi:10.6342/NTU202500264

the orange-fleshed family were homozygous as the recurrent parent for the marker showing heterozygous genotype at the first recombinant selection. Among these individuals, one green-fleshed and two orange-fleshed ones were heterozygous genotypes at the foreground markers F1.2 and F2. The average RPG recovery rate was 88.3% in green-fleshed individuals and 79.6% in orange-fleshed individuals. Individuals with the highest RPG recovery, i.e., 88.3% for the green-fleshed background (#102) and 84.8% for the orange-fleshed background (#087) were selected and crossed with the respective recurrent parent to form the BC_3F_1 generation. At the BC_3F_1 generation, a two-step selection was applied since the recombinant selection was done at BC_2F_1. Among the 89 green-fleshed BC_3F_1 progeny, 37 carried the resistance allele at the heterozygous state, while 50 out of 91 orange-fleshed individuals carried heterozygous genotypes. The background selection revealed an average RPG recovery of 94.0% for the green-fleshed family and 91.1% for the orange-fleshed family. The individuals with the highest RPG recovery, i.e., 95.8% and 94.8% for green- and orange-fleshed backgrounds, respectively, were self-pollinated to generate the BC_3F_2 generation. Two BC_3F_2 families were maintained for both green- and orange-fleshed backgrounds: #051 and #080 for green-fleshed families, and #016 and #088 for orange-fleshed families. At the BC_3F_2 generation, only foreground markers were used to select the individuals carrying homozygous resistance alleles from 60 individuals in each family. For the green- and orange-fleshed background, selected individuals were evaluated for the DI of powdery mildew. The average DI for green- and orange-fleshed near-isogenic lines were 1.0 and 0.0, indicating that the foreground marker regions effectively improved powdery mildew resistance to the elite backgrounds.

3.2.2.4. MABC process for developing powdery mildew resistance NILs carrying *qPM12*

For the recombinant selection of *qPM12* at the BC_1F_1 generation, 29 out of 181 (16.0%) green-fleshed individuals and 29 out of 127 (22.8%) orange-fleshed individuals showed one marker as homozygous for the recurrent parents allele and the other as heterozygous (Table 20, Figure 23 and Figure 24). Among these individuals, five green-fleshed and seven orange-fleshed ones carried heterozygous genotypes for foreground markers F1 and F2. The average RPG recovery for green- and orange-fleshed individuals was 67.5% and 72.6%, respectively. Individuals with the highest RPG recovery for green-fleshed background (#159) and orange-fleshed background (#078) were selected and

doi:10.6342/NTU202500264

crossed with the respective recurrent parent to form the BC_2F_1 generation. At the second recombinant selection performed at the BC_2F_1 generation, 92 out of 182 green-fleshed individuals from the green-fleshed family and 75 out of 176 individuals from the orange-fleshed family were homozygous for the recurrent parent allele. Among these individuals, five green-fleshed and 15 orange-fleshed ones carried heterozygous genotypes for the foreground markers F1 and F2. The average RPG recovery rate was 87.0% in green-fleshed individuals and 83.6% in orange-fleshed individuals. Individuals with the highest RPG recovery, 88.2% for the green-fleshed background (#065) and 90.0% for the orange-fleshed background (#064), were selected and crossed with respective parents to form the BC_3F_1 generation. Among the 67 green-fleshed BC_3F_1 progeny, 32 carried the resistance allele at the heterozygous state, while 45 out of 90 orange-fleshed individuals carried heterozygous genotypes. The background selection revealed an average RPG recovery of 92.6% for the green-fleshed family and 93.5% for the orange-fleshed family. The individuals with the highest RPG recovery, nearing 97.1% and 96.1% for green- and orange-fleshed backgrounds, respectively, were self-pollinated to generate the BC_3F_2 families. Two BC_3F_2 families were maintained for both green- and orange-fleshed backgrounds: #022 and #044 for green-fleshed families, and #002 and #090 for orange-fleshed families. At the BC_3F_2 generation, only foreground markers were used to select the individuals carrying homozygous resistance alleles from 60 individuals in each family. For the green- and orange-fleshed background, selected individuals were evaluated for the DI of powdery mildew. The average DI for green-fleshed near-isogenic lines were 0.3 and 0.1, and each line showed a DI of zero in orange-fleshed backgrounds indicating that the resistance alleles of *qPM12* was effectively introgressed into the elite backgrounds.

3.2.3. Discussion

In this MABC study, we first focused on recombinant selection, followed by foreground and background selection. Performing recombinant selection before foreground selection at the BC_1F_1 generation optimized the use of low-throughput genotyping platforms. While this strategy was efficient, it could not entirely avoid the loss of target foreground genotypes due to potential double crossovers. Therefore, foreground selection was performed at each generation. The ddRAD sequencing system facilitated high-throughput genotyping for background selection in later generations, significantly increasing the intensity of selection with enhanced recovery of the RPG. Our MABC scheme at the BC_1F_1 generation reduces the genotyping effort by 17% to 28%

doi:10.6342/NTU202500264

due to the selection intensity for recombinant selection ranging from 8.7% to 24.7% in the *qPM5* orange-fleshed and *qPM2* green-fleshed MABC processes. The whole strategy was also time-effective: creating the three F_2 populations, including the powdery mildew resistance QTL mapping, took one and a half years, followed by two years of MABC. This led to the production of near-isogenic powdery mildew resistance lines from creating the population for QTL mapping in 3.5 years.

The near-isogenic lines developed in this study are valuable resources for melon resistance breeding and genetic studies. On one hand, they can be used as parental lines to confer resistance in other elite backgrounds or allow the pyramiding of major QTL. On the other hand, they could be used for fine mapping of the three powdery mildew resistance QTL. Future research could explore the durability of resistance related to multiple minor QTL and identify additional genetic resources for resistance breeding in melon.

4. Conclusion

The NGS-based marker development and genotyping methods have been widely applied to genetic analysis and MAS. The ddRADseq contained high flexibility to different experimental requirements. We conducted a series of *in silico* and empirical experiments and developed analysis tools for an optimal implementation of ddRADseq. Appropriate restriction enzyme combination was the first step to be considered. Our results demonstrated that an optimized ddRADseq could drastically increase the ability to construct a reliable high-density linkage map and identify the genetic architecture of traits of interest. Based on the results of a single enzyme and two enzyme combinations, the ratio of the number of cutting sites between common cutters and rare cutters would be recommended to be 10 to 30-fold different. However, it always requires taking the biological features of the target species into account. The cost of the restriction enzymes and their compatibility were to be considered. That is the reason we selected the *Hind*III, *Eco*RI, *Xba*I, *Sac*I, *Pst*I, and *Kpn*I as the candidate enzyme set for facing the new target species and new experimental requirements. One essential element for a flexible adjustment of library preparation is the dual-index multiplex system, which can drastically reduce the investment of sample multiplexing. Therefore, the ddRADseq system as a regular wet lab work is possible. Given the target species, the ddRADseq can not only be the solution for genetic analysis but also be integrated into molecular breeding.

doi:10.6342/NTU202500264

References

Aamir M, Karmakar P, Singh VK, et al (2021) A novel insight into transcriptional and epigenetic regulation underlying sex expression and flower development in melon (*Cucumis melo* L.). Physiologia Plantarum 173:1729–1764. https://doi.org/10.1111/ppl.13357

Aballay MM, Aguirre NC, Filippi CV, et al (2021) Fine-tuning the performance of ddRAD-seq in the peach genome. Sci Rep 11:6298. https://doi.org/10.1038/s41598-021-85815-0

Argyris JM, Ruiz-Herrera A, Madriz-Masis P, et al (2015) Use of targeted SNP selection for an improved anchoring of the melon (*Cucumis melo* L.) scaffold genome assembly. BMC Genomics 16:4. https://doi.org/10.1186/s12864-014-1196-3

Baird NA, Etter PD, Atwood TS, et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376. https://doi.org/10.1371/journal.pone.0003376

Bellundagi A, Ramya KT, Krishna H, et al (2022) Marker-assisted backcross breeding for heat tolerance in bread wheat (*Triticum aestivum* L.). Front Genet 13:1056783. https://doi.org/10.3389/fgene.2022.1056783

Beraldo-Hoischen P, Gomez-Guillamon ML, Lopez-Sese AI, et al (2012) QTL associated with one recessive gene for powdery mildew resistance in the melon genotype TGR-1551. Cucurbitaceae 2012: Proceedings of the Xth Eucarpia meeting on genetics and breeding of cucurbitaceae

Branham SE, Kousik C, Mandal MK, Wechter WP (2021) Quantitative trait loci mapping of resistance to powdery mildew race 1 in a recombinant inbred line population of melon. Plant Dis 105:3809–3815. https://doi.org/10.1094/PDIS-12-20-2643-RE

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890. https://doi.org/10.1093/bioinformatics/btg112

Campos-Martin R, Schmickler S, Goel M, et al (2023) Reliable genotyping of recombinant genomes using a robust hidden Markov model. Plant Physiology 192:821–836. https://doi.org/10.1093/plphys/kiad191

Cao Y, Diao Q, Chen Y, et al (2021) Development of KASP markers and identification of a QTL underlying powdery mildew resistance in melon (*Cucumis melo* L.) by bulked segregant analysis and RNA-Seq. Front Plant Sci 11:593207. https://doi.org/10.3389/fpls.2020.593207

Castanera R, Ruggieri V, Pujol M, et al (2020) An improved melon reference genome with single-molecule sequencing uncovers a recent burst of transposable elements with potential impact on genes. Front Plant Sci 10:. https://doi.org/10.3389/fpls.2019.01815

doi:10.6342/NTU202500264

Chafin TK, Martin BT, Mussmann SM, et al (2018) FRAGMATIC: in silico locus prediction and its utility in optimizing ddRADseq projects. Conservation Genetics Resources 10:325–328. https://doi.org/10.1007/s12686-017-0814-1

Chandrasekharan N, Ramanathan N, Pukalenthy B, et al (2022) Development of *β*-carotene, lysine, and tryptophan-rich maize (*Zea mays*) inbreds through marker-assisted gene pyramiding. Sci Rep 12:8551. https://doi.org/10.1038/s41598-022-11585-y

Christiansen H, Heindler FM, Hellemans B, et al (2021) Facilitating population genomics of non-model organisms through optimized experimental design for reduced representation sequencing. BMC Genomics 22:625. https://doi.org/10.1186/s12864-021-07917-3

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences 363:557–572. https://doi.org/10.1098/rstb.2007.2170

Cui H, Fan C, Ding Z, et al (2022) *CmPMRl* and *CmPMrs* are responsible for resistance to powdery mildew caused by *Podosphaera xanthii* race 1 in Melon. Theor Appl Genet. https://doi.org/10.1007/s00122-021-04025-4

Davey JW, Hohenlohe PA, Etter PD, et al (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12:499–510. https://doi.org/10.1038/nrg3012

DePristo MA, Banks E, Poplin R, et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43:491–498. https://doi.org/10.1038/ng.806

Egel DS, Adkins ST, Wintermantel WM, et al (2022) Diseases of cucumbers, melons, pumpkins, squash, and watermelons. In: Elmer WH, McGrath M, McGovern RJ (eds) Handbook of vegetable and herb diseases. Springer International Publishing, Cham, pp 1–105

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379. https://doi.org/10.1371/journal.pone.0019379

FAOSTAT (2022) Crops and livestock products. https://www.fao.org/faostat/en/#data/QCL. Accessed 16 Apr 2024

Fazza AC, Dallagnol LJ, Fazza AC, et al (2013) Mapping of resistance genes to races 1, 3 and 5 of *Podosphaera xanthii* in melon PI 414723. Crop Breed Appl Biotechnol 13:349–355. https://doi.org/10.1590/S1984-70332013000400005

doi:10.6342/NTU202500264

Fragoso CA, Heffelfinger C, Zhao H, Dellaporta SL (2016) Imputing Genotypes in Biallelic Populations from Low-Coverage Sequence Data. Genetics 202:487–495. https://doi.org/10.1534/genetics.115.182071

Frisch M (2004) Breeding strategies: optimum design of marker-assisted backcross programs. In: Molecular marker systems in plant breeding and crop improvement. Springer, pp 319–334

Frisch M, Bohn M, Melchinger AE (1999a) Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. Crop Sci 39:967–975

Frisch M, Bohn M, Melchinger AE (1999b) Comparison of selection strategies for marker-assisted backcrossing of a gene. Crop Sci 39:1295–1301

Fu Y-B, Peterson GW, Dong Y (2016) Increasing Genome Sampling and Improving SNP Genotyping for Genotyping-by-Sequencing with new Combinations of Restriction Enzymes. G3: Genes, Genomes, Genetics g3.115.025775. https://doi.org/10.1534/g3.115.025775

Fukino N, Ohara T, Monforte AJ, et al (2008) Identification of QTLs for resistance to powdery mildew and SSR markers diagnostic for powdery mildew resistance genes in melon (*Cucumis melo* L.). Theor Appl Genet 118:165–175. https://doi.org/10.1007/s00122-008-0885-1

Furuta T, Yamamoto T, Ashikari M (2023) GBScleanR: robust genotyping error correction using a hidden Markov model with error pattern recognition. Genetics 224:iyad055. https://doi.org/10.1093/genetics/iyad055

Gao J, Gileta AF, Bimschleger HV, et al (2019) Adapting genotyping-by-sequencing and variant calling for heterogeneous stock rats. bioRxiv 523043. https://doi.org/10.1101/523043

Garcia-Mas J, Benjak A, Sanseverino W, et al (2012) The genome of melon (*Cucumis melo* L.). Proc Natl Acad Sci U S A 109:11872–11877. https://doi.org/10.1073/pnas.1205415109

Gonda I, Ashrafi H, Lyon DA, et al (2019) Sequencing-Based Bin Map Construction of a Tomato Mapping Population, Facilitating High-Resolution Quantitative Trait Loci Detection. The Plant Genome 12:0. https://doi.org/10.3835/plantgenome2018.02.0010

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Hamblin MT, Rabbi IY (2014) The Effects of Restriction-Enzyme Choice on Properties of Genotyping-by-Sequencing Libraries: A Study in Cassava (Manihot esculenta). Crop Science 54:2603. https://doi.org/10.2135/cropsci2014.02.0160

Haonan C, Zhuo D, Chao F, et al (2020) Genetic mapping and nucleotide diversity of two powdery mildew resistance loci in melon (*Cucumis melo*). Phytopathology® 110:1970–1979. https://doi.org/10.1094/PHYTO-03-20-0078-R

doi:10.6342/NTU202500264

Heffelfinger C, Fragoso CA, Moreno MA, et al (2014) Flexible and scalable genotyping-by-sequencing strategies for population studies. BMC genomics 15:979

Hoffberg SL, Kieran TJ, Catchen JM, et al (2016) RAD cap: sequence capture of dual-digest RAD seq libraries with identifiable duplicates and reduced missing data. Molecular Ecology Resources 16:1264–1278. https://doi.org/10.1111/1755-0998.12566

Huang JH, Wang YH (2007) The races of *Podosphaera xanthii* causing melon powdery mildew in Taiwan. J Taiwan Agric Res 56:307–315

Huang JH, Wang YH, Lo CT (2002) Development of leaf-disk method for screening melon varieties resistant to *Sphaerotheca fuliginea* race 1. J Agric Res China 51:49–56. https://doi.org/10.29951/JARC.200212.0005

Jiang G-L (2015) Molecular Marker-Assisted Breeding: A Plant Breeder's Review. In: Al-Khayri JM, Jain SM, Johnson DV (eds) Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools. Springer International Publishing, Cham, pp 431–472

Jobes DV, Hurley DL, Thien LB (1995) Plant DNA isolation: A method to efficiently remove polyphenolics, polysaccharides, and RNA. Taxon 44:379–386. https://doi.org/10.2307/1223408

Kim HT, Park JI, Robin AHK, et al (2016) Identification of a new race and development of DNA markers associated with powdery mildew in melon. Plant Breeding and Biotechnology 4:225–233. https://doi.org/10.9787/PBB.2016.4.2.225

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357–359. https://doi.org/10.1038/nmeth.1923

Lawrence M, Huber W, Pagès H, et al (2013) Software for Computing and Annotating Genomic Ranges. PLOS Comput Biol 9:e1003118. https://doi.org/10.1371/journal.pcbi.1003118

Lepais O, Weir JT (2014) SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. Mol Ecol Resour 1314–1312. https://doi.org/10.1111/1755-0998.12273

Li B, Zhao Y, Zhu Q, et al (2017) Mapping of powdery mildew resistance genes in melon (*Cucumis melo* L.) by bulked segregant analysis. Scientia Horticulturae 220:160–167. https://doi.org/10.1016/j.scienta.2017.04.001

Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li L-L, Huang S, Hou J, et al (2020) Construction of a high-density genetic map for melon using ddRAD-Seq technology from a population derived from flexuosus

doi:10.6342/NTU202500264

and reticulatus botanical groups. Scientia Horticulturae 272:109531. https://doi.org/10.1016/j.scienta.2020.109531

López-Martín M, Pérez-de-Castro A, Picó B, Gómez-Guillamón ML (2022) Advanced genetic studies on powdery mildew resistance in TGR-1551. IJMS 23:12553. https://doi.org/10.3390/ijms232012553

Lorieux M, Gkanogiannis A, Fragoso C, Rami J-F (2019) NOISYmputer: genotype imputation in bi-parental populations for noisy low-coverage next-generation sequencing data. bioRxiv. https://doi.org/10.1101/658237

Lübberstedt T, Beavis W, Bhattacharyya M, et al (2023) Molecular Plant Breeding. Iowa State University Digital Press

Lübberstedt T, Varshney RK (eds) (2013) Diagnostics in Plant Breeding. Springer Netherlands, Dordrecht

Manichaikul A, Moon JY, Sen Ś, et al (2009) A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. Genetics 181:1077–1086. https://doi.org/10.1534/genetics.108.094565

Marè C, Zampieri E, Cavallaro V, et al (2023) Marker-assisted introgression of the salinity tolerance locus *Saltol* in temperate *Japonica* Rice. Rice 16:2. https://doi.org/10.1186/s12284-023-00619-2

McCreight JD (2006) Melon-powdery mildew interactions reveal variation in melon cultigens and *Podosphaera xanthii* races 1 and 2. J Amer Soc Hort Sci 131:59–65

McKenna A, Hanna M, Banks E, et al (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20:1297–1303. https://doi.org/10.1101/gr.107524.110

Metzker ML (2010) Sequencing technologies — the next generation. Nat Rev Genet 11:31–46. https://doi.org/10.1038/nrg2626

Miao C, Fang J, Li D, et al (2018) Genotype-Corrector: improved genotype calls for genetic mapping in F_2 and RIL populations. Scientific Reports 8:. https://doi.org/10.1038/s41598-018-28294-0

Mora-Márquez F, García-Olivares V, Emerson BC, López de Heredia U (2017) ddradseqtools: a software package for in silico simulation and testing of double-digest RADseq experiments. Molecular Ecology Resources 17:230–246. https://doi.org/10.1111/1755-0998.12550

Morgan M, Pagès H, Obenchain V, Hayden N (2024) Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. R package version 2.22.0, https://bioconductor.org/packages/Rsamtools.

Neeraja CN, Maghirang-Rodriguez R, Pamplona A, et al (2007) A marker-assisted backcross approach for developing submergence-tolerant rice cultivars. Theor Appl Genet 115:767–776. https://doi.org/10.1007/s00122-007-0607-0

doi:10.6342/NTU202500264

Ning X, Wang X, Gao X, et al (2014) Inheritances and location of powdery mildew resistance gene in melon Edisto47. Euphytica 195:345–353. https://doi.org/10.1007/s10681-013-1000-5

Nishimura K, Kokaji H, Motoki K, et al (2024) Degenerate oligonucleotide primer MIG-seq: an effective PCR-based method for high-throughput genotyping. The Plant Journal 118:2296–2317. https://doi.org/10.1111/tpj.16708

Pagès H, Aboyoun P, Gentleman R, DebRoy S (2024) Biostrings: Efficient manipulation of biological strings. R package version 2.74.0, https://bioconductor.org/packages/Biostrings.

Pagès H, Kakopo AK (2024) BSgenomeForge: Forge your own BSgenome data package. R package version 1.6.0, https://bioconductor.org/packages/BSgenomeForge

Perchepied L, Bardin M, Dogimont C, Pitrat M (2005) Relationship between loci conferring downy mildew and powdery mildew resistance in melon assessed by quantitative trait loci mapping. Phytopathology 95:556–565. https://doi.org/10.1094/PHYTO-95-0556

Pértille F, Guerrero-Bosagna C, Silva VH da, et al (2016) High-throughput and Cost-effective Chicken Genotyping Using Next-Generation Sequencing. Scientific Reports 6:. https://doi.org/10.1038/srep26929

Peterson BK, Weber JN, Kay EH, et al (2012) Double Digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135. https://doi.org/10.1371/journal.pone.0037135

Pitrat M (2016) Melon genetic resources: Phenotypic diversity and horticultural taxonomy. In: Genetics and Genomics of Cucurbitaceae. Springer, Cham, pp 25–60

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253. https://doi.org/10.1371/journal.pone.0032253

R Core Team (2023) R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. https://www.r-project.org/

Rodríguez-Moreno L, González VM, Benjak A, et al (2011) Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. BMC Genomics 12:424. https://doi.org/10.1186/1471-2164-12-424

Ruggieri V, Alexiou KG, Morata J, et al (2018) An improved assembly and annotation of the melon (Cucumis melo L.) reference genome. Scientific Reports 8:8088. https://doi.org/10.1038/s41598-018-26416-2

doi:10.6342/NTU202500264

Scheben A, Batley J, Edwards D (2017) Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant Biotechnol J 15:149–161. https://doi.org/10.1111/pbi.12645

Schubert M, Lindgreen S, Orlando L (2016) AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Research Notes 9:. https://doi.org/10.1186/s13104-016-1900-2

Semagn K, Babu R, Hearne S, Olsen M (2014) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol Breeding 33:1–14. https://doi.org/10.1007/s11032-013-9917-x

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. https://doi.org/10.1093/dnares/dsw004

Siberchicot A, Bessy A, Guéguen L, Marais GA (2017) MareyMap online: a user-friendly web application and database service for estimating recombination rates using physical and genetic maps. Genome Biology and Evolution 9:2506–2509. https://doi.org/10.1093/gbe/evx178

Sim S-C, Durstewitz G, Plieske J, et al (2012) Development of a Large SNP Genotyping Array and Generation of High-Density Genetic Maps in Tomato. PLoS ONE 7:e40563. https://doi.org/10.1371/journal.pone.0040563

Singh J, Sharma S, Kaur A, et al (2021) Marker-assisted pyramiding of *lycopene-ε-cyclase*, *β-carotene hydroxylase1* and *opaque2* genes for development of biofortified maize hybrids. Sci Rep 11:12642. https://doi.org/10.1038/s41598-021-92010-8

Sinn BT, Simon SJ, Santee MV, et al (2022) ISSRseq: An extensible method for reduced representation sequencing. Methods Ecol Evol 13:668–681. https://doi.org/10.1111/2041-210X.13784

Teixeira APM, Barreto FA da S, Camargo LEA (2008) An AFLP marker linked to the *Pm-1* gene that confers resistance to *Podosphaera xanthii* race 1 in *Cucumis melo*. Genetics and Molecular Biology 31:547–550. https://doi.org/10.1590/S1415-47572008000300023

Thermo Fisher Scientific (2024) Custom TaqMan® assay design tool. https://www.thermofisher.com/order/custom-genomic-products/tools/genotyping/. Accessed 29 May 2024

Van der Auwera GA, Carneiro MO, Hartl C, et al (2013) From fastq data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. In: Current Protocols in Bioinformatics. John Wiley & Sons, Inc., p 11.10.1-11.10.33

doi:10.6342/NTU202500264

Wang J, Li L, Qi H, et al (2016a) RestrictionDigest: A powerful Perl module for simulating genomic restriction digests. Electronic Journal of Biotechnology 21:36–42. https://doi.org/10.1016/j.ejbt.2016.02.003

Wang X, Li G, Gao X, et al (2011) Powdery mildew resistance gene *Pm-AN* located in a segregation distortion region of melon LGV. Euphytica 180:421–428. https://doi.org/10.1007/s10681-011-0406-1

Wang YH (2016) Mapping quantitative trait loci for fruit traits and powdery mildew resistance in melon (*Cucumis melo* L.). Doctoral Dissertation, National Taiwan University

Wang YH, Wu DH, Huang JinH, et al (2016b) Mapping quantitative trait loci for fruit traits and powdery mildew resistance in melon (*Cucumis melo*). Bot Stud 57:19. https://doi.org/10.1186/s40529-016-0130-1

Xu L, He Y, Tang L, et al (2022) Genetics, Genomics, and Breeding in Melon. Agronomy 12:2891. https://doi.org/10.3390/agronomy12112891

Xu Yunbi (2010) Molecular plant breeding. CABI, Cambridge, MA

Yuste-Lisbona FJ, Capel C, Gómez-Guillamón ML, et al (2011a) Codominant PCR-based markers and candidate genes for powdery mildew resistance in melon (*Cucumis melo* L.). Theor Appl Genet 122:747–758. https://doi.org/10.1007/s00122-010-1483-6

Yuste-Lisbona FJ, Capel C, Sarria E, et al (2011b) Genetic linkage map of melon (*Cucumis melo* L.) and localization of a major QTL for powdery mildew resistance. Mol Breeding 27:181–192. https://doi.org/10.1007/s11032-010-9421-5

Yuste-Lisbona FJ, López-Sesé AI, Gómez-Guillamón ML (2010) Inheritance of resistance to races 1, 2 and 5 of powdery mildew in the melon TGR-1551. Plant Breeding 129:72–75. https://doi.org/10.1111/j.1439-0523.2009.01655.x

Zhang C, Ren Y, Guo S, et al (2013) Application of comparative genomics in developing markers tightly linked to the *Pm-2F* gene for powdery mildew resistance in melon (*Cucumis melo* L.). Euphytica 190:157–168. https://doi.org/10.1007/s10681-012-0828-4

Zhang T, Cui H, Luan F, et al (2023) A recessive gene *Cmpmr2F* confers powdery mildew resistance in melon (*Cucumis melo* L.). Theor Appl Genet 136:4. https://doi.org/10.1007/s00122-023-04269-2

Zhang Y, Yang Z, Ma H, et al (2021) Pyramiding of fusarium head blight resistance quantitative trait loci, *Fhb1*, *Fhb4*, and *Fhb5*, in modern Chinese wheat cultivars. Front Plant Sci 12:694023. https://doi.org/10.3389/fpls.2021.694023

doi:10.6342/NTU202500264

Table 1. Single-enzyme *in silico* digestion.

Enzyme combinations	Cutting site	Counts	Total Frags		Selected Frags (160 – 460 bp)			
			Counts	Ave.	Counts	Mb	%	S/T
Taq$^{\alpha}$I	T/CGA	852,395	852,410	483	249,820	71.56	17.5	29.3
*Msp*I	C/CGG	163,048	163,063	2,515	23,399	6.75	1.7	14.4
*Ape*KI	G/CWGC	229,248	229,263	1,791	37,976	11.24	2.7	16.6
*Hind*III	A/AGCTT	146,155	146,170	2,808	18,737	5.55	1.4	12.8
*Eco*RI	G/AATTC	102,104	102,119	4,017	8,656	2.64	0.6	8.5
*Xba*I	T/CTAGA	84,958	84,973	4,827	6,756	2.07	0.5	8.0
*Sac*I	GAGCT/C	39,361	39,376	10,412	2,376	0.74	0.2	6.0
*Pst*I	CTGCA/G	34,419	34,434	11,906	1,839	0.53	0.1	5.3
*Kpn*I	GGTAC/C	24,338	24,353	16,832	735	0.22	0.1	3.0

Total Frags indicates the total number of fragments predicted *in silico*.
Selected Frags indicates the predicted fragment with size ranging from 160 to 460 bp.
Ave. indicates the average fragment length (bp) for the predicted fragments.
Mb indicates the sum of the selected fragment length.
"%" indicates the genome coverage by selected fragment.
S/T indicates the percentage of selected fragments relative to the total fragments.

doi:10.6342/NTU202500264

Table 2. Two-enzyme *in silico* digestion.

Enzyme combinations	Total Frags			Selected Frags (160 – 460 bp)			
	Counts	Mb	%	Counts	Mb	%	S/T
*Hind*III-*Taq*$^{\alpha}$I	234,982	103.88	25.4	76,142	21.87	5.3	32.4
*Eco*RI-*Taq*$^{\alpha}$I	177,626	81.31	19.8	54,614	15.77	3.9	30.8
*Xba*I-*Taq*$^{\alpha}$I	147,433	72.08	17.6	48,476	14.08	3.4	32.9
*Sac*I-*Taq*$^{\alpha}$I	73,494	32.37	7.9	20,817	5.99	1.5	28.3
*Pst*I-*Taq*$^{\alpha}$I	62,170	32.11	7.8	19,130	5.46	1.3	30.8
*Kpn*I-*Taq*$^{\alpha}$I	46,840	26.67	6.5	15,373	4.58	1.1	32.8

Total Frags indicates the total number of fragments predicted by *in silico*.
Selected Frags indicates the predicted fragment with size ranging from 160 to 460 bp.
Mb indicates the sum of the selected fragment length.
"%" indicates the genome coverage by selected fragment.
S/T indicates the percentage of selected fragments relative to the total fragments.

doi:10.6342/NTU202500264

Table 3. Alignment results of size-selected fragments by two-enzyme *in silico* digestion

Enzyme combinations	Total reads	Unmapped		Exactly one match		More than one matches		MAPQ<30	
	Counts	Counts	%	Counts	%	Counts	%	Counts	%
*Hind*III-*Taq*$^{\alpha}$I	76,142	576	0.8	62,433	82.0	13,133	17.3	358	0.5
*Eco*RI-*Taq*$^{\alpha}$I	54,614	730	1.3	44,135	80.8	9,749	17.9	444	0.8
*Xba*I-*Taq*$^{\alpha}$I	48,476	293	0.6	39,932	82.4	8,251	17.0	206	0.4
*Sac*I-*Taq*$^{\alpha}$I	20,817	342	1.6	16,724	80.3	3,751	18.0	182	0.9
*Pst*I-*Taq*$^{\alpha}$I	19,130	132	0.7	15,150	79.2	3,848	20.1	148	0.8
*Kpn*I-*Taq*$^{\alpha}$I	15,373	129	0.8	10,452	68.0	4,792	31.2	112	0.7

MAPQ<30 indicates the number of aligned reads with mapping quality score less than 30.

Total reads indicates the *in silico* predicted fragment after the size selection 160 – 460 bp.

doi:10.6342/NTU202500264

Table 4. Average alignment results of PE read of two-enzyme combinations by mapping categories, origin, and MAPQ levels.

PE reads	*Hind*III-*Taq*$^{\alpha}$I		*Eco*RI-*Taq*$^{\alpha}$I		*Xba*I-*Taq*$^{\alpha}$I		*Sac*I-*Taq*$^{\alpha}$I		*Pst*I-*Taq*$^{\alpha}$I		*Kpn*I-*Taq*$^{\alpha}$I	
Paired-end reads												
Sample 1	1,092,905		1,088,788		1,317,848		1,052,778		1,048,743		959,424	
Sample 2	1,113,158		1,165,588		1,272,990		1,209,537		804,702		949,125	
Sample 3	1,136,282		1,035,543		1,112,029		988,207		1,008,661		906,010	
max/min	1.04		1.13		1.19		1.22		1.30		1.06	
Average	1,114,115		1,096,640		1,234,289		1,083,507		954,035		938,186	
Alignment summary												
Unique	288,679	(25.9)	401,781	(36.4)	733,083	(59.3)	195,745	(18.2)	561,326	(59.6)	285,698	(30.4)
Organelle	11,372	(1.0)	22,468	(2.1)	25,589	(2.1)	14,717	(1.4)	26,170	(2.7)	29,963	(3.2)
Nucleus	277,307	(24.9)	379,313	(34.4)	707,494	(57.2)	181,028	(16.8)	535,157	(56.9)	255,735	(27.2)
Non-unique	738,688	(66.3)	566,334	(51.9)	338,357	(27.5)	813,150	(74.9)	281,204	(28.5)	576,023	(61.4)
Organelle	619,502	(55.6)	347,157	(32.0)	143,436	(11.7)	752,796	(69.3)	237,857	(23.9)	247,623	(26.4)
Nucleus	119,186	(10.7)	219,177	(19.8)	194,921	(15.8)	60,354	(5.6)	43,347	(4.6)	328,400	(35.0)
Unmapped	86,748	(7.8)	128,525	(11.7)	162,849	(13.3)	74,612	(6.9)	111,505	(12.0)	76,465	(8.2)
MAPQ levels												
0-9	649,945	(63.2)	312,047	(32.6)	264,200	(24.5)	768,409	(76.0)	236,792	(27.0)	323,305	(37.5)
10-19	30,158	(2.9)	50,198	(5.2)	38,844	(3.6)	16,544	(1.7)	23,720	(2.7)	124,984	(14.5)
20-29	49,094	(4.8)	153,492	(15.7)	96,983	(9.0)	27,858	(2.8)	50,953	(6.4)	101,426	(11.8)
30-42	298,169	(29.0)	452,378	(46.5)	671,412	(62.8)	196,084	(19.6)	531,066	(63.8)	312,006	(36.2)

max/min represents the ratio of PE reads between the samples with the highest and lowest read counts, reflecting the consistency of sequencing output across samples.
Average indicates the average number of PE reads in three samples.
Non-unique indicates the PE reads have more than one aligned location across the genome.
Values in parentheses indicate corresponding ratios to Average.

doi:10.6342/NTU202500264

Table 5. Average read quality of empirical two-enzyme ddRADseq.

PE reads	*Hind*III-*Taq*$^{\alpha}$I		*Eco*RI-*Taq*$^{\alpha}$I		*Xba*I-*Taq*$^{\alpha}$I		*Sac*I-*Taq*$^{\alpha}$I		*Pst*I-*Taq*$^{\alpha}$I		*Kpn*I-*Taq*$^{\alpha}$I	
Total aligned	1,027,367		968,115		1,071,440		1,008,895		842,530		861,722	
ISR	392,997	(38.3)	743,183	(76.9)	782,030	(73.2)	283,918	(28.5)	754,504	(89.5)	690,873	(80.2)
EMP	597,734	(58.1)	201,446	(20.7)	268,646	(24.9)	677,937	(66.8)	75,626	(9.0)	144,525	(16.8)
MISS	36,635	(3.6)	23,485	(2.5)	20,764	(1.9)	47,039	(4.7)	12,401	(1.5)	26,324	(3.0)
MAPQ≧20												
ISR	307,254	(88.5)	489,155	(81.0)	648,514	(84.4)	202,561	(90.4)	552,209	(94.9)	373,284	(90.3)
EMP	40,009	(11.5)	116,715	(19.0)	119,882	(15.6)	21,380	(9.6)	29,810	(5.1)	40,149	(9.7)
MAPQ<20	680,104	(66.2)	362,245	(37.4)	303,044	(28.3)	784,953	(77.8)	260,511	(30.9)	448,289	(52.0)
High Freq	578,732	(56.3)	357,231	(36.9)	193,059	(18.0)	685,782	(68.0)	271,123	(32.2)	281,933	(32.7)
MAPQ<20 ∩ High Freq	565,112	(55.0)	203,801	(21.1)	157,250	(14.7)	654,300	(64.9)	204,186	(24.2)	226,609	(26.3)
Final	333,644	(32.5)	452,440	(46.7)	732,587	(68.4)	192,459	(19.1)	515,082	(61.1)	358,109	(41.6)

ISR indicates the number of PE reads matched the fragments predicted *in silico*.
EMP indicates the number of PE reads did not match *in silico* fragments but can be assigned as empirical fragments.
MISS indicates the rest of the PE reads cannot be classified into fragment categories and MAPQ<10.
MAPQ indicates the number of PE read processed by MAPQ with the threshold after the number behind.
High Freq. indicates The PE reads are located within the high-frequency fragments listed in Supplementary Table 3.
Final indicates total aligned reads after removing the intersection of MAPQ<20 and High Freq.
Values in parentheses indicate corresponding ratios within total aligned reads.

doi:10.6342/NTU202500264

Table 6. The average numbers of fragment and depth summary statistics from PE reads with MAPQ>20.

	*Hind*III-*Taq*$^{\alpha}$I	*Eco*RI-*Taq*$^{\alpha}$I	*Xba*I-*Taq*$^{\alpha}$I	*Sac*I-*Taq*$^{\alpha}$I	*Pst*I-*Taq*$^{\alpha}$I	*Kpn*I-*Taq*$^{\alpha}$I	*Xba*I-*Taq*$^{\alpha}$I-*Cvi*AII	*Xba*I-*Taq*$^{\alpha}$I-*Mse*I	*Xba*I-*Taq*$^{\alpha}$I-*Sph*I
Fragment Type									
--	480	463	1,364	182	168	256	470	536	1,158
C-	3,419	3,797	9,098	1,804	1,949	2,046	3,992	4,441	6,428
R-	4,170	3,673	11,673	1,287	1,644	1,779	3,089	3,905	11,340
RC	7,193	3,868	3,816	2,145	1,738	1,667	15,691	37,675	5,640
AB	62,175	45,651	40,470	17,513	14,562	14,724	19,512	8,667	42,378
Total	77,437	57,452	66,422	22,932	20,061	20,472	42,754	55,224	66,944
Fragment Depth									
Q25	1.0	2.0	1.0	1.3	1.0	2.0	1.8	1.9	1.4
Q50	3.0	5.3	4.7	4.7	14.7	11.0	4.1	5.0	2.3
Mean	4.5	10.5	11.6	9.8	29.0	20.2	10.1	12.0	4.5

The fragment type of --, C-, R- and RC indicate the empirical fragment types corresponding to the fragment border closed to rare (R), common (C), or none (-) cutters, and the AB indicates the *in silico* digested fragments.
Q25, Q50, and Mean indicate the first quantile, median and mean of the fragment depth.

doi:10.6342/NTU202500264

Table 7. Average alignment result of the empirical third-enzyme ddRADseq

Name	*Xba*I-*Taq*$^{\alpha}$I-*Cvi*AII		*Xba*I-*Taq*$^{\alpha}$I-*Mse*I		*Xba*I-*Taq*$^{\alpha}$I-*Sph*I	
Paired-end reads						
Sample 1	1,076,343		955,868		874,316	
Sample 2	1,167,182		1,095,880		882,214	
Sample 3	1,088,362		924,853		946,920	
max/min	1.07		1.17		1.15	
Average	1,110,629		992,200		901,150	
Alignment summary						
Unique	467,236	(42.1)	469,723	(47.4)	442,156	(49.0)
Organelle	26,729	(2.4)	20,434	(2.1)	16,138	(1.8)
Nucleus	440,507	(39.7)	449,289	(45.3)	426,018	(47.2)
Not unique	528,073	(47.5)	418,127	(42.1)	255,952	(28.4)
Organelle	354,963	(31.9)	250,429	(25.2)	135,507	(15.1)
Nucleus	173,110	(15.6)	167,698	(16.9)	120,445	(13.4)
Unmapped	115,320	(10.4)	104,351	(10.5)	203,042	(22.6)
MAPQ levels						
0-9	323,181	(32.5)	258,843	(29.1)	169,034	(24.3)
10-19	42,671	(4.3)	36,467	(4.1)	27,165	(3.9)
20-29	102,088	(10.2)	130,679	(14.7)	77,523	(11.1)
30-42	527,370	(53.0)	461,860	(52.0)	424,385	(60.7)

max/min represents the ratio of PE reads between the samples with the highest and lowest read counts, reflecting the consistency of sequencing output across samples.
Average indicates the average of the PE reads in three samples.
Non-unique indicates the PE reads have more than one aligned location across the genome.
Values in parentheses indicate corresponding ratios to Average.

doi:10.6342/NTU202500264

Table 8. PE reads summary for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I of the parental lines and F2 population.

Name	*Pst*I-*Taq*$^{\alpha}$I-*Sph*I						*Xba*I-*Taq*$^{\alpha}$I-*Mse*I					
PE reads	*Parental lines*				*F2 population*		*Parental lines*				*F2 population*	
	Female		Male		Average		Female		Male		Average	
	3,586,511		3,683,169		698,856		3,103,754		2,429,661		647,953	
Alignment summary												
Unique	2,120,737	(59.1)	1,951,150	(53.0)	412,001	(58.9)	1,725,433	(55.6)	1,317,258	(54.2)	348,780	(53.8)
Organelle	62,085	(1.7)	81,265	(2.2)	13,205	(1.9)	194,806	(6.3)	217,335	(8.9)	43,590	(6.7)
Nucleus	2,058,652	(57.4)	1,869,885	(50.8)	398,796	(57.0)	1,530,627	(49.3)	1,099,923	(45.3)	305,190	(47.2)
Non-unique	795,769	(22.2)	1,187,655	(32.2)	189,203	(27.1)	923,754	(29.8)	752,428	(31.0)	197,847	(30.5)
Organelle	693,364	(19.3)	1,090,388	(29.6)	169,017	(24.3)	479,445	(15.4)	447,750	(18.4)	108,380	(16.7)
Nucleus	102,405	(2.9)	97,267	(2.6)	20,186	(2.9)	444,309	(14.3)	304,678	(12.5)	89,467	(13.8)
Unmapped	660,803	(18.4)	520,964	(14.1)	97,040	(13.8)	450,100	(14.5)	355,751	(14.6)	101,019	(15.6)

Values in parentheses indicate corresponding ratios to total PE read for parental lines, and to Average of the F_2 population

doi:10.6342/NTU202500264

Table 9. The fragment polymorphism of the parental lines for the two enzyme combinations

Fragment types	*Pst*I-*Taq*$^{\alpha}$I-*Sph*I				*Xba*I-*Taq*$^{\alpha}$I-*Mse*I			
	Counts	Poly.	Ratio	Variant	Counts	Poly.	Ratio	Variant
AB	12,445	4,049	32.5	6,457	7,084	2,322	32.8	4,540
RC	981	234	23.9	338	31,133	11,862	38.1	24,291
R-	619	171	27.6	281	2,714	735	27.1	1,512
C-	418	56	13.4	95	2,348	418	17.8	832
--	46	8	17.4	14	292	46	15.8	83
	14,509	4,518	31.1	(66)	43,571	15,383	35.3	(711)

Counts indicate the number for the different types of fragments present in F_2 population with missing rate<0.5.
Poly. indicates the number of polymorphic fragments which the variants located within the fragments.
Ratio indicates the percentage of the polymorphic fragments within the types of the fragments.
Variant indicates the number of variants located within the fragment.
The fragment type of AB indicates the *in silico* digested fragments, and the RC, R-, C- and -- indicate the empirical fragment types corresponding to the fragment border closed to rare (R), common (C), or none (-) cutters.
Values in parentheses indicates the variant did not locate at any fragments.

doi:10.6342/NTU202500264

Table 10. The raw variant summary of the *Pst*I-*Taq*$^{\alpha}$I-*Sph*I and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I.

Variant matrices	*Pst*I-*Taq*$^{\alpha}$I-*Sph*I	*Xba*I-*Taq*$^{\alpha}$I-*Mse*I
No. of total variants	86,941	184,442
SNV	71,668	156,191
InDel	14,418	26,221
MultiAllele	855	2,030
Missing<0.5	12,749	54,450
QUAL>200 and QD>10	11,185	47,774
MAF>0.01	8,188	34,796
QUAL>200 and QD>10 ∩ MAF>0.01	7,252	31,969
R/qtl (in Chr1 – 12)	7,141	31,350
Missing<0.25	6,066	17,713
Segregation Distortion	5,181	14,383
Remove markers in RF	5,156	14,126
Final markers with GQ>0 for linkage map	5,146	13,832

SNV, InDel, and MultiAllele indicate the variant types as Single Nucleotide Variant (SNV), Insertion or Deletion (InDel), Multiple alternate alleles (MultiAllele)
Missing indicates the missing rate of markers.
Quality filter indicates the variant with QUAL>200 and QD>10.
QUAL>200 and QD>10 ∩ MAF>0.01 were under Missing<0.5
R/qtl inficate the genotype data set input into R/qtl for analysis.
MAF indicates the minor allele frequency.
RF indicates the recombination fraction plot.

doi:10.6342/NTU202500264

Table 11. Number of markers for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I (PTS) and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I (XTM) linkage maps built with different dataset

	Total		Missing>0.25		SDL		ODD		Final		GQ0	
Chr	PTS	XTM	PTS	XTM	PTS	XTM	PTS	XTM	PTS	XTM	PTS	XTM
1	693	3,183	111	1,363	78	328	0	7	504	1,485	1	37
2	580	2,683	78	1,128	83	315	1	73	418	1,167	0	23
3	578	2,445	99	1,112	42	244	0	9	437	1,080	0	18
4	696	2,678	90	1,120	87	300	0	12	519	1,246	3	34
5	496	2,484	92	1,074	67	258	6	16	331	1,136	0	29
6	760	3,175	133	1,421	74	317	0	15	553	1,422	0	14
7	617	2,545	85	1,080	78	339	14	70	440	1,056	0	15
8	791	2,981	116	1,343	100	276	1	11	574	1,351	3	30
9	453	1,584	64	687	60	173	0	5	329	719	1	14
10	463	2,739	75	1,148	60	221	1	35	327	1,335	2	28
11	611	3,158	92	1,432	81	380	2	3	436	1,343	0	23
12	403	1,695	40	729	75	179	0	1	288	786	0	29
Total	7,141	31,350	1,075	13,637	885	3,330	25	257	5,156	14,126	10	294

Missing>0.25 indicates the markers filtered out by a missing rate larger than 0.25.
SDL indicates the markers filtered out by the segregation distortion test.
ODD indicates the markers filtered out by observing the correlation between flanking markers in the recombination plot.
Final indicates the remaining markers after removing the Missing>0.25, SDL, and ODD markers.
GQ0 indicates the markers filtered out by a missing rate larger than 0.25 after removing the genotype with GQ less or equal to zero.

doi:10.6342/NTU202500264

Table 12. Summary of *Pst*I-*Taq*$^{\alpha}$I-*Sph*I (PTS) and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I (XTM) linkage maps based on markers with GQ>0.

	No. of markers		Length (cM)		Max interval (cM)	
Chr.	PTS	XTM	PTS	XTM	PTS	XTM
1	503	1,448	599.2	6,197.4	19.4	25.4
2	418	1,144	447.4	5,577.1	12.7	460.5
3	437	1,062	474.0	4,654.6	12.7	460.5
4	516	1,212	846.4	5,923.2	29.4	22.9
5	331	1,107	499.3	4,559.6	23.5	21.7
6	553	1,408	691.7	6,031.8	14.1	22.1
7	440	1,041	548.0	3,902.4	18.3	30.6
8	571	1,321	940.6	5,717.2	25.3	22.3
9	328	705	456.0	2,953.2	13.1	23.5
10	325	1,307	488.5	5,299.1	14.6	21.5
11	436	1,320	579.6	6,088.9	16.8	22.2
12	288	757	369.4	3,770.0	13.0	21.8
Overall	5,146	13,832	6,939.9	60,674.6	29.4	460.5

doi:10.6342/NTU202500264

Table 13. Summary of *Pst*I-*Taq*$^{\alpha}$I-*Sph*I (PTS) and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I (XTM) linkage maps based on markers with GQ>50.

Chr.	No. of markers		Length (cM)		Max interval (cM)	
	PTS	XTM	PTS	XTM	PTS	XTM
1	376	191	139.7	155.8	8.4	25.3
2	318	171	94.0	98.6	6.0	12.9
3	327	140	80.3	85.2	8.4	14.5
4	373	143	136.9	144.7	14.7	31.7
5	231	109	109.1	108.4	10.9	33.7
6	376	202	116.0	108.6	6.0	9.0
7	324	149	112.6	113.3	7.0	14.0
8	420	199	123.0	119.5	6.1	18.0
9	238	89	103.8	106.2	7.8	20.2
10	228	168	85.8	66.7	10.4	8.4
11	321	185	120.6	122.0	6.0	14.9
12	216	122	75.1	85.6	5.9	24.6
Overall	3,748	1,868	1,296.9	1,314.6	14.7	33.7

doi:10.6342/NTU202500264

Table 14. The number of markers remains after different overall GQ-filtered thresholds.

	*Pst*I-*Taq*$^{\alpha}$I-*Sph*I		*Xba*I-*Taq*$^{\alpha}$I-*Mse*I	
GQ	Total	Diff (%)	Total	Diff (%)
0	5,146	-	13,832	-
10	4,816	6.4	7,261	47.5
20	4,441	7.8	4,450	38.7
30	4,201	5.4	3,027	32.0
40	3,950	6.0	2,185	27.8
50	3,748	5.1	1,868	14.5
60	3,576	4.6	1,694	9.3
70	3,332	6.8	1,507	11.0
80	3,112	6.6	1,363	9.5

Diff (%) indicates the percentage reduction in the number of markers compared to the marker count at the previous GQ level.

doi:10.6342/NTU202500264

Table 15. The genotype quality improvement results are represented by the length of the linkage map in chromosome 1 (cM).

	*Pst*I-*Taq*$^{\alpha}$I-*Sph*I						*Xba*I-*Taq*$^{\alpha}$I-*Mse*I					
Method	GQ0	GQ10	GQ20	GQ30	GQ40	GQ50	GQ0	GQ10	GQ20	GQ30	GQ40	GQ50
Overall	940.6	389.8	179.2	156.7	144.9	139.7	6197.4	1230.4	429.8	229.4	163.9	155.8
W5	864.0	843.4	760.8	722.9	668.0	661.7	3584.8	1682.5	967.5	625.3	469.4	415.6
W7	1205.0	1145.2	1075.4	982.0	959.3	926.5	3769.4	1934.4	1194.3	840.2	599.6	516.7
W9	427.3	407.1	416.6	380.1	337.2	333.9	1866.4	773.6	496.6	307.9	276.2	226.3
W11	193.2	174.0	188.1	167.5	169.3	158.9	926.7	341.3	226.7	174.4	145.8	137.5
W13	133.1	129.4	122.7	130.1	122.7	124.7	572.7	232.7	154.8	132.3	124.6	124.2
W15	146.2	144.0	135.3	154.7	133.0	131.8	525.7	237.3	154.1	155.5	131.5	132.2
W17	124.8	116.9	120.1	119.2	124.6	117.1	366.0	187.2	134.5	135.1	123.2	129.9
W19	122.1	122.4	114.9	117.8	117.9	117.9	299.8	169.1	126.9	128.7	116.3	126.0
W21	127.7	121.3	120.3	116.6	120.9	115.3	290.1	169.6	126.9	130.0	122.3	120.4
XOquality	136.8	133.6	133.6	133.6	133.7	133.7	200.4	187.0	176.1	147.6	138.8	138.8

GQ with numbers indicates the overall GQ-filtered genotype data.
Overall indicates the map length estimated by the overall GQ-filtered genotype data.
W5 to W21 indicates the map length estimated after Genotype-Corrector corrected the genotype data by setting window size in 5 to 21 markers from the overall GQ-filtered genotype data.
XOquality indicates the map length estimated after filtering for the genotypes with GQ flanking at the crossover event from the overall GQ-filtered genotype data.

doi:10.6342/NTU202500264

Table 16. The parental lines of the F_2 population and the marker-assisted backcrossing breeding programs.

Population	Female parent	Male parent
F_2 population for powdery mildew resistance QTL mapping		
A6	TARI-18-437	TARI-18-494
B2	TARI-18-410	TARI-18-491
C4	TARI-18-491	TARI-18-449
Marker-assisted backcrossing		
qPM2	TARI-18-432	TARI-18-437
	TARI-18-431	
qPM5 and *qPM12*	TARI-18-432	TARI-18-449
	TARI-18-431	

TARI-18-432 and TARI-18-431 are the recurrent parents with green (G) and orange (O) flesh colors. The TARI-18-437 and TARI-18-449 are the powdery mildew resistance donor parents.

doi:10.6342/NTU202500264

Table 17. The average disease index of powdery mildew for the parental lines.

Line	Powdery mildew disease index	
	2017	2018
TARI-18-437	0.0	0.0
TARI-18-494	6.8	7.3
TARI-18-410	1.3	0.0
TARI-18-491	7.5	6.8
TARI-18-449	0.0	0.0
TARI-18-432	6.5	7.5
TARI-18-431	2.5	5.8

doi:10.6342/NTU202500264

Table 18. Summary of linkage maps used in this study.

Chr.	No. of markers			Length (cM)			Max interval (cM)		
	A6	B2	C4	A6	B2	C4	A6	B2	C4
1	294	761	873	157.4	155.4	142.1	5.8	4.2	5.1
2	250	199	296	126.0	113.8	128.8	8.2	13.3	11.2
3	345	518	601	138.3	126.0	130.4	11.8	4.2	4.8
4	401	644	721	179.1	162.5	181.4	7.9	5.1	6.5
5	281	334	430	131.0	123.3	123.0	9.7	24.3	7.2
6	309	488	588	154.5	149.2	163.3	6.8	9.0	4.2
7	324	428	511	126.7	116.3	112.9	9.9	3.4	4.5
8	283	414	543	166.4	150.5	164.2	15.5	8.4	9.3
9	245	399	425	127.7	129.3	121.6	10.9	7.1	5.4
10	169	277	327	93.4	106.4	85.8	7.8	10.9	5.4
11	267	485	595	137.3	127.4	141.5	9.5	8.1	8.9
12	298	502	498	122.0	125.7	114.3	7.2	5.1	5.7
Overall	3,466	5,449	6,408	1,659.8	1,586.0	1,609.2	15.5	24.3	11.2

doi:10.6342/NTU202500264

Table 19. QTL for Disease index resistance identified in the tree F_2 populations.

Population	QTL	Chr	Pos (cM)	a	d	LOD	R^2	Confidence interval (cM)	Flanking markers (bp)
A6	*qPM2*	2	8.9	-2.79	-2.79	94.5	93.5	8.34 – 10.43 [2.1]	778,333 – 1,142,653
B2	*qPM5*	5	92.7	-2.17	-1.28	62.8	80.1	91.63 – 94.67 [3.0]	25,276,104 – 25,725,099
C4	*qPM5*	5	89.0	-1.58	-1.10	27.7	29.3	86.45 – 90.58 [4.1]	25,203,821 – 25,678,875
	qPM12	12	77.4	-1.52	-1.12	28.9	31.0	75.15 – 78.44 [3.3]	22,493,040 – 22,879,440
	Total model					49.2	71.6		

Chr and Pos indicate the chromosome and peak position of QTL.
Confidence interval (cM) indicates the QTL confidence interval, defined using the 1.5-LOD drop method, with flanking positions reported in centimorgans (cM) and the interval length specified in [ ].
Flanking markers (bp) is the confidence interval of QTL extending to the closest markers.
a and d are the additive and the dominant effects of the QTL, respectively.
R^2 is the phenotypic variance explained by the QTL.

doi:10.6342/NTU202500264

Table 20. The designed TaqMan assays used in this study.

Target	Flesh color	Code	Assay Name	Pd.cM	BC_n	Context Sequence
qPM2	G	R1	2_382892_C-A	7.47	1	TAATAAATGTTTTAACTTTTCTAGT[C/A]TTTTTGGAAGAGGAATTGAAAGAAA
	G	F1	2_621190_A-G	9.73	1+2	CGTCCCTAGTTACTTCTGTTTGTTT[A/G]CCTCATTATACCACTGAAGCAACAA
	G	F2	2_1603309_A-T	19.32	1+2	TCGGTTCTGATAAGCAGGATGCTTC[A/T]GGAAATAAGCTTCTTCAAGATATTG
	G	R2	2_2351587_A-C	28.67	1	GGCTTGTGGAGAAGATCAACAATTC[A/C]ATTGGGCAAGACCCAAATTCTAAAT
	O	R1	2_103153_C-A	1.40	1	TCTTCTCACCCTCCATCTCATTAAC[C/A]AACTGTCAGCACCCTAAGATTCCCT
	O	F1.2	2_627533_C-T	9.78	2	GAATANTATCGTTGCATTTGATGAA[C/T]TTAAGGCTCTCCTAAGTTTAGAGCT
	O	F1.1	2_778333_T-A	10.72	1	ACGAAAATGTCACATAGCCTACAAC[T/A]CCTCCATTGACGGCATCATTCTCCC
	O	F2	2_816421_C-G	11.06	1+2	CCATTCTCAAATCCATCCAAAACAC[C/G]AATCCAACAAAGAACAAATAAAACC
	O	R2	2_1410652_T-C	17.03	1	CTGGTATGCTCAGTTATGCAAAGAG[T/C]GTTGCAGCTGATAACTGGTTGGCAT
qPM5	G/O	R1	5_24520366_C-T	86.51	1	CAACTTTCCATTTTAACTATTGGAA[C/T]GTGAGTTACTATTATTCCTTTCTTT
	G/O	F1.2	5_24941008_G-C	89.02	2	GCTTGTTCAATCTTGTGAGTTGGAT[G/C]AGATGATTTAGTTGAAGCGTGANTG
	G/O	F1.1	5_25264851_C-T	89.31	1	TTCTCCTAGTCACTTGTTTCTTACA[C/T]GAGATCTTTTGTTTCACTAGGAGGA
	G/O	F2	5_25862218_T-C	96.52	1+2	AACATGAAACACATCTCAATGATGA[T/C]GACGACGATGAATACCAGGCTGATT
	G/O	R2	5_26105586_T-A	99.27	1	AGGGACATCTTCATGGTTTGGTGGC[T/A]ACTTTAGCAAACTAAATTTTAGACA
qPM12	G	R1	12_21800139_A-G	71.69	1+2	CAACAAAAAATTTGTTTGATGGCCG[A/G]TAGGATTGTAACCTAATAAAAAAAC
	O	R1	12_21116734_T-G	63.61	1+2	GGACTATTATTCTGAACCTACCCAG[T/G]TCACAAGAGGCCCTTGTTCTTCTGA
	G/O	F1	12_22418198_A-G	78.87	1+2	TGAAGTAAAGTGAAAAGAAATAATC[A/G]TGTATGGCTCGCTGTTGACGTTTGT
	G/O	F2	12_23222171_G-A	84.38	1+2	TGATCGCCAAAGACGAGAAACTTGC[G/A]ACTTCATACCGTGATGAGAAATGGA
	G/O	R2	12_23491653_A-G	87.74	1	CCGTGGTTTGTCTATCAAGTTTTTA[A/G]AATCAACCTATGTGACAATCGCATT

The flesh colors G and O indicate the assays used for the green- or orange-fleshed recurrent parents, and G/O indicates the assay used for both recurrent parents in marker-assisted backcrossing programs. Codes R and F indicate the assays used for recombinant or foreground selection. The R1, R2, F1, and F2 indicate the physical position order of the TaqMan assays. The F1.1 and F1.2 indicate the F1 TaqMan assays but the different versions for their slight adjustment of the physical positions. Pd.cM indicates the predicted genetic position (cM) of the TaqMan assays through the LOESS regression models within each chromosome. BC_n filled with 1, 2, and 1+2 values indicate that the TaqMan assays were used during BC_1, BC_2, or both BC_1 and BC_2 generations.

doi:10.6342/NTU202500264

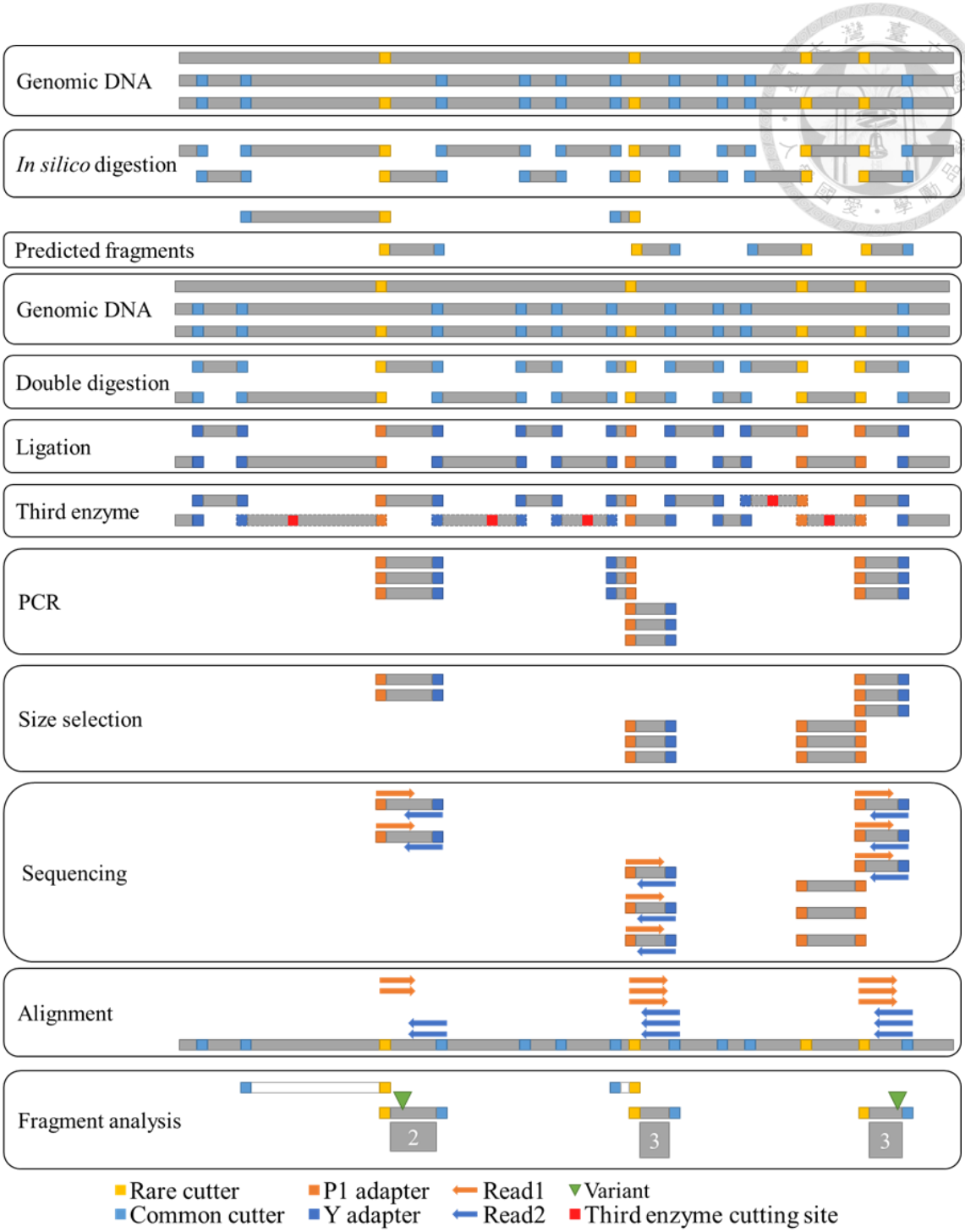


Figure 1. The *in silico* digestion, ddRADseq library preparation process, and the concept of fragment analysis in this study.

doi:10.6342/NTU202500264

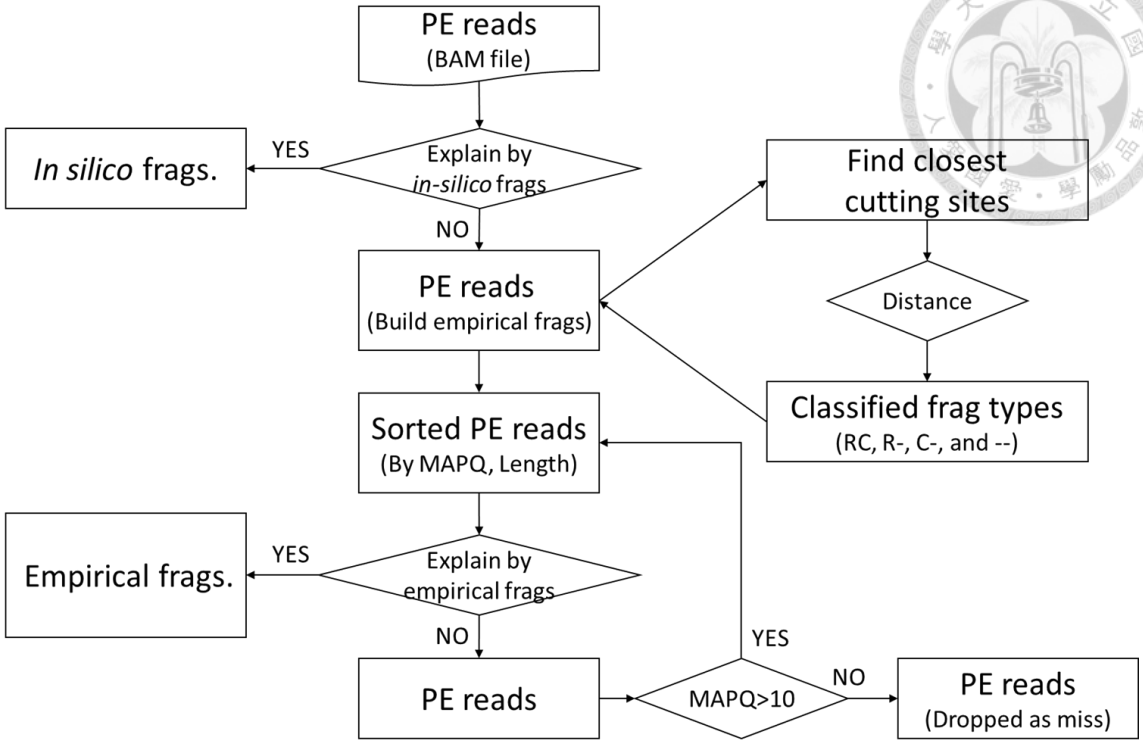


Figure 2. Data analysis flowchart of Fragment analysis.

Aligned paired-end (PE) reads were queried against *in silico*-predicted fragments (AB type, 100–1000 bp) using findOverlaps. PE reads that did not match the predicted fragments were categorized as empirical fragments (RC, R-, C-, or -- types) based on their relationship to rare (R) and common (C) cutter sites. The initial empirical fragment was selected from sorted PE reads (by MAPQ and insert size) as the subject, and PE reads were iteratively grouped using findOverlaps. The remaining ungrouped PE reads were labeled as “miss”.

doi:10.6342/NTU202500264

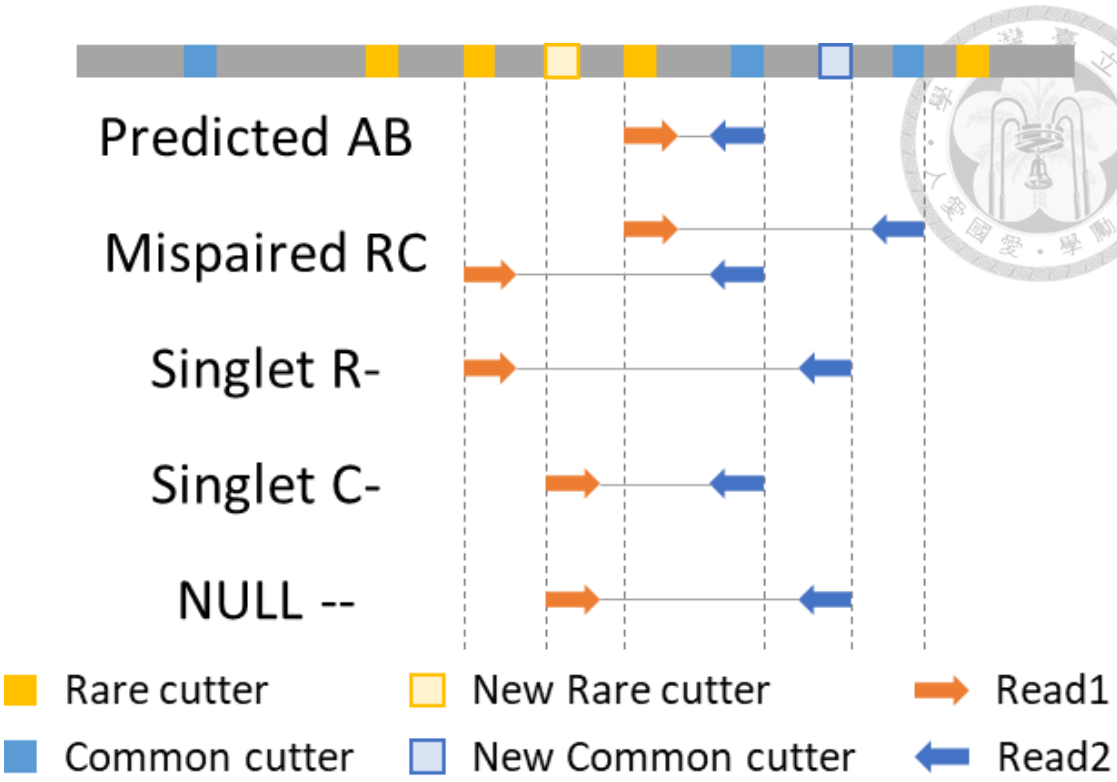


Figure 3. The Fragment types to classify the paired-end reads are inspired by Heffelfinger et al. (2014).

Yellow and blue squares indicate the rare and common cutters, and the lighter color indicates that the cutting sites were not included in the reference genome. The orange and blue color arrows indicate the paired-end sequence reads.

doi:10.6342/NTU202500264

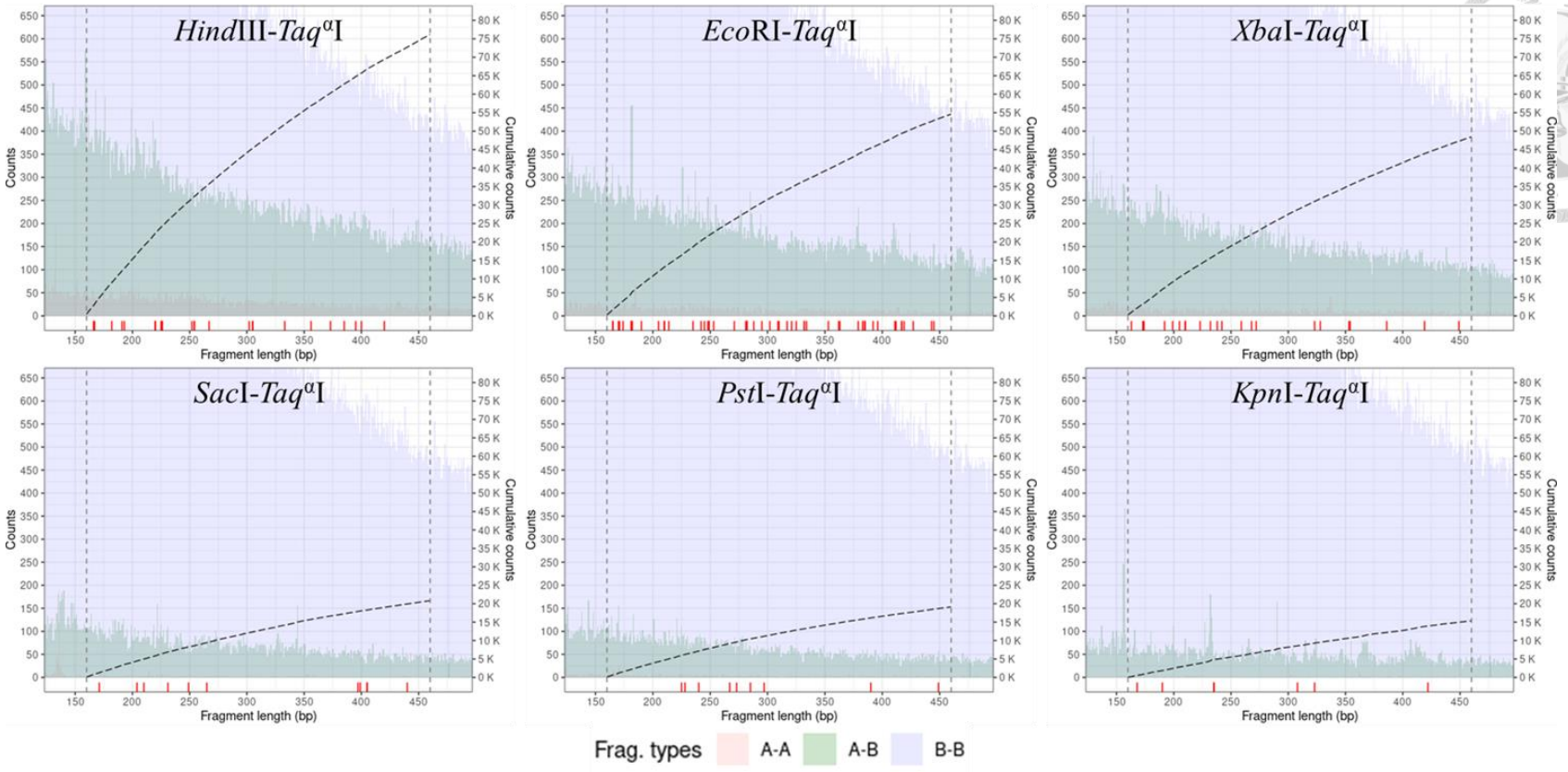


Figure 4. The fragment length distribution of six enzyme combinations.
The left y-axis indicates the fragment count of the histogram, the right y-axis indicates the cumulative fragment counts, and the x-axis indicates the fragment length (bp). The histogram indicates the three fragment types, AA, AB, and BB, in red, green, and blue colors. The black dashed curve indicates that the cumulative fragment counts of AB type were estimated between the two vertical gray dashed lines. The red rug indicates the AB fragments from the chloroplast.

doi:10.6342/NTU202500264

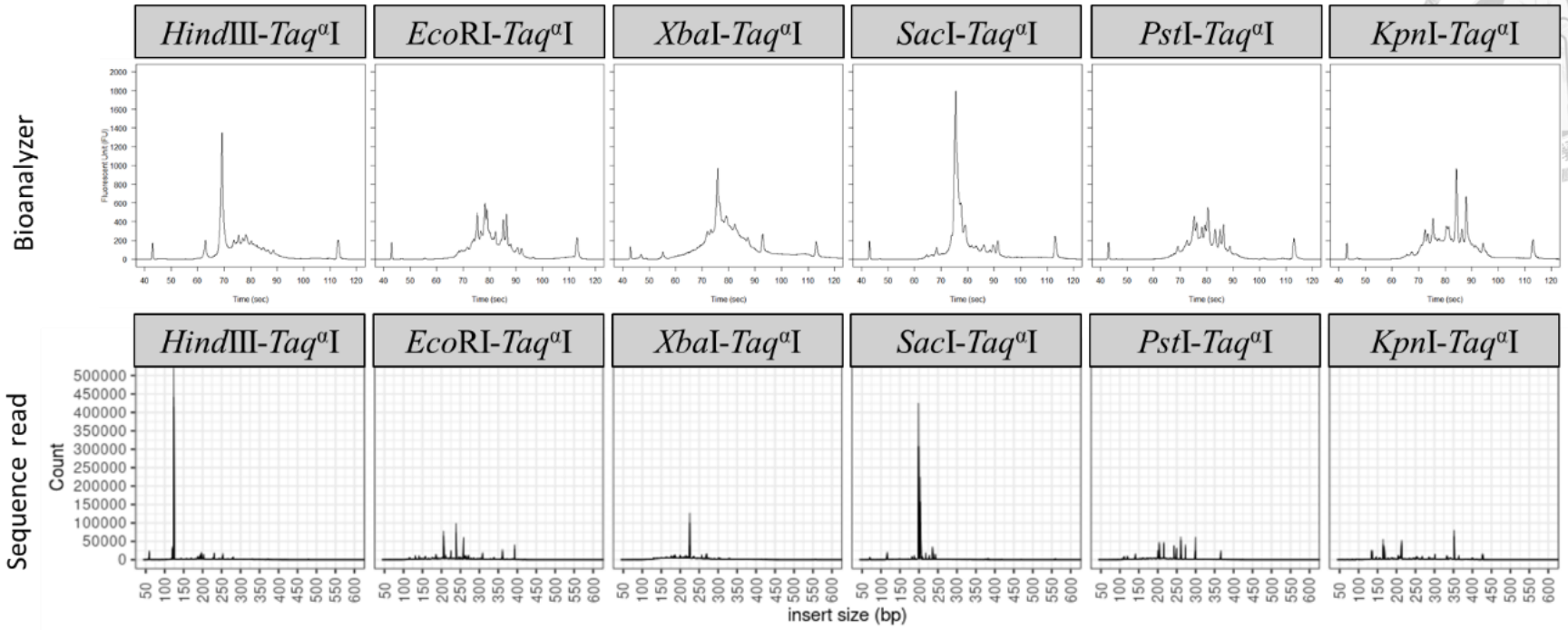


Figure 5. The electrophoresis of ddRAD library in Bioanalyzer and insert DNA size distribution of PE reads from sequencing reads for the enzyme combinations.

doi:10.6342/NTU202500264

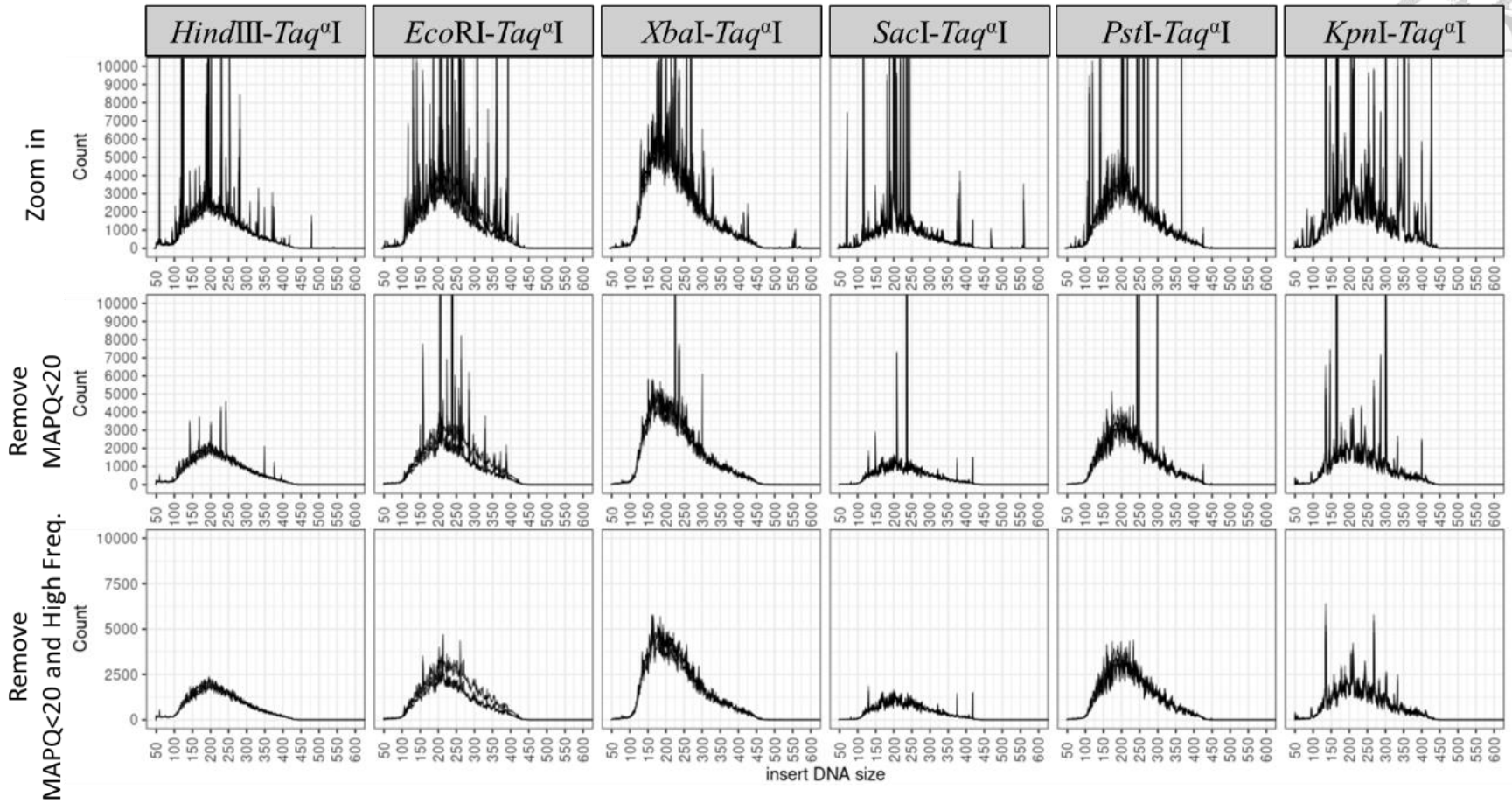


Figure 6. The insert DNA size distribution of PE reads for the enzyme combinations.

From top to bottom, the insert size distribution in 10,000 counts (Zoom in), the distribution after removing PE reads of MAPQ<20, and additional removing the PE reads for both MAPQ<20 and the PE reads within high-frequency fragments.

doi:10.6342/NTU202500264

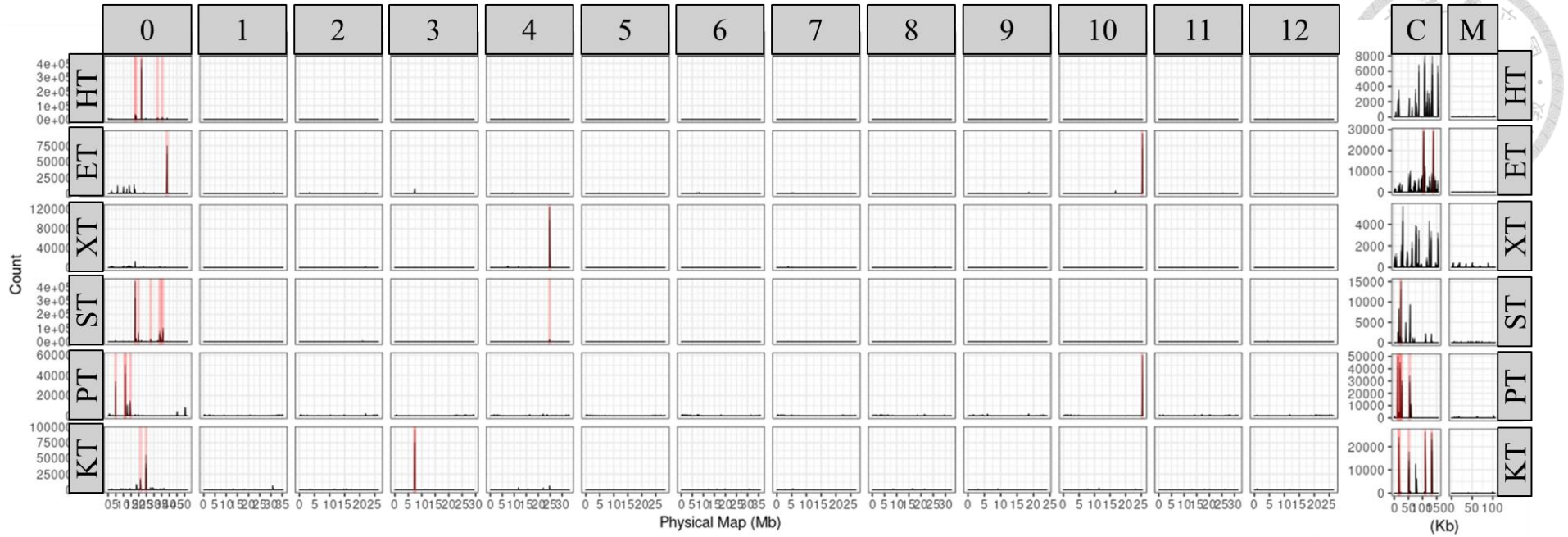


Figure 7. The empirical distribution of raw PE read locations across the genome for each enzyme combination.

The strips on top of the figures were the chromosomes 0 to 12, chloroplast (C), and mitochondria (M). The strips on both sides of the figures were the enzyme combinations in HT indicates *Hind*III-*Taq*$^{\alpha}$I as well as ET, *Eco*RI-*Taq*$^{\alpha}$I; XT, *Xba*I-*Taq*$^{\alpha}$I; ST, *Sac*I-*Taq*$^{\alpha}$I; PT, *Pst*I-*Taq*$^{\alpha}$I; and KT, *Kpn*I-*Taq*$^{\alpha}$I.
The red lines indicate at least two samples contained the high-frequency fragments.

doi:10.6342/NTU202500264

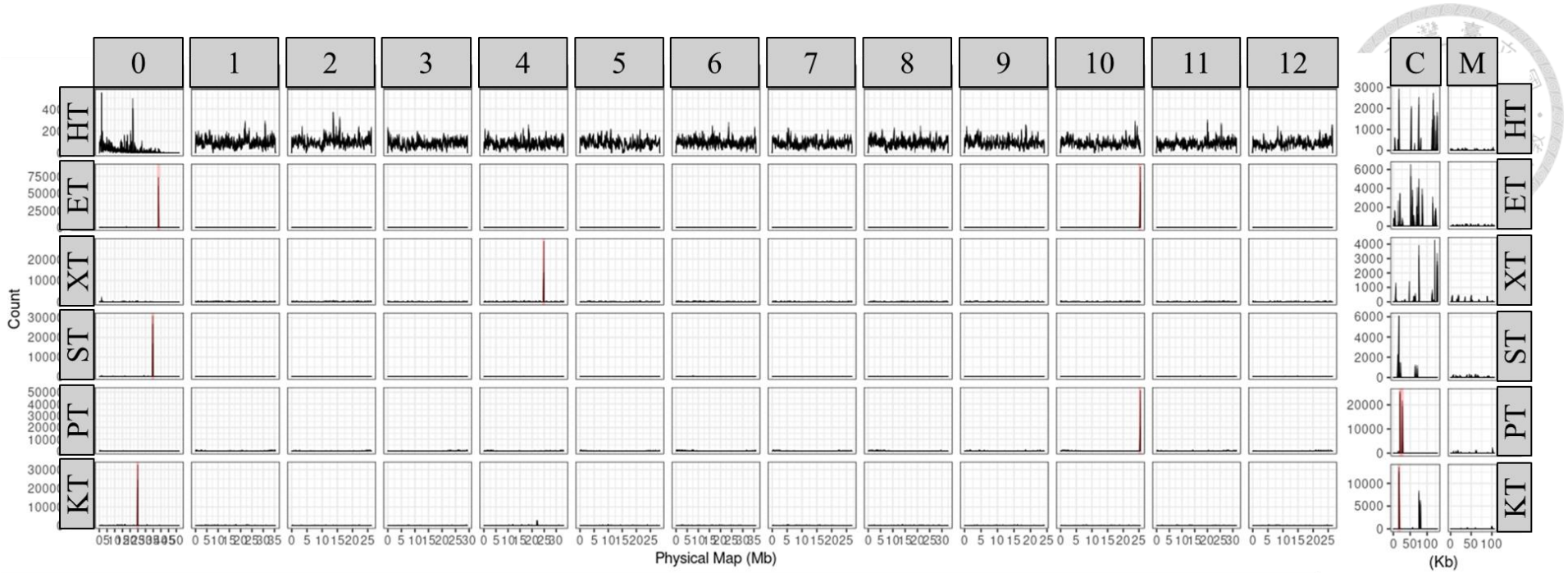


Figure 8. The empirical distribution of MAPQ≧20 PE read locations across the genome for each enzyme combination.

The strips on top of the figures were the chromosomes 0 to 12, chloroplast (C), and mitochondria (M). The strips on both sides of the figures were the enzyme combinations in HT indicates *Hind*III-*Taq*$^{\alpha}$I as well as ET, *Eco*RI-*Taq*$^{\alpha}$I; XT, *Xba*I-*Taq*$^{\alpha}$I; ST, *Sac*I-*Taq*$^{\alpha}$I; PT, *Pst*I-*Taq*$^{\alpha}$I; and KT, *Kpn*I-*Taq*$^{\alpha}$I.
The red lines indicate at least two samples contained the high-frequency fragments.

doi:10.6342/NTU202500264

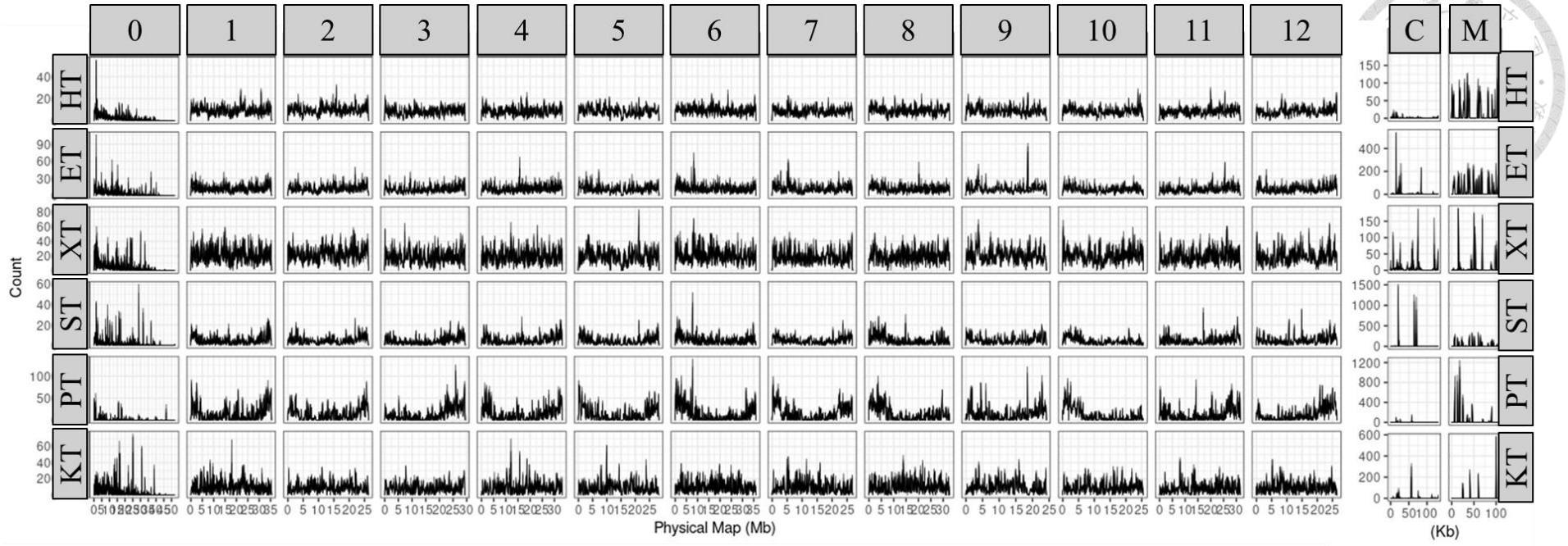


Figure 9. The empirical distribution of MAPQ≧20 and removed high-frequency fragments PE read locations across the genome for each enzyme combination.

The strips on top of the figures were the chromosomes 0 to 12, chloroplast (C), and mitochondria (M). The strips on both sides of the figures were the enzyme combinations in HT indicates *Hind*III-*Taq*$^{\alpha}$I as well as ET, *Eco*RI-*Taq*$^{\alpha}$I; XT, *Xba*I-*Taq*$^{\alpha}$I; ST, *Sac*I-*Taq*$^{\alpha}$I; PT, *Pst*I-*Taq*$^{\alpha}$I; and KT, *Kpn*I-*Taq*$^{\alpha}$I.

doi:10.6342/NTU202500264

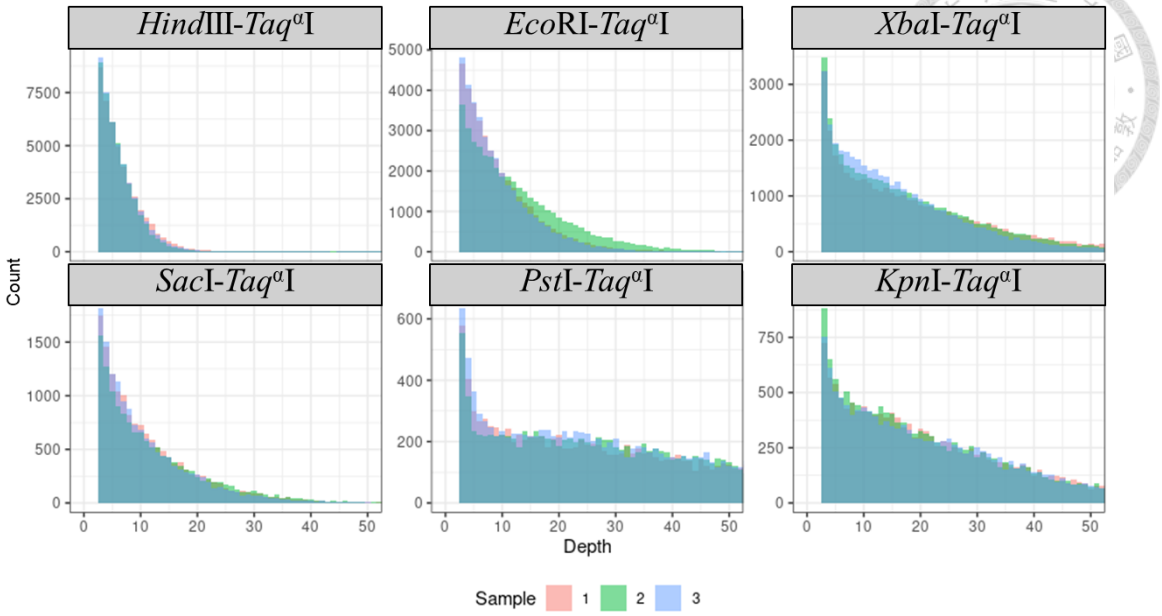


Figure 10. Distribution of fragment depth across enzyme combinations.

doi:10.6342/NTU202500264

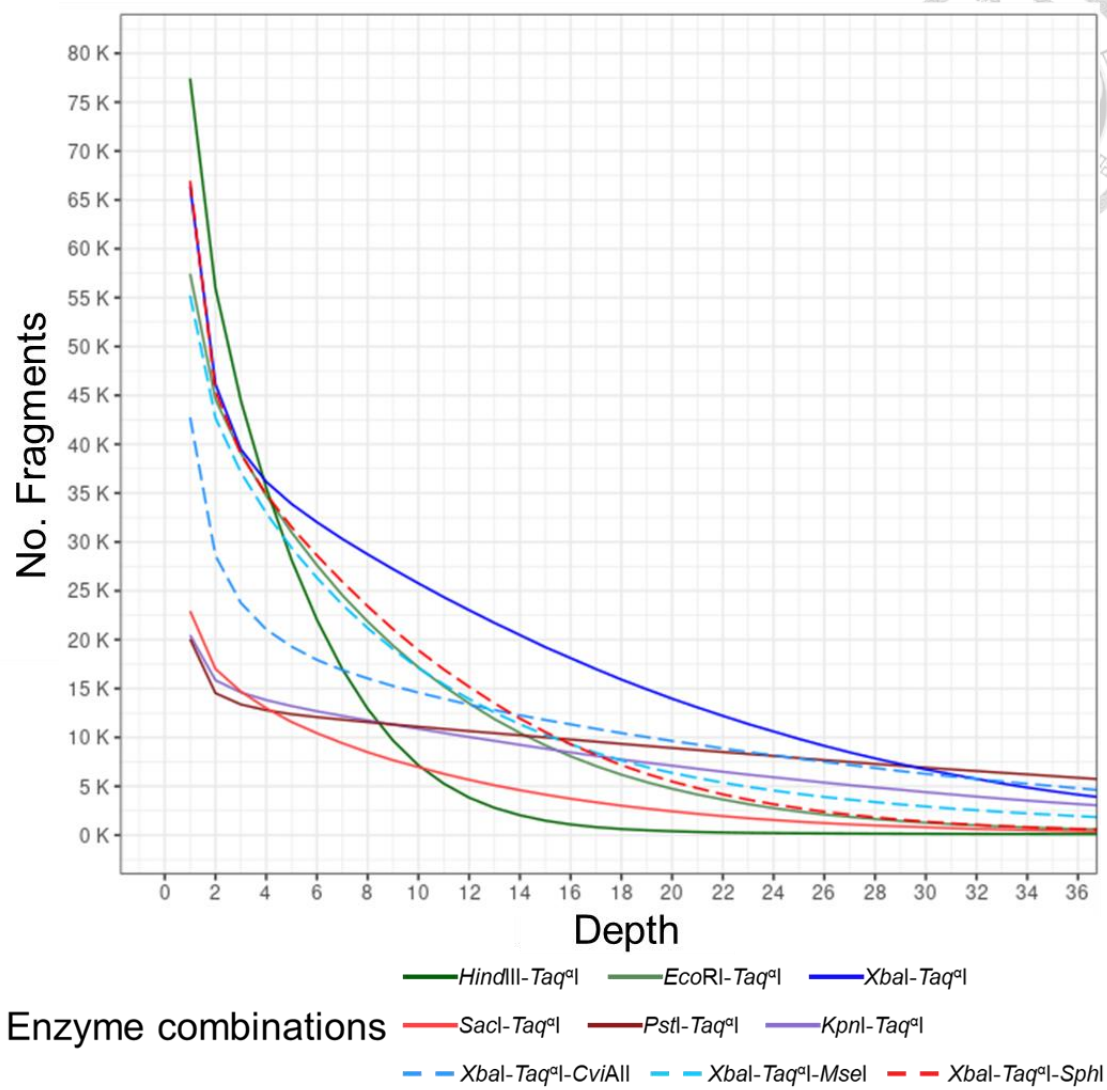


Figure 11. The number of fragments from given depth thresholds across the enzyme combinations.

doi:10.6342/NTU202500264

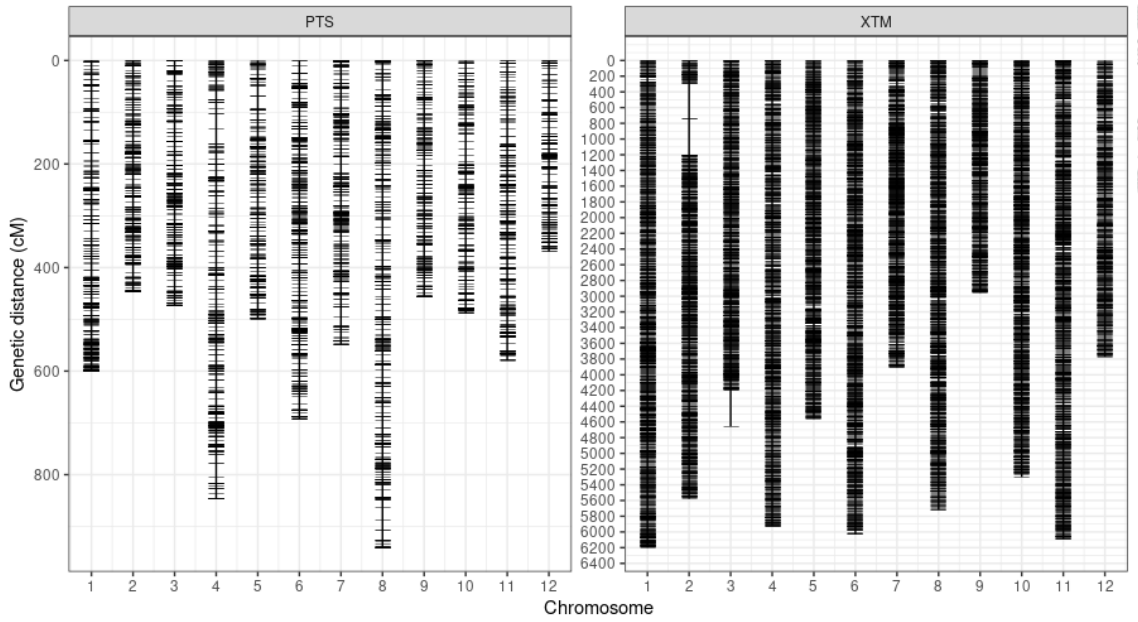


Figure 12. Linkage maps built with 5,146 markers for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I (PTS) and with 13,832 markers for *Xba*I-*Taq*$^{\alpha}$I-*Mse*I (XTM) with GQ>0.

doi:10.6342/NTU202500264

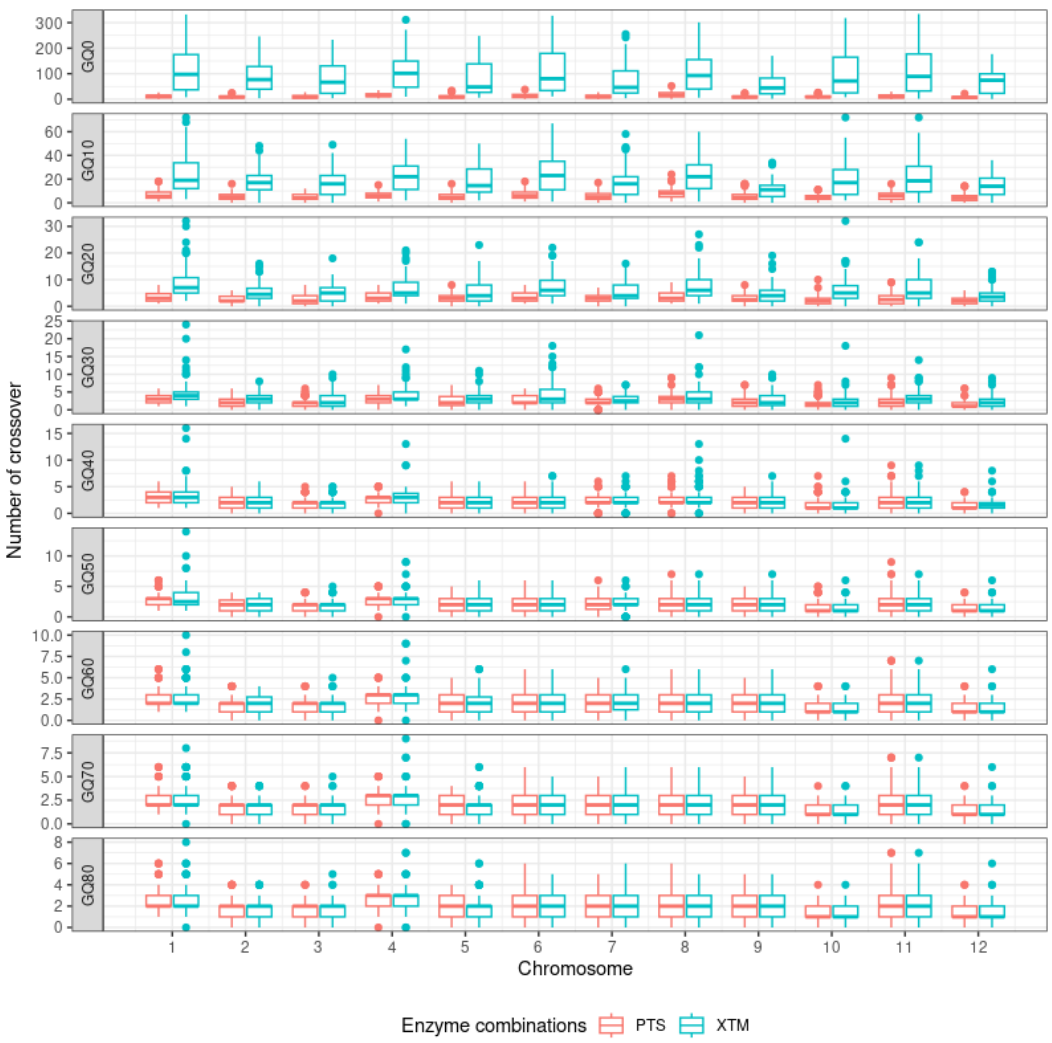


Figure 13. The boxplot for the number of crossovers across different GQ thresholds. The enzyme combinations are *Pst*I-*Taq*$^{\alpha}$I-*Sph*I (PTS) and *Xba*I-*Taq*$^{\alpha}$I-*Mse*I (XTM).

doi:10.6342/NTU202500264

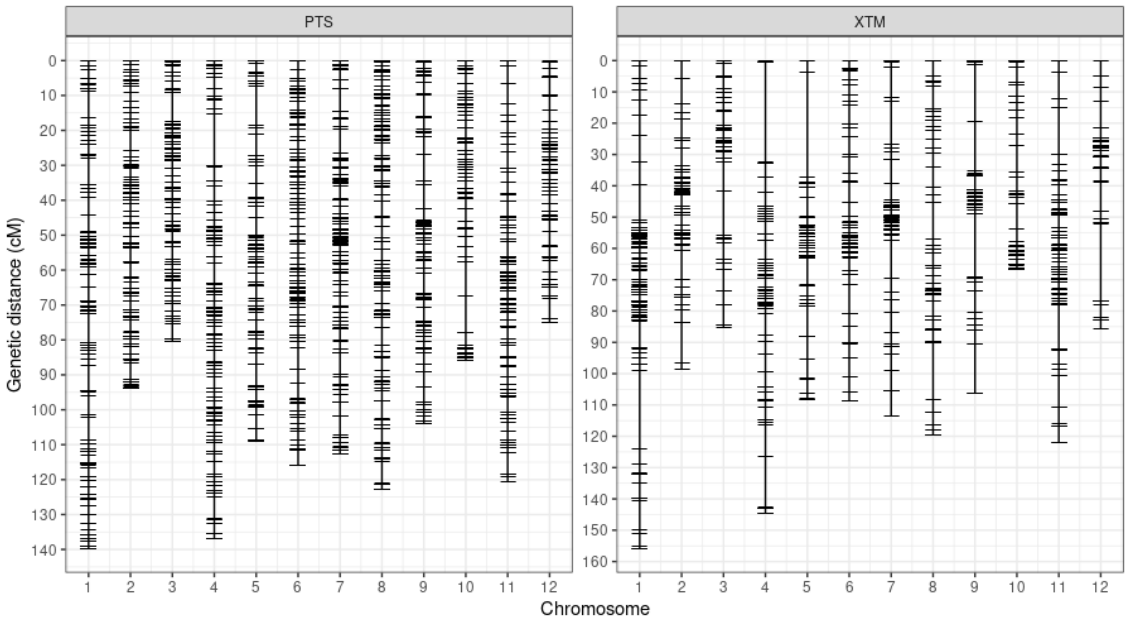


Figure 14. Linkage maps from 3,748 markers for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I (PTS) and 1,868 markers for *Xba*I-*Taq*$^{\alpha}$I-*Mse*I (XTM) based on markers with GQ>50.

doi:10.6342/NTU202500264

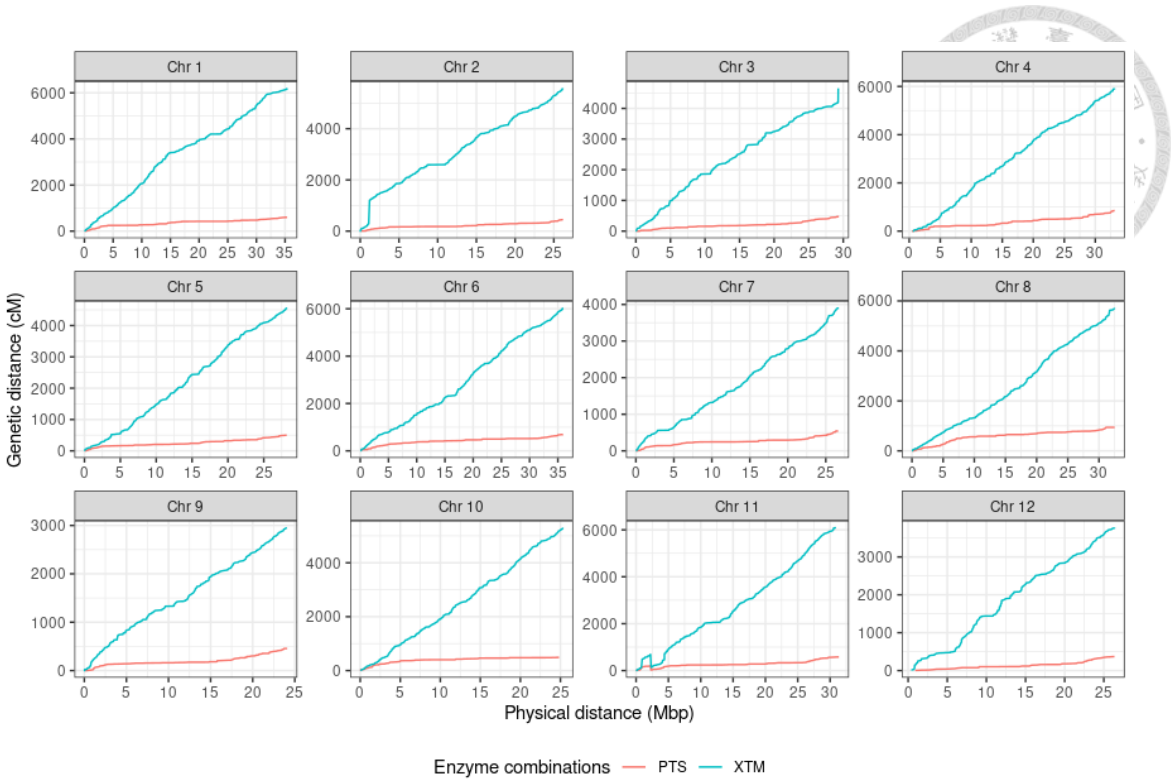


Figure 15. The relationship between the genetic map and the physical map of the linkage maps in 5,146 markers for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I (PTS) and 13,832 markers for *Xba*I-*Taq*$^{\alpha}$I-*Mse*I (XTM) in GQ>0.

doi:10.6342/NTU202500264

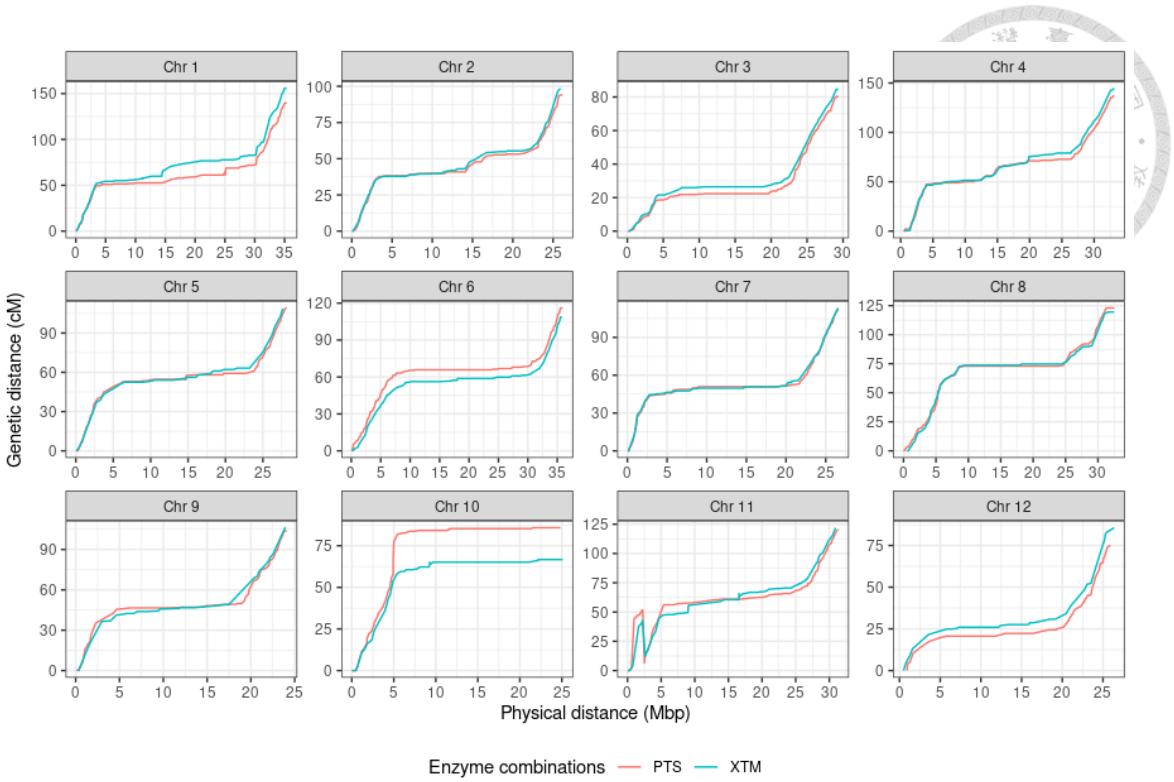


Figure 16. The relationship between the genetic map and the physical map in 3,748 markers for *Pst*I-*Taq*$^{\alpha}$I-*Sph*I (PTS) and 1,868 markers for *Xba*I-*Taq*$^{\alpha}$I-*Mse*I (XTM) built with markers GQ>50.

doi:10.6342/NTU202500264

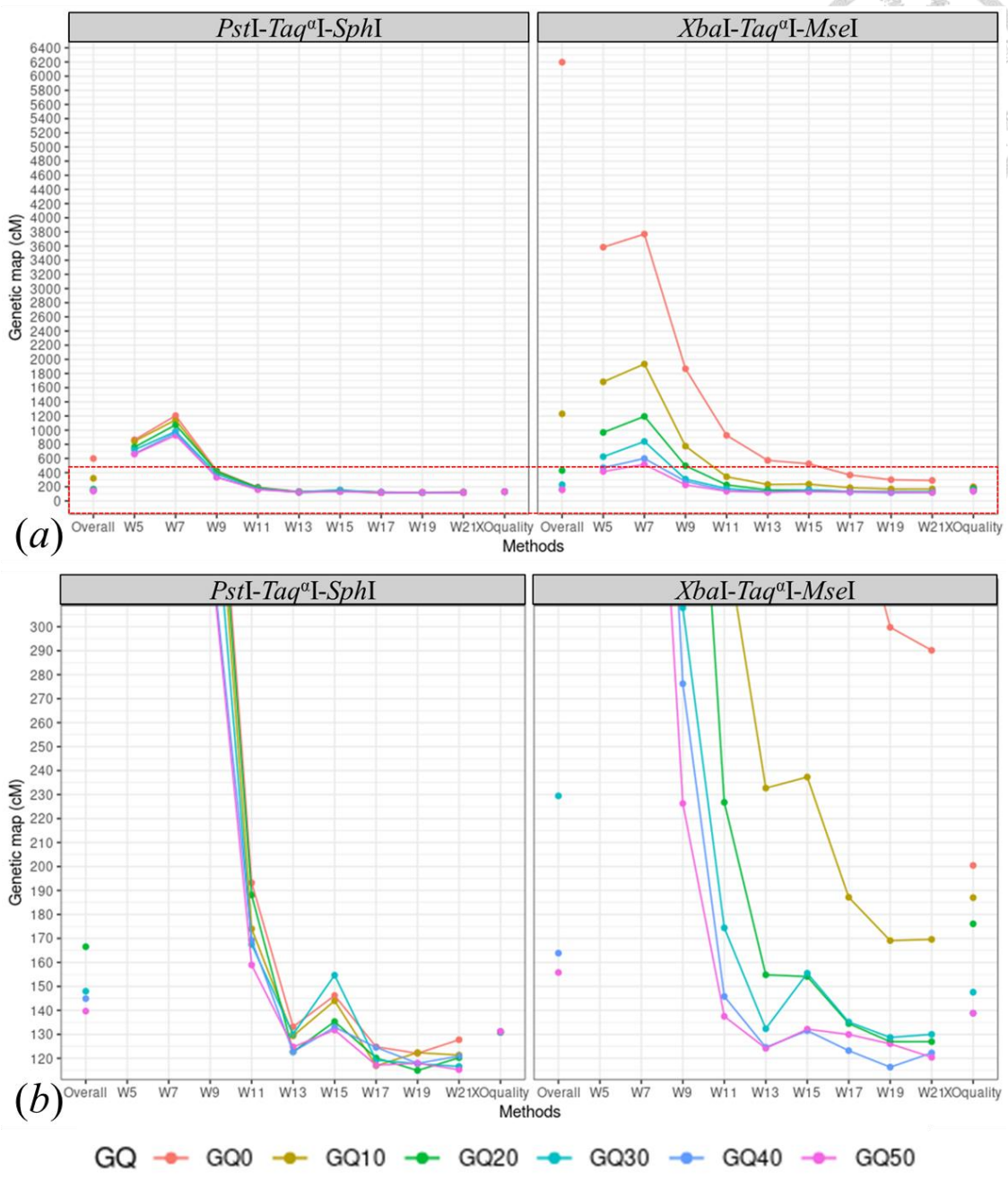


Figure 17. Genotype data improvement for map length of chromosome 1.

GQ with numbers indicates the overall GQ-filtered genotype data. Overall indicates the map length estimated by the overall GQ-filtered genotype data. W5 to W21 indicates the map length estimated after Genotype-Corrector corrected the genotype data by setting window size in 5 to 21 markers from the overall GQ-filtered genotype data. XOquality indicates the map length estimated after filtering for the genotypes with GQ flanking at the crossover event from the overall GQ-filtered genotype data. (a) is the full scale for the y-axis, and (b) is the zoom-in for the y-axis within the 120 - 300 cM range.

doi:10.6342/NTU202500264

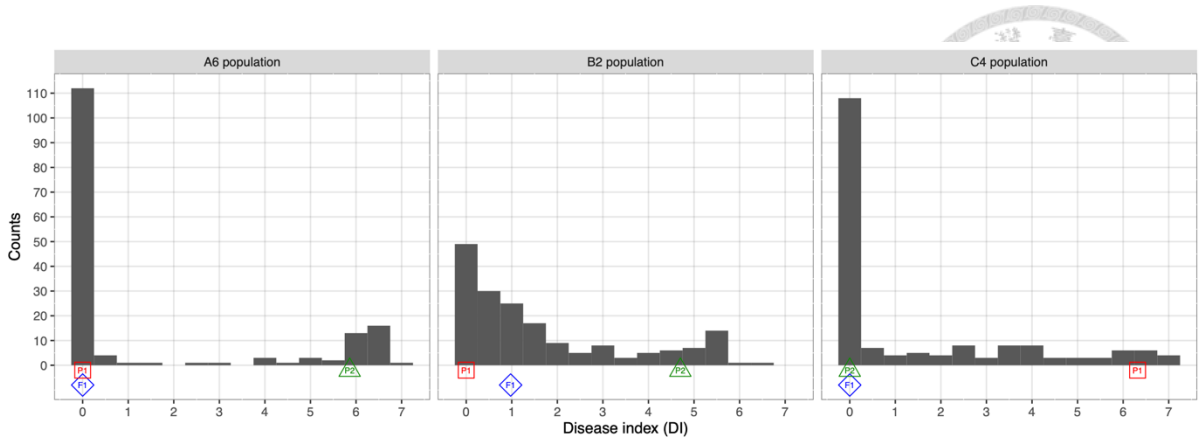


Figure 18. The powdery mildew disease index distribution of the three F_2 populations. The average disease index of the two parents and F_1 are indicated below the histogram using red square, green triangle, and blue diamond, respectively.

doi:10.6342/NTU202500264

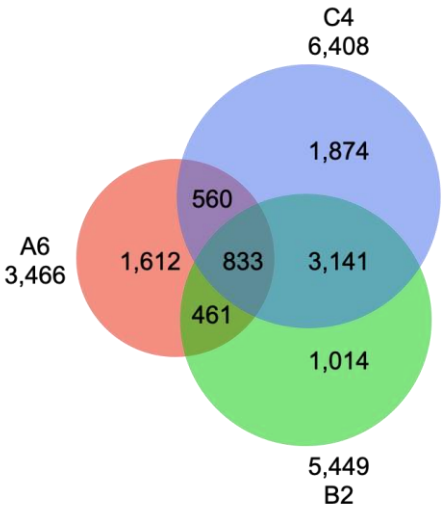


Figure 19. Number of SNPs across the three F_2 populations

doi:10.6342/NTU202500264

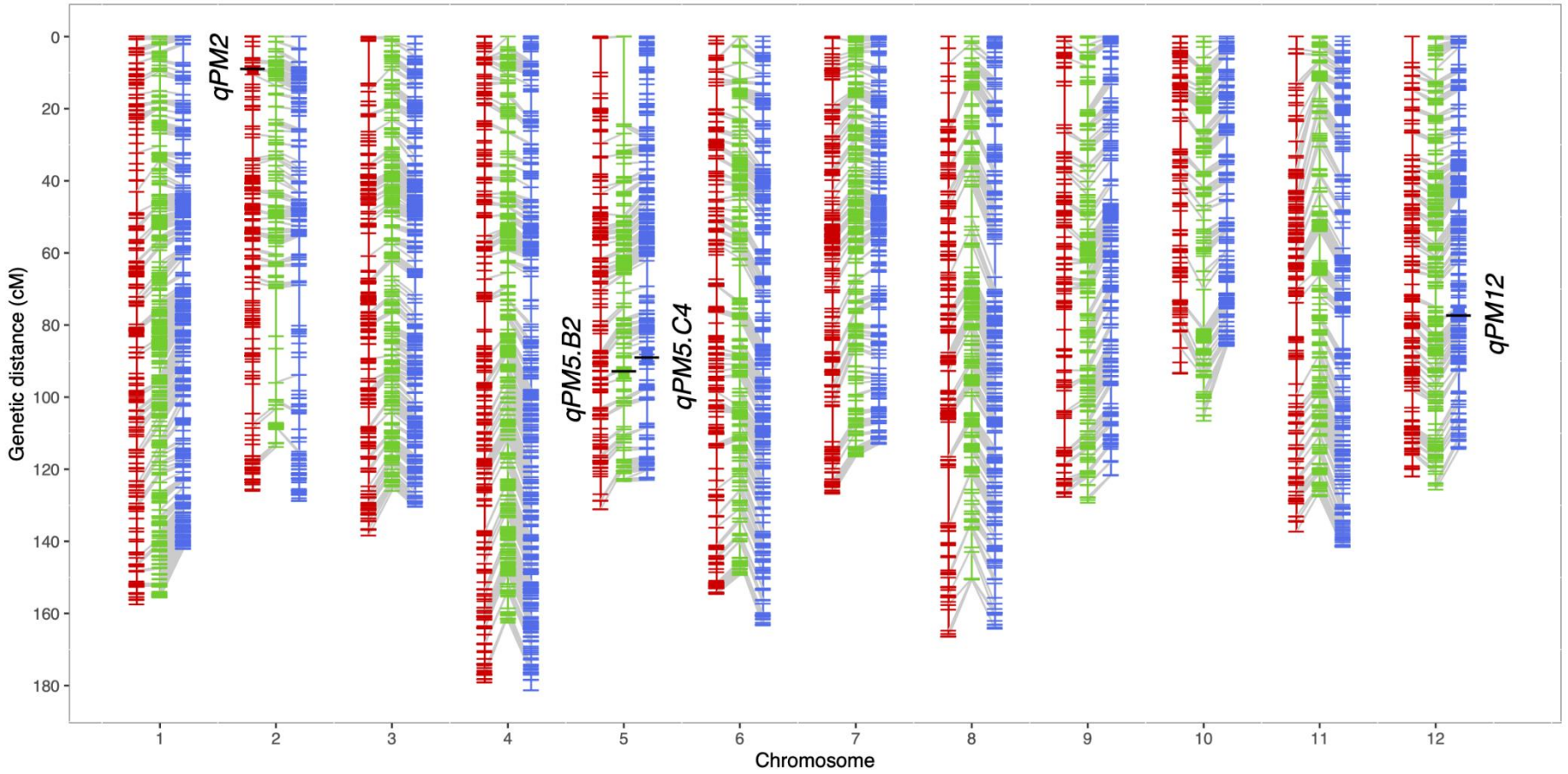
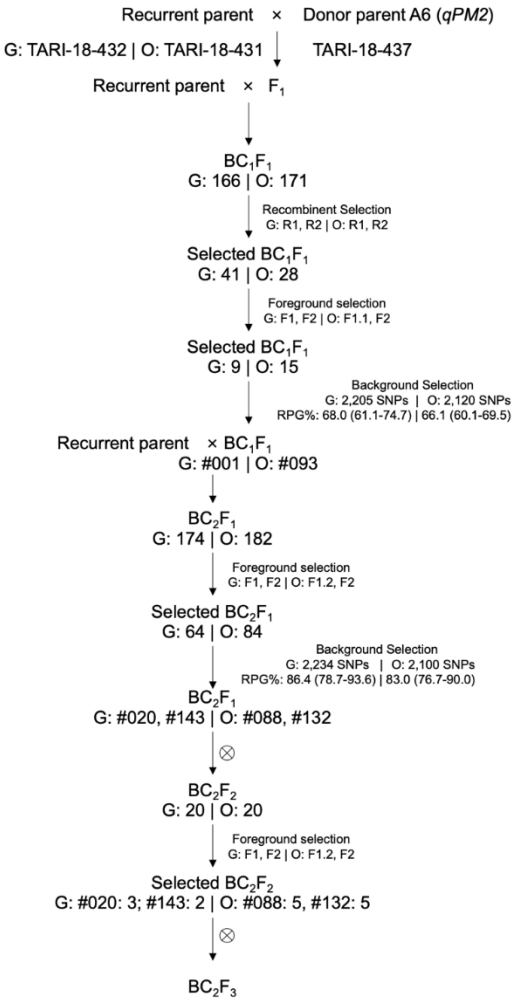


Figure 20. The linkage maps of the three F_2 populations.
The three F_2 populations, A6, B2, and C4, are shown in red, green, and blue, respectively. The common markers between populations are linked using gray lines, and the black bars on the linkage groups are the QTL identified in this study. *qPM5.B2* and *qPM5.C4* represent the *qPM5* detected in B2 and C4 population, respectively.

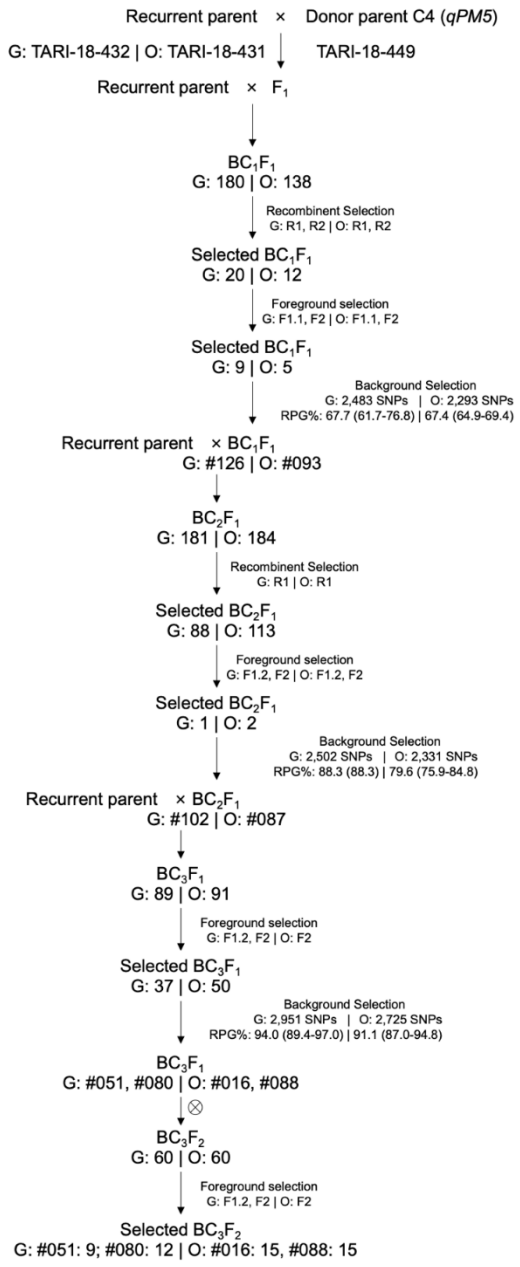
doi:10.6342/NTU202500264



Line Name	Green		Orange	
	#020	#143	#088	#132
No. of individuals	20	20	20	20
Average DI	0.0	0.0	0.0	0.0

Figure 21. Marker-assisted backcrossing breeding scheme for *qPM2*.
Crosses were made between the donor TARI-18-437 and recurrent parents TARI-18-432 (green-fleshed, G) and TARI-18-431 (orange-fleshed, O). G: Number | O: Number indicates the number of individuals for the green- and orange-fleshed background at each step. R1, R2, F1, F1.1, F1.2, and F2 are markers used for recombinant and foreground selection (Table 20). #Numbers indicate the selected line names derived from the selected individuals. Background selection text boxes contain the number of SNPs for RPG evaluation and average RPG recovery (%) with their ranges in brackets under green- and orange-fleshed backgrounds.

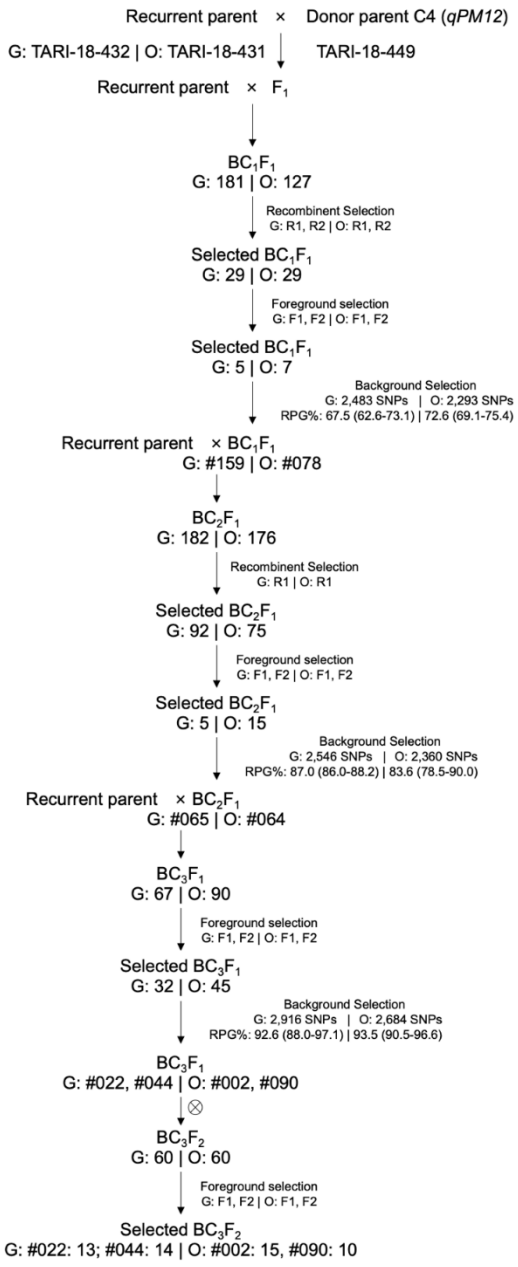
doi:10.6342/NTU202500264



Line Name	Green		Orange	
	#051	#080	#016	#088
No. of individuals	9	12	15	15
Average DI	1.0	1.0	0.0	0.0

Figure 22. Marker-assisted backcrossing breeding scheme for *qPM5*.
Crosses were made between the donor TARI-18-449 and recurrent parents TARI-18-432 (green-fleshed, G) and TARI-18-431 (orange-fleshed, O). G: Number | O: Number indicates the number of individuals for the green- and orange-fleshed background at each step. R1, R2, F1, F1.1, F1.2, and F2 are markers used for recombinant and foreground selection (Table 20). #Numbers indicate the selected line names derived from the selected individuals. Background selection text boxes contain the number of SNPs for RPG evaluation and average RPG recovery (%) with their ranges in brackets under green- and orange-fleshed backgrounds.

doi:10.6342/NTU202500264



Line Name	Green		Orange	
	#022	#044	#002	#090
No. of individuals	13	14	15	10
Average DI	0.3	0.1	0.0	0.0

Figure 23. Marker-assisted backcrossing breeding scheme for *qPM12*.
Crosses were made between the donor TARI-18-449 and recurrent parents TARI-18-432 (green-fleshed, G) and TARI-18-431 (orange-fleshed, O). G: Number | O: Number indicates the number of individuals for the green- and orange-fleshed background at each step. R1, R2, F1, F1.1, F1.2, and F2 are markers used for recombinant and foreground selection (Table 20). #Numbers indicate the selected line names derived from the selected individuals. Background selection text boxes contain the number of SNPs for RPG evaluation and average RPG recovery (%) with their ranges in brackets under green- and orange-fleshed backgrounds.

doi:10.6342/NTU202500264

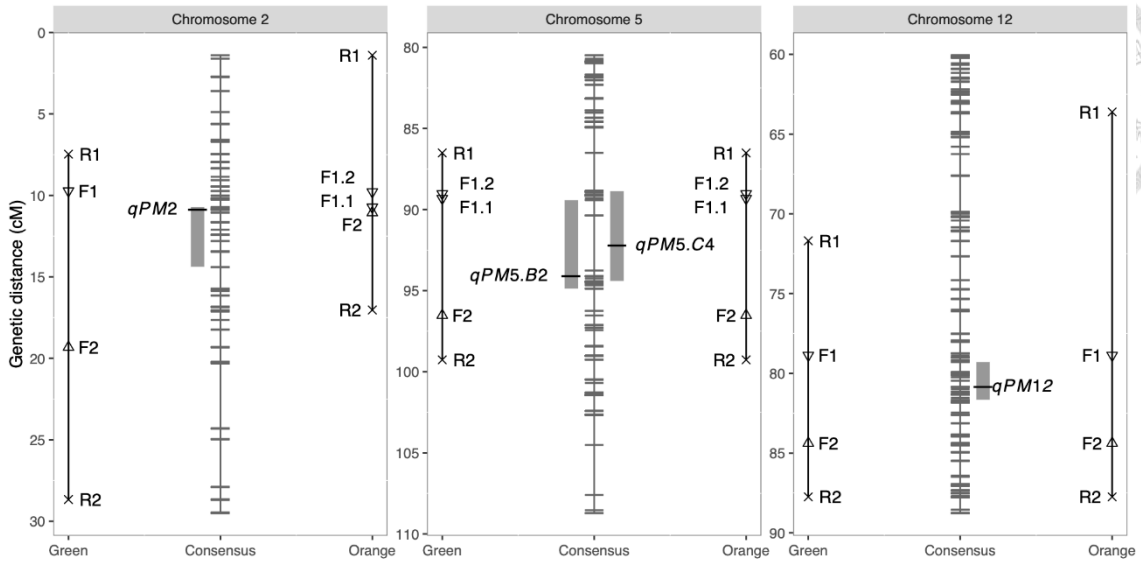


Figure 24. Position of the TaqMan assays used for foreground and recombinant selection.

The consensus linkage groups built from A6, B2, and C4 are shown in the middle. *qPM5.B2* and *qPM5.C4* represent the *qPM5* detected in the B2 and C4 populations, respectively. Marker positions for green- or orange-fleshed parents are indicated at the left and right sides, respectively. F1, F1.1, F1.2, F2, R1, R2 are marker names. Please refer to Table 20 and Supplementary Table 4 for detailed marker information.

doi:10.6342/NTU202500264

Supplementary Tables

Supplementary Table 1. The formula for input adapters for ligation in ddRAD library preparation.

Steps	P1 adapter						Y adapter
	*Hind*III	*Eco*RI	*Xba*I	*Sac*I	*Pst*I	*Kpn*I	*Taq*$^{\alpha}$I
Cut frequency (bp)	2,808	4,017	4,827	10,412	11,906	16,832	483
Fragment mass (g/mole)	1,853,280	2,651,220	3,185,820	6,871,920	7,857,960	11,109,120	318,780
Sample mass (g)	3.00E-7	3.00E-7	3.00E-7	3.00E-7	3.00E-7	3.00E-7	3.00E-7
Fragments/sample (moles)	1.62E-13	1.13E-13	9.42E-14	4.37E-14	3.82E-14	2.70E-14	9.41E-13
Ends/sample (moles)	3.24E-13	2.26E-13	1.88E-13	8.73E-14	7.64E-14	5.40E-14	1.88E-12
Ends/sample (pmoles)	0.3238	0.2263	0.1883	0.0873	0.0764	0.0540	1.8822

Initial DNA mass = 300 ng= 0.3 μg

doi:10.6342/NTU202500264

Supplementary Table 2. Bioanalyzer result of two enzyme combination libraries.

Enzyme combinations	Size [bp]	Conc. [pg/μl]	Molarity [pmol/l]	Area	Aligned Migration Time [s]	Peak Height	Peak Width	% of Total	Time corrected area
*Hind*III-*Taq*$^{\alpha}$I	290	1,806.3	9,450.8	2,007.8	69.2	1351.9	7.0	41.9	2,844.0
*Sac*I-*Taq*$^{\alpha}$I	367	1,662.6	6,858.5	2,844.8	75.5	1803.2	2.9	46.1	3,770.4
*Kpn*I-*Taq*$^{\alpha}$I	514	732.1	2,158.9	1,210.7	84.3	1190.8	2.3	18.0	1,469.3
	590	426.9	1,096.2	791.9	88.0	816.2	1.9	11.3	919.7
*Xba*I-*Taq*$^{\alpha}$I	385	1,304.1	5,132.3	1,891.4	75.8	970.8	3.7	23.8	2,470.8
*Eco*RI-*Taq*$^{\alpha}$I	366	358.1	1,482.1	594.0	75.4	499.8	1.8	11.5	770.9
	403	336.9	1,265.6	604.2	78.2	599.8	1.4	11.2	754.3
	418	318.6	1,156.2	577.8	79.0	538.3	1.4	10.6	713.8
	556	216.3	589.5	455.9	86.4	485.0	1.5	7.7	513.4
*Pst*I-*Taq*$^{\alpha}$I	331	337.5	1,544.4	454.7	72.6	204.0	3.3	9.5	620.0
	365	504.8	2,093.1	736.5	75.3	462.2	2.6	14.9	966.5
	404	316.2	1,186.5	499.8	78.3	350.1	1.8	9.7	630.3
	443	455.7	1,557.0	743.2	80.5	551.1	2.1	14.0	910.2
	492	292.5	900.1	495.2	83.2	322.8	2.3	9.0	585.7
	556	240.7	656.1	446.6	86.4	372.6	2.2	7.8	508.4

Conc indicates the peak concentration. “% of total” indicates the percentage of the area of the individual fragment compared to the total area above the baseline.

doi:10.6342/NTU202500264

Supplementary Table 3. High-frequency fragments under all PE reads and MAPQ filtered PE reads across enzyme combinations.

Enzyme combinations	PE read types	Chr	Start	End	Length	Frag. Types	Ave. PE reads	%
*Hind*III-*Taq*$^{\alpha}$I	all_reads	0	21,831,735	21,831,858	124	C-	377,401	36.7
	all_reads	0	17,566,366	17,566,486	121	C-	35,906	3.5
	all_reads	0	18,165,693	18,165,816	124	C-	21,276	2.1
	all_reads	0	35,529,284	35,529,536	253	--	12,988	1.3
	all_reads	0	32,314,489	32,314,612	124	C-	11,436	1.1
*Eco*RI-*Taq*$^{\alpha}$I	all_reads	10	25,358,758	25,358,996	239	C-	69,021	7.1
	MQ20_filter	10	25,358,758	25,358,996	239	C-	64,917	17.9
	all_reads	0	38,485,095	38,485,324	230	A-B	56,522	5.8
	MQ20_filter	0	38,485,095	38,485,324	230	A-B	55,670	15.4
	all_reads	C	138,325	138,606	282	A-B	22,269	2.3
	all_reads	C	103,746	104,027	282	A-B	22,233	2.3
*Xba*I-*Taq*$^{\alpha}$I	all_reads	4	24,752,204	24,752,428	225	R-	84,580	7.9
	MQ20_filter	4	24,752,204	24,752,428	225	R-	17,184	5.7
*Sac*I-*Taq*$^{\alpha}$I	all_reads	0	17,566,288	17,566,486	199	C-	341,854	33.9
	all_reads	0	35,844,077	35,844,281	205	RC	84,397	8.4
	all_reads	0	33,732,463	33,732,665	203	C-	59,263	5.9
	all_reads	0	19,725,288	19,725,490	203	R-	54,081	5.4
	all_reads	0	34,718,493	34,718,750	258	A-B	27,805	2.8
	MQ20_filter	0	34,718,493	34,718,750	258	A-B	24,902	3.2
	all_reads	0	27,865,563	27,865,765	203	--	16,172	1.6
	all_reads	4	24,764,536	24,764,775	240	A-B	10,630	1.1
	all_reads	C	22,654	22,918	265	A-B	10,399	1.0
*Pst*I-*Taq*$^{\alpha}$I	all_reads	10	25,353,624	25,353,946	323	A-B	34,587	4.1
	MQ20_filter	10	25,353,624	25,353,946	323	A-B	29,779	11.4
	all_reads	C	20,512	20,739	228	A-B	31,315	3.7
	all_reads	C	20,249	20,515	267	A-B	26,396	3.1
	MQ20_filter	C	20,249	20,515	267	A-B	18,494	7.1
	all_reads	C	26,664	26,936	273	A-B	20,975	2.5
	MQ20_filter	C	26,664	26,936	273	A-B	15,301	5.9

doi:10.6342/NTU202500264

Enzyme combinations	PE read types	Chr	Start	End	Length	Frag. Types	Ave. PE reads	%
	all_reads	0	11,181,601	11,181,885	285	A-B	20,483	2.4
	all_reads	C	11,811	12,095	285	A-B	20,408	2.4
	all_reads	C	54,998	55,237	240	A-B	15,070	1.8
	all_reads	0	4,860,505	4,860,744	240	A-B	14,819	1.8
	all_reads	0	14,537,374	14,537,514	141	R-	14,717	1.7
	all_reads	0	11,181,306	11,181,602	297	A-B	12,915	1.5
	all_reads	C	11,516	11,812	297	A-B	12,903	1.5
*Kpn*I-*Taq*$^{\alpha}$I	all_reads	3	7,387,133	7,387,506	374	A-B	67,098	7.8
	all_reads	0	24,795,040	24,795,226	187	A-B	43,230	5.0
	MQ20_filter	0	24,795,040	24,795,226	187	A-B	25,735	5.7
	all_reads	C	109,575	109,809	235	A-B	22,378	2.6
	all_reads	C	132,543	132,777	235	A-B	22,341	2.6
	all_reads	3	7,386,688	7,387,136	449	A-B	14,342	1.7
	all_reads	0	21,241,567	21,241,756	190	A-B	14,268	1.7
	all_reads	C	52,036	52,225	190	A-B	13,966	1.6
	all_reads	C	16,952	17,274	323	A-B	12,714	1.5
	MQ20_filter	C	16,952	17,274	323	A-B	12,696	2.8
*Xba*I-*Taq*$^{\alpha}$I-*Cvi*AII	all_reads	C	123,274	123,596	323	A-B	30,142	3.0
	MQ20_filter	C	123,274	123,596	323	A-B	29,928	8.2
	all_reads	C	87,109	87,318	210	A-B	28,260	2.8
	all_reads	0	10,043,857	10,044,066	210	A-B	28,036	2.8
	all_reads	C	155,034	155,243	210	A-B	26,294	2.6
	all_reads	C	76,298	76,556	259	A-B	22,155	2.2
	MQ20_filter	C	76,298	76,556	259	A-B	21,903	6.0
	all_reads	C	129,819	129,992	174	A-B	19,767	2.0
	MQ20_filter	C	129,819	129,992	174	A-B	19,619	5.4
	all_reads	0	2,854,464	2,854,662	199	A-B	13,868	1.4
	all_reads	C	76,553	76,751	199	A-B	13,819	1.4
	all_reads	0	2,083,605	2,083,836	232	A-B	13,450	1.4
	all_reads	C	123,593	123,824	232	A-B	12,926	1.3
*Xba*I-*Taq*$^{\alpha}$I-*Mse*I	all_reads	4	24,752,204	24,752,428	225	R-	49,406	5.6

doi:10.6342/NTU202500264

Enzyme combinations	PE read types	Chr	Start	End	Length	Frag. Types	Ave. PE reads	%
	MQ20_filter	4	24,752,204	24,752,428	225	R-	45,923	15.6
	all_reads	C	155,034	155,243	210	A-B	18,688	2.1
	all_reads	C	87,109	87,318	210	A-B	18,313	2.1
	all_reads	0	10,043,857	10,044,066	210	A-B	17,832	2.0
	all_reads	C	76,298	76,556	259	A-B	14,353	1.6
	MQ20_filter	C	76,298	76,556	259	A-B	14,183	4.8
	all_reads	C	129,819	129,992	174	A-B	12,731	1.4
	MQ20_filter	C	129,819	129,992	174	A-B	12,621	4.3
	all_reads	C	29,735	29,976	242	A-B	5,642	0.8
	all_reads	C	123,274	123,596	323	A-B	5,083	0.7
	MQ20_filter	C	123,274	123,596	323	A-B	5,023	2.6
	all_reads	C	76,298	76,556	259	A-B	4,690	0.7
*Xba*I-*Taq*$^{\alpha}$I-*Sph*I	MQ20_filter	C	76,298	76,556	259	A-B	4,620	2.4
	all_reads	C	79,160	79,397	238	A-B	3,831	0.5
	all_reads	0	13,754,632	13,754,869	238	A-B	3,780	0.5
	all_reads	0	10,043,857	10,044,066	210	A-B	2,744	0.4

PE read types indicate the PE reads for fragment analysis with or without the MAPQ filtering, denoted as all_reads and MQ20_filter.
Chr, Start, End, and Length indicate the fragment originated from the chromosome, the starting position, the end position, and its length.
Frag. Types indicate the fragment types.
Ave. PE reads indicate the average number of PE reads of the three samples.
"%" indicates the proportion of the PE reads represented by the fragment

doi:10.6342/NTU202500264

Supplementary Table 4. The sequence information of the TaqMan assays in this study.

Assay Name	Primer Sequence (5' - 3')	Probe Sequence (5' - 3')
2_382892_C-A	F: GCTCAACAGTCAACAACACTACT	VIC: TTTTAACTTTTCTAGTCTTTTTG
	R: GTTTCTCTTATGTGTCTTTGCTTGTAGATTATAAAT	FAM: AACTTTTCTAGTATTTTTG
2_621190_A-G	F: GGTTTACATGGAGAGTTCATTAATAGCTCTTT	VIC: ACTTCTGTTTGTTTACCTCATT
	R: GCTGCTATCACCTGTTTACAAGAAGTA	FAM: CTGTTTGTTTGCCTCATT
2_1603309_A-T	F: TTCTGGAAGCTTCGGTTCTGATAAG	VIC: AAGCTTATTTCCTGAAGCAT
	R: TCTCTGAGATAGCCATAGCCCAAT	FAM: AGCTTATTTCCAGAAGCAT
2_2351587_A-C	F: CTGTAGGCTTGTGGAGAAGATCA	VIC: TGCCCAATTGAATTGT
	R: TCAAATCCATAAATATCCAAAACTCCAATCAATG	FAM: CCCAATGGAATTGT
2_103153_C-A	F: CCCTTCTTCTCACCCTCCATCT	VIC: CTGACAGTTGGTTAATG
	R: TCGTGGAAATATATGGTAGGAGTAGTAAGG	FAM: CTGACAGTTTGTTAATG
2_627533_C-T	F: ATTTATCGTACTCTTTCCACTTCTTGTGAAT	VIC: AGAGCCTTAAGTTCATC
	R: CTAATTGAGAATGGATTCTAGCTCTAAACTTAGG	FAM: AGAGCCTTAAATTCATC
2_778333_T-A	F: CCCCAAGCTCTTACGAAAATGTCA	VIC: CATAGCCTACAACTCCTCCA
	R: GGGAGAATGATGCCGTCAA	FAM: TAGCCTACAACACCTCCA
2_816421_C-G	F: CCCACTTCCCATTCTCAAATCCAT	VIC: TTTGTTGGATTGGTGTTTT
	R: GCTGTTGGGTTTGTTGGGTTTTATT	FAM: TTGTTGGATTCGTGTTTT
2_1410652_T-C	F: TCCAATAATTTTTCTCTGGTATGCTCAGTT	VIC: CTGCAACACTCTTTG
	R: ATCCATCAATGCCAACCAGTTATCA	FAM: TGCAACGCTCTTTG
5_24520366_C-T	F: ATAACAAATATTCTTGAACTTTTCAACTTTCCA	VIC: AATAGTAACTCACGTTCCAAT
	R: TTTTTATAAAATTAAGTGTTTGATCAGGAAAGAAAGGAA	FAM: AATAGTAACTCACATTCCAAT
5_24941008_G-C	F: CCTGTTGTAGGCTTGTTCAATCTTG	VIC: TGAGTTGGATGAGATGAT
	R: CCAAACAAGCCCATTTTGCA	FAM: TGAGTTGGATCAGATGAT
5_25264851_C-T	F: AAAATCTTGGTTTTCTCCTAGTCACTTGT	VIC: CAAAAGATCTCGTGTAAGAA

doi:10.6342/NTU202500264

Assay Name	Primer Sequence (5' - 3')	Probe Sequence (5' - 3')
	R: CTCTTTTCTTCTTTCATCCTCCTAGTGA	FAM: ACAAAAGATCTCATGTAAGAA
5_25862218_T-C	F: CTCAACCCTCTGATTCTGATGAACA	VIC: ATCGTCGTCATCATCA
	R: TTCATTGTTTTGTGCATCTTCTTGGT	FAM: ATCGTCGTCGTCATCA
5_26105586_T-A	F: GGTAGTGTGTAATAAACCTAGGGACATC	VIC: TGGTTTGGTGGCTACTTT
	R: CCTGGGATCTTCCAATCAAAGTTGT	FAM: TGGTGGCAACTTT
12_21800139_A-G	F: TGATTCTTGGAGTTGTGGTAAAGCA	VIC: TGATGGCCGATAGGATT
	R: GTAATCTATATTACAAATCGGATACAACTTATTGAAATTCAAAT	FAM: ATGGCCGGTAGGATT
12_21116734_T-G	F: TGTTTGTTTTAGTTCTTGAGGTTTGGACTA	VIC: AACCTACCCAGTTCACAAG
	R: GAAAGCTCCCATACTCAGAAGAACA	FAM: CCTACCCAGGTCACAAG
12_22418198_A-G	F: GCTCAGAATCGGGTATTGTTGTTGA	VIC: AGCCATACATGATTATTT
	R: CGTGGAACAGTTAACATGCATATTTTGA	FAM: CCATACACGATTATTT
12_23222171_G-A	F: TGACAACATTTTGATCGCCAAAGAC	VIC: AGAAACTTGCGACTTCA
	R: CTCAACTTCAAACTTTGGCTTCCAT	FAM: AAACTTGCAACTTCA
12_23491653_A-G	F: ACCTTTCATAAATGATAGTCCGTGGTT	VIC: CATAGGTTGATTTTAAAAAC
	R: GCTACATGCTAGAATGCGATTGTCA	FAM: TAGGTTGATTCTAAAAAC

F and R indicate the forward and reverse primer; VIC and FAM indicate the two fluorescent dyes labeled on the minor groove-binding probes.

doi:10.6342/NTU202500264

Supplementary Table 5. Candidate genes listed for *qPM2*

Candidate gene ID	Description
MELO3C015296	26S proteasome regulatory subunit 6A homolog
MELO3C015297	translationally-controlled tumor protein homolog
MELO3C015298	RING-type domain-containing protein
MELO3C015299	Serine/arginine-rich splicing factor
MELO3C015300	purine permease 11
MELO3C015301	Phosphoacetylglucosamine mutase
MELO3C015302	HIG1 domain-containing protein
MELO3C015303	SOSS complex subunit B-like protein
MELO3C015304	60S ribosomal protein L22-2
MELO3C015305	ER membrane protein complex subunit 6
MELO3C015306	dormancy-associated protein homolog 3-like isoform X2
MELO3C015307	Tafazzin
MELO3C015308	T-box protein 41
MELO3C015310	REF/SRPP-like protein At3g05500
MELO3C015311	Heavy metal-associated isoprenylated plant protein 36
MELO3C015312	Respiratory burst oxidase, putative
MELO3C015313	Formin-like protein
MELO3C015314	protein SRC2-like
MELO3C015315	protein indeterminate-domain 12-like
MELO3C015316	Pentatricopeptide repeat-containing protein
MELO3C015317	S-adenosyl-L-methionine-dependent methyltransferase
MELO3C015319	Eukaryotic translation initiation factor 6
MELO3C015321	Beta-galactosidase
MELO3C015322	Protein AUXIN RESPONSE 4
MELO3C015324	Beta-mannosidase, putative
MELO3C015325	Protein YIP
MELO3C015326	NAD(P)-binding Rossmann-fold superfamily protein
MELO3C015327	isoaspartyl peptidase/L-asparaginase
MELO3C015328	amidase 1-like isoform X1
MELO3C015329	Amidase 1-like
MELO3C015330	Root phototropism protein 3-like isoform X1

doi:10.6342/NTU202500264

Candidate gene ID	Description
MELO3C015331	root phototropism protein 3-like isoform X1
MELO3C015332	Nuclear transcription factor Y subunit C-9
MELO3C015334	cation/calcium exchanger 5
MELO3C015335	Signal peptidase complex subunit 3
MELO3C015336	heavy metal-associated isoprenylated plant protein 34
MELO3C015337	NADPH-dependent pterin aldehyde reductase-like
MELO3C015338	Ribosome biogenesis protein BRX1-like
MELO3C015339	UDP-N-acetylglucosamine transferase subunit ALG13 homolog
MELO3C015340	Protein kinase domain-containing protein
MELO3C015341	BHLH domain-containing protein
MELO3C015342	L10-interacting MYB domain-containing protein-like isoform X4
MELO3C015343	39S ribosomal protein L45, mitochondrial isoform X1
MELO3C015344	Unknown protein
MELO3C015345	Glutathione synthetase
MELO3C015346	Protein of unknown function (DUF1068)
MELO3C015347	cell division cycle protein 123 homolog
MELO3C015348	Phosphoglycerate mutase-like protein AT74
MELO3C015350	Sugar transporter
MELO3C015351	Superoxide dismutase
MELO3C015352	Dihydroorotate dehydrogenase (quinone), mitochondrial
MELO3C015353	CNL-II [*Pm2.1* (Haonan et al. 2020)]
MELO3C015354	Disease resistance protein RGA2-like [*Pm2.1* (Haonan et al. 2020)]
MELO3C015355	NAC domain-containing protein
MELO3C015357	NAC domain-containing protein
MELO3C015358	Autophagy 18H like protein
MELO3C015359	histidine-containing phosphotransfer protein 1-like
MELO3C015360	transcription factor IIIB 60 kDa subunit
MELO3C015361	Pyrimidine-specific ribonucleoside hydrolase RihA
MELO3C015362	gibberellin receptor GID1C
MELO3C015363	Histone H2A
MELO3C015364	F-box domain-containing protein

doi:10.6342/NTU202500264

Candidate gene ID	Description
MELO3C015365	LOW QUALITY PROTEIN: uncharacterized protein LOC110415317
MELO3C015366	J domain-containing protein
MELO3C015367	Non-specific serine/threonine protein kinase
MELO3C015368	trafficking protein particle complex subunit 6B
MELO3C015369	vesicle-associated protein 1-3-like
MELO3C015371	vesicle-associated protein 1-3-like
MELO3C015372	transmembrane protein adipocyte-associated 1
MELO3C015373	40S ribosomal protein S25
MELO3C015374	Superoxide dismutase [Cu-Zn]
MELO3C015375	Sodium/hydrogen exchanger
MELO3C015376	Protein kinase domain-containing protein
MELO3C015377	Light-inducible protein CPRF2-like
MELO3C015378	Sec-independent protein translocase protein TATA
MELO3C015379	transcription factor UNE12-like
MELO3C015382	RNA cytidine acetyltransferase
MELO3C015383	kinesin-like protein KIFC3
MELO3C015384	protein translocase subunit SecA
MELO3C015385	Asparagine--tRNA ligase
MELO3C015386	N-acetyltransferase domain-containing protein
MELO3C015387	Unknown protein
MELO3C015388	cytochrome P450 CYP72A219-like
MELO3C015389	Protein kinase domain-containing protein
MELO3C015391	protein MON2 homolog isoform X1
MELO3C015395	late embryogenesis abundant protein 47
MELO3C015396	Cysteine synthase
MELO3C015397	Unknown protein
MELO3C015398	Cysteine synthase
MELO3C015399	Cysteine synthase
MELO3C015400	DNA-binding storekeeper protein-related transcriptional regulator
MELO3C015402	guanine deaminase
MELO3C015403	DUF761 domain-containing protein

doi:10.6342/NTU202500264

Candidate gene ID	Description
MELO3C015406	RNA-dependent RNA polymerase
MELO3C015407	Malic enzyme
MELO3C015408	Protein-serine/threonine phosphatase
MELO3C015409	ABC transporter
MELO3C015411	Ceramide glucosyltransferase
MELO3C015412	SpoU_sub_bind domain-containing protein
MELO3C015413	mitochondrial import inner membrane translocase subunit TIM17-2-like
MELO3C015414	receptor-like protein kinase HSL1
MELO3C015415	receptor-like protein kinase 5
MELO3C015416	Phosphotransferase
MELO3C015417	Protein DETOXIFICATION
MELO3C015418	protein IQ-DOMAIN 14-like
MELO3C015419	Pectinesterase
MELO3C015420	Non-specific serine/threonine protein kinase
MELO3C015421	Programmed cell death protein 4-like
MELO3C015422	Arginyl-tRNA synthetase
MELO3C015423	Oxidative stress 3, putative isoform 1
MELO3C015424	(1->3)-beta-glucan endohydrolase
MELO3C015425	Derlin
MELO3C015426	Leucine-rich repeat family protein / extensin family protein
MELO3C015427	NAC domain-containing protein
MELO3C015428	ATP-dependent 6-phosphofructokinase
MELO3C029287	Cysteine synthase
MELO3C029290	Nudix hydrolase
MELO3C029293	Retrovirus-related Pol polyprotein from transposon TNT 1-94
MELO3C029294	Unknown protein
MELO3C029300	Pentatricopeptide repeat-containing protein
MELO3C029302	Sugar transporter ERD6-like 5 isoform X5
MELO3C029303	Unknown protein
MELO3C029307	Defensin-like family protein, putative
MELO3C029311	Pentatricopeptide repeat-containing protein DOT4

doi:10.6342/NTU202500264

Candidate gene ID	Description
MELO3C029313	Translation initiation factor 3 subunit I
MELO3C029314	Unknown protein
MELO3C029316	Unknown protein
MELO3C030402	Disease resistance protein RGA2-like

The [ ] indicates candidate gene ID were shown in previous studies.

doi:10.6342/NTU202500264

Supplementary Table 6. Candidate genes listed for *qPM5*

Candidate gene ID	Description
MELO3C000062	LRR and NB-ARC domains-containing disease resistance protein
MELO3C000887	Kunitz-type trypsin inhibitor
MELO3C001812	Terpene cyclase/mutase family member
MELO3C004274	peroxisomal adenine nucleotide carrier 1-like
MELO3C004275	Nucleolar GTPase
MELO3C004276	Glycoprotein membrane precursor GPI-anchored
MELO3C004277	GPI-anchored protein-like protein
MELO3C004278	Mevalonate kinase
MELO3C004279	RING-type domain-containing protein
MELO3C004280	Transmembrane protein
MELO3C004281	Mevalonate kinase
MELO3C004283	alpha-crystallin domain-containing protein 22.3 isoform X2
MELO3C004284	increased DNA methylation 2
MELO3C004285	Signal peptidase complex subunit 3
MELO3C004286	RING-type domain-containing protein
MELO3C004287	transcription factor bHLH94-like
MELO3C004288	TMV resistance protein N-like isoform X1
MELO3C004289	TMV resistance protein N-like
MELO3C004291	LOW QUALITY PROTEIN: TMV resistance protein N-like
MELO3C004292	TMV resistance protein N-like isoform X1
MELO3C004294	TMV resistance protein N-like isoform X1
MELO3C004296	kunitz trypsin inhibitor 1-like
MELO3C004297	Branched-chain amino acid aminotransferase [López-Martín et al. (2022)]
MELO3C004298	Unknown protein
MELO3C004299	GDSL esterase/lipase
MELO3C004300	GDSL esterase/lipase
MELO3C004301	TMV resistance protein N-like isoform X1
MELO3C004302	TMV resistance protein N-like isoform X1
MELO3C004303	TIR-NBS-LRR disease resistance protein
MELO3C004305	Pre-mRNA-splicing factor SLU7

doi:10.6342/NTU202500264

Candidate gene ID	Description
MELO3C004306	GDSL esterase/lipase
MELO3C004307	UPF0603 protein At1g54780, chloroplastic-like
MELO3C004308	Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP family protein
MELO3C004309	TMV resistance protein N-like
MELO3C004311	TMV resistance protein N-like isoform X2 [López-Martín et al. (2022)]
MELO3C004313	TMV resistance protein N-like
MELO3C004314	protein FANTASTIC FOUR 1
MELO3C004315	receptor-like cytosolic serine/threonine-protein kinase RBK2
MELO3C004316	polyamine-modulated factor 1-binding protein 1
MELO3C004317	NB-ARC domain-containing disease resistance protein
MELO3C004318	NB-ARC domain-containing protein
MELO3C004319	NB-ARC domain-containing disease resistance protein
MELO3C004320	NB-ARC domain-containing disease resistance protein
MELO3C004321	NB-ARC domain-containing disease resistance protein
MELO3C004323	LRR and NB-ARC domains-containing disease resistance protein
MELO3C004329	Terpene cyclase/mutase family member
MELO3C004332	Unknown protein
MELO3C004333	DCD domain-containing protein
MELO3C004334	ERAD-associated E3 ubiquitin-protein ligase HRD1B-like
MELO3C004335	50S ribosomal protein L18
MELO3C004336	phosphate deficiency response 2
MELO3C004337	vesicle-associated protein 2-2 isoform X1
MELO3C004338	Acyl carrier protein
MELO3C004339	coiled-coil domain-containing protein SCD2 isoform X1
MELO3C004340	Ervatamin-B-like
MELO3C004341	Pentatricopeptide repeat-containing protein
MELO3C004342	Superoxide dismutase [Cu-Zn]
MELO3C004343	ATP-dependent RNA helicase, putative
MELO3C004345	eIF-2B GDP-GTP exchange factor subunit epsilon
MELO3C004347	protein HASTY 1

doi:10.6342/NTU202500264

Candidate gene ID	Description
MELO3C004349	Protein kinase domain-containing protein
MELO3C004350	Nucleolar protein 58
MELO3C004351	Nodulation-signaling pathway protein isoform 2
MELO3C004352	acyltransferase-like protein At1g54570, chloroplastic
MELO3C004353	WD repeat-containing protein 48
MELO3C004354	Cc-nbs-lrr resistance protein
MELO3C004355	Cysteine synthase
MELO3C004356	5'-3' exoribonuclease
MELO3C004358	Rubredoxin
MELO3C004359	DNA-binding storekeeper protein-related transcriptional regulator
MELO3C020729	Cysteine synthase
MELO3C027376	TMV resistance protein N-like
MELO3C027385	NB-ARC domain-containing disease resistance protein
MELO3C027425	TMV resistance protein N-like
MELO3C027615	Vat protein
MELO3C028480	CACTA en-spm transposon protein
MELO3C031324	NB-ARC domain-containing protein
MELO3C031325	NB-ARC domain-containing disease resistance protein
MELO3C031326	NB-ARC domain-containing protein
MELO3C031327	DUF4283 domain-containing protein
MELO3C031329	Unknown protein
MELO3C031330	NB-ARC domain-containing protein
MELO3C031332	NB-ARC domain-containing protein
MELO3C031544	Selenium binding protein-like protein
MELO3C031545	Pentatricopeptide repeat (PPR) superfamily protein
MELO3C031546	Pentatricopeptide repeat (PPR) superfamily protein
MELO3C031547	Unknown protein
MELO3C031548	Mevalonate kinase-like
MELO3C031550	Unknown protein
MELO3C031551	LINE-1 retrotransposable element ORF2 protein
MELO3C031552	Disease resistance protein RPP4-like
MELO3C031553	ULP_PROTEASE domain-containing protein

doi:10.6342/NTU202500264

Candidate gene ID	Description
MELO3C031554	NB-ARC domain-containing protein
MELO3C031555	Unknown protein
MELO3C031556	NB-ARC domain-containing protein
MELO3C031557	Unknown protein
MELO3C031558	Unknown protein
MELO3C031799	Pre-mRNA-processing factor 39 isoform X3

The [ ] indicates candidate gene ID were shown in previous studies.

doi:10.6342/NTU202500264

Supplementary Table 7. Candidate genes listed for *qPM12*

Candidate gene ID	Description
MELO3C002388	calnexin homolog
MELO3C002389	LRR receptor-like serine/threonine-protein kinase GSO1
MELO3C002390	BEST Arabidopsis thaliana protein match is: root hair specific 4 .
MELO3C002391	Peroxidase
MELO3C002392	LRRNT_2 domain-containing protein [*qPx1-12* (Branham et al. 2021)]
MELO3C002393	leucine-rich repeat receptor protein kinase EMS1-like [*qPx1-12* (Branham et al. 2021)]
MELO3C002394	LRRNT_2 domain-containing protein
MELO3C002396	protein PLASTID TRANSCRIPTIONALLY ACTIVE 14 isoform X1
MELO3C002397	28S ribosomal protein S33, mitochondrial
MELO3C002398	Cationic amino acid transporter like
MELO3C002399	GDSL esterase/lipase
MELO3C002400	Glycosyltransferase
MELO3C002401	Protein UNUSUAL FLORAL ORGANS-like
MELO3C002402	C-type lectin receptor-like tyrosine-protein kinase
MELO3C002403	allantoate deiminase isoform X1 [*Cmpmr2F* (Zhang et al. 2023)]
MELO3C002404	PLATZ transcription factor family protein
MELO3C002405	50S ribosomal protein L7/L12
MELO3C002406	AICAR transformylase
MELO3C002407	Protein kinase domain-containing protein
MELO3C002408	Potassium transporter
MELO3C002409	Potassium transporter
MELO3C002411	Polyadenylate-binding protein 1-B-binding protein
MELO3C002412	Ankyrin repeat-containing protein
MELO3C002413	Ankyrin repeat-containing protein
MELO3C002414	Ankyrin repeat-containing protein
MELO3C002416	Ankyrin repeat-containing protein
MELO3C002417	Ankyrin repeat-containing protein
MELO3C002418	Ankyrin repeat-containing protein

doi:10.6342/NTU202500264

Candidate gene ID	Description
MELO3C002419	Dirigent protein
MELO3C002420	Retrovirus-related Pol polyprotein from transposon RE2
MELO3C002421	alpha/beta-Hydrolases superfamily protein
MELO3C002422	stigma-specific STIG1-like protein 1
MELO3C002423	stigma-specific STIG1-like protein 1
MELO3C002424	Ankyrin repeat-containing protein
MELO3C002425	Ankyrin repeat-containing protein
MELO3C002426	Ankyrin repeat-containing protein
MELO3C002428	Early nodulin-20-like
MELO3C002429	Ankyrin repeat-containing protein
MELO3C002430	GUB_WAK_bind domain-containing protein
MELO3C002431	Pectate lyase
MELO3C002433	stigma-specific STIG1-like protein 1
MELO3C002434	Ankyrin repeat-containing protein [*qPx1-12* (Branham et al. 2021)]
MELO3C002435	Retrovirus-related Pol polyprotein from transposon RE1
MELO3C002436	p55
MELO3C002437	p55 [*BPm12.1* (Li et al. 2017)]
MELO3C002438	Ankyrin repeat-containing protein [*pm12.1* (Haonan et al. 2020)]
MELO3C002439	Ankyrin repeat-containing protein [*BPm12.1* (Li et al. 2017)] and *pm12.1* (Haonan et al. 2020)]
MELO3C002441	Ankyrin repeat-containing protein [*BPm12.1* (Li et al. 2017)], *pm12.1* (Haonan et al. 2020), and *CmPMRl* (Cui et al. 2022)]
MELO3C002442	Aspartic proteinase [*pm12.1* (Haonan et al. 2020), and *CmPMRl* (Cui et al. 2022)]
MELO3C002443	F-box family protein with a domain of Uncharacterized protein function, putative [*BPm12.1* (Li et al. 2017)], and *CmPMRl* (Cui et al. 2022)]
MELO3C002444	Aminomethyltransferase [*CmPMRl* (Cui et al. 2022)]
MELO3C002445	glycine-rich cell wall structural protein 1-like [*BPm12.1* (Li et al. 2017)] and *CmPMRl* (Cui et al. 2022)]
MELO3C002446	glycine-rich protein 5-like [*CmPMRl* (Cui et al. 2022)]

doi:10.6342/NTU202500264

Candidate gene ID	Description
MELO3C002447	L-ascorbate oxidase homolog [*CmPMRl* (Cui et al. 2022)]
MELO3C002448	tyrosyl-DNA phosphodiesterase 2 [*CmPMRl* (Cui et al. 2022)]
MELO3C002449	Glucan endo-1,3-beta-glucosidase 1 [*CmPMRl* (Cui et al. 2022)]
MELO3C002450	5'-adenylylsulfate reductase 1
MELO3C002451	Glyoxalase I
MELO3C002452	Autophagy-related protein
MELO3C002453	Hexosyltransferase
MELO3C002454	Protein of unknown function, DUF584
MELO3C002455	Starch synthase 3
MELO3C002456	ATP synthase gamma chain
MELO3C002457	Peroxidase
MELO3C002458	Homer protein isoform 2
MELO3C002459	Protein of unknown function, DUF584
MELO3C002460	DUF4228 domain-containing protein
MELO3C002461	SPARK domain-containing protein
MELO3C002462	Unknown protein
MELO3C002463	UPF0051 protein ABCI8
MELO3C002464	polyadenylate-binding protein RBP45-like
MELO3C002465	15.4 kDa class V heat shock protein
MELO3C002466	Abnormal spindle-like microcephaly-associated protein-like protein
MELO3C002468	tobamovirus multiplication protein 1
MELO3C002469	Patatin
MELO3C002470	Patatin
MELO3C002471	RNA pseudouridine synthase 6, chloroplastic-like
MELO3C002473	RNA pseudouridine synthase 6, chloroplastic-like
MELO3C002474	beta-glucosidase 47
MELO3C002475	Tryptophan synthase beta chain 1
MELO3C002476	Tryptophan synthase
MELO3C002477	homeobox-leucine zipper protein PROTODERMAL FACTOR 2
MELO3C002478	CRIB domain-containing protein
MELO3C002479	nuclear transport factor 2

doi:10.6342/NTU202500264

Candidate gene ID	Description
MELO3C002480	Xyloglucan endotransglucosylase/hydrolase
MELO3C002482	chaperone protein dnaJ 50
MELO3C002483	gibberellin 2-beta-dioxygenase 8
MELO3C002484	pyruvate, phosphate dikinase regulatory protein 1, chloroplastic
MELO3C002485	Unknown protein
MELO3C002486	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
MELO3C002488	Cysteine-rich receptor-like protein kinase
MELO3C002489	cysteine-rich receptor-like protein kinase 10
MELO3C002491	Cysteine-rich receptor-like protein kinase
MELO3C002492	cysteine-rich receptor-like protein kinase 25
MELO3C002493	Aminoacyl-tRNA ligase
MELO3C002495	Cysteine-rich receptor-like protein kinase 29
MELO3C002496	cysteine-rich receptor-like protein kinase 29
MELO3C002499	cysteine-rich repeat secretory protein 38-like
MELO3C002500	Sulfite exporter TauE/SafE family protein
MELO3C002501	cysteine-rich receptor-like protein kinase 28
MELO3C002504	cysteine-rich receptor-like protein kinase 28 [López-Martín et al. (2022)]
MELO3C002506	Cysteine-rich receptor-like protein kinase 28
MELO3C002507	transcription factor ILR3-like
MELO3C002508	thioredoxin-like protein CXXS1
MELO3C002509	Kinesin-like protein
MELO3C002510	oxygen-evolving enhancer protein 3-2, chloroplastic
MELO3C002511	Protein of unknown function (DUF1218)
MELO3C035396	Floral homeotic protein DEFICIENS-like
MELO3C035529	polyadenylate-binding protein RBP45-like
MELO3C035729	Cysteine-rich receptor-kinase-like protein
MELO3C035730	Cysteine-rich receptor-like protein kinase 10
MELO3C035731	Pentatricopeptide repeat-containing protein
MELO3C035732	ankyrin repeat-containing protein At3g12360-like isoform X2
MELO3C035733	ANK_REP_REGION domain-containing protein

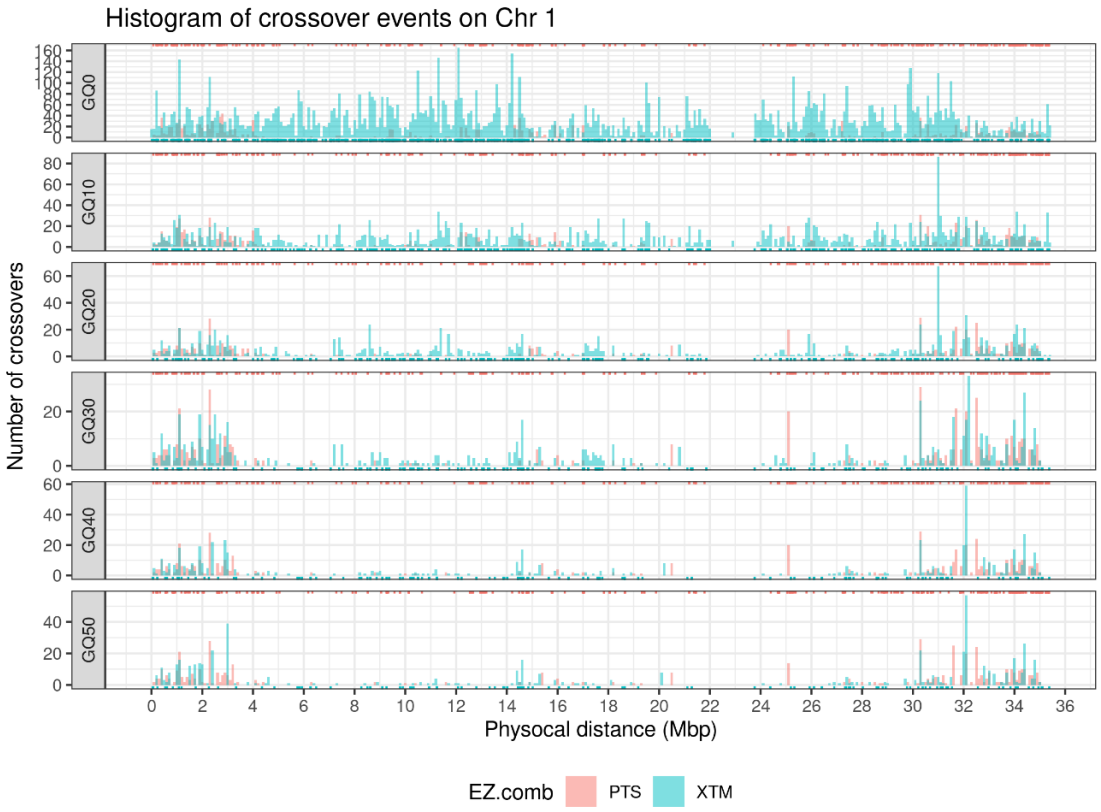
doi:10.6342/NTU202500264

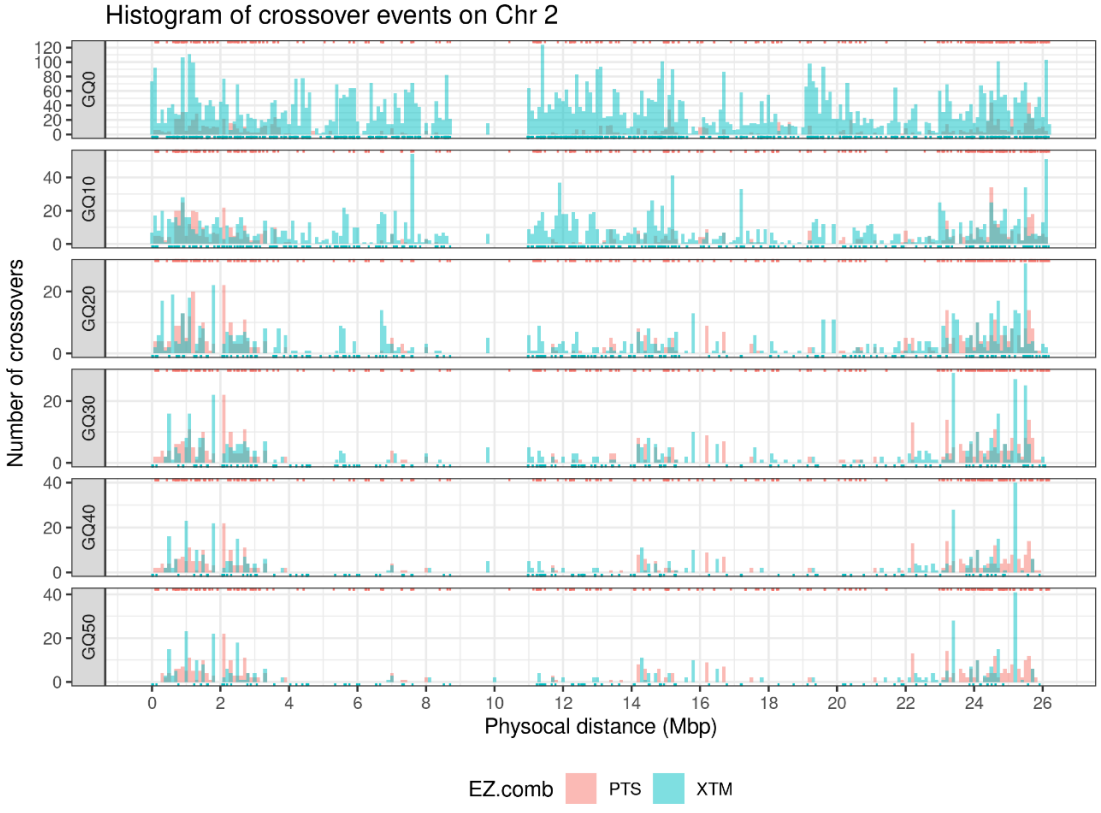
Candidate gene ID	Description
MELO3C035734	Ankyrin repeat-containing protein
MELO3C035735	Protein FAR1-RELATED SEQUENCE

The [ ] indicates candidate gene ID were shown in previous studies.

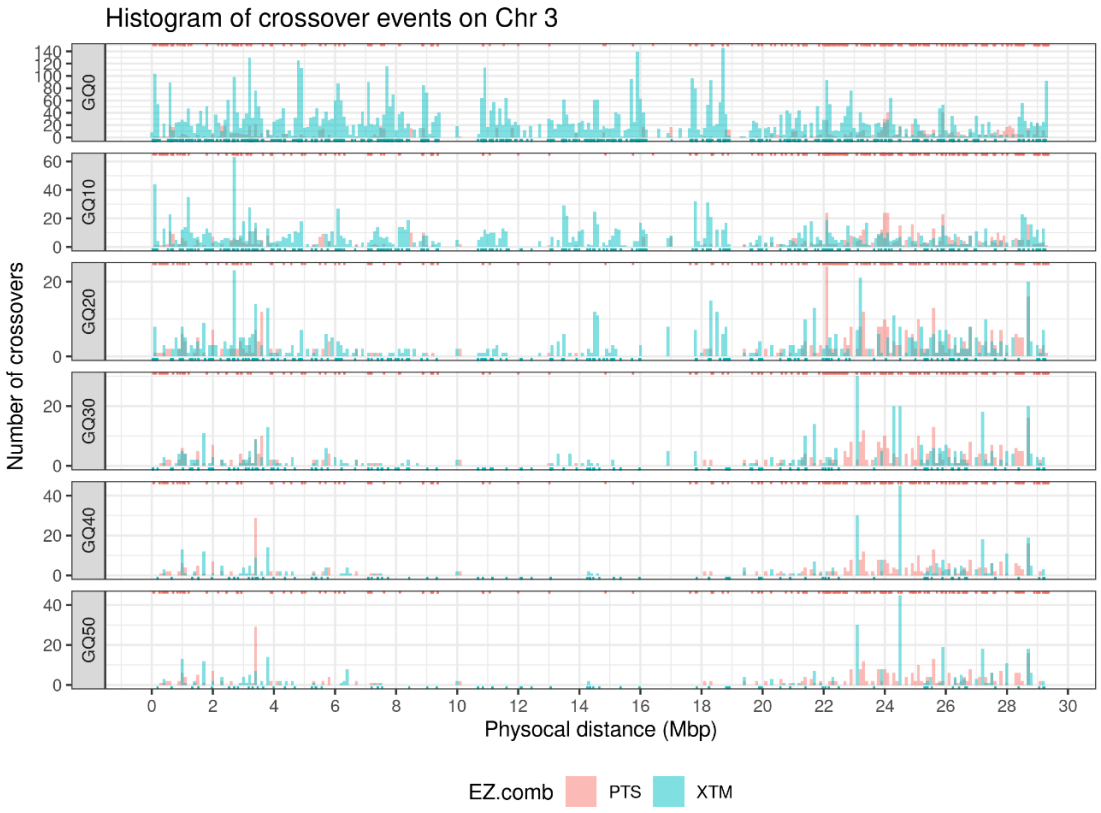
doi:10.6342/NTU202500264

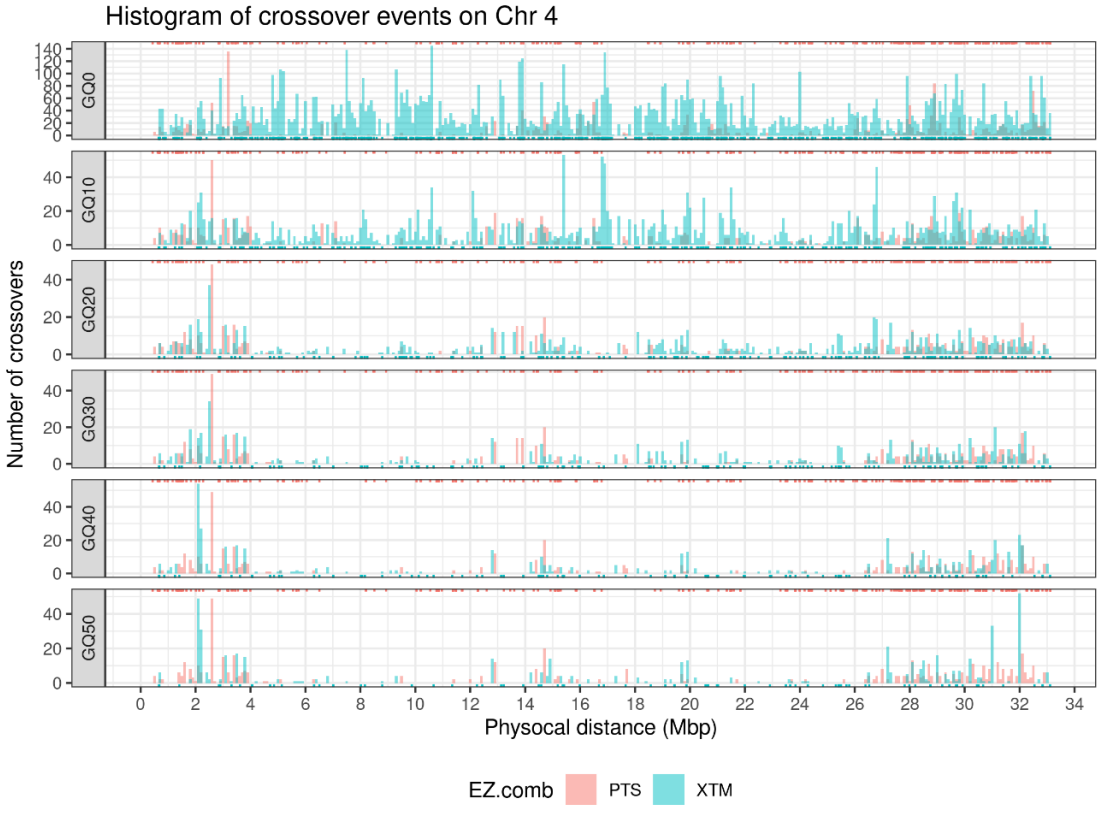
Supplementary Figures



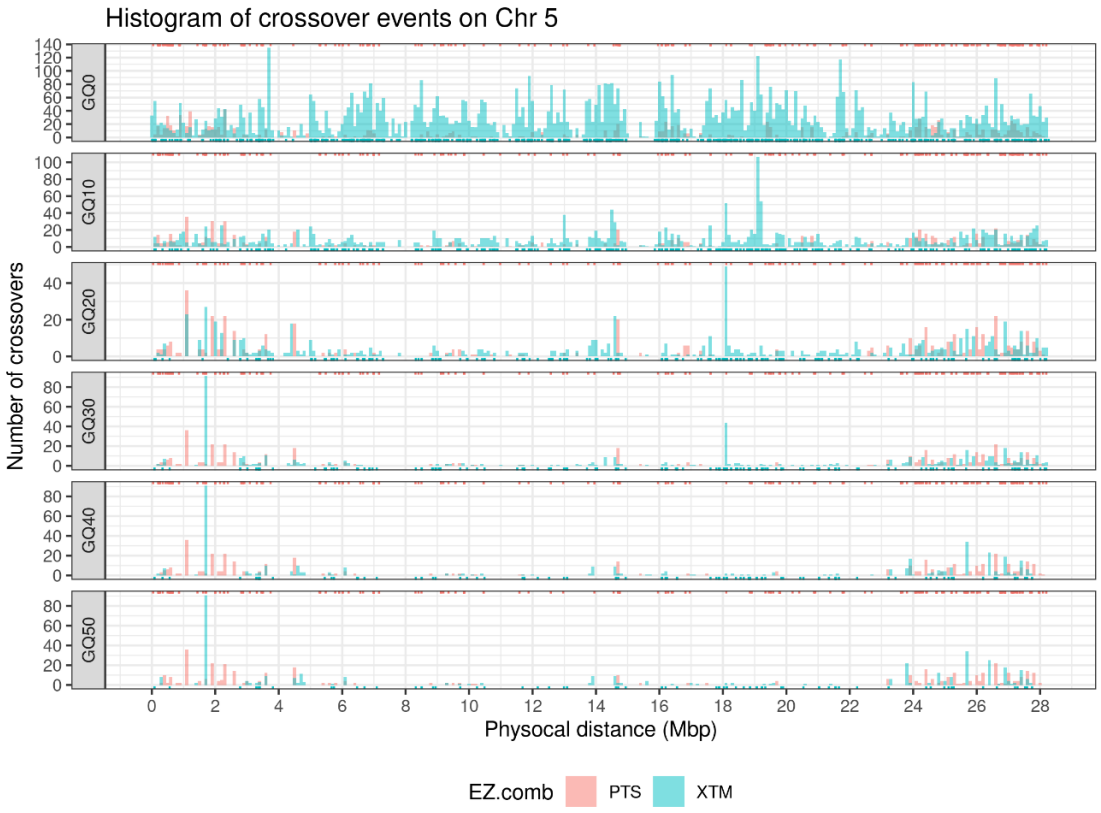


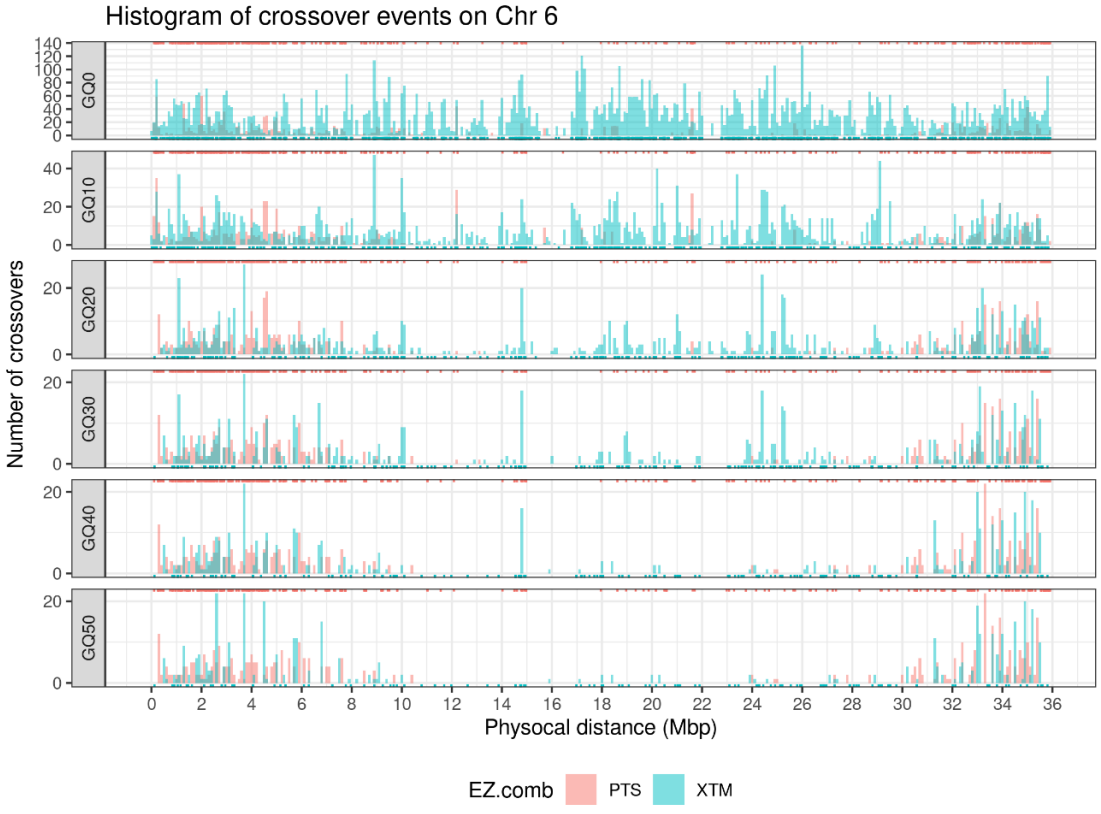
doi:10.6342/NTU202500264



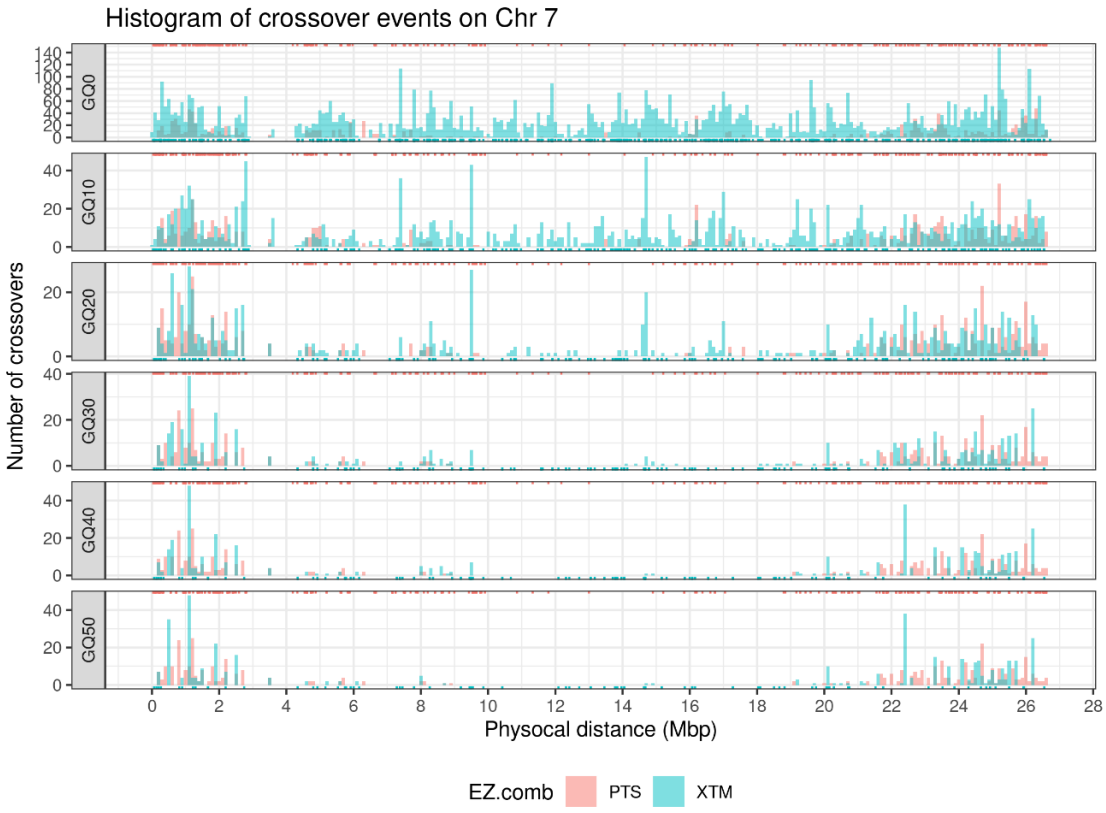


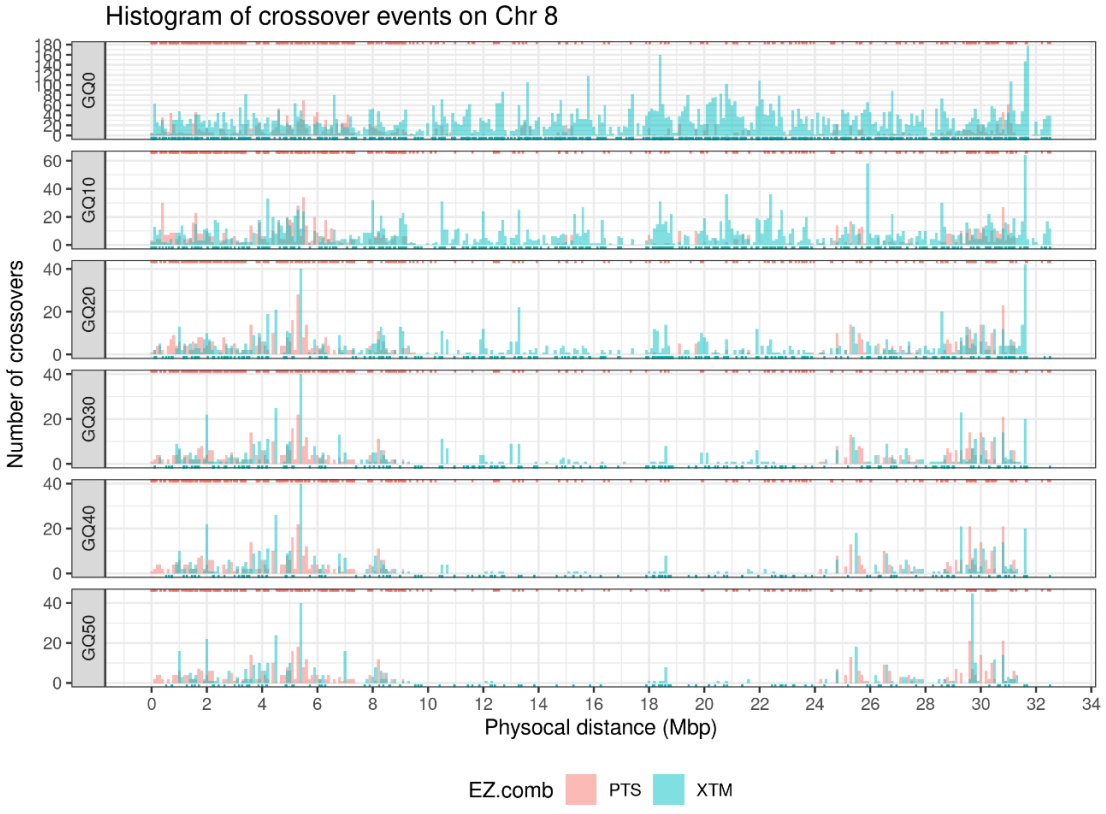
doi:10.6342/NTU202500264



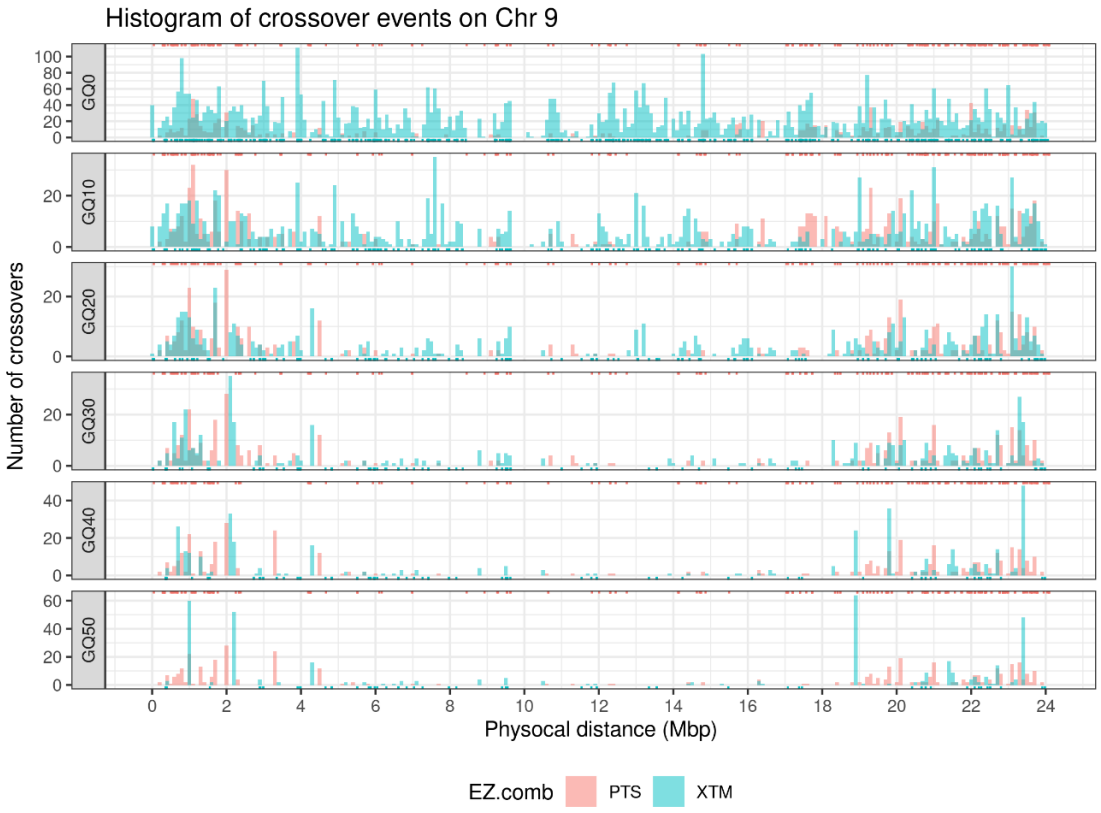


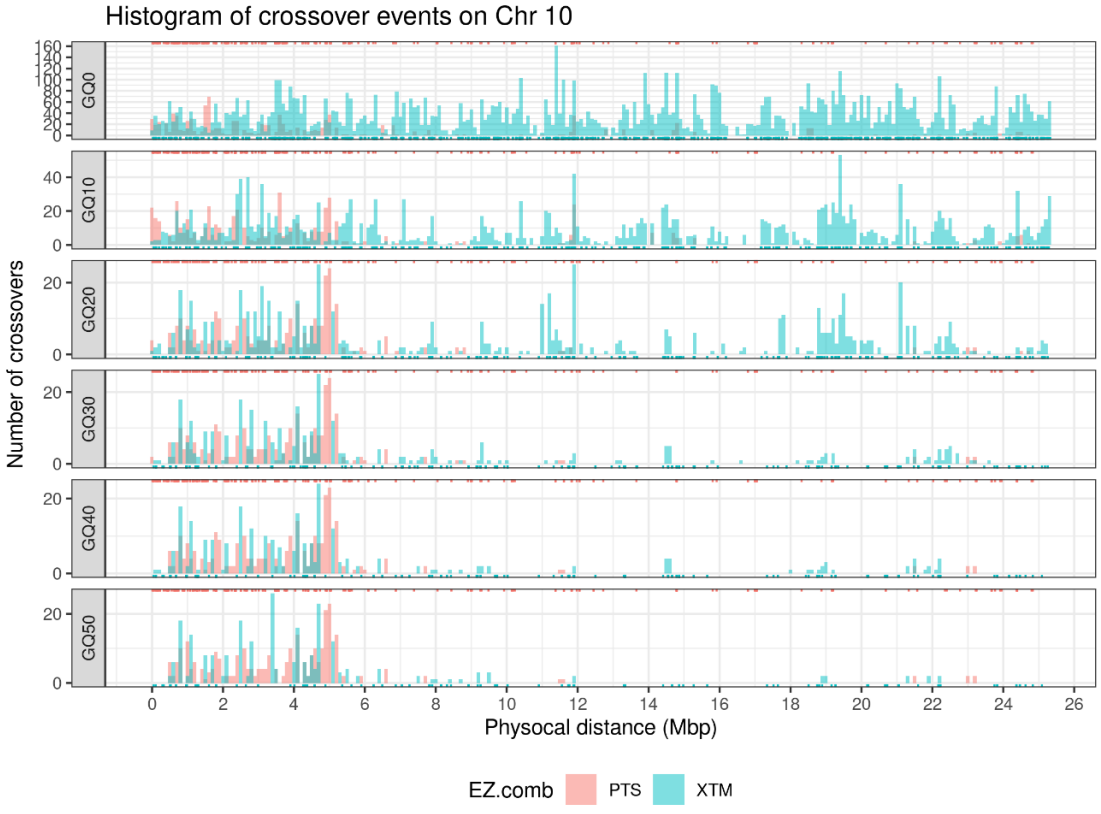
doi:10.6342/NTU202500264



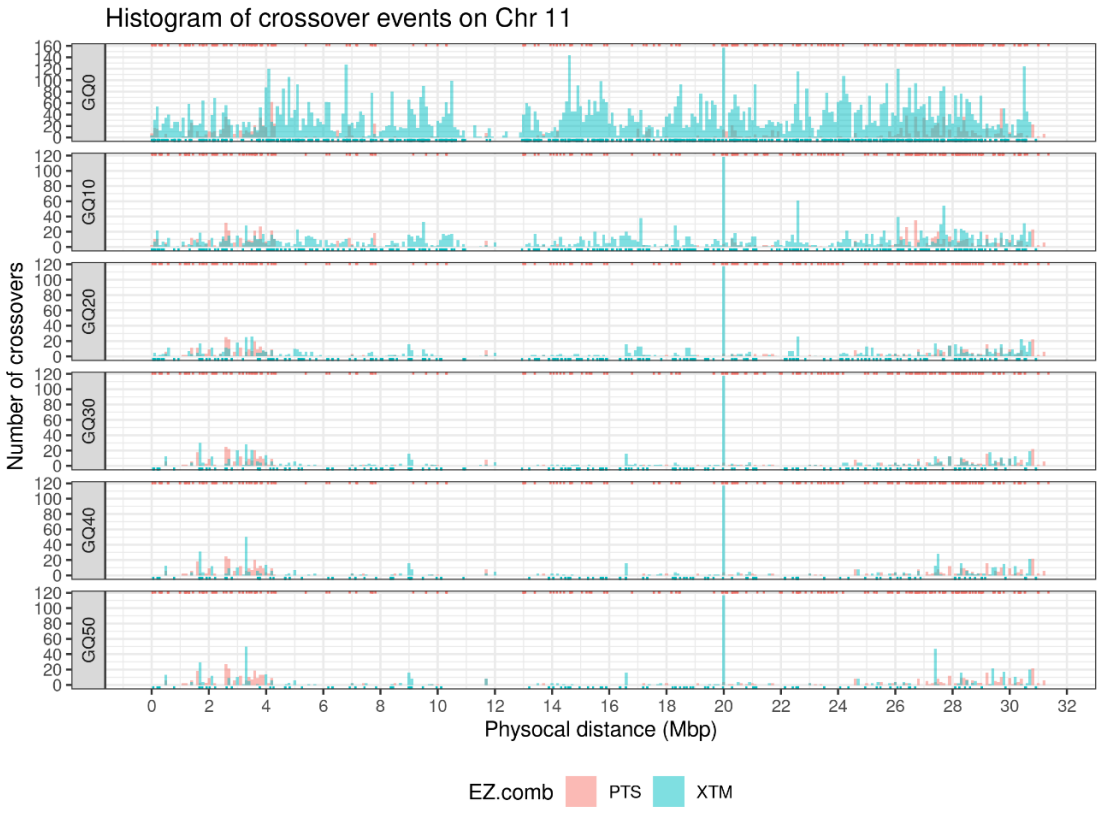


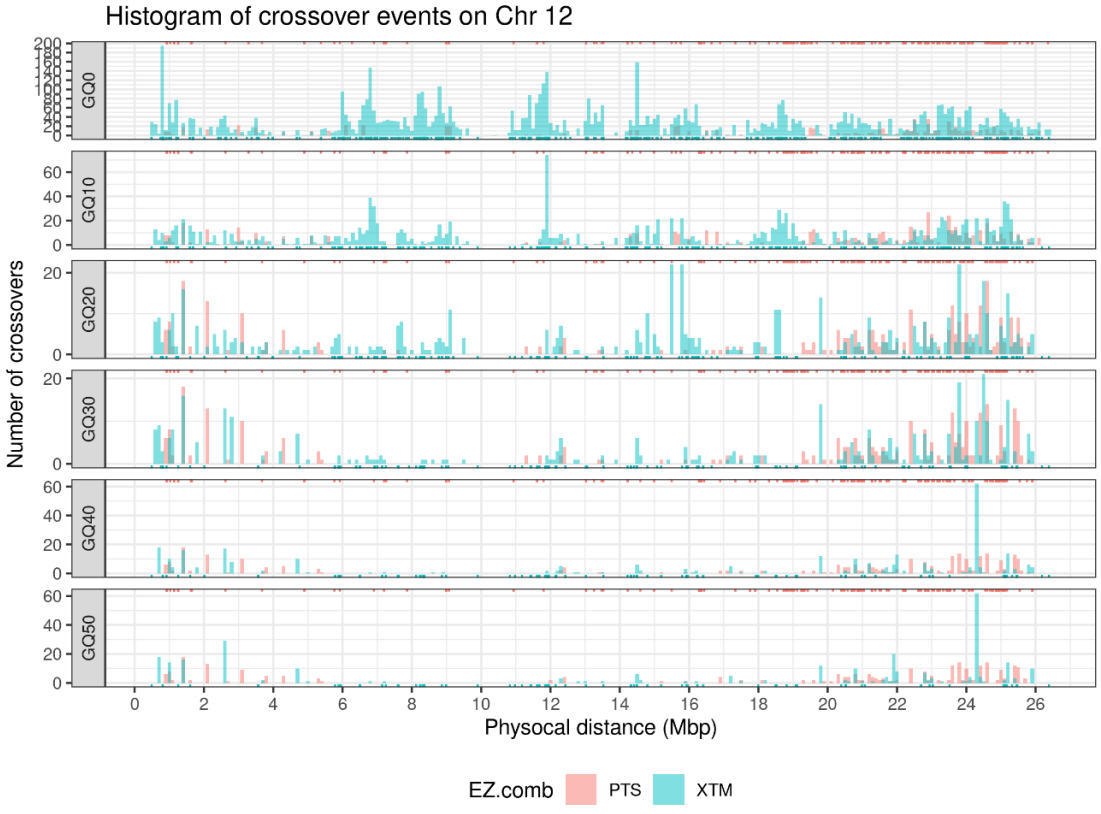
doi:10.6342/NTU202500264





doi:10.6342/NTU202500264





doi:10.6342/NTU202500264