



OPEN

Construction of a dense genetic map of the *Malus fusca* fire blight resistant accession MAL0045 using tunable genotyping-by-sequencing SNPs and microsatellites

Ofere Francis Emeriewen^{1✉}, Klaus Richter², Thomas Berner³, Jens Keilwagen³, Patrick S. Schnable^{4,5}, Mickael Malnoy⁶ & Andreas Peil^{1✉}

Although, the Pacific crabapple, *Malus fusca*, is a hardy and disease resistant species, studies relating to the genetics of its unique traits are very limited partly due to the lack of a genetic map of this interesting wild apple. An accession of *M. fusca* (MAL0045) of Julius Kühn-Institut collection in Germany is highly resistant to fire blight disease, incited by different strains of the causative pathogen—*Erwinia amylovora*. This is the most destructive bacterial disease of *Malus* of which most of the domesticated apples (*Malus domestica*) are susceptible. Using a scarcely dense genetic map derived from a population of 134 individuals of MAL0045 × 'Idared', the locus (*Mfu10*) controlling fire blight resistance mapped on linkage group 10 (LG10) and explained up to 66% of the phenotypic variance with different strains. Although the development of robust and tightly linked molecular markers on LG10 through chromosome walking approach led to the identification of a major candidate gene, any minor effect locus remained elusive possibly due to the lack of marker density of the entire genetic map. Therefore, we have developed a dense genetic map of *M. fusca* using tunable genotyping-by-sequencing (tGBS) approach. Of thousands of de novo SNPs identified, 2677 were informative in *M. fusca* and 90.5% of these successfully mapped. In addition, integration of SNP data and microsatellite (SSR) data resulted in a final map comprising 17 LGs with 613 loci spanning 1081.35 centi Morgan (cM). This map will serve as a template for mapping using different strains of the pathogen.

The genus *Malus* Mill. is classed under the Rosaceae family of flowering plants and comprise wild apples and their domesticated relative. The wild apple species *Malus fusca* (Raf.) C.K. Schneid., also called Pacific or Oregon crabapple, is a small deciduous tree that naturally occurs on mesic habitats along the pacific coast of North America, specifically from the northern part of California to the Kenai Peninsula in Alaska¹. Based on AFLP analyses² and nuclear ribosomal and chloroplast DNA³, the Pacific crabapple was categorized within the group of species that are native to central Asia. *M. fusca* is a very hardy and disease resistant species, which tolerates moist soils better than other crabapples and severe frosts up to -57°C ⁴.

¹Julius Kühn Institute (JKI) – Federal Research Centre for Cultivated Plants, Institute for Breeding Research on Fruit Crops, Pillnitzer Platz 3a, 01326 Dresden, Germany. ²Julius Kühn Institute (JKI) – Federal Research Centre for Cultivated Plants, Institute for Resistance Research and Stress Tolerance, Erwin-Baur-Str. 27, 06484 Quedlinburg, Germany. ³Julius Kühn Institute (JKI) – Federal Research Centre for Cultivated Plants, Institute for Biosafety in Plant Biotechnology, Erwin-Baur-Str. 27, 06484 Quedlinburg, Germany. ⁴Data2Bio LLC, Ames, IA 50011-3650, USA. ⁵Plant Sciences Institute, Iowa State University, 2035B Carver, Ames, IA 50011-3650, USA. ⁶Research and Innovation Centre, Genomics and Biology of Fruit Crops Department, Fondazione Edmund Mach, Via E. Mach, 1, 38010 San Michele all'Adige (Trentino), Italy. ✉email: ofere.emeriewen@julius-kuehn.de; andreas.peil@julius-kuehn.de

The domesticated apple (*Malus domestica* Borkh.), is an important fruit crop consumed worldwide and one of the most studied members of the Rosaceae family. It is believed that the progenitors of apple are the wild apple of Central Asia and its close relatives, with Kazakhstan being a critical location in apple domestication^{5,6}. Seed dispersal from Central Asia through to West Europe following human migration through trade routes allowed for hybridization and introgression with wild apple species such as the Siberian crabapple (*M. baccata*), Caucasian crabapple (*M. orientalis*) and the European crabapple (*M. sylvestris*)⁷, important factors leading to the domestication of apple. Thus, the domesticated apple is a complex hybrid with genetic and phenotypic variations across cultivars⁸. Apple consists of $x = 17$ chromosomes, and although most apple genotypes are diploid ($2n = 34$), other ploidy levels have been identified in *Malus*⁹. The use of structural and functional genomics vis-à-vis mapping of genetic markers and genes to specific chromosomes, as well as the sequencing of the apple genomes has increased our understanding of the domesticated apple and identified *M. sieversii* as the main progenitor of *M. domestica*^{10,11}. An approximate sequence size of 643 Mb was reported for the reference cultivar ‘Golden Delicious’¹¹ and around 651 Mb for the doubled haploid genotype GDDH13¹⁰.

Since the first sets of apple genetic linkage maps based on isozymes, RAPDs (Randomly Amplified Polymorphic DNA) and AFLPs (Amplified Fragment Length Polymorphism) were published^{12–14}, many other linkage maps have now been established from bi-parental crosses of cultivars, crabapples and rootstocks, and for several traits, using robust and reproducible DNA marker technologies such as simple sequence repeat (SSR or microsatellites) markers and single nucleotide polymorphism (SNP) markers (e.g.^{15–22}). Genetic linkage maps of apple are accessible at <https://www.rosaceae.org>.

The apple genome has an abundance of highly polymorphic and co-dominant microsatellites^{18,23}. This PCR-based marker approach is suitable for genotyping of large populations due to their easy transferability and ability for multiplexing several in a single PCR reaction. Numerous SSRs were developed and mapped and form the core of published genetic maps in apple^{15,17,19–21,23}. However, SSRs require huge investment of time for their development and mapping²⁰. In the last decade, a generic, hybridization-based fingerprinting method culminated in the development of DARt (Diversity Arrays Technology) SNP markers²⁴. DARts, first applied in rice, were applied successfully in other plant species including members of the Rosaceae^{25–28}.

Genome-wide SNP identification in *Malus* has been achieved with whole genome genotyping (WGG) array. Thus, thousands of SNP markers have been developed following re-sequencing data derived from 27 apple cultivars from the International RosBREED SNP Consortium (IRSC)²⁹ as well as from the genomes of 13 apple cultivars and a crabapple species (*M. micromalus*)³⁰. An 480 k Axiom apple array was developed after high-depth resequencing of 63 cultivars representing most of the genetic diversity in cultivated apple³¹. Furthermore, a high-density SNP integrated genetic linkage map (iGLMap) was developed in apple and includes a large number of markers, families and individuals³². Although in recent times costs for high-throughput next-generation sequencing (NGS) have reduced tremendously, the application of NGS to genetic mapping in apple is still insufficient^{33,34}. Nevertheless, genotyping-by-sequencing (GBS³⁵) has gradually become prominent. GBS, a targeted sequencing technique based on reducing genome complexity with restriction enzymes, is a cost-effective method that allows for robust and reliable de novo SNP identification^{35–37} and has now been applied in some members of Rosaceae including apple and pear^{33,34,38}. Although conventional GBS (cGBS) allows for robust SNP identification, major limitations of this technique are high rates of missing data and genotyping errors³⁹. Tunable GBS (tGBS) addresses some of these limitations by employing two restriction enzymes to generate overhangs in opposite orientations to which single-strand oligos rather than double-strand adaptors are ligated, thereby ensuring that only double-digested fragments are amplified and sequenced³⁹. Unlike cGBS, during polymerase chain reaction (PCR) amplification in tGBS, selective nucleotides at the 3′-end of PCR primers achieve additional genome reduction.

Dense genetic linkage maps are critical for molecular breeding, identifying quantitative trait loci and genes controlling important agronomic traits in plant species. DARt and *M. domestica* SNPs¹¹ markers applied to *M. fusca* accession MAL0045 × ‘Idared’ population provided insights in to a specific genetic locus controlling fire blight resistance in this wild apple accession, with DARt markers significantly linked with resistance²⁵. Since QTL mapping relies on the correct ordering of markers along any chromosome³², already published apple SSR markers and SSRs developed from the ‘Golden Delicious’ sequences¹¹ were applied to the *M. fusca* × ‘Idared’ population as anchor markers to develop a draft genetic map and to precisely determine the position of the fire blight resistance locus of *M. fusca* mapped on linkage group 10 (Mfu10)^{16,40}. Only 1% of ‘Golden Delicious’ SNPs¹¹ were transferable to *M. fusca* and although DARt and SSR markers were applied, the draft map remained scarcely dense with a few linkage groups of apple reference maps, still missing¹⁶. Thus, it was imperative to construct a dense map of this fire blight resistant accession, which would be comparable to reference *Malus* maps, and serve as template for QTL mapping with different strains of *Erwinia amylovora*—the causal pathogen of fire blight. Another rationale for developing a dense and complete map of *M. fusca* was to investigate if minor QTLs contributing to resistance to fire blight could be detected. Here, we report the use of tunable genotyping-by-sequencing (tGBS) markers to facilitate the construction of a dense genetic linkage map of *M. fusca* and the subsequent merging with the initial sets of SSR marker data²⁵. Using previous data for three strains of *E. amylovora*^{41,42} and a new strain together with the constructed tGBS map as template, all previously detected fire blight resistance QTLs were confirmed. However, no minor locus was detected with all the aforementioned strains.

Material and methods

Plant materials. Progenies of crosses between *M. fusca* (acc. No. MAL0045) and ‘Idared’, and both parents were investigated. The progenies were composed of 116 individuals from the cross MAL0045 × ‘Idared’ (population number 05210²⁵) established in 2005, 45 genotypes resulting from a cross ‘Idared’ × MAL0045 (population number 09260) done in 2008, and 21 selected recombinants from the same population planted in 2012 (population number 12229, planted for the identification of recombinants for the fire blight resistance QTL on LG10⁴⁰

of *M. fusca*). Additionally, four more *M. fusca* accessions were used. All genotypes are being maintained in the experimental fields of Julius Kühn-Institut, Dresden-Pillnitz, Germany. Young leaves of the material described above including three replicates of each parent were harvested from trees, lyophilized and sent to Data2Bio, LLC (Ames, Iowa, USA) for DNA extraction and tGBS analyses.

tGBS SNP-type genotyping. Samples were sequenced using five runs on an Ion Proton Instrument by Data2Bio LLC, (Ames, Iowa, USA).

Trimming of sequencing reads. Prior to read mapping, raw reads were preprocessed with the trimming software Lucy^{43,44} according to default parameters. Thereby the nucleotides of each raw read was scanned for low quality regions and bases with PHRED quality value < 15, i.e., error rates of at least 3%, were removed^{45,46}. In addition, the remaining nucleotides were scanned using overlapping windows of 10 bp and sequences beyond the last window with an average quality value less than the specified threshold were truncated.

Read mapping to reference genome assembly. The reference genome *Malus domestica* Whole Genome v3.0.a1 Assembly was downloaded from https://www.rosaceae.org/species/malus/malus_x_domestica/genome_v3.0.a1. Trimmed reads were mapped to the reference genome assembly using GSNAP⁴⁷. A confidently mapped read (≤ 2 mismatches every 36 bp and less than five bases for every 75 bp as tails) was retained for subsequent analyses only if it mapped uniquely in the genome.

SNP calling. These uniquely mapped reads were used for SNP discovery. Potential SNP sites were carefully examined and putative homozygous and heterozygous SNPs were identified in each sample separately using different criteria for homozygous and heterozygous SNP calling, respectively. Data2Bio has previously generated custom scripts for SNP calling bioinformatics pipeline available at <https://schnablelab.plantgenomics.iastate.edu/software/123SNP/>⁴⁸, although the pipeline has evolved over time. Thus:

A SNP site was classified as homozygous in a given sample if at least five reads supported the major common allele at that site and at least 90% of all mapped reads covering that site shared the same nucleotide.

A SNP was classified as heterozygous in a given sample if at least one read supported each of at least two different alleles and each of the two alleles comprised more than 20% of the reads mapped to that site. In addition, it was required that the sum of the number of reads supporting those two alleles were at least equal to five and comprised at least 90% of all reads covering that site.

Furthermore, the generated SNPs were filtered to define a specific minimum call rate of 70% across all samples (MCR70). Additional filtering criteria were: allele number = 2; number of genotypes ≥ 2 ; $10 \leq$ minor allele frequency (MAF) $\leq 40\%$; and $35\% \leq$ heterozygosity rate $\leq 65\%$. Similar to Li et al.⁴⁹ and Zheng et al.⁵⁰, these tGBS parameters were established empirically with a focus on minimizing false positive and false negative SNP calls.

In addition, SNPs identified in the other four wild *Malus fusca* samples (MAL0200, MAL0289, MAL0357, and MAL0768) and the two parent samples (MAL0045 and 'Idared') were used to build a phylogenetic tree. SNP calling was performed using uniquely mapped reads from the six samples independently. A phylogenetic tree was built with these SNPs using the maximum likelihood based PhyML3.0⁵¹.

Construction of genetic linkage maps. The *M. fusca* tGBS SNP genetic map was constructed with software package R/qtl at an LOD of 18⁵² using MRC70 SNPs described initially with only individuals of populations 05210 and 09260.

For the integration of the SSR data established on the 05210 population¹⁶ to the tGBS SNP map, individuals with missing SSR data were genotyped with such SSR markers. Furthermore, individuals of the 09260 population that were included in the tGBS analyses possessed no original SSR data and hence were genotyped with 53 SSR markers already found to be polymorphic in the 05210 population¹⁶. Genotyping was performed by multiplex PCR with up to 5 SSR markers using the Type-It Kit (Qiagen, Hilden, Germany) with the following conditions: 95 °C for 5 min, followed by 30 cycles of 95 °C for 1 min, 60 °C for 1 min 30 s, and 72 °C for 30 s, and a final extension for 30 min at 60 °C. PCR fragments were then analysed either on a CEQ 2000XL DNA Sequencer (Beckman Coulter, Germany) or on a 3500/3500xL Genetic Analyzer (Applied Biosystems, Germany). Sample preparations for both systems of fragment analyses are as previously reported in Emeriewen et al.¹⁶.

For mapping, SNP markers showing identical genotype profiles were reduced to one and markers expressing unreliable double recombinants were removed from analysis. Mapping was performed using JoinMap 4.0⁵³ at an LOD threshold of 12.0–18.0. Linkage groups were assigned by comparing results with the location of SSR markers in the reference genetic linkage maps of apple.

Mapping of markers informative in *M. fusca* to physical chromosomes of *M. domestica*. Positions of SNP markers informative in *M. fusca* on physical *M. domestica* chromosomes were determined by blasting the merged reads containing the respective SNPs (sequence length from 220 to 401 bp, Table S1) to the Golden Delicious Doubled Haploid genome (GDDH13). Only the best hits were taken into account. The position of SSRs on GDDH13 was determined by blasting both forward and reverse primers to GDDH13. If the distance of both primers was reasonable, the SSR could be placed on GDDH13, otherwise the SSR was discarded. These positions of the SSRs were used for the alignment of the genetic linkage map to the physical map.

Corresponding apple chromosome	<i>M. fusca</i> tGBS SNP map			<i>M. fusca</i> integrated map			
	LG	Number of SNPs	Length (cM)	LG	Number of markers	Length (cM)	Largest gap (cM)
Chr1	16	94	135.7	1	41	73.56	9.98
Chr2	10	134	171.6	2	47	69.80	7.60
Chr3	9	143	115.5	3	52	73.99	4.20
Chr4	17	94	176.6	4	37	60.81	7.35
Chr5	1	205	203.6	5	39	70.62	7.80
Chr6	15	98	107.5	6	30	55.21	6.49
Chr7	7	152	187.4	7	34	70.47	17.98
Chr8	13	118	121.0	8	33	64.74	8.65
Chr9	6	154	122.5	9	33	49.27	4.28
Chr10	4*	182	159.9	10	37	69.35	8.05
Chr11	2	191	218.2	11	34	72.82	9.13
Chr12	5	155	167.2	12	34	62.04	7.16
Chr13	12	134	105.2	13	33	69.45	10.39
Chr14	14	98	146.2	14	25	34.68	4.18
Chr15	8	149	180.7	15	31	73.54	19.79
Chr16	11	137	107.6	16	27	53.44	14.49
Chr17	3	186	134.4	17	46	57.56	4.98
Overall	17	2,424	2,560.8	17	613	1081.35	

Table 1. tGBS SNPs mapped across 17 linkage groups of *M. fusca* and 17 linkage groups representing the *M. fusca* integrated map. LG linkage group. *LG in the tGBS SNP map where *Mfu10* is located.

Phenotypic evaluation and QTL-mapping. Artificial shoot inoculation of individuals with *E. amylovora* strains Ea222, Ea3049, Ea1189, and ZYRKD3-1 (an AvrRpt2_{EA} effector deletion mutant of Ea1189), as well as disease evaluation was performed as described by Peil, et al.⁵⁴. Briefly, up to 10 replicates of each individual were inoculated with each strain, and 28 days post inoculation (dpi), fire blight lesion length (cm) was measured and converted to percent lesion length (PLL) by dividing the necrotic shoot by the total shoot length and multiplying by 100. Phenotypic data for percent lesion length (PLL) of 05210 individuals for strains Ea222 (data from¹⁶), Ea3049 (data from Emeriewen et al.⁴¹) and ZYRKD3-1 (data from Emeriewen et al.⁴²), were complemented by data obtained from inoculation of 05210 and 09260 individuals with Ea1189, and inoculation of 09260 individuals with Ea222, Ea3049, and ZYRKD3-1. Phenotypic data of individuals for each strain as well as the average of all strains and the created map as template were used for QTL-mapping, which was performed with MapQTL5⁵⁵.

Results

tGBS SNP identification. From 188 samples submitted to Data2Bio, 460,552,947 Quality Trimmed tGBS reads, with an average length of 127 base pairs (bp), were generated, and 46.4% of these could be uniquely aligned (i.e., aligned to a single location) to the ‘Golden Delicious’ reference genome. Reads that aligned to more than one location were excluded in subsequent analyses. From uniquely aligned reads, 696,798 polymorphic sites were identified. SNP filtering led to the identification of 39,496 SNPs. Excessive rates of heterozygous SNPs were observed for 9 individuals which were excluded from subsequent analyses, involving 154 genotypes of populations 05210 and 09260 and 19 recombinants. Subsequently, further filtering of SNP set led to the identification of a subset of 11,998 SNPs (Table S1) called in at least 70% of samples (described as MCR70 SNPs in methods). Figure S1 shows the mapping population SNP classification.

Genetic map of *M. fusca* based on tGBS SNPs. Genetic mapping was performed with only individuals of populations 05210 and 09260. SNPs found to be informative in *M. fusca* were 2677 of which 109 SNPs showing distorted segregation were discarded. Six individuals were removed from analysis due to an excessive number of cross overs, i.e., in the end 148 progeny were used for map construction. In total 2424 SNPs, representing 90.5% were successfully mapped. The resulting genetic map (Fig. S2) consists of 17 linkage groups (LGs) with a minimum SNP number of 94, the highest with 205 SNPs and an overall genetic length of 2560.8 cM (Table 1).

Linkage mapping to integrate tGBS and microsatellite data. Prior to mapping with all marker data (i.e. tGBS SNP and SSRs), redundant SNPs as well as SNPs raising questionable double recombination were discarded. Redundancy in this regard was defined as SNPs clustering together with same segregating pattern. In such cases, only one SNP was chosen. These led to the exclusion of 1864 of 2424 SNPs.

Analysis with JoinMap4.0⁵³ was performed using the Kosambi’s function and grouping independence LOD of 12–18. The final integrated map of *M. fusca* is comprised of 560 tGBS SNPs and 53 previously mapped SSR markers, 613 loci in total. All markers could be assigned to the 17 linkage groups of apple, with a total genetic length of 1081.35 cM (Fig. 1, Table 1). The total length is 1479.45 cM less than the overall genetic distance with

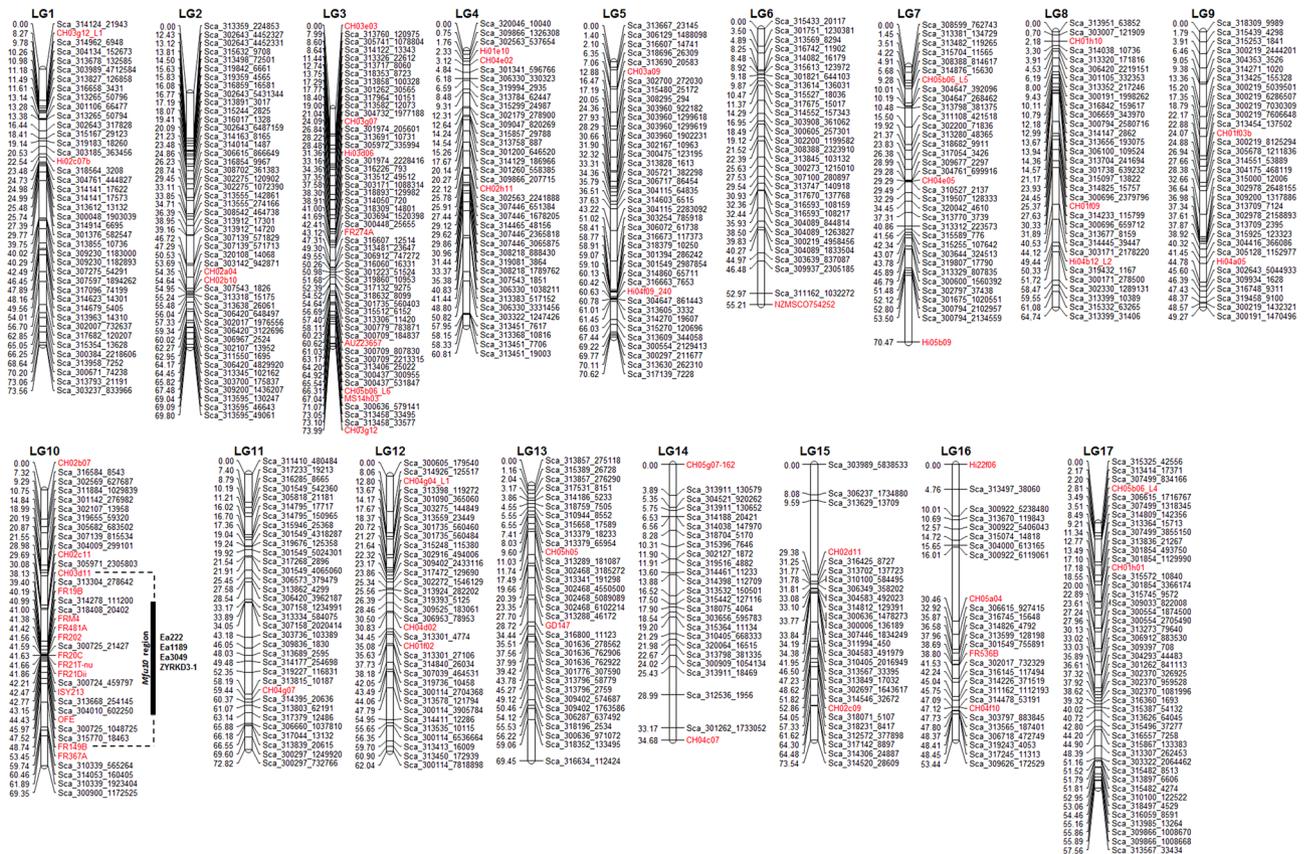


Figure 1. Integrated dense genetic map of MAL0045 comprising 560 tGBS SNPs and 53 SSR markers (highlighted in red). Total length is 1081.35 cM.

only the 2424 SNP markers. Two SSR markers were found to be multi-locus: CH03g12 mapped on LG1 and LG3 whilst CH05b06 mapped on LG3, LG7 and LG17. Figure S3 shows the comparison of the position of markers of each *M. fusca* LG to their physical position on the respective chromosome of GDDH13¹⁰. The alignment of the marker sequences, i.e. both the forward and the reverse primers of the SSRs in an appropriate distance, to the GDDH13 physical chromosomes revealed 116 markers for which the best alignment hit either did not match the corresponding GDDH13 chromosome or gave no hit.

The order of most markers on the 17 LGs is in agreement with their physical order on the respective chromosomes (Fig. S3). Additionally, a homoelogous to the fire blight resistant gene *Mfu10* was mapped on chromosome 10 of GDDH13 (Fig. S3). A possible inversion can be seen in between the upper part of LG5 and some possible rearrangements on LG10 and LG14 (Fig. S3). The largest physical distance in Mb for the first marker of a LG to the start of the chromosome was observed for LG15 with 4.7 Mb, for the last marker of a LG to the end of the chromosome the distance is 16.62 Mb for LG13 and the largest distance between two markers is 16.61 Mb on LG11.

Comparative fire blight mapping using tGBS SNP map. Previously, a strong fire blight resistance QTL (*Mfu10*) was identified on LG10 of *M. fusca* after phenotypic evaluation of the 05210 mapping population with *Erwinia amylovora* strain Ea22¹⁶. *Mfu10* was found to be stable after subsequent analyses with the highly aggressive Canadian strain Ea3049⁴¹ as well as with a mutant strain ZYRKD3-1⁴² and its wild type strain Ea1189. Pairwise correlations and scatter plots of the four strains and the overall PLL average are presented in Fig. 2a. However, previous QTL mapping were not performed with a dense genetic map as the currently developed map. As a proof of concept, phenotypic data determined on 05210 and 09260 for the three strains and the wild type strain of ZYRKD3-1, i.e. Ea1189, were used for QTL analyses. QTL mapping for percent lesion length with Ea22, Ea3049, ZYRKD3-1, and Ea1189 as well as an average of all strains were thus performed. QTLs with significant LOD scores above the chromosome wide threshold of LOD=3.0 were detected only on LG10 of the SNP-SSR map of *M. fusca*. Interval mapping resulted in QTLs for all *E. amylovora* strains and the overall average at the same position on LG10 (Fig. 2b).

Table 2 shows the markers with the highest LOD of the QTL, the position of the marker, the respective phenotypic variance explained, and the PLL of progenies with and without the resistance allele. For all strains, progenies with the least lesion were those that inherited the resistance alleles with the highest difference (65.5%) between progenies that possessed the resistance allele and those that did not observed for mutant strain ZYRKD3-1. The highest correlation of PLL between two strains was determined for ZYRKD3-1 and its wild type strain Ea1189 (Fig. 2a) with $r = 0.93$.

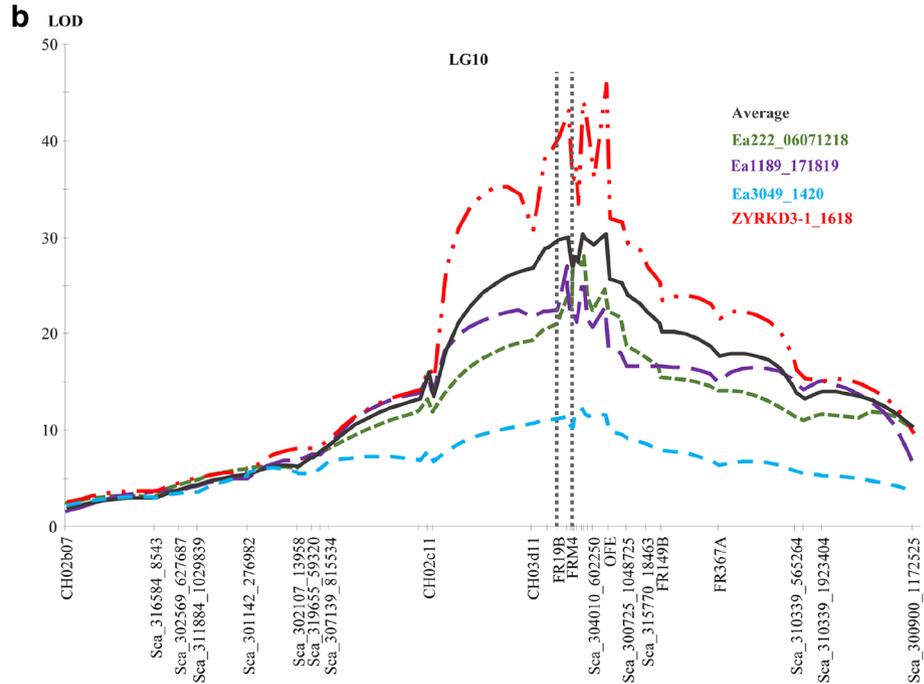
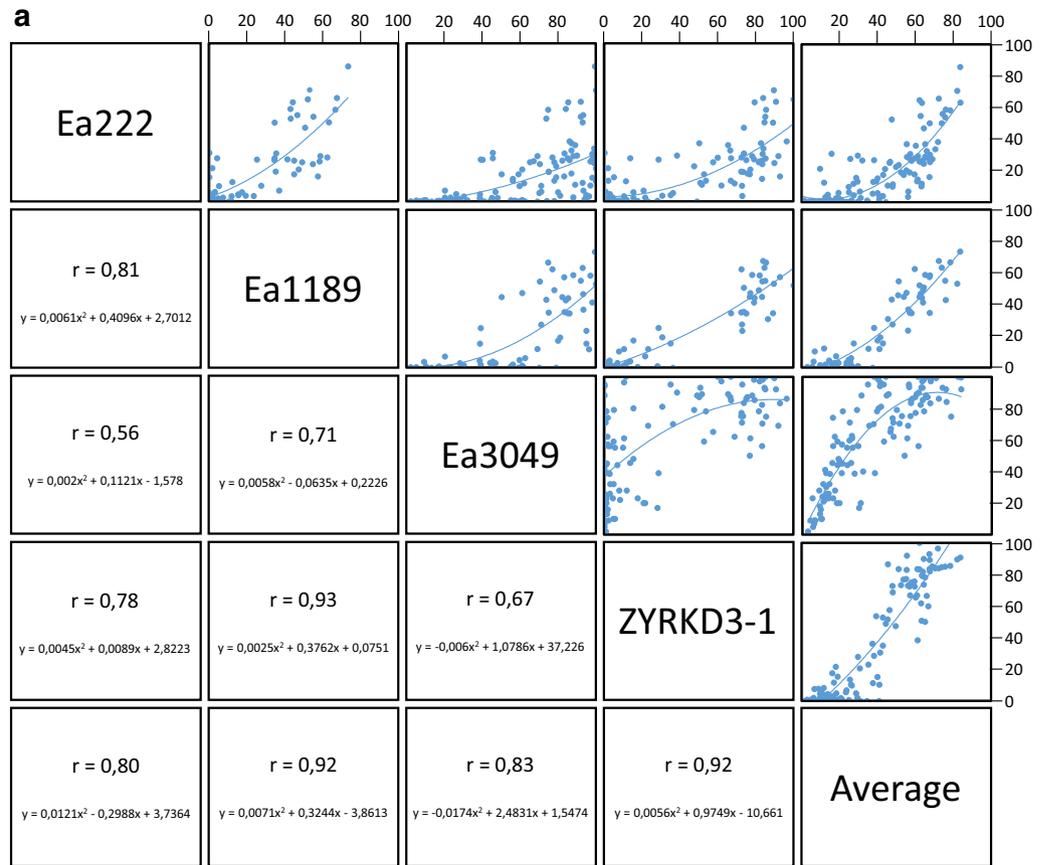


Figure 2. (a) Pairwise correlations and scatter plots for the phenotypic values of MAL0045-derived progeny. The average values for each individual for the four strains and the overall average were also used for QTL analyses. The lines represent the binominal trend line. Axes show percent lesion length (PLL). (b) LOD score plot of LG10 of MAL0045 for the four *E. amylovora* strains and the average value of all strains. The numbers attached after each strain name indicate the year(s) of phenotyping (e.g. Ea3049_1420 means that individuals were inoculated in 2014 and 2020).

<i>E. amylovora</i> strain	Marker with highest LOD	cM	LOD	% expl	Average PLL of progenies	
					With <i>R</i> -allele	Without <i>R</i> -allele
Ea222	FR202	41.555	28.06	62.8	3.5	32.3
Ea1189	Sca_314278_111200	40.994	27.09	87.1	4.0	47.3
ZYRKD3-1	Sca_300724_459797	42.207	44.13	82.4	9.1	74.7
Ea3049	Sca_300724_459797	42.207	12.13	37.3	46.4	81.0
Average	Sca_300724_459797	42.207	30.32	62.6	19.4	56.9

Table 2. Summary of QTL characteristics showing markers with the highest LOD of a QTL obtained by inoculation with different *E. amylovora* strains, the position on LG10, the LOD and the phenotypic variance explained (%), respectively. LOD logarithm of the odd, *R*-allele resistance allele.

Phylogenetic tree construction with wild apples and Idared-specific SNPs. A mini diversity panel study of four wild *Malus fusca* samples (MAL0200, MAL0289, MAL0357, and MAL0768) and the two parent samples (MAL0045 and ‘Idared’) was performed with 64,664 SNPs identified for these samples. The SNPs could be genotyped in at least four of the six samples and had a minor allele frequency of $\geq 1/12$ allele as well as a genotype number of ≥ 2 and an allele number equal to two. The phylogenetic tree shows that all wild accessions clustered together and are much different from the domesticated parent ‘Idared’ (Fig. S4).

Discussion

High-density marker platforms are capable of enhancing fire blight resistance breeding through fast and precise identification of fire blight QTLs⁵⁶. In the present study, we constructed a dense genetic map of the wild apple *M. fusca* using tunable genotyping-by-sequencing SNP-type technology and subsequently integrated previously reported SSR data¹⁶ as anchor markers. Although, a high density SNP apple array, the Axiom 480 k apple array³¹, is available for apple, SNP discovery by tGBS was preferred. The array was developed to cover the genetic diversity of the cultivated apple³¹ but due to the genetic distance between *M. fusca* accessions and *M. domestica* as shown in Fig. S4, only a low percentage of apple SNPs were transferrable to *M. fusca*¹⁶. The thousands of SNPs detected with tGBS technology in the current study further demonstrates the high throughput and effectiveness for generating genome-wide SNP data which are suitable for genetic mapping studies³⁶. The number of SNPs successfully mapped here (2424) is similar and comparable to other studies in apple e.g. 2436³³ and 2590³⁷. However, the high degree of unexplainable double recombinations found in the tGBS SNP map data (not shown) is a disadvantage and the question arise what the reason could be. This situation leads to the overestimation of the sizes of linkage groups (LGs) allowing them to be partly incomparable to reference maps in apple¹⁹. Besides, redundancy (no recombination of SNPs) was also high leading to clustering of scaffolds however; this is expected in a population size as in the current study. Nevertheless after the removal of such SNPs and the integration of SSR data, the resultant map (Fig. 1) spanned 1081.35 cM (1479.45 cM less than the original tGBS SNP map) (Fig. S2). The initial map reported by¹⁶ spanned 889 cM with markers spread across 20 linkage groups. In contrast, markers in this newly integrated map spread across the recognized 17 linkage groups of apple and is more comparable to 1140 cM and 1450 cM for apple reference maps of ‘Fiesta’ and ‘Discovery’, respectively, reported by¹⁹. Even after enriching the ‘Fiesta’ and ‘Discovery’ maps with more markers, Silberberg-Dilworth et al.²⁰ reported genetic lengths of 1145.3 cM and 1417.1 cM respectively for both maps. Moreover, using GBS technology, Norelli et al.³⁴ reported a total genetic length ranging from 1230.2 to 1722.9 cM for a map of *M. sieversii* comprising GBS markers and SSRs whilst GBS markers spanned 1271 cM for a ‘Golden Delicious’ × ‘Scarlet Spur’ population³³ and 1350.1 cM for ‘Hongro’ × ‘Alps otome’ apple population⁵⁷. The order of the SSRs on the 17 LGs of *M. fusca* corresponds to their positions on the apple-integrated map³². Furthermore, the order of most markers of *M. fusca* LGs is in agreement with their physical position on the respective chromosome of GDDH13. Only few possible inversions have been observed. Nevertheless, in all LGs markers have been mapped for which no position could be mapped on the respective chromosome. This could be because only the best hit was taken into account, to typing errors in marker analysis, to errors in genetic mapping or to the genetic distance of *M. domestica* and *M. fusca* (Fig. S4). Duan et al.⁵⁸ analysed the population structure of 117 apple cultivars and wild species accessions and constructed phylogenetic tree showing the large genetic distance of *M. domestica* and *M. fusca*. By comparing the LGs to the chromosomes, large parts from the start or the end of some chromosome end are not identified by a marker. Additionally, some large gaps in between two adjacent markers on a chromosome were identified. Large gaps are explainable by the relatively low density of markers and the smaller nuclear DNA content of *M. fusca*. The median nuclear DNA content of five *M. fusca* accessions was determined as 1.464 pg compared to the median nuclear DNA content of 1.514 pg for *M. domestica*⁹.

The *M. fusca* (MAL0045) × ‘Idared’ 05210 population was initially developed for mapping the fire blight resistance locus of this crabapple²⁵. The major locus (*Mfu10*), which confers resistance to fire blight, caused by the bacterium, *Erwinia amylovora*, was detected following artificial shoot inoculation with *E. amylovora* strain, Ea222. Subsequent studies found *Mfu10* to be stable following phenotypic evaluation with the highly aggressive Canadian strain, Ea3049⁴¹ and the AvrRpt2_{EA} mutant ZYRKD3-1⁴². Ea222 and Ea3049 possess the C-allele and the S-allele, respectively, of AvrRpt2_{EA} effector of *E. amylovora* whereas the effector is deleted in strain ZYRKD3-1^{59,60}. The findings of Vogt et al.⁶⁰ imply that the S-allele of AvrRpt2_{EA} is responsible for aggressive virulence in *Malus* genotypes and the deletion of the effector gene (ZYRKD3-1) leads to a gene-for-gene relationship in *Malus* × *robusta* 5 (Mr5)—*E. amylovora* system. Although the highly aggressive Canadian strain does not affect

MAL0045, several of its progeny individuals were highly susceptible with only a few showing strong resistance to this strain⁴¹, resulting in the low LOD compared to the other strains in Fig. 2b. In similar studies with Ea3049, Peil et al.⁶¹ and Wöhner et al.²² reported the breakdown of the major QTL on LG3 of Mr5 but identified minor QTLs in different linkage groups. Previous mapping studies using MAL0045-derived progeny and the three strains mentioned above did not identify any minor QTLs leading to the assumption that the lack of density of MAL0045 genetic map could be the hindrance especially as the development of an improved map of Mr5 facilitated the detection of more minor fire blight QTLs with Ea3049²².

Yet, QTL mapping with the dense map developed in the current study did not detect any minor locus. With the knowledge gained from Mr5 fire blight resistance and results from the current study, we therefore hypothesize that detection of minor QTLs does not only depend on a dense genetic map, but also largely depends on the *E. amylovora* strain. Fire blight resistance is strain dependent and in Mr5, no minor QTLs were detected with Mr5-avirulent strains^{54,62–64}, but only with strains virulent to Mr5^{22,61}. It is thus plausible that the failure to detect a minor locus with the strains used in the current study is down to their interactions with MAL0045 and MAL0045-derived progeny. Nevertheless, it was worthwhile to construct such a map for this highly resistant wild species accession, as future-mapping studies with different pathogen strains, including mutant strains, will then eliminate the lack of a dense genetic map. To our knowledge, this is the first report of QTL mapping in *Malus* using the average data of four different strains of *E. amylovora*. Further genetic studies would focus on if bacterial effector mutants will be able to overcome MAL0045 resistance and if a second resistance factor will be identified, which due to previous results^{16,41,42}, must be present in this accession. A resistance not overcome by any strain of the pathogen is valuable to breeders to develop resistant cultivars.

Received: 9 July 2020; Accepted: 15 September 2020

Published online: 01 October 2020

References

- Viereck, L. A. & Little Elbert, L. J. *Alaska Trees and Shrubs* 205–207 (University of Alaska Press, Fairbanks, 2007).
- Qian, G.-Z., Liu, L.-F. & Tang, G.-G. A new selection of *Malus* (Rosaceae) from China. *Ann. Bot. Fenn.* **1**, 68–73 (2006).
- Robinson, J., Harris, S. & Juniper, B. Taxonomy of the genus *Malus* Mill. (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. *Plant Syst. Evol.* **226**, 35–58 (2001).
- Fiala, J. L. *Flowering Crabapples: The Genus Malus* (Timber Press Inc, Portland, 1994).
- Vavilov, N. I. Studies on the origin of cultivated plants. *Bull. Appl. Bot.* **16**, 1–10 (1926).
- Zohary, D. & Hopf, M. *Domestication of Plants in the Old World: The Origin and Spread of Cultivated Plants in West Asia, Europe and the Nile Valley* (Oxford University Press, Oxford, 2000).
- Cornille, A., Giraud, T., Smulders, M. J., Roldán-Ruiz, I. & Gladieux, P. The domestication and evolutionary ecology of apples. *Trends Genet.* **30**, 57–65 (2014).
- Hanke, M.-V., Flachowsky, H., Peil, A. & Emeriewen, O. F. in *Biotechnology of Fruit and Nut Crops Biotechnology in Agricultural Series* (eds R. Litz, F. Pliego-Alfaro, & J. I. Hormaza) 440–473 (CAB International, Cambridge, 2020).
- Höfer, M. & Meister, A. Genome size variation in *Malus* species. *Journal of Botany* <https://doi.org/10.1155/2010/480873> (2010).
- Daccord, N. et al. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat. Genet.* **49**, 1099–1106 (2017).
- Velasco, R. et al. The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat. Genet.* **42**, 833–839. <https://doi.org/10.1038/ng.654> (2010).
- Conner, P. J., Brown, S. K. & Weeden, N. F. Randomly amplified polymorphic DNA-based genetic linkage maps of three apple cultivars. *J. Am. Soc. Hortic. Sci.* **122**, 350–359 (1997).
- Hemmat, M., Weedon, N., Manganaris, A. & Lawson, D. Molecular marker linkage map for apple. *J. Hered.* **85**, 4–11 (1994).
- Maliepaard, C. et al. Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theor. Appl. Genet.* **97**, 60–73 (1998).
- Celton, J. M., Tustin, D. S., Chagne, D. & Gardiner, S. E. Construction of a dense genetic linkage map for apple rootstocks using SSRs developed from *Malus* ESTs and *Pyrus* genomic sequences. *Tree Genet. Genomes* **5**, 93–107. <https://doi.org/10.1007/s11295-008-0171-z> (2009).
- Emeriewen, O. F. et al. Identification of a major quantitative trait locus for resistance to fire blight in the wild apple species *Malus fusca*. *Mol. Breeding* **34**, 407–419. <https://doi.org/10.1007/s11032-014-0043-1> (2014).
- Han, Y. et al. Integration of physical and genetic maps in apple confirms whole-genome and segmental duplications in the apple genome. *J. Exp. Bot.* **62**, 5117–5130 (2011).
- Liebhart, R. et al. Development and characterisation of 140 new microsatellites in apple (*Malus × domestica* Borkh.). *Mol. Breeding* **10**, 217–241 (2002).
- Liebhart, R., Koller, B., Gianfranceschi, L. & Gessler, C. Creating a saturated reference map for the apple (*Malus × domestica* Borkh.) genome. *Theor. Appl. Genet.* **106**, 1497–1508. <https://doi.org/10.1007/s00122-003-1209-0> (2003).
- Silfverberg-Dilworth, E. et al. Microsatellite markers spanning the apple (*Malus × domestica* Borkh.) genome. *Tree Genet. Genomes* **2**, 202–224. <https://doi.org/10.1007/s11295-006-0045-1> (2006).
- Wang, A. et al. EST contig-based SSR linkage maps for *Malus × domestica* cv. Royal Gala and an apple scab resistant accession of *M. sieversii*, the progenitor species of domestic apple. *Mol. Breeding* **29**, 379–397 (2012).
- Wöhner, T. W. et al. QTL mapping of fire blight resistance in *Malus xrobusta* 5 after inoculation with different strains of *Erwinia amylovora*. *Mol. Breeding* **34**, 217–230. <https://doi.org/10.1007/s11032-014-0031-5> (2014).
- Guilford, P. et al. Microsatellites in *Malus × domestica* (apple): abundance, polymorphism and cultivar identification. *Theor. Appl. Genet.* **94**, 249–254 (1997).
- Jaccoud, D., Peng, K., Feinstein, D. & Kilian, A. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res.* **29**, e25–e25 (2001).
- Emeriewen, O. et al. Evidence of a major QTL for fire blight resistance in the apple wild species *Malus fusca*. *Acta Hortic.* **1056**, 289–294 (2014).
- Sánchez-Sevilla, J. F. et al. Diversity Arrays Technology (DArT) marker platforms for diversity analysis and linkage mapping in a complex crop, the octoploid cultivated strawberry (*Fragaria × ananassa*). *PLoS ONE* **10**, e0144960 (2015).
- Schouten, H. J. et al. Diversity arrays technology (DArT) markers in apple for genetic linkage maps. *Mol. Breeding* **29**, 645–660. <https://doi.org/10.1007/s11032-011-9579-5> (2012).

28. Soriano, J. M. *et al.* Identification and mapping of the novel apple scab resistance gene Vd3. *Tree Genet. Genomes* **5**, 475–482. <https://doi.org/10.1007/s11295-009-0201-5> (2009).
29. Chagne, D. *et al.* Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. *PLoS ONE* **7**, e31745 (2012).
30. Bianco, L. *et al.* Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus × domestica* Borkh). *PLoS ONE* **9**, e110377 (2014).
31. Bianco, L. *et al.* Development and validation of the Axiom Apple480K SNP genotyping array. *Plant J.* **86**, 62–74 (2016).
32. Di Piero, E. A. *et al.* A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. *Hortic. Res.* **3**, 16057 (2016).
33. Gardner, K. M. *et al.* Fast and cost-effective genetic mapping in apple using next-generation sequencing. *G3 Genes Genomes Genet.* **4**, 1681–1687 (2014).
34. Norelli, J. L. *et al.* Genotyping-by-sequencing markers facilitate the identification of quantitative trait loci controlling resistance to *Penicillium expansum* in *Malus sieversii*. *PLoS ONE* **12**, e0172949 (2017).
35. Elshire, R. J. *et al.* A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379 (2011).
36. Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J.-L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* **7**, e32253 (2012).
37. Poland, J. A. & Rife, T. W. Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* **5**, 92–102 (2012).
38. Kumar, S. *et al.* Genotyping-by-sequencing of pear (*Pyrus* spp.) accessions unravels novel patterns of genetic diversity and selection footprints. *Hortic. Res.* **4**, 17015 (2017).
39. Ott, A. *et al.* tGBS genotyping-by-sequencing enables reliable genotyping of heterozygous loci. *Nucleic Acids Res.* **45**, e178–e178 (2017).
40. Emeriewen, O. F. *et al.* Towards map-based cloning of FB_Mfu10: identification of a receptor-like kinase candidate gene underlying the *Malus fusca* fire blight resistance locus on linkage group 10. *Mol. Breeding* **38**, 106. <https://doi.org/10.1007/s11032-018-0863-5> (2018).
41. Emeriewen, O. F., Richter, K., Hanke, M. V., Malnoy, M. & Peil, A. The fire blight resistance QTL of *Malus fusca* (Mfu10) is affected but not broken down by the highly virulent Canadian *Erwinia amylovora* strain E2002A. *Eur. J. Plant Pathol.* **141**, 631–635. <https://doi.org/10.1007/s10658-014-0565-8> (2015).
42. Emeriewen, O., Richter, K., Hanke, M.-V., Malnoy, M. & Peil, A. Further insights into *Malus fusca* fire blight resistance. *J. Plant Pathol.* **99**, 45–49 (2017).
43. Chou, H., Sutton, G., Glodek, A. & Scott, J. in *Proceedings of the Tenth Annual Genome Sequencing and Annotation Conference (GSAC X)*.
44. Li, S. & Chou, H.-H. LUCY2: an interactive DNA sequence quality trimming and vector removal tool. *Bioinformatics* **20**, 2865–2866 (2004).
45. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
46. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using Phred I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
47. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).
48. Li, X. *et al.* Genic and nongenic contributions to natural variation of quantitative traits in maize. *Genome Res.* **22**, 2436–2444 (2012).
49. Li, T. *et al.* Genetic characterization of inbred lines from Shaan A and B groups for identifying loci associated with maize grain yield. *BMC Genet.* **19**, 63 (2018).
50. Zheng, Z. *et al.* Genetic diversity, oopulation structure, and botanical variety of 320 global peanut accessions revealed through tunable genotyping-by-sequencing. *Sci. Rep.* **8**, 14500. <https://doi.org/10.1038/s41598-018-32800-9> (2018).
51. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
52. Broman, K., Wu, H., Sen, S. & Churchill, G. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
53. Van Ooijen, J. *JoinMap 4, software for the calculation of genetic linkage maps in experimental population* (Kyzama B.V, Cambridge, 2006).
54. Peil, A. *et al.* Strong evidence for a fire blight resistance gene of *Malus × robusta* located on linkage group 3. *Plant Breeding* **126**, 470–475. <https://doi.org/10.1111/j.1439-0523.2007.01408.x> (2007).
55. Van Ooijen, J. W. *MapQTL5, Software for the mapping of quantitative trait loci in experimental populations* (Kyazma B.V., Cambridge, 2004).
56. Peil, A., Emeriewen, O. F., Khan, A., Kostick, S. & Malnoy, M. Status of fire blight resistance breeding in Malus.
57. Ban, S. H. & Choi, C. Development of an apple F1 segregating population genetic linkage map using genotyping-by-sequencing. *Plant Breeding Biotechnol.* **6**, 434–443 (2018).
58. Duan, N. *et al.* Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nat. Commun.* **8**, 249. <https://doi.org/10.1038/s41467-017-00336-7> (2017).
59. Emeriewen, O. F., Wöhner, T., Flachowsky, H. & Peil, A. Malus Hosts-Erwinia amylovora interactions: strain pathogenicity and resistance mechanisms. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2019.00551> (2019).
60. Vogt, I. *et al.* Gene-for-gene relationship in the host-pathogen system *Malus xrobusta* 5-*Erwinia amylovora*. *New Phytol.* **197**, 1262–1275. <https://doi.org/10.1111/nph.12094> (2013).
61. Peil, A., Flachowsky, H., Hanke, M.-V., Richter, K. & Rode, J. Inoculation of *Malus × robusta* 5 progeny with a strain breaking resistance to fire blight reveals a minor QTL on LG5. *Acta Hort.* **896**, 357–362 (2011).
62. Gardiner, S. E. *et al.* Putative resistance gene markers associated with quantitative trait loci for fire blight resistance in *Malus 'Robusta* 5' accessions. *BMC Genet.* **13**, 25. <https://doi.org/10.1186/1471-2156-13-25> (2012).
63. Peil, A. *et al.* Confirmation of the fire blight QTL of *Malus × robusta* 5 on linkage group 3. *Acta Hort.* **793**, 297–303 (2008).
64. Peil, A. *et al.* Mapping of fire blight resistance in *Malus × robusta* 5 flowers following artificial inoculation. *BMC Plant Biol.* **19**, 532. <https://doi.org/10.1186/s12870-019-2154-7> (2019).

Author contributions

P.S.S. developed the SNPs, established the SNP map and calculated the phylogenetic relationships; O.F.E. did the SSR work; O.F.E. and A.P. established the integrated map, the QTL-mapping, prepared the figures and wrote the main manuscript text; M.M. analysed results; M.M., J.K. and P.S.S. and revised the manuscript; T.B. and J.K. did the physical mapping; K.R. performed the phenotyping; all authors reviewed the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. The Deutsche Forschungsgemeinschaft (DFG) financially supported projects on *Malus fusca* fire blight resistance (Project numbers: AOBJ-619970, AOBJ-661177).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-73393-6>.

Correspondence and requests for materials should be addressed to O.F.E. or A.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020