

Chapter 8

Genomic Designing for Genetic Improvement of Biotic Stress Resistance in Flax



Frank M. You, Khalid Y. Rashid, and Sylvie Cloutier

Abstract Biotic stresses attributable to various pathogens such as fungi, bacteria and viruses are external threats to plant growth, development and ultimately productivity. To date, genetic improvement of varieties continues to be the most powerful, sustainable and eco-friendly way to overcome these continuously evolving biotic-based threats in plants. The development of genomewide molecular markers and the identification of quantitative trait loci and genes which are linked to biotic stress resistance have the potential to efficiently and genetically enhance the biotic stress resistance of varieties by marker-assisted selection, genomic selection and precision breeding via genome editing. Powdery mildew, Fusarium wilt, pasmo, and rust are major fungal diseases threatening flax production. This chapter briefly reviews the genomic designing for genetic improvement of biotic stress related traits in flax, with a particular emphasis on genomic studies of pasmo resistance, including methodology, outcomes and potential application in breeding.

Keywords Flax · Biotic stress · Genomewide association study (GWAS) · Genomic selection (GS) · PasmO resistance · Precision breeding · Quantitative trait loci (QTLs)

8.1 Introduction

Flax (*Linum usitatissimum* L.) is a valuable source of linseed and stem fiber. Linseed, also known as flaxseed is rich in omega(ω)-3 essential fatty acids (α -linolenic acid

F. M. You (✉) · S. Cloutier

Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, 960 Carling Avenue, Ottawa, ON K10C6, Canada

e-mail: frank.you@agr.gc.ca

S. Cloutier

e-mail: sylviej.cloutier@agr.gc.ca

K. Y. Rashid

Morden Research and Development Centre, Agriculture and Agri-Food Canada, 101 Route 100, Morden, MB R6M 1Y5, Canada

e-mail: khalid.rashid@agr.gc.ca

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

C. Kole (ed.), *Genomic Designing for Biotic Stress Resistant Oilseed Crops*,

https://doi.org/10.1007/978-3-030-91035-8_8

or ALA), lignans, and soluble and insoluble fibers, making it one of the most healthy plant foods (Fofana et al. 2010; Touré and Xueming 2010; Kim and Ilich 2011; Leyva et al. 2011). Linseed oil also has various industrial uses such as soap, vehicle paints, linoleum, printing inks, oil clothing, textiles, patent leather, shoe polish and others (Juita et al. 2012). Flax fiber extracted from the skin of the flax stem is mainly used for linen, the manufacture of twine and rope and as raw materials for some high quality paper products (Deyholos 2006).

Flax has been grown worldwide, but primarily in temperate and subtropical regions, such as Canada (linseed), China (fiber and linseed), USA (linseed), India (linseed), Russia and Europe (fiber and linseed) and Kazakhstan (linseed) (Fouk et al. 2004; Liu et al. 2011; Worku et al. 2015; You et al. 2016b). In these growing regions, the biotic stresses primarily involve various diseases produced by fungi, viruses and mycoplasma like organisms, with fungal diseases including rust (*Melampsora lini*), anthracnose (*Colletotrichum lini*), pasmoor sphaerella linorium (*Septorialinicola*, *Mycosphaerella linicola*), wilt (soil-borne fungus *Fusarium oxysporum* f.sp. lini), seedling blight and root rot, and stem break and browning (*Aureobasidium pullulan* var. lini or *Polyspora lini*) (seedborne and soil-borne fungi *Rhizoctonia solani*, *Fusarium* spp., or *Pythium* spp. etc.) being predominant (<https://flaxcouncil.ca/growing-flax/chapters/diseases/>). These diseases damage flax plants, affect plant growth and development, and ultimately reduce seed and fiber yield and quality. To control these biotic stresses, rotations with other crops such as cereals (spring and winter wheats, barley and oat), oilseed (canola and mustard) and pulse (peas, lentils and soybean) crops are an effective agronomic practice in Canada. Seed treatment with suitable fungicides is another useful practice to kill seed borne pathogens (Bradley et al. 2007).

Incorporating genetic differences to develop agronomic characteristics and add long-term disease tolerance in flax has traditionally been done by conventional breeding methods (You et al. 2016b). A successful example is the genetic improvement against flax rust which has the potential to be the most destructive disease affecting flax. The rapidity with which rust races can evolve represents a challenge in breeding new resistant varieties. Over the last 70 years, more than 500 flax rust races have been recorded. Fortunately, flax rust resistance to different races is controlled by several major genes (Lawrence et al. 1995; Anderson et al. 1997; Ellis et al. 1999; Dodds et al. 2001a, b; Lawrence et al. 2010) that have been successfully pyramided in elite varieties by conventional breeding in Canada. Currently, all Canadian modern cultivars are immune to the locally existing rust races.

However, resistance to other major diseases such as wilt, pasmo, and powdery mildew is quantitative and controlled mostly by minor-effect polygenes (You et al. 2017a; He et al. 2019b), which poses a challenge to the widely used conventional breeding methods. To date, all flax cultivars registered in Canada are moderately resistant to powdery mildew, wilt, and pasmo (You et al. 2016b). The development of advanced genomic tools, such as quantitative trait locus (QTL) mapping, genomewide association study (GWAS) and genomic selection (GS) allows the rapid identification of QTLs that control complex quantitative traits, contributing to more efficient offspring selection and assisting candidate gene isolation whose validation

can now be accurately performed via gene editing (GE), all of which contribute to accelerating crop genetic improvement.

This chapter briefly introduces genomic design strategies for genetic improvement of biotic stresses with special emphasis on pasmo as an example to describe methodology, outcomes and potential applications in breeding.

8.2 Genomic Design for Genetic Improvement of Biotic Stress Traits

With the development of QTL markers associated with biotic stress resistance, including functional markers, conventional breeding techniques are being revolutionized. Marker-assisted selection (MAS) has been used for traits controlled by major genes such as rust (Kumar et al. 2011; Miedaner and Korzun 2012). GS has been used for complex quantitative traits controlled by numerous polygenes such as resistance to Fusarium wilt, powdery mildew and pasmo (He et al. 2019a), and precision breeding using GE has been used for improving traits controlled by known genes (Nekrasov et al. 2017). Therefore, the identification and characterization of QTLs and causal genes are now an integral part of modern flax breeding programs.

8.2.1 Identification of QTLs

While classical quantitative or statistical genetics is capable of estimate genetic variances of polygenes for quantitative traits at the phenotypic level (Falconer and Mackay 1996), combining suitable genomic design with molecular markers provide a precise way to identify individual polygenic loci or QTLs on chromosomes, estimate their effects and predict co-located candidate genes related to the traits.

Two types of the QTL mapping strategies have been developed and successfully used for QTL identification: linkage mapping (LM) and GWAS (Sehgal et al. 2016). LM use segregating biparental populations, such as F₂, backcross (BC), recombinant inbred line (RIL), and doubled haploid (DH) populations, to create a recombination-based genetic map using molecular makers that is suitable to find QTLs responsible for the characteristics that segregate in the population (Price 2006). The statistical methods and software tools for QTL mapping in biparental populations have been well developed (Kulwal 2018). The major statistical methods to detect additive, dominant and epistatic QTLs include simple interval mapping (SIM), composite or inclusive composite interval mapping (CIM/ICIM), multiple interval mapping (MIM), Bayesian interval mapping (BIM), and multiple trait mapping (MTM) (Kulwal 2018). These methods are implemented in many software tools, such as R/qtl (Arends et al. 2010), MAPMAKER/QTL (Lander et al. 1987), and QGene (Joehanes and Nelson 2008). QTL Ici Mapping may be mostly recommended because it provides

functions of both construction of genetic maps and QTL mapping for additive, dominant, and digenic epistasis as well as interaction of QTLs with environments for various biparental and nested association mapping (NAM) populations (Meng et al. 2015). Traditional statistical methods primarily detect large-effect QTLs and have limited power to identify small-effect and linked QTLs. Recently, Zhang et al. (2020c) proposed a genomewide composite interval mapping (GCIM) for segregating biparental populations and developed a corresponding R package with a command line version called QTL.gCIMapping (v3.2) and a graphical user interface version named QTL.gCIMapping.GUI (v2.0). This method has been effective in identifying small-effect and associated QTLs in biparental populations (Wang et al. 2016b; Wen et al. 2019, 2020).

GWAS is based on linkage disequilibrium (LD) between molecular markers and QTLs in a diverse genetical panel, as opposed to biparental populations, in order to overcome the limitations of the latter. Many population types can be used for GWAS, including natural germplasm collections, diversity panels of both genetic germplasm and breeding lines, and multi-parent breeding populations such as nested association mapping (NAM) (Yu et al. 2008; Monir and Zhu 2018; Ren et al. 2018) and multi-parent advanced generation intercross (MAGIC) populations (Mackay and Powell 2007; Cavanagh et al. 2008; Camargo et al. 2018; Ongom and Ejeta 2018).

GWAS advantages over linkage-based QTL mapping include high genetic variation among individuals, high density molecular markers, and high resolution of QTLs and causal genes on chromosomes (Goutam et al. 2015; Ogura and Busch 2015). Many statistical models have been developed to identify large- and small-effect QTLs that can simply be grouped into two categories: single- and multi-locus models. General Linear Model (GLM) (Price et al. 2006) and Mixed Linear Model (MLM) (Yu et al. 2006) are two traditional single-locus statistical models implemented in many software tools such as TASSEL (Bradbury et al. 2007) for example. Single-locus approaches search the genome in one dimension and measure the significant marker-trait associations one by one. To control for false positives, the stringent Bonferroni correction for multiple tests (P value divided by the number of markers in the model) is frequently used, usually resulting in the exclusion of many false negative loci. This drawback can be particularly acute in crop genetics for traits measured from field experiments that are often plagued by large inherent experimental errors (Zhang et al. 2019). Thus, these types of methods have a restricted capability to detect polygenes with small effects that control the bulk of quantitative traits.

Multi-locus statistical methods that simultaneously test multiple markers include early proposed models such as Multi-Locus Mixed-Model (MLMM) (Segura et al. 2012), and more recent powerful methods to identify quantitative trait nucleotides (QTNs) with small effects. The latter include mrMLM (Wang et al. 2016a; Li et al. 2017), FASTmrMLM (Zhang and Tamba 2018), FASTmrEMMA (Wen et al. 2018), pLARmEB (Zhang et al. 2017a), ISIS EM-BLASSO (Tamba et al. 2017), and pKWmEB, which have been implemented in the R package “mrMLM”, thus called “mrMLM models” (Table 8.1). These multi-locus models use LOD score (≥ 3), rather than the stringent Bonferroni correction to identify significant QTNs, which substantially increases the statistical power to detect small effect QTNs and reduces Type 1

Table 8.1 Some single- and multi-locus statistical methods for genomewide association study (GWAS)

Statistical model	Threshold for significant QTNs	R package	Reference
<i>Single-locus models</i>			
GLM	Bonferroni	MVP v1.0.1	Price (2006)
MLM	Bonferroni	MVP v1.0.1	Yu and Buckler (2006)
GEMMA	Bonferroni	GEMMA v0.96	Zhou and Stephens (2012)
<i>Multi-locus models</i>			
mrMLM	LOD \geq 3	mrMLM v3.0	Wang et al. (2016a)
FASTmrEMMA	LOD \geq 3	mrMLM v3.0	Wen et al. (2018)
ISIS EM-BLASSO	LOD \geq 3	mrMLM v3.0	Tamba et al. (2017)
pLARmEB	LOD \geq 3	mrMLM v3.0	Zhang et al. (2017a)
pKWmEB	LOD \geq 3	mrMLM v3.0	Ren et al. (2017)
FASTmrMLM	LOD \geq 3	mrMLM v3.0	https://cran.r-project.org/web/packages/mrMLM/index.html
RTM-GWAS	Default ^b or Bonferroni	RTM-GWAS ^a	He et al. (2017)
FarmCPU	Bonferroni	MVP v1.0.1	Liu et al. (2016)

Bonferroni: $P = 0.05$ followed by Bonferroni correction for multiple tests. ^a A standalone software tool. ^b Default threshold for significant QTNs is $P = 0.05$ for preselection of markers using single-locus model and $P = 0.01$ for multi-locus and multi-allele models

errors and running time (Wang et al. 2016a; Li et al. 2017; Ren et al. 2017; Tamba et al. 2017; Zhang et al. 2017a; Wen et al. 2018). FarmCPU, a multi-locus model implemented in the MVP R package, is an exception because it still relies on the Bonferroni correction to declare significance of association (Liu et al. 2016).

The haplotype block based multi-locus GWAS method RTM-GWAS (He et al. 2017) is implemented in a standalone software (<https://github.com/njau-sri/rtm-gwas>). This two-step method first groups highly correlated SNPs into LD blocks (called SNPLDBs) to define bi- or multi-allelic haplotypes. This is then followed by a two-stage association analysis to identify QTNs: (1) pre-screening haplotype markers using a single-locus model, and (2) identifying significant QTNs using a multi-locus and multi-allele model with stepwise regression (He et al. 2017).

We have evaluated these single and multi-locus models in several studies for agronomic traits, abiotic and biotic traits in flax and wheat (He et al. 2019b; Fatima et al. 2020; Lan et al. 2020; Sertse et al. 2020). Our results demonstrate that the single-locus models detected mostly large-effect QTNs, while the multi-locus models were capable of identifying QTNs with smaller effects. Some QTNs were identified by multiple models, but, generally speaking, the models identified different subsets of QTNs, indicative of the uniqueness and complementarity of these algorithms (He et al. 2019b). Therefore, both single and multi-locus models resulted in the

identification of a more comprehensive set of QTNs that has been shown to increase prediction ability of GS, and hence is recommended (Lan et al. 2020).

In flax, several biparental populations have been developed to identify QTLs for biotic stress resistance. For Fusarium wilt resistance, a DH population of 143 lines was developed from a cross between the resistant variety Linola and the susceptible Australian flax variety Glenelg, from which two independent and additive genes were identified under greenhouse and field conditions (Spielmeyer et al. 1998). Using a RIL population of 160 lines derived from the resistant cultivar Aurore and the susceptible cultivar Oliver, two independent and recessive genes were also identified for wilt resistance (Edirisinghe 2016). For powdery mildew resistance, three QTLs were detected from F₃ and F₄ families derived from an F₂ population of a cross between the susceptible cultivar NorMan and the resistant cultivar Linda (Asgarinia et al. 2013). Additional biparental populations have also been developed for QTL mapping of flax biotic stress resistance, for example, a Bison/Novelty population of 704 RILs segregating for Fusarium wilt and a Linda/Norman (LNm) population of 160 RILs segregating for powdery mildew (unpublished). These populations have been evaluated for field resistance in multiple years and locations and also re-sequenced using a genotyping-by-sequencing method.

GWAS have been successful in identifying QTLs for agronomic and seed quality traits in flax (Soto-Cerda et al. 2014a, 2014b; Xie et al. 2017; You et al. 2018b). The strength and effectiveness of GWAS using the flax core collection (You et al. 2017a) to detect QTNs for biotic stress traits have been shown for pasmo (He et al. 2019b), powdery mildew (unpublished) and Fusarium wilt (You et al. 2017b).

8.2.2 Candidate Gene Prediction

QTL mapping and GWAS are used to find causal genes underlying traits of interest. Prediction of candidate genes linked to QTNs first requires genomewide gene scans along chromosomes to pinpoint the co-located genes. Although QTNs can be located within coding regions, QTL mapping and GWAS do not provide sufficient resolution to pin the QTLs to accurate intragenic locations or genetic features responsible for controlling the traits. Most QTNs are located in intergenic regions. To infer causal genes linked to a QTN, a logically reasonable method is to check whether the LD correlation (r^2 or D') between the QTN and the markers on neighboring genes is sufficiently high (e.g., >0.8) or, alternatively, to partition the whole genome into haplotype/LD blocks based on the genomewide markers of the diversity panel (Purcell et al. 2007; He et al. 2017; Kim et al. 2019) and then perform candidate gene searches within haplotype blocks harboring significant QTNs. An obvious limitation of this method is that LD blocks or correlations depend on the genetic diversity and the structure of a population. For example, the size of LD blocks in the diversity panel for GWAS are much smaller than that of a biparental population because the former represents a greater number of historical recombination events of the GWAS panel. Thus, GWAS may find a candidate gene of a higher resolution.

A straight forward approach for prediction of candidate genes is to find related genes on the fixed-size flanking regions within a QTL, such as a window of 100–200 kb downstream and upstream of a QTL (Kumar et al. 2015; He et al. 2019b; Sertse et al. 2019; You and Cloutier 2019). The fixed window size may be estimated through analysis of LD decay curve (You et al. 2018b). However, this method has a disadvantage in that the fixed block size does not reflect the differential recombination rates across the genome. Therefore, no matter the methods used to identify candidate genes, all must be validated through functional genomics.

Resistance gene analogs (RGAs) are candidates of resistance genes in plants. They can be identified based on known structural features. RGAs can be clustered as either nucleotide-binding site leucine-rich repeat (NBS-LRR) or transmembrane leucine-rich repeat (TM-LRR) (Hammond-Kosack and Jones 1997). NBS-LRR can be further divided into toll/interleukin receptor (TIR)-NBS-LRR (TNL) or non-TNL/coiled coil-NBS-LRR (CNL) (Hammond-Kosack and Jones 1997). Similarly, TM-LRRs could be classified into two classes: receptor-like kinases (RLKs) and receptor-like proteins (RLPs) (Hammond-Kosack and Jones 1997). Genome-wide RGAs can be identified through software tools (Li et al. 2016) or manually using basic local alignment search tool (BLAST) against annotated gene sequences (You et al. 2018a). Using these approaches, we identified 1327 RGAs in the flax genome which constitute a useful subset to investigate co-localized QTLs associated with biotic stresses (You et al. 2018a).

8.2.3 Genomic Selection

Genomic selection (GS) is a promising breeding selection method that employs prediction models constructed using a training population that is both genotyped with genomewide markers and phenotyped, to predict genomic estimated breeding values (GEBVs) of genotyped but unphenotyped breeding lines. GS promises to increase selection accuracy, shorten breeding cycles, and reduce breeding cost. To date, GS has been implemented in many breeding programs to improve yield, quality, abiotic and biotic stresses, in a wide-range of crop plants such as wheat (Rutkoski et al. 2012, 2014, 2015; Daetwyler et al. 2014), rice (Spindel et al. 2015), flax (You et al. 2016a; He et al. 2019a; Lan et al. 2020), and others. GS is most often used for progeny selection in a breeding program but it can also be applied to evaluation of germplasm and parents, and to predict general combining ability (GCA) and specific combining ability (SCA) of crosses (Bernardo 2015; Lado et al. 2017; Yao et al. 2018). However, the performance of GS depends on (1) choosing a proper statistical model to construct a prediction model; (2) choosing a proper marker set to construct the prediction model; and (3) choosing a proper training population closely related to the test populations.

To evaluate the prediction accuracy or ability of GS models, cross-validation schemes which randomly split the whole population into several subsets (or folds) are frequently used, e.g., five subsets would be called five-fold cross-validation scheme.

For a given random sample, each subset or fold is in turn used for a test data set, and the remaining four subsets are merged to be a training data set. This process is iterated multiple times, e.g. 100 to generate a set of random samples. In this case, a total of 500 permutations of training data sets are generated to construct GS models, which are then used to predict GEBVs. Each of these random sample data sets is used for GS modeling and GEBV prediction. The prediction accuracy or ability is defined using a Pearson's correlation between the GEBVs and the observed phenotypes (You et al. 2016a).

Various genomic models have been developed to optimize prediction models for numerous complex traits. These models include classical parametric statistics based models such as best linear unbiased prediction (RR-BLUP) (Henderson 1975), and genomic BLUP (GBLUP) (Daetwyler et al. 2014); Bayesian statistics based parametric methods such as Bayesian LASSO (BL) (Park and Casella 2008), Bayesian ridge regression (BRR) (Campos et al. 2009), BayesA, BayesB and BayesC; and machine learning based non-parametric methods such as support vector machine (SVM), random forest (RF), radial basis function neural network (RBFNN) and some deep learning methods (Gonzalez-Camacho et al. 2018; Montesinos-Lopez et al. 2018; Fukuoka 2019; Lo-Ciganic et al. 2019; Grinberg et al. 2020; Gupta et al. 2020). These models have been implemented in some popular software tools, especially in some useful R packages (Table 8.2).

GS parametric statistical models are usually built on additive genetic models and their prediction abilities differ depending on genetic architecture of the traits that are examined. However, because some non-additive effects such as dominance and epistasis interactions are common in quantitative traits, these effects are also considered in some GS models (Varona et al. 2018). Besides genomic prediction for individual traits, multi-trait models in GS have been evaluated (Covarrubias-Pazaran et al. 2018; Fernandes et al. 2018; Montesinos-Lopez et al. 2019b). Providing significant genetic correlation between the target traits, the multi-trait GS models outperform those for individual traits. Nevertheless, construction of multi-trait models is computation-intensive, especially for a large molecular marker and phenotypic data set. Recently, some computation-efficient GS models and R packages have been developed for modeling of multiple traits (Montesinos-Lopez et al. 2019a).

Although many GS models have been implemented and evaluated in a variety of crops and traits, RR-BLUP is the most widely used because of its high-caliber capability (Arruda et al. 2015; Rutkoski et al. 2015; Poland and Rutkoski 2016; Dong et al. 2018; Liabeuf et al. 2018). For example, RR-BLUP effectively identified complicated patterns with additive effects and conveyed effective genomic prediction in wheat disease resistance (Ornella et al. 2012). RR-BLUP also has a distinct benefit as well in the performance of computing compared with most of the alternative statistical models (Piepho 2009; Endelman 2011; Arruda et al. 2015; Liabeuf et al. 2018).

GS was initially suggested by Meuwissen et al. (2001). The main idea behind GS is the use of genomewide markers to train statistical models without prior knowledge of genes or QTLs associated with the traits. With the development of high-throughput genotyping technology, high-density genomewide molecular markers can be readily

Table 8.2 Some popular R packages for modeling of genomic selection

R package	Features	Reference
rrBLUP	A classical and efficient maximum—likelihood algorithm based model	Endelman (2011)
BGLR	Bayesian regression models; GBLUP; continuous and categorical traits	Perez and Campos (2014)
BWGS	A pipeline wrapper package integrating random cross-validation, imputation and 15 statically models	Charmet et al. (2020)
BMTME	Bayesian models for multi-trait and multi-environment	Montesinos-Lopez et al. (2019a)
sommer	GBLUP, rrBLUP, faster algorithms	Covarrubias-Pazaran (2016)
G2P	A wrapper package integrating 16 statistical models (BayesA, BayesB, BayesC, BRR, BL, RKHS, RR, rrBLUP, SPLS, LASSO, BRNN, AI, NR, EM, EMM and bigRR) and four machine learning models (RFC, RFR, SVC and SVR) which are provided by other R genomic selection packages; cross-validation	https://github.com/cma2015/G2P
BGGE	Genomic selection for genotype by environment	Granato et al. (2018)
DeepGP	A deep learning pipeline implementing deep learning models of multilayer perceptron networks (MLP) and Convolutional neural network (CNN)	https://github.com/lauzingaretti/DeepGP
DeepGS	A deep learning model	https://github.com/cma2015/DeepGS

obtained and breeding populations can be genotyped at low costs. Several popular genotyping methods are available, such as genotyping by sequencing (GBS), array-based genotyping (e.g., iSelect 90 K array for wheat), and target sequence based genotyping (Bekele et al. 2020; Zhang et al. 2020a). To date, most GS models are constructed based on genomewide random markers. Though some studies have discussed the use of QTLs as markers, only major QTLs were used and the outcome was only a minor improvement in prediction accuracy. Our recent studies revealed that combining single and multi-locus GWAS methods can effectively detect both large and minor effect QTLs that can be used to build GS models, thereby significantly improving genomic prediction accuracy (He et al. 2019a, b; Lan et al. 2020).

8.2.4 Genome Editing (GE) and Precision Breeding

GE is a genome-engineering technology that facilitates precise and efficient targeted modification of genomes to characterize the functions of genes and create novel genetic resources for the genetic improvement of plants (Langner et al. 2018; Chen et al. 2019). GE starts with the creation of site-specific double-strand breaks (DSBs) at the target loci by sequence-specific nucleases. Then the DSBs are repaired by the plants endogenous DNA repair mechanisms, either error-prone non-homologous end joining (NHEJ) or homology-dependent recombination (HDR). NHEJ generates small random insertions, deletions and substitutions, preferably causing a gene knockout, whereas HDR is able to generate accurate point mutations, deletions, or gene knock-in especially useful for plant precision breeding but with low editing frequencies (Langner et al. 2018). Broad-sense genome editing techniques include reverse genetic tools such as induced mutagenesis (Rowland 1991; Chantreau et al. 2013; Fofana et al. 2017), oligonucleotide directed mutagenesis (Sauer et al. 2016), epigenome editing (Migliani et al. 2020), transposons, RNA interference (RNAi), and typical genome editing tools such as zinc-finger nucleases (Bibikova et al. 2002; Shukla et al. 2009; Osakabe et al. 2010), Transcriptional Activator Like Effector Nucleases (TALENs) (Malzahn et al. 2017), and Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR/Cas9) systems (Langner et al. 2018; Chen et al. 2019). In particular, the CRISPR/Cas9 system with CRISPR-associated protein 9 (Cas9) is presently the most commonly used approach for plant genome editing due to its ease and robustness.

GE has been successfully applied to improve disease resistance against various plant pathogens, such as in rice (Li et al. 2012), wheat (Wang et al. 2014; Zhang et al. 2017b), tomato (Nekrasov et al. 2017), citrus (Peng et al. 2017), watermelon (Zhang et al. 2020b) and virus (Chandrasekaran et al. 2016).

Fusarium wilt (*Fusarium oxysporum*) and powdery mildew are destructive diseases in many crops, including flax. Examples of GE applications for these two diseases are listed (Table 8.3). *F. oxysporum* is a soil-borne fungus that exists as pathogenic and non-pathogenic strains (Leslie and Summerell 2006). Three *Fusarium* mitogen-activated protein kinase (MAPK) signaling genes (*FMK1*, *HOG1* and *PBS2*) are associated with plant surface hydrophobicity (sensing) and pathogenesis (Di Pietro et al. 2001). The RNAi-mediated silencing of these three genes in *F. oxysporum* resulted in decreased mycelial growth on tomato fruits, leading to reduced pathogenicity compared to the unsilenced fungus (Pareek and Rajam 2017). The *F. oxysporum* species complex (FOSC) is an economically important group of pathogenic filamentous fungi that are able to infect both animals and plants. Wang et al. (2018) developed an *F. oxysporum*-optimized Cas9 ribonucleoprotein (RNP) and a protoplast transformation method to generate a mutant *bik1* of *BIK1*, a gene in a secondary metabolite biosynthetic cluster, confirming that this polyketide synthase was involved in the synthesis of the red pigment bikaverin.

Mildew resistance locus O (*Mlo*) harbors a gene associated with powdery mildew resistance. Its wild-type alleles confer susceptibility to fungi resulting in the powdery

Table 8.3 Some applications of genome editing in improving biotic stress resistance

Crop	Enhanced disease resistance	GE method	Target genes	Type of modification	Reference
Tomato	Fusarium wilt against <i>F. oxysporum</i>	RNAi-mediated silencing	<i>FMK1, HOG1, PBS2</i>	Silencing	Pareek and Rajam (2017)
Cotton	Fusarium wilt against <i>F. oxysporum</i> f. sp. <i>vasinfectum</i>	CRISPR-Cas9	<i>URA5, URA3, BIK1</i>	Insertion	Wang et al. (2018)
Wheat	Powdery mildew against <i>Blumeria graminis</i> f. sp. <i>tritici</i>	TALEN, CRISPR-Cas9	<i>TaMLO-A1, TaMLO-B1, TaMLO-D1</i>	Knockout	Wang et al. (2014)
Tomato	Powdery mildew against <i>O. neolyopersici</i>	CRISPR-Cas9	<i>Mlo</i>	Knockout, deletion	Nekrasov et al. (2017)
Tomato	Powdery mildew against <i>O. neolyopersici</i>	CRISPR-Cas9	<i>SIPMR4</i>	Knockout	Koseoglou (2017)
Tomato	Powdery mildew against <i>O. neolyopersici</i>	RNAi-mediated silencing	<i>SIPMR4</i>	Silencing	Huibers et al. (2013)

mildew disease (Acevedo-Garcia et al. 2014), while its homozygous knockout mutations (*mlo*) lead to resistance to powdery mildew. Nekrasov et al. (2017) reported a non-transgenic tomato variety resistant to powdery mildew (*Oidium neolyopersici*) using the CRISPR/Cas9 technology to edit the *Mlo* gene (*SlMlo1*) which is based on the Cas9 DNA nuclease guided to a specific DNA target by a single guide-RNA (sgRNA). *PMR4* encodes a callose synthase and its loss-of-function mutants are resistant to powdery mildew in *Arabidopsis* and tomato. The CRISPR/Cas9-mediated knockout mutants of the *PMR4* ortholog (*SIPMR4*) in tomato showed partial resistance against the powdery mildew pathogen *O. neolyopersici* (Koseoglou 2017). RNA silencing of *SIPMR4* also enhanced the resistance to powdery mildew in tomato (Huibers et al. 2013).

The new technology represented by the CRISPR/Cas-based GE opens a new era in plant precision breeding and is expected to drive the second green evolution (Chen et al. 2019). This technology is considered a novel plant breeding technique that could provide an alternative to the strict regulations applied to ‘genetically modified organisms’ (GMOs). Technically, GE can be employed in precision breeding in many ways (Chen et al. 2019): (1) knocking out genes that confer undesirable traits; (2)

knock-in and replacement to introduce new favorable alleles without linkage drag or generating allelic variants that do not exist naturally; (3) nucleotide editing to alter SNPs in either coding or noncoding regions; (4) fine-tuning gene regulation by altering gene expression, mRNA processing, and mRNA translation; and (5) development of high-throughput mutant libraries for functional genomics and genetic improvement.

In flax, the first application of GE aimed to develop an herbicide tolerant version of CDC Bethune, the most popular flax variety in Western Canada, by precisely editing the *ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE* (*EPSPS*) genes using single-stranded oligonucleotides (ssODNs) and CRISPR/Cas9 (Sauer et al. 2016). Attempts to create a new flax variety tolerant to the herbicide glyphosate are being made by CIBUS (<https://www.cibus.com/>), a precision gene-editing company located in San Diego, using their proprietary GE method.

8.3 QTL Identification and Genetic Improvement for PasmO Resistance in Flax

PasmO disease affects flax production worldwide. This fungal disease caused by *Septoria linicola* (Speg.) Garassini is widespread through all flax growing regions and infects flax plants during the entire growth season (Halley et al. 2004). Rainfall accumulation from June to August increases the incidence and severity of the disease (Halley et al. 2004). High humidity and high temperature conditions during ripening mostly promote disease incidence. The major symptoms are brown circular lesions on leaves and brown or black banding patterns interspersed with green healthy tissues on stems. PasmO negatively impacts both seed yield and fiber quality (Hall et al. 2016).

PasmO resistance is a quantitatively heritable trait. The genetic improvement of pasmo resistance is hindered by the scarcity of highly resistant germplasm and a poor understanding of its complex genetic architecture. To date, no flax cultivars are truly highly resistant to pasmo (Diederichsen et al. 2008). Current flax cultivars developed in Canada are only moderately resistant and show a narrow genetic base (You et al. 2016b). To broaden the genetic base of flax cultivars, a core collection of 407 flax accessions has been assembled from a world collection of approximately 3,500 accessions of cultivated flax maintained by Plant Gene Resources of Canada (PGRC) (Diederichsen et al. 2012; Soto-Cerda et al. 2013). We previously evaluated pasmo resistance of the flax core collection and found significant variation associated with the geographical origin (You et al. 2017a). The most pasmo-susceptible accessions originate from India and Pakistan, whereas the accessions from Europe possessed the highest levels of resistance. Of the accessions from North America, most were moderately susceptible and susceptible. Even though CN101536 was evaluated as the most resistant Canadian linseed breeding line in the flax core collection, it was

just moderately resistant to pasmo with a rating of 4.4 (You et al. 2017a). Therefore, pyramiding additional favorable alleles into current elite varieties is considered an efficient first step to develop highly resistant varieties. The *a priori* identification of QTLs associated with pasmo resistance is not only a prerequisite to perform such marker-assisted backcrossing but could also be applied to screen advanced flax breeding germplasm.

8.3.1 Genetic Panel and SNP Set for GWAS

The flax core collection of 407 accessions is a diverse genetic panel. The entire collection was re-sequenced using GBS methodology and generated 100-bp Illumina paired-end reads to an average of $17 \times$ genome coverage using the Illumina HiSeq 2000 platform (Illumina Inc., San Diego, USA). The reads were mapped to the CDC Bethune reference (Wang et al. 2012) using BWA v0.6.1. The mapped reads were analyzed as described (He et al. 2019b) and 1.7 M SNPs were obtained. These SNPs were remapped to the chromosome-scale reference (You et al. 2018a; You and Cloutier 2019). From this unfiltered SNP data set, 258,873 SNPs were extracted using the following filtering criteria: minor allele frequency (MAF) ≥ 0.05 , genotyping rate $\geq 60\%$ and pairwise correlation coefficients (r^2) among neighboring SNPs > 0.8 (International HapMap Consortium et al. 2005; Huang et al. 2010). Imputation was performed utilizing Beagle v.4.2 with default parameters (Browning and Browning 2007) to predict some of the 14.13% missing SNPs.

8.3.2 PasmO Field Resistance of the Core Collection

Evaluation of flax accessions to pasmo resistance was carried out in a pasmo nursery that was established in the 1960s. To assure sufficient pasmo infection in the nursery, additional pasmo-infested chopped straw was spread from past growth periods as inoculum between rows when plants were roughly 30-cm tall. In addition, a misting system was applied to spread water for five minutes every half hour for four weeks, except on rainy days, to ensure conidia dispersal and disease infection and development. The 391 accessions were rated for pasmo resistance in the same nursery for five consecutive years from 2012 to 2016 at the farm of Agriculture and Agri-Food Canada, Morden Research and Development Centre, Morden, Manitoba, Canada. The field trial data was adjusted using a type-2 modified augmented design (MAD2) (Lin and Poushinsky 1985).

PasmO severity, rated on a 0–9 scale, was evaluated based on symptoms on leaves and stems of all plants in a single row plot. Evaluation was conducted at four growth stages, i.e., the early (P1) and late flowering stages (P2), the green boll stage (P3), and the early brown boll stage (P4). To group the resistance of accessions, a rating of 0–2 is categorized as resistant (R), 3–4 as moderately resistant (MR), 5–6 as moderately

Table 8.4 Field evaluation criteria for pasmo severity on a scale of 0–9

Severity score	Criteria	Resistance class
0	No symptom of pasmo	R
1	<10% leaf or/and stem area affected	R
2	10–20% leaf or/and stem area affected	R
3	21–30% leaf or/and stem area affected	MR
4	31–40% leaf or/and stem area affected	MR
5	41–50% leaf or/and stem area affected	MS
6	51–60% leaf or/and stem area affected	MS
7	61–70% leaf or/and stem area affected	S
8	71–80% leaf or/and stem area affected	S
9	>80% leaf or/and stem area affected	S

susceptible (MS), and 7–9 as susceptible (S) (Table 8.4). Statistical analyses for pasmo ratings were previously described in You et al. (2013).

We observed that pasmo infection increased with growth stages and peaked at the final evaluation stage every year, which followed a nearly normal distribution (Fig. 8.1) (You et al. 2017a); thus, only the data observed at the final growth stages (P3 or P4) was used for GWAS. Although significant correlation existed among years, significant differences between years and significant genotype \times year interactions were also observed, indicating that the individual year data sets could be used for GWAS to identify environment-specific QTLs.

8.3.3 QTL Identification

A total of 370 accessions from the 391 pasmo evaluated accessions, which have both quality SNP and phenotype data, were used for GWAS. We employed three single-locus models (GLM, MLM and GEMMA) and seven multi-locus models (six implemented in mrMLM and one in FarmCPU) (Table 8.1) to identify QTNs from the 370 accessions with 258,873 SNPs. Six pasmo rating data sets were independently analyzed for GWAS: five individual year data set and the 5-year average dataset. Significant QTNs associated with the traits were detected at $\alpha = 0.05$ followed by a Bonferroni correction ($1.93 \times 10^{-7} = 0.05/258,873$ SNPs) for GLM, MLM and

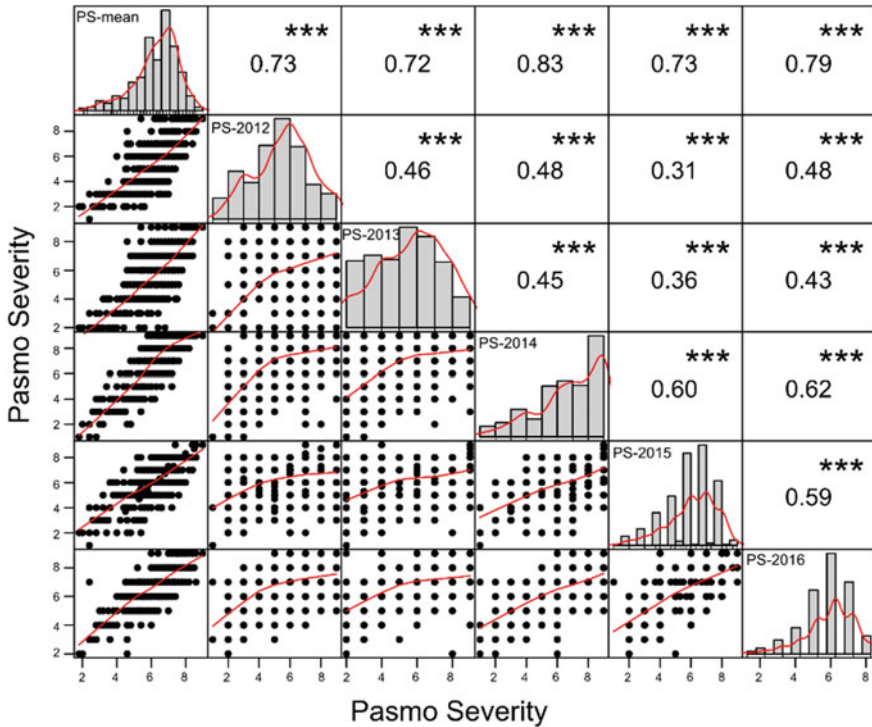


Fig. 8.1 Pearson correlations (upper triangle), scatter plots (lower triangle), and histograms (diagonal) between six pasmo severity datasets. Fitted curves are displayed in scatter plots and histograms. *** represents significance at the <0.001 probability level. *Source* He et al. (2019a)

FarmCPU, and a log of odds (LOD) score threshold of 3.0 for the remaining models. The pipeline for QTL identification and annotation is described in Fig. 8.2.

There were a total of 719 QTNs detected using the ten statistical models for the six pasmo rating datasets. These QTNs were further filtered by removing the QTNs for which the allele effect was not significant, and then grouped into 500 QTN clusters or QTLs based on LDs of contiguous markers as shown in Fig. 8.3. When there was more than one QTN in a cluster, the tag QTN with the largest QTL effect among all QTNs in the cluster was chosen to represent the QTLs. Hereafter QTN and QTL are interchangeably used.

Of these 500 QTNs, 14.4% (72) had large QTN effects ($R^2 > 10\%$), i.e., QTNs explaining a major portion of the phenotypic variance, while 24% (120) had minor effect ($R^2 < 1\%$). Several notably large-effect QTNs were identified, including Lu1-9232234 ($R^2 = 16.17\%$), Lu8-23104696 ($R^2 = 16.53\%$), Lu9-1896658 ($R^2 = 17.12\%$), and Lu9-4333365 ($R^2 = 23.39\%$).

QTN detection power varies depending on statistical models used. Single-locus models mostly identified large-effect QTLs. Of the three single-locus models, MLM identified only one large-effect QTN with $R^2 = 15.02\%$, GEMMA identified six with

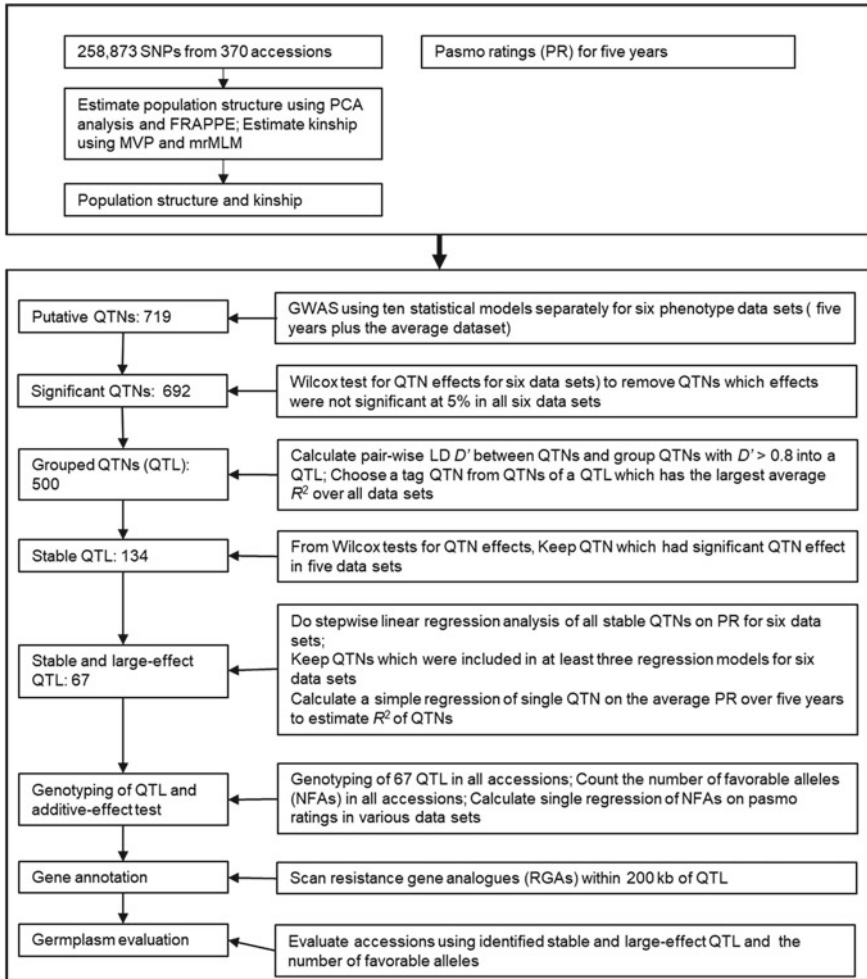


Fig. 8.2 Pipeline of quantitative trait loci (QTLs) identification using genomewide association study (GWAS) and annotation for flax pasmo resistance. *Source* Modified from He et al. (2019b)

an average R^2 of 11.13%, whereas GLM detected 209 QTNs that had an average R^2 of 5.57% and a range from 0.48 to 15.02%. Multi-locus models identified more small-effect QTNs than single-locus data models. In addition, the six mrMLM models detected more QTNs with smaller effects (average R^2 of 2.80%) than FarmCPU (average R^2 of 5.09%), because the high stringency of the Bonferroni correction was applied to FarmCPU.

The stability and reliability of the QTNs identified correlated with the number of statistical models (NSMs) and the number of pasmo phenotype datasets (NPDs) to display significant allele effects for the QTNs (Fig. 8.4). A total of 127 QTNs were identified by two or more statistical models, but most of them (373) were detected by

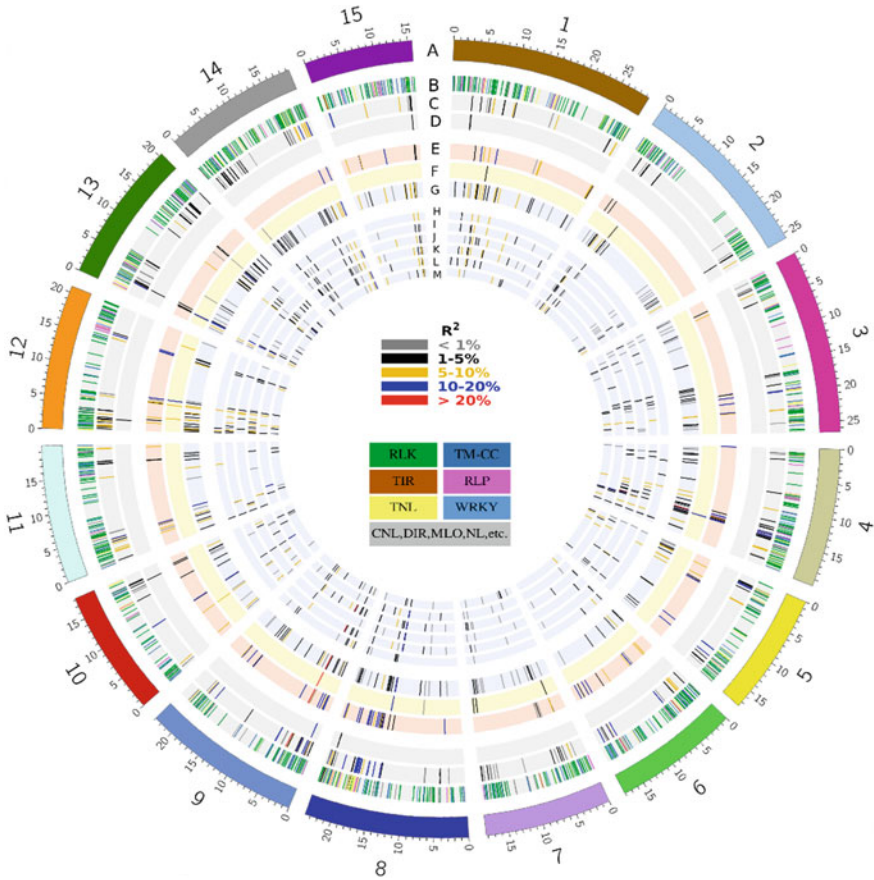


Fig. 8.3 Circos map of 500 QTNs associated with pasmo resistance measured in the field for five consecutive years and identified using ten single- and multi-locus models. Track A: flax genome chromosomes; B: 1599 resistance gene analogs (RGAs); C: 372 putative candidate RGAs for pasmo resistance; D: 8 RGAs co-located with identified QTNs; E: 209 QTNs identified by GLM; F: 22 QTNs identified by FarmCPU; G: 281 QTNs identified by all six “mrMLM models” (from H to M); H: 60 QTNs identified by FASTmrEMMA; I: 125 QTNs identified by FASTmrMLM; J: 97 QTNs identified by ISIS-EM-BLASSO; K: 97 QTNs identified by mrMLM; L: 95 QTNs identified by pKWmeB; M: 118 QTNs identified by pLARmEB

a single model. However, the effect size of QTNs was not necessarily associated with the NSMs (Fig. 8.4a), though the large-effect QTNs Lu4-14738243, Lu9-4333365 and Lu8-14317356 were all detected by more than five or all six models (Fig. 8.4a).

Nevertheless, the effect size of QTNs significantly correlated with NPDs (Fig. 8.4b), indicating that QTNs detected by a greater number of data sets were more reliable and associated with larger effect than QTNs identified in fewer data sets. Inversely, small-effect QTNs were usually identified in only one or two phenotypic datasets (or environments), indicative of their environment-specific associations.

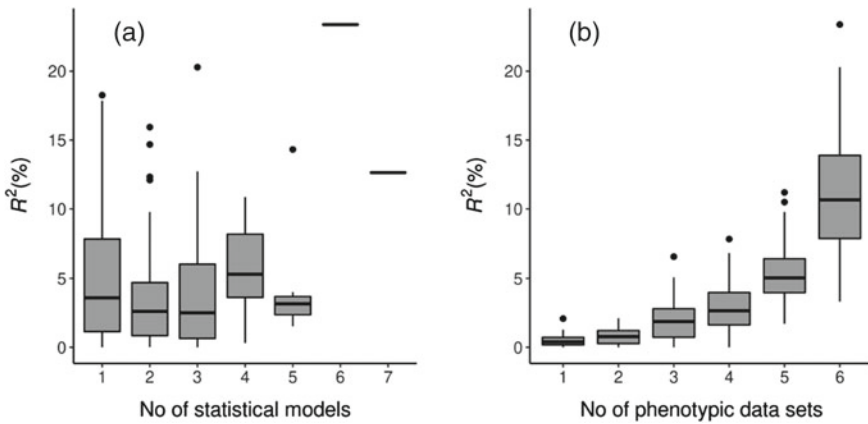


Fig. 8.4 Relationship between R^2 (phenotypic variance explained by a QTL, %) with the number of statistical models that detected the QTLs (a) and the number of pasmo phenotypic datasets that showed significant allele effects for the QTLs (b)

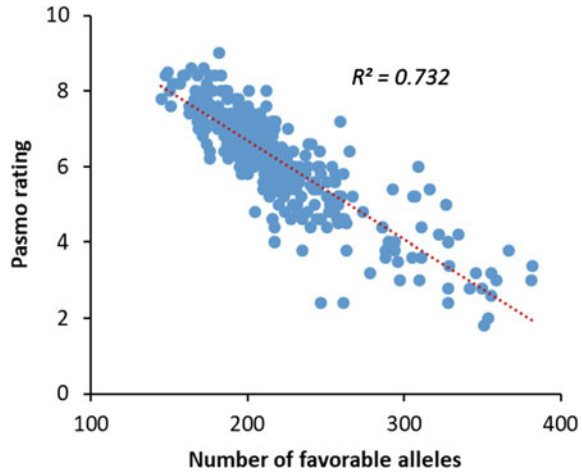
Based on the QTN effect size and the number of pasmo phenotypic datasets that showed significant QTN effect, two QTN subsets were generated from the 500 QTN set associated with pasmo resistance in flax. The first subset was defined based on 134 stable QTNs that have significant QTN effects in all six phenotypic datasets and explained 27.4–60.9% of the total variation. The second subset of 67 QTNs represented the non-redundant and stable QTN subset, which were identified by the construction of forward stepwise multiple regression models and retained in at least three models. This subset contributed 31.5–64.2% of the total variation in the six phenotypic datasets, a range comparable or moderately larger than that of the 134 QTL subset, indicating that the latter retained redundant markers.

The 500 QTN set appeared to be primarily additive for pasmo resistance. Significant negative correlation between the number of favorable alleles (NFAs) and pasmo ratings were observed ($R^2 = 0.73$) (Fig. 8.5), signifying that NFA is a good indicator or criterion to evaluate pasmo resistance of accessions.

8.3.4 Candidate Genes

To find candidate resistance genes that are co-localized with the detected QTNs, we first identified 1599 RGAs on the 15 chromosomes (Fig. 8.3, Track B), including the 1327 initially detected in the flax pseudomolecule (You et al. 2018a). We then performed genomewide scans along chromosomes to locate all the RGAs within a 200-kb window of the QTN's flanking regions. A total of 372 RGAs co-locating with 314 QTNs were thus detected. Among them, Lu1-3420323, Lu2-23730537, Lu8-22525597, Lu9-1067536, Lu10-16054459, Lu12-1874446, Lu13-2227366 and Lu15-14719354 were located in the following RGAs per se: *Lus10042324* (RLK), *Lus10030634* (RLK), *Lus10015350* (TNL), *Lus10028975* (TM-CC), *Lus10022900*

Fig. 8.5 Relationship between the number of favorable alleles and the average pasmo ratings of 370 flax accessions evaluated in the field for five consecutive years



(CNL), *Lus10023329* (TN), *Lus10026988* (RLK), and *Lus10014810* (RLK), respectively (Table 8.5, Fig. 8.3).

We further analyzed the 67 stable and large-effect QTN subset and found that 45 QTNs co-localized with 85 RGAs (Table 8.5), representing all four types, i.e., RLP, RLK, NBS coding genes, and those encoding transmembrane coiled-coil proteins (TM-CC) (Sekhwal et al. 2015). RLKs accounted for 36.47% of RGAs, while TNLs for 22.35% (He et al. 2019b).

Table 8.5 Quantitative trait nucleotides (QTNs) and putative candidate genes associated with pasmo resistance

Tag QTN	Chr	Position	SNP	Favorable allele	Effect	R ²	Gene/annotation
Lu1-3420323	1	3420323	G/A	A	0.28	2.89	<i>Lus10042324</i> /RLK ^a
Lu1-28707496	1	28707496	G/A	G	-0.54	5.7	<i>Lus10006052</i> /RLK, <i>Lus10006056</i> /RLK, <i>Lus10006057</i> /RLK, <i>Lus10006067</i> /RLK
Lu2-23730537	2	23730537	A/T	T	0.56	1.24	<i>Lus10030634</i> /RLK ^a
Lu3-19643168	3	19643168	G/A	G	-1.97	12.82	<i>Lus10008221</i> /TNL, <i>Lus10008222</i> /TNL, <i>Lus10008230</i> /RLP
Lu3-22688547	3	22688547	C/G	C	-0.89	8.98	<i>Lus10033291</i> /RLK
Lu4-14576826	4	14576826	A/G	G	0.42	7.99	<i>Lus10041509</i> /RLK, <i>Lus10041512</i> /TM-CC
Lu4-14615685	4	14615685	A/T	A	-0.65	10.85	<i>Lus10041509</i> /RLK, <i>Lus10041512</i> /TM-CC

(continued)

Table 8.5 (continued)

Tag QTN	Chr	Position	SNP	Favorable allele	Effect	R^2	Gene/annotation
Lu4-17204590	4	17204590	C/A	A	0.64	5.17	<i>Lus10004040/RLK</i> , <i>Lus10009107/TNL</i> , <i>Lus10009108/TX</i> , <i>Lus10009109/NBS</i> , <i>Lus10020794/TM-CC</i>
Lu4-17214936	4	17214936	G/T	T	0.7	5.81	<i>Lus10004040/RLK</i> , <i>Lus10009107/TNL</i> , <i>Lus10009108/TX</i> , <i>Lus10009109/NBS</i> , <i>Lus10020779/CNL</i> , <i>Lus10020794/TM-CC</i>
Lu5-1554121	5	1554121	T/A	T	-0.67	7.75	<i>Lus10004719/TNL</i> , <i>Lus10004726/CNL</i> , <i>Lus10004727/TN</i>
Lu5-1650980	5	1650980	C/G	C	-0.81	6.61	<i>Lus10004719/TNL</i> , <i>Lus10008486/RLK</i> , <i>Lus10008491/RLK</i>
Lu5-4604607	5	4604607	A/G	A	-0.56	6.58	<i>Lus10034787/TM-CC</i> , <i>Lus10034790/RLK</i> , <i>Lus10034795/RLK</i>
Lu5-13500692	5	13500692	G/A	G	-1.4	11.9	<i>Lus10029802/RLK</i> , <i>Lus10029810/TX</i>
Lu6-2081466	6	2081466	T/C	C	0.68	8.3	<i>Lus10017611/RLK</i>
Lu6-14738507	6	14738507	C/T	C	-2.01	13.34	<i>Lus10014441/RLP</i>
Lu6-15455712	6	15455712	A/G	A	-1.42	9.63	<i>Lus10021003/RLK</i> , <i>Lus10021022/RLK</i>
Lu6-15506450	6	15506450	A/G	A	-1.81	12.62	<i>Lus10021022/RLK</i>
Lu7-2452981	7	2452981	C/T	C	-0.53	6.3	<i>Lus10012159/RLK</i>
Lu7-2453965	7	2453965	T/C	T	-0.56	7.03	<i>Lus10012159/RLK</i>
Lu7-2491132	7	2491132	G/A	G	-0.56	8.05	<i>Lus10012159/RLK</i>
Lu8-14317356	8	14317356	A/T	A	-0.98	14.32	<i>Lus10016620/RLK</i> , <i>Lus10016612/RLP</i>
Lu8-16366918	8	16366918	C/T	C	-1.38	10.9	<i>Lus10022340/RLK</i> , <i>Lus10022345/RLK</i> , <i>Lus10022351/CNL</i>
Lu8-17270785	8	17270785	C/G	C	-1.08	9.59	<i>Lus10000591/TM-CC</i>
Lu8-17749357	8	17749357	G/A	G	-1.23	10.16	<i>Lus10011039/RLP</i> , <i>Lus10011064/RLP</i>

(continued)

Table 8.5 (continued)

Tag QTN	Chr	Position	SNP	Favorable allele	Effect	R ²	Gene/annotation
Lu8-18251174	8	18251174	G/A	G	-1.45	10.38	<i>Lus10007812/TNL</i> , <i>Lus10007813/TNL</i> , <i>Lus10007814/TNL</i> , <i>Lus10007821/TNL</i> , <i>Lus10007822/TNL</i> , <i>Lus10007823/OTHER</i> , <i>Lus10007825/TNL</i> , <i>Lus10007826/TNL</i> , <i>Lus10007828/TNL</i> , <i>Lus10007829/OTHER</i> , <i>Lus10007830/NL</i> , <i>Lus10007831/TNL</i> , <i>Lus10007836/TNL</i> , <i>Lus10007852/TX</i>
Lu8-18447612	8	18447612	T/C	T	-1.41	11.66	<i>Lus10007790/TNL</i> , <i>Lus10007795/TM-CC</i> , <i>Lus10007808/TNL</i> , <i>Lus10007809/NL</i> , <i>Lus10007810/TNL</i> , <i>Lus10007811/TNL</i> , <i>Lus10007812/TNL</i> , <i>Lus10007813/TNL</i> , <i>Lus10008540/RLK</i>
Lu8-22525597	8	22525597	T/C	T	-0.3	2.74	<i>Lus10015350/TNL</i> ^a
Lu8-23104696	8	23104696	C/A	C	-1.8	16.53	<i>Lus10018470/TX</i>
Lu8-23142500	8	23142500	T/C	T	-1.56	13.34	<i>Lus10018459/RLK</i> , <i>Lus10018470/TX</i>
Lu9-1067536	9	1067536	A/C	A	-0.67	5.06	<i>Lus10028975/TM-CC</i> ^a
Lu9-1430465	9	1430465	G/C	G	-0.69	10.76	<i>Lus10004333/RLK</i>
Lu9-4333365	9	4333365	C/A	C	-2.22	23.39	<i>Lus10040315/TM-CC</i>
Lu9-6270376	9	6270376	A/G	A	-0.81	14.34	<i>Lus10031043/RLK</i> , <i>Lus10031058/TM-CC</i>
Lu9-19857367	9	19857367	G/A	G	-1.7	12.67	<i>Lus10011917/RLK</i>
Lu10-8700793	10	8700793	A/G	A	-0.53	12.1	<i>Lus10039958/RLP</i>
Lu10-16054459	10	16054459	A/G	G	0.31	1.2	<i>Lus10022900/CNL</i> ^a
Lu11-3330783	11	3330783	A/T	A	-1.11	7.09	<i>Lus10042097/TM-CC</i>
Lu12-474480	12	474480	C/T	T	0.51	8.33	<i>Lus10020016/CNL</i>
Lu12-1621325	12	1621325	T/A	T	-1.9	9.41	<i>Lus10023391/RLK</i>
Lu12-1874446	12	1874446	G/A	A	0.34	4.3	<i>Lus10023329/TN</i> ^a
Lu12-2719326	12	2719326	C/T	C	-0.62	9.9	<i>Lus10006971/TM-CC</i>
Lu12-5795458	12	5795458	A/G	G	0.54	9.67	<i>Lus10037786/TM-CC</i>
Lu12-5819991	12	5819991	C/G	G	0.35	6.9	<i>Lus10037786/TM-CC</i>

(continued)

Table 8.5 (continued)

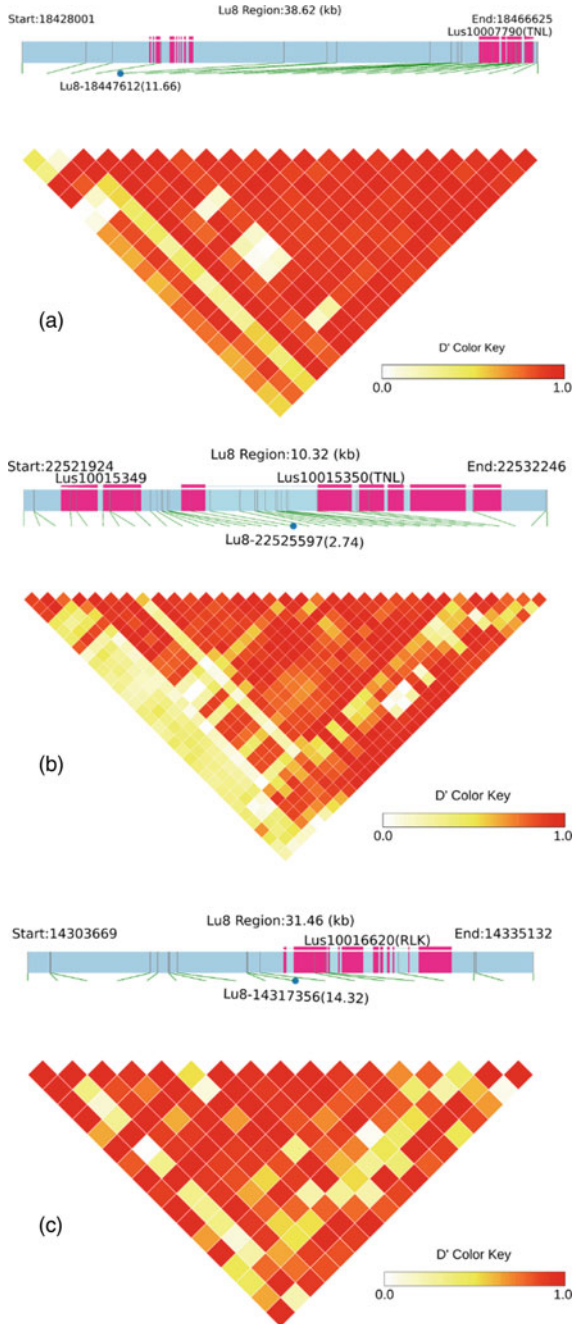
Tag QTN	Chr	Position	SNP	Favorable allele	Effect	R ²	Gene/annotation
Lu12-16056974	12	16056974	A/C	A	-1.26	11.26	<i>Lus10043083/RLK</i>
Lu13-1919638	13	1919638	G/A	G	-1.55	13.67	<i>Lus10026845/TX</i>
Lu13-2227366	13	2227366	T/C	C	0.41	1.21	<i>Lus10026988/RLK^a</i>
Lu13-14299019	13	14299019	A/G	G	0.39	8.28	<i>Lus10034637/RLK</i> , <i>Lus10034642/RLK</i>
Lu15-976617	15	976617	T/A	T	-1.65	16.08	<i>Lus10011216/TX</i> , <i>Lus10011223/RLK</i> , <i>Lus10011229/TM-CC</i>
Lu15-995626	15	995626	T/A	T	-0.44	6.27	<i>Lus10011216/TX</i> , <i>Lus10011223/RLK</i> , <i>Lus10011229/TM-CC</i>
Lu15-14719354	15	14719354	T/C	C	0.33	4.07	<i>Lus10014810/RLK^a</i>

RLK: receptor-like protein kinase; RLP: receptor-like protein; TM-CC: transmembrane coiled-coil protein; NBS: nucleotide-binding site domain; LRR: leucine-rich repeat; TIR: Toll/interleukin-1 receptor-like domain; TNL: TIR-NBS-LRRs; TN, TIR-NBS; TX, TIR-unknown.^a QTNs in genes

Of note, Chr 8 contains an important genomic region associated with pasmo resistance. A total of 49 QTNs were identified on Chr 8, and nine of them were classified stable and major QTNs with nearby candidate genes (Table 8.5). QTNs Lu8-18251174 ($R^2 = 10.38\%$) and Lu8-18447612 ($R^2 = 11.66\%$) both co-located with TNL gene clusters. Lu8-18251174 had high LD correlations with both *Lus10007830* (NL) and *Lus10007831* (TNL), while Lu8-18447612 was significantly correlated with *Lus10007790* (TNL) (Fig. 8.6a). In addition, QTN Lu8-22525597 ($R^2 = 2.74\%$) is located within TNL gene *Lus10015350* (Table 8.5 and Fig. 8.6b). Besides TNL genes in this genomic region, the RLK gene *Lus10016620* was also found to be significantly correlated with QTN Lu8-14317356 ($R^2 = 14.32\%$) (Fig. 8.6c).

Lus10031043 (RLK) and *Lus10020016* (CNL) are two candidate genes which co-locate with the major QTNs Lu9-6270375 and Lu12-474480, respectively. These two genes are orthologous to *Arabidopsis* resistance genes *AT5G20480.1* and *AT3G07040.1* (*RPM1*), respectively (Xiang et al. 2008; Saijo et al. 2009). *AT5G20480.1* encodes a leucine-rich repeat receptor kinase (LRR-RLK) and behaves as the receptor for bacterial pathogen-associated molecular patterns (PAMPs) EF-Tu (EFR). The LRR-RLK EFR can recognize the bacterial epitopes elf18 that is derived from elongation factor-Tu, and then activates the plant immune response (Saijo et al. 2009). The *Pseudomonas syringae* effector *AvrPto* has been shown to bind receptor kinases, including *Arabidopsis* LRR-RLK EFR, inhibit plant PAMP-triggered immunity and elicit strong immune responses (Xiang et al. 2008). *RPM1* has a tripartite nucleotide binding site at the N-terminal and a tandem array of leucine-rich repeats at the C-terminal, and it conveys resistance to *P. syringae* strains that carry the avirulence genes *avrB* and *avrRpm1*. The *RPM1* gene confers dual pathogen specificity that expresses either of the two unrelated *P. syringae* avirulence genes (Grant et al.

Fig. 8.6 Linkage disequilibrium plots for three QTNs associated with pasmo resistance (a). QTN Lu8-18447612 ($R^2 = 11.66\%$) co-located with the gene *Lus10007790* (TNL); (b). QTN Lu8-22525597 ($R^2 = 2.74\%$) located within the gene *Lus10015350* (TNL); (c). QTN Lu8-14317356 ($R^2 = 14.32\%$) co-located with the gene *Lus10016620* (RLK). The values in parentheses after QTN names are R^2 values



1995). Therefore, *Lus10031043* and *Lus10020016* are two additional candidate genes deserving further functional analyses.

8.3.5 Genomic Evaluation of the Resistance Germplasm

Flax has two morphotypes: seed and fiber. PasmO resistance correlates with these morphotypes. Significant correlations between morphotype and pasmo ratings ($r = 0.49$, $p < 0.00001$) as well as between morphotype and NFAs ($r = -0.65$, $p < 0.00001$) were observed in the diversity panel which comprised 80 fiber and 290 linseed accessions (Fig. 8.7). Fiber accessions generally appeared to be more resistant to pasmo than linseed accessions. This is likely an indication that fiber flax breeders have expended greater effort into breeding for pasmo resistance than linseed breeders because fiber flax quality can be greatly affected by high pasmo incidence. Aside from artificial selection by breeders, long term natural selection and probably independent domestication of the fiber flax may also account for the differential in pasmo resistance between the morphotypes (Fu et al. 2012).

A variety of pasmo resistance was observed in the core collection (You et al. 2017a), allowing further investigations on a genomic scale. Making use of the QTN information of the genotypes, we identified 14 accessions with resistant phenotypes and high numbers of favorable alleles (Table 8.6). For instance, the fiber accession CN19001 from the Netherlands and the linseed accession CN101367 from Georgia, have average pasmo ratings of 2.0 and 1.8 and 354 and 351 favorable alleles, respectively. Netherlands's accessions CN40081 and CN33390 had the most favorable alleles but slightly higher pasmo ratings than the previous two. It is also notable that ten of the 14 resistant accessions are fibers. These fiber and linseed accessions are good parents to further improve flax resistance to pasmo through direct cross breeding through the pyramiding of favorable alleles into elite varieties.

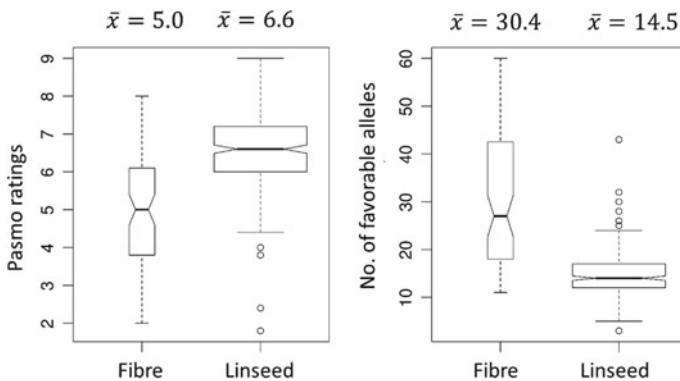


Fig. 8.7 Boxplots of flax morphotypes in terms of flax pasmo ratings and number of favorable alleles in the accessions. *Source* Modified from He et al. (2019b)

Table 8.6 Genetic resources resistant to pasmo disease identified by genomic and phenotypic evaluation

Accession	Country	Morphotype	Pasmo rating	No. of favorable alleles ^a
CN40081	Netherlands	Fiber	3.4	382
CN33390	Netherlands	Fiber	3	381
CN101053	China	Fiber	3	359
CN100929	Netherlands	Fiber	2.6	356
CN18982	France	Fiber	3.2	356
CN19001	Netherlands	Fiber	2	354
CN101367	Georgia	Linseed	1.8	351
CN18983	Netherlands	Fiber	2.8	350
CN18988	France	Fiber	3.2	346
CN101298	Russian	Linseed	2.8	342
CN100939	Russian	Linseed	2.4	328
CN101419	China	Fiber	2.8	328
CN101230	China	Fiber	3	310
CN101299	Russian	Linseed	3	297

^a Out of 500 QTNs

8.3.6 Evaluation of Genomic Selection (GS)

For complex quantitatively heritable traits, the major purpose of genomewide QTL identification is to provide molecular markers for breeding selection. Some large-effect QTLs such as Lu9-4333365, Lu4-14213405, Lu5-14838893, Lu4-13813266 and Lu9-1896658, have R^2 values exceeding 17%, which could be useful for MAS, but most of the QTNs identified have small allele effects, which would not be considered for MAS but could be valuable for GS. To explore the values of these QTNs in GS, we first assessed the efficiency of various GS models to ascertain the best model for GS of pasmo resistance. The GS models RR-BLUP, GBLUP, BL, BRR, BayesA, BayesB, BayesC, RFR, RKHS and SVR were evaluated using the 500 QTN subset as marker input and the five-year average pasmo rating dataset as the phenotype. The five-fold cross-validation results revealed the same prediction ability (r) of 0.92 for 9/10 models, exception being RFR which had a prediction ability of 0.79 (Fig. 8.8).

We further evaluated GS models with different marker sets to determine the best marker set in the development of GS model for pasmo resistance. Six different marker sets were tested with the six pasmo phenotype datasets using the random five-fold cross-validation scheme. The marker sets were three SNP data sets (SNP-66723, SNP-9415 and SNP-3057) and three QTL data sets (QTL-500, QTL-134 and QTL-67). SNP-66723 was selected from the 258,873 SNP data set by a Pearson's χ^2 test with Yate's continuity correction to identify all SNPs related to pasmo ratings. SNP-9415 and SNP-3057 are two subsets of SNP-65723 that were selected with probability value thresholds of 0.01 and 0.001, respectively. QTL-67, QTL-134 and QTL-500

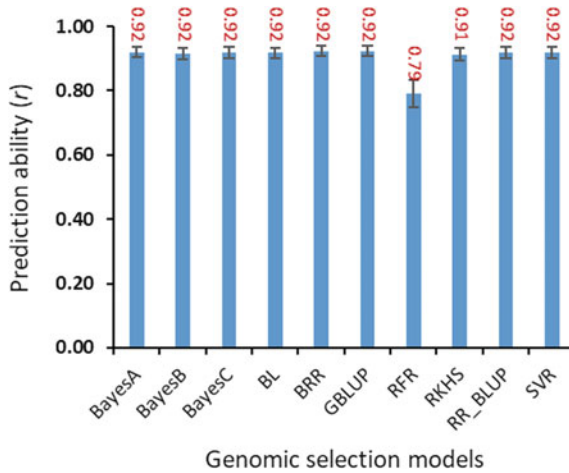


Fig. 8.8 Comparison of prediction ability (r) of ten genomic selection (GS) models. The GWAS-derived 500 QTN subset (QTL-500) with the five-year average pasmo rating dataset were used for GS model construction

represent the 500 GWAS-derived unique QTLs, the 134 statistically stable QTLs and 67 non-redundant and stable QTL subsets, respectively. QTL-67 is contained in QTL-134, which is in turn contained in QTL-500. RR-BLUP was used to construct the GS models. Results showed that the GS models with QTL markers consistently outperformed those with SNP markers for all pasmo phenotypic datasets (Fig. 8.9), similarly to our previous results on seven breeding target traits (Lan et al. 2020).

In the three QTL marker based GS models, GS models built from QTL-500 significantly outperformed those from QTL-134 and QTL-67, indicating that at least a portion of the minor-effect QTNs contribute positively to the development of the GS models. The similar prediction ability of the two smaller marker sets was anticipated since QTL-67 is fundamentally a non-redundant set of QTL-134. These GS prediction results indirectly serve as a validation of the QTL identified via GWAS. In addition, a prediction ability as high as 0.92, seen in the GS models clearly illustrates the effectiveness of genomic prediction for pasmo resistance by employing a comprehensive range of stable or environment-specific QTLs with large- and small-effect QTLs.

8.4 Future Perspectives

Resistance to diseases such as pasmo, Fusarium wilt and powdery mildew is a complex quantitative trait in flax. The conventional approach to flax genetic improvement still involves cross breeding through hybridization of two parents followed by offspring segregation and phenotypic selection. In such conventional approach,

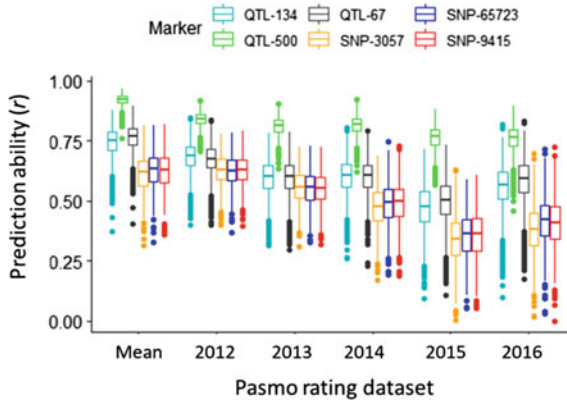


Fig. 8.9 Comparison of prediction ability (r) of RR-BLUP prediction models constructed using six different marker sets and the five-year average pasmo rating dataset using a random five-fold cross-validation scheme. SNP-66723 is a SNP subset selected from 258,873 SNPs by a Pearson's χ^2 test with Yate's continuity correction to identify all SNPs statistically correlated with pasmo ratings. SNP-9415 and SNP-3057 are two subsets of SNP-65723 that were selected at different probability thresholds. QTL-67, QTL-134 and QTL-500 represent the 500 unique QTL, the 135 stable QTL and the 67 non-redundant QTL subsets identified by GWAS, respectively. QTL-67 is comprised within QTL-134, which is in turn comprised within QTL-500

the quantitative inheritance nature of these disease resistances impedes the rapid pyramiding of desirable or resistant alleles/genes from donor parents into a single plant, resulting in slow advance in resistance breeding for these biotic stresses in flax. To date, the majority of registered flax varieties are moderately resistant to pasmo, Fusarium wilt and powdery mildew. However, large-scale QTL identification through linkage-based QTL mapping and GWAS has already identified a large number of QTLs associated with biotic stresses in flax, including large- and minor effect QTLs. QTL markers identified from the flax core collection offer the potential to enhance selection accuracy and efficiency of cross breeding through GS. In addition, QTL markers of parents can be combined with genetic simulation to generate virtual crosses and their offspring populations (Khan et al. 2022). Then GS can be applied to predict GCA of parents and SCA of the virtual crosses, which facilitate parent selection and cross making to make best crosses.

The “breeding by design” was proposed by Peleman and Voort (2003), aiming to gather favorable alleles or QTLs associated with breeding target traits from potential genetic resources to develop superior varieties. We have identified an array of QTNs related to the traits of interest, including biotic stresses, and deciphered the distribution of the favorable alleles on the genetic resources. We also found that the identified QTNs were primarily additive. Therefore, this offers a genomic approach to evaluate all genetic resources based on their genomewide QTN content. Furthermore, based on complementarity of favorable alleles among parents, suitable parents can be selected to “design” potential superior varieties. Such varieties may contain

all favorable alleles in one variety and can be implemented through conventional breeding, MAS and GS.

Some candidate genes have been predicted for some of the significant QTNs, but validation and characterization of these candidate genes via functional genomic approaches remain challenging. Once their functions are validated and functional markers are developed, precision breeding through gene editing technologies is expected to be a revolutionary strategy towards rapid and accurate pyramiding of multiple resistant genes into elite flax varieties. The impending first successful application of GE in flax has the potential to accelerate the deployment of precision breeding technologies in flax genetic improvement.

Acknowledgements We thank Ms. Chunfang Zheng for her assistance in preparation of some figures.

References

- Acevedo-Garcia J, Kusch S, Panstruga R (2014) Magical mystery tour: MLO proteins in plant immunity and beyond. *New Phytol* 204:273–281
- Anderson PA, Lawrence GJ, Morrish BC, Ayliffe MA, Finnegan EJ et al (1997) Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeat coding region. *Plant Cell* 9:641–651
- Arends D, Prins P, Jansen RC, Broman KW (2010) R/qtl: high-throughput multiple QTL mapping. *Bioinformatics* 26:2990–2992
- Arruda MP, Brown PJ, Lipka AE, Krill AM, Thurber C et al (2015) Genomic selection for predicting *Fusarium* head blight resistance in a wheat breeding program. *Plant Genome* 8. <https://doi.org/10.3835/plantgenome2015.3801.0003>
- Asgarinia P, Cloutier S, Duguid S, Rashid K, Mirlohi A et al (2013) Mapping quantitative trait loci for powdery mildew resistance in flax (*Linum usitatissimum* L.). *Crop Sci* 53:2462–2472
- Bekele WA, Itaya A, Boyle B, Yan W, Mitchell Fetch J et al (2020) A targeted genotyping-by-sequencing tool (Rapture) for genomics-assisted breeding in oat. *Theor Appl Genet* 133:653–664
- Bernardo R (2015) Genomewide selection of parental inbreds: classes of loci and virtual biparental populations. *Crop Sci* 55:2586–2595
- Bibikova M, Golic M, Golic KG, Carroll D (2002) Targeted chromosomal cleavage and mutagenesis in drosophila using zinc-finger nucleases. *Genetics* 161:1169–1175
- Bradley CA, Halley S, Henson RA (2007) Evaluation of fungicide seed treatments on flax cultivars differing in seed color. *Ind Crops Prod* 25:301–304
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y et al (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635
- Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 81:1084–1097
- Camargo AV, Mackay I, Mott R, Han J, Doonan JH et al (2018) Functional mapping of quantitative trait loci (QTLs) associated with plant performance in a wheat MAGIC mapping population. *Front Plant Sci* 9:887
- Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr Opin Plant Biol* 11:215–221

- Chandrasekaran J, Brumin M, Wolf D, Leibman D, Klap C et al (2016) Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol Plant Pathol* 17:1140–1153
- Chantreau M, Grec S, Gutierrez L, Dalmais M, Pineau C et al (2013) PT-Flax (phenotyping and TILLinG of flax) development of a flax (*Linum usitatissimum* L.) mutant population and TILLinG platform for forward and reverse genetics. *BMC Plant Biol* 13:159
- Charmet G, Tran LG, Auzanneau J, Rincet R, Bouchet S (2020) BWGS: a R package for genomic selection and its application to a wheat breeding programme. *PLoS One* 15:e0222733
- Chen K, Wang Y, Zhang R, Zhang H, Gao C (2019) CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu Rev Plant Biol* 70:667–697
- Covarrubias-Pazarán G (2016) Genome-assisted prediction of quantitative traits using the R package *sommer*. *PLoS One* 11:e0156744
- Covarrubias-Pazarán G, Schlautman B, Diaz-García L, Grygleski E, Polashock J et al (2018) Multi-variate GBLUP improves accuracy of genomic selection for yield and fruit weight in biparental populations of *Vaccinium macrocarpon* Ait. *Front Plant Sci* 9:1310
- Daetwyler HD, Bansal UK, Bariana HS, Hayden MJ, Hayes BJ (2014) Genomic prediction for rust resistance in diverse wheat landraces. *Theor Appl Genet* 127:1795–1803
- de los Campos G, Naya H, Gianola D, Crossa J, Legarra A et al (2009) Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182:375–385
- Deyholos MK (2006) Bast fiber of flax (*Linum usitatissimum* L.): biological foundations of its ancient and modern uses. *Israel J Plant Sci* 54:273–280
- Di Pietro A, Garcia-Maceira FI, Meglecz E, Roncero MI (2001) A MAP kinase of the vascular wilt fungus *Fusarium oxysporum* is essential for root penetration and pathogenesis. *Mol Microbiol* 39:1140–1152
- Diederichsen A, Kusters PM, Kessler D, Baines Z, Gugel RK (2012) Assembling a core collection from the flax world collection maintained by Plant Gene Resources of Canada. *Genet Resour Crop Evol* 60:1479–1485
- Diederichsen A, Rozhmina TA, Kudrjavceva LP (2008) Variation patterns within 153 flax (*Linum usitatissimum* L.) genebank accessions based on evaluation for resistance to fusarium wilt, anthracnose and pasmo. *Plant Genet Resour* 6:22–32
- Dodds PN, Lawrence GJ, Ellis JG (2001a) Contrasting modes of evolution acting on the complex *N* locus for rust resistance in flax. *Plant J* 27:439–453
- Dodds PN, Lawrence GJ, Ellis JG (2001b) Six amino acid changes confined to the leucine-rich repeat β -strand/ β -turn motif determine the difference between the *P* and *P2* rust resistance specificities in flax. *Plant Cell* 13:163–178
- Dong H, Wang R, Yuan Y, Anderson J, Pumphrey M et al (2018) Evaluation of the potential for genomic selection to improve spring wheat resistance to Fusarium head blight in the Pacific Northwest. *Front Plant Sci* 9:911
- Edirisinghe VP (2016) Characterization of flax germplasm for resistance to Fusarium wilt caused by *Fusarium oxysporum* f. sp. *lini*. Department of Plant Science. University of Saskatchewan, p 120
- Ellis JG, Lawrence GJ, Luck JE, Dodds PN (1999) Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. *Plant Cell* 11:495–506
- Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package *rrBLUP*. *Plant Genome* 4:250–255
- Falconer DS, Mackay FC (1996) Introduction to quantitative genetics. Benjamin-Cummings Pub Co
- Fatima F, McCallum BD, Pozniak CJ, Hiebert CW, McCartney CA et al (2020) Identification of new leaf rust resistance loci in wheat and wild relatives by array-based SNP genotyping and association genetics. *Front Plant Sci* 11:583738
- Fernandes SB, Dias KOG, Ferreira DF, Brown PJ (2018) Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. *Theor Appl Genet* 131:747–755

- Fofana B, Ghose K, Somalraju A, McCallum J, Main D et al (2017) Induced mutagenesis in *UGT74S1* gene leads to stable new flax lines with altered secoisolariciresinol diglucoside (SDG) profiles. *Front Plant Sci* 8:1638
- Fofana B, Ragupathy R, Cloutier S (2010) Flax lipids: classes, biosynthesis, genetics and the promise of applied genomics for understanding and altering of fatty acids. In: Gilmore PL (ed) *Lipids: categories, biological functions and metabolism, nutrition, and health*. Nova Science Publishers Inc., New York, pp 71–98
- Foulk JA, Akin DE, Dodd RB, Frederick JR (2004) Optimising flax production in the South Atlantic region of the USA. *J Sci Food Agri* 84:870–876
- Fu Y-B, Diederichsen A, Allaby RG (2012) Locus-specific view of flax domestication history. *Ecol Evol* 2:139–152
- Fukuoka Y (2019) Machine learning approach for predicting new uses of existing drugs and evaluation of their reliabilities. *Methods Mol Biol* 1903:269–279
- Gonzalez-Camacho JM, Ornella L, Perez-Rodriguez P, Gianola D, Dreisigacker S et al (2018) Applications of machine learning methods to genomic selection in breeding wheat for rust resistance. *Plant Genome* 11. <https://doi.org/10.3835/plantgenome2017.3811.0104>
- Goutam U, Kukreja S, Yadav R, Salaria N, Thakur K et al (2015) Recent trends and perspectives of molecular markers against fungal diseases in wheat. *Front Microbiol* 6:861
- Granato I, Cuevas J, Luna-Vazquez F, Crossa J, Montesinos-Lopez O et al (2018) BGGE: a new package for genomic-enabled prediction incorporating genotype x environment interaction models. *G3 (Bethesda)* 8:3039–3047
- Grant MR, Godiard L, Straube E, Ashfield T, Lewald J et al (1995) Structure of the Arabidopsis *RPM1* gene enabling dual specificity disease resistance. *Science* 269:843–846
- Grinberg NF, Orhobor OI, King RD (2020) An evaluation of machine-learning for predicting phenotype: studies in yeast, rice, and wheat. *Mach Learn* 109:251–277
- Gupta S, Ko DT, Azizi P, Bouadjenek MR, Koh M et al (2020) Evaluation of machine learning algorithms for predicting readmission after acute myocardial infarction using routinely collected clinical data. *Can J Cardiol* 36:878–885
- Hall LM, Booker H, Siloto RMP, Jhala AJ, Weselake RJ (2016) Flax (*Linum usitatissimum* L.). In: McKeon TA, Hayes DG, Hidebrand DF, Weselake RJ (eds) *Industrial oil crops*. AOCs Press, Urbana, IL, pp 157–194
- Halley S, Bradley CA, Lukach JR, McMullen M, Knodel JJ et al (2004) Distribution and severity of pasmo on flax in North Dakota and evaluation of fungicides and cultivars for management. *Plant Dis* 88:1123–1126
- Hammond-Kosack KE, Jones JDG (1997) Plant disease resistance genes. *Annu Rev Plant Physiol Plant Mol Biol* 48:575–607
- He J, Meng S, Zhao T, Xing G, Yang S et al (2017) An innovative procedure of genome-wide association analysis fits studies on germplasm population and plant breeding. *Theor Appl Genet* 130:2327–2343
- He L, Xiao J, Rashid KY, Jia G, Li P et al (2019a) Evaluation of genomic prediction for pasmo resistance in flax. *Intl J Mol Sci* 20:359
- He L, Xiao J, Rashid KY, Yao Z, Li P et al (2019b) Genome-wide association studies for pasmo resistance in flax (*Linum usitatissimum* L.). *Front Plant Sci* 9:1982
- Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–447
- Huang X, Wei X, Sang T, Zhao Q, Feng Q et al (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* 42:961–967
- Huibers RP, Loonen AE, Gao D, Van den Ackerveken G, Visser RG et al (2013) Powdery mildew resistance in tomato by impairment of *SIPMR4* and *SIDMRI*. *PLoS One* 8:e67467
- International HapMap Consortium, Altshuler D, Donnelly PA (2005) A haplotype map of the human genome. *Nature* 437:1299–1320
- Joehanes R, Nelson JC (2008) QGene 4.0, an extensible Java QTL-analysis platform. *Bioinformatics* 24:2788–2789

- Juita DBZ, Kennedy EM, Mackie JC (2012) Low temperature oxidation of linseed oil: a review. *Fiv Sci Rev* 1:3
- Khan N, You FM, Cloutier S (2022) Designing genomic solutions to enhance abiotic Stress resistance in flax. In: Kole C (ed) Genomic designing for abiotic stress resistant oilseed crops. https://doi.org/10.1007/978-3-030-90044-1_8
- Kim SA, Brossard M, Roshandel D, Paterson AD, Bull SB et al (2019) gpart: human genome partitioning and visualization of high-density SNP data by identifying haplotype blocks. *Bioinformatics* 35:4419–4421
- Kim Y, Ilich JZ (2011) Implications of dietary α -linolenic acid in bone health. *Nutrition* 27:1101–1107
- Koseoglou E (2017) The study of SIPMR4 CRISPR/Cas9-mediated tomato allelic series for resistance against powdery mildew. Wageningen University, Wageningen, The Netherlands, p 29
- Kulwal PL (2018) Trait mapping approaches through linkage mapping in plants. *Adv Biochem Eng Biotechnol* 164:53–82
- Kumar J, Choudhary AK, Solanki RK, Pratap A (2011) Towards marker-assisted selection in pulses: a review. *Plant Breed* 130:297–313
- Kumar S, You FM, Duguid S, Booker H, Rowland G et al (2015) QTL for fatty acid composition and yield in linseed (*Linum usitatissimum* L.). *Theor Appl Genet* 128:965–984
- Lado B, Battenfield S, Guzman C, Quincke M, Singh RP et al (2017) Strategies for selecting crosses using genomic prediction in two wheat breeding programs. *Plant Genome* 10. <https://doi.org/10.3835/plantgenome2016.3812.0128>
- Lan S, Zheng C, Hauck K, McCausland M, Duguid SD et al (2020) Genomic prediction accuracy of seven breeding selection traits improved by QTL identification in flax. *Int J Mol Sci* 21:1577
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ et al (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174–181
- Langner T, Kamoun S, Belhaj K (2018) CRISPR crops: plant genome editing toward disease resistance. *Annu Rev Phytopathol* 56:479–512
- Lawrence GJ, Dodds PN, Ellis JG (2010) Transformation of the flax rust fungus, *Melampsora lini*: selection via silencing of an avirulence gene. *Plant J* 61:364–369
- Lawrence GJ, Finnegan EJ, Ayliffe MA, Ellis JG (1995) The *L6* gene for flax rust resistance is related to the Arabidopsis bacterial resistance gene *RPS2* and the tobacco viral resistance gene *N*. *Plant Cell* 7:1195–1206
- Leslie JF, Summerell BA (2006) The *Fusarium* laboratory manual. Blackwell Publishing, Ames, IA
- Leyva DR, Zahradka P, Ramjiawan B, Guzman R, Aliani M et al (2011) The effect of dietary flaxseed on improving symptoms of cardiovascular disease in patients with peripheral artery disease: rationale and design of the FLAX-PAD randomized controlled trial. *Contemp Clin Trials* 32:724–730
- Li H, Zhang L, Hu J, Zhang F, Chen B et al (2017) Genome-wide association mapping reveals the genetic control underlying branch angle in rapeseed (*Brassica napus* L.). *Front Plant Sci* 8:1054
- Li P, Quan X, Jia G, Xiao J, Cloutier S et al (2016) RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC Genomics* 17:852
- Li T, Liu B, Spalding MH, Weeks DP, Yang B (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat Biotechnol* 30:390–392
- Liabeuf D, Sim SC, Francis DM (2018) Comparison of marker-based genomic estimated breeding values and phenotypic evaluation for selection of bacterial spot resistance in tomato. *Phytopathology* 108:392–401
- Lin CS, Poushinsky G (1985) A modified augmented design (type 2) for rectangular plots. *Can J Plant Sci* 65:743–749

- Liu FH, Chen X, Long B, Shuai RY, Long CL (2011) Historical and botanical evidence of distribution, cultivation and utilization of *Linum usitatissimum* L. (flax) in China. *Veg Hist Archaeobotany* 20:561
- Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet* 12:e1005767
- Lo-Ciganic WH, Huang JL, Zhang HH, Weiss JC, Wu Y et al (2019) Evaluation of machine-learning algorithms for predicting opioid overdose risk among medicare beneficiaries with opioid prescriptions. *JAMA Netw Open* 2:e190968
- Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci* 12:57–63
- Malzahn A, Lowder L, Qi Y (2017) Plant genome editing with TALEN and CRISPR. *Cell Biosci* 7:21
- Meng L, Li H, Zhang L, Wang J (2015) QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J* 3:269–283
- Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Miedaner T, Korzun V (2012) Marker-assisted selection for disease resistance in wheat and barley breeding. *Phytopathology* 102:560–566
- Miglani GS, Kaur A, Kaur L (2020) Plant gene expression control using genome- and epigenome-editing technologies. *J Crop Improv* 34:1–63
- Monir MM, Zhu J (2018) Dominance and epistasis interactions revealed as important variants for leaf traits of maize NAM population. *Front Plant Sci* 9:627
- Montesinos-Lopez OA, Montesinos-Lopez A, Crossa J, Gianola D, Hernandez-Suarez CM et al (2018) Multi-trait, multi-environment deep learning modeling for genomic-enabled prediction of plant traits. *G3 (Bethesda)* 8:3829–3840
- Montesinos-Lopez OA, Montesinos-Lopez A, Luna-Vazquez FJ, Toledo FH, Perez-Rodriguez P et al (2019a) An R package for Bayesian analysis of multi-environment and multi-trait multi-environment data for genome-based prediction. *G3 (Bethesda)* 9:1355–1369
- Montesinos-Lopez OA, Montesinos-Lopez A, Tuberosa R, Maccaferri M, Sciarra G et al (2019b) Multi-trait, multi-environment genomic prediction of durum wheat with genomic best linear unbiased predictor and deep learning methods. *Front Plant Sci* 10:1311
- Nekrasov V, Wang C, Win J, Lanz C, Weigel D et al (2017) Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Sci Rep* 7:482
- Ogura T, Busch W (2015) From phenotypes to causal sequences: using genome wide association studies to dissect the sequence basis for variation of plant development. *Curr Opin Plant Biol* 23:98–108
- Ongom PO, Ejeta G (2018) Mating design and genetic structure of a multi-parent advanced generation intercross (MAGIC) population of sorghum (*Sorghum bicolor* (L.) Moench). *G3 (Bethesda)* 8:331–341
- Ornella L, Singh S, Perez P, Burgueño J, Singh R et al (2012) Genomic prediction of genetic values for resistance to wheat rusts. *Plant Genome* 5:136–148
- Osakabe K, Osakabe Y, Toki S (2010) Site-directed mutagenesis in *Arabidopsis* using custom-designed zinc finger nucleases. *Proc Natl Acad Sci USA* 107:12034–12039
- Pareek M, Rajam MV (2017) RNAi-mediated silencing of MAP kinase signalling genes (*Fmk1*, *Hog1*, and *Pbs2*) in *Fusarium oxysporum* reduces pathogenesis on tomato plants. *Fungal Biol* 121:775–784
- Park T, Casella G (2008) The Bayesian Lasso. *J Am Stat Assoc* 103:681–686
- Peleman JD, van der Voort JR (2003) Breeding by design. *Trends Plant Sci* 8:330–334
- Peng A, Chen S, Lei T, Xu L, He Y et al (2017) Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene *CsLOB1* promoter in citrus. *Plant Biotechnol J* 15:1509–1519
- Perez P, de los Campos G (2014) Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198:483–495

- Piepho HP (2009) Ridge regression and extensions for genomewide selection in maize. *Crop Sci* 49:1165–1176
- Poland J, Rutkoski J (2016) Advances and challenges in genomic selection for disease resistance. *Annu Rev Phytopathol* 54:79–98
- Price AH (2006) Believe it or not, QTLs are accurate! *Trends Plant Sci* 11:213–216
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA et al (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904–909
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575
- Ren D, Fang X, Jiang P, Zhang G, Hu J et al (2018) Genetic architecture of nitrogen-deficiency tolerance in wheat seedlings based on a nested association mapping (NAM) population. *Front Plant Sci* 9:845
- Ren WL, Wen YJ, Dunwell JM, Zhang YM (2017) pKWmEB: integration of Kruskal-Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. *Heredity (Edinb)* 120:208–218
- Rowland GG (1991) An EMS-induced low-linolenic-acid mutant in McGregor flax (*Linum usitatissimum* L.). *Can J Plant Sci* 71:393–396
- Rutkoski J, Benson J, Jia Y, Brown-Guedira G, Jannink J-L et al (2012) Evaluation of genomic prediction methods for fusarium head blight resistance in wheat. *Plant Genome* 5:51–61
- Rutkoski J, Singh RP, Huerta-Espino J, Bhavani S, Poland J et al (2015) Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat. *Plant Genome* 8. <https://doi.org/10.3835/plantgenome2014.3810.0074>
- Rutkoski JE, Poland JA, Singh RP, Huerta-Espino J, Bhavani S et al (2014) Genomic selection for quantitative adult plant stem rust resistance in wheat. *Plant Genome* 7. <https://doi.org/10.3835/plantgenome2014.3802.0006>
- Saijo Y, Tintor N, Lu X, Rauf P, Pajeroska-Mukhtar K et al (2009) Receptor quality control in the endoplasmic reticulum for plant innate immunity. *EMBO J* 28:3439–3449
- Sauer NJ, Narvaez-Vasquez J, Mozoruk J, Miller RB, Warburg ZJ et al (2016) Oligonucleotide-mediated genome editing provides precision and function to engineered nucleases and antibiotics in plants. *Plant Physiol* 170:1917–1928
- Segura V, Vilhjalmsón BJ, Platt A, Korte A, Seren U et al (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet* 44:825–830
- Sehgal D, Singh R, Rajpal VR (2016) Quantitative trait loci mapping in plants: concepts and approaches. In: Rajpal VR, Rao SR, Raina SN (eds) *Molecular breeding for sustainable crop improvement*, vol 2. Springer, Cham, pp 31–59
- Sekhwil MK, Li P, Lam I, Wang X, Cloutier S et al (2015) Disease resistance gene analogs (RGAs) in plants. *Int J Mol Sci* 16:19248–19290
- Sertse D, You FM, Ravichandran S, Cloutier S (2019) The complex genetic architecture of early root and shoot traits in flax revealed by genome-wide association analyses. *Front Plant Sci* 10:1483–1483
- Sertse D, You FM, Ravichandran S, Soto-Cerda BJ, Duguid S et al (2020) Loci harboring genes with important role in drought and related abiotic stress responses in flax revealed by multiple GWAS models. *Theor Appl Genet*. <https://doi.org/10.1007/s00122-020-03691-0>
- Shukla VK, Doyon Y, Miller JC, DeKaveler RC, Moehle EA et al (2009) Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* 459:437–441
- Soto-Cerda B, Diederichsen A, Ragupathy R, Cloutier S (2013) Genetic characterization of a core collection of flax (*Linum usitatissimum* L.) suitable for association mapping studies and evidence of divergent selection between fiber and linseed types. *BMC Plant Biol* 13:78
- Soto-Cerda BJ, Duguid S, Booker H, Rowland G, Diederichsen A et al (2014a) Association mapping of seed quality traits using the Canadian flax (*Linum usitatissimum* L.) core collection. *Theor Appl Genet* 127:881–896

- Soto-Cerda BJ, Duguid S, Booker H, Rowland G, Diederichsen A et al (2014b) Genomic regions underlying agronomic traits in linseed (*Linum usitatissimum* L.) as revealed by association mapping. *J Integr Plant Biol* 56:75–87
- Spielmeier W, Lagudah ES, Mendham N, Green AG (1998) Inheritance of resistance to flax wilt (*Fusarium oxysporum* f.sp. *lini* Schlecht) in a doubled haploid population of *Linum usitatissimum* L. *Euphytica* 101:287–291
- Spindel J, Begum H, Akdemir D, Virk P, Collard B et al (2015) Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet* 11:e1004982
- Tamba CL, Ni YL, Zhang YM (2017) Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PLoS Comput Biol* 13:e1005357
- Touré A, Xueming X (2010) Flaxseed lignans: source, biosynthesis, metabolism, antioxidant activity, bio-active components, and health benefits. *Compr Rev Food Sci Food Saf* 9:261–269
- Varona L, Legarra A, Toro MA, Vitezica ZG (2018) Non-additive effects in genomic selection. *Front Genet* 9:78
- Wang Q, Cobine PA, Coleman JJ (2018) Efficient genome editing in *Fusarium oxysporum* based on CRISPR/Cas9 ribonucleoprotein complexes. *Fungal Genet Biol* 117:21–29
- Wang SB, Feng JY, Ren WL, Huang B, Zhou L et al (2016a) Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci Rep* 6:19444
- Wang SB, Wen YJ, Ren WL, Ni YL, Zhang J et al (2016b) Mapping small-effect and linked quantitative trait loci for complex traits in backcross or DH populations via a multi-locus GWAS methodology. *Sci Rep* 6:29951
- Wang Y, Cheng X, Shan Q, Zhang Y, Liu J et al (2014) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat Biotechnol* 32:947–951
- Wang Z, Hobson N, Galindo L, Zhu S, Shi D et al (2012) The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. *Plant J* 72:461–473
- Wen Y, Zhang Y, Zhang J, Feng J, Zhang Y (2020) The improved FASTmrEMMA and GCIM algorithms for genome-wide association and linkage studies in large mapping populations. *Crop J* 8:723–732
- Wen YJ, Zhang H, Ni YL, Huang B, Zhang J et al (2018) Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Briefings Bioinform* 19:700–712
- Wen YJ, Zhang YW, Zhang J, Feng JY, Dunwell JM et al (2019) An efficient multi-locus mixed model framework for the detection of small and linked QTLs in F2. *Briefings Bioinform* 20:1913–1924
- Worku N, Heslop-Harrison JS, Adugna W (2015) Diversity in 198 Ethiopian linseed (*Linum usitatissimum*) accessions based on morphological characterization and seed oil characteristics. *Genet Resour Crop Evol* 62:1037–1053
- Xiang T, Zong N, Zou Y, Wu Y, Zhang J et al (2008) *Pseudomonas syringae* effector *AvrPto* blocks innate immunity by targeting receptor kinases. *Curr Biol* 18:74–80
- Xie D, Dai Z, Yang Z, Sun J, Zhao D et al (2017) Genome-wide association study identifying candidate genes influencing important agronomic traits of flax (*Linum usitatissimum* L.) using SLAF-seq. *Front Plant Sci* 8:2232
- Yao J, Zhao D, Chen X, Zhang Y, Wang J (2018) Use of genomic selection and breeding simulation in cross prediction for improvement of yield and quality in wheat (*Triticum aestivum* L.). *Crop J* 6:353–365
- You FM, Booker HM, Duguid SD, Jia G, Cloutier S (2016a) Accuracy of genomic selection in biparental populations of flax (*Linum usitatissimum* L.). *Crop J* 4:290–303
- You FM, Cloutier S (2019) Mapping quantitative trait loci onto chromosome-scale pseudomolecules in flax. *Methods Protoc* 3:28

- You FM, Duguid DS, Thambygala D, Cloutier S (2013) Statistical analysis and field evaluation of the type 2 modified augmented design (MAD) in phenotyping of flax (*Linum usitatissimum*) germplasms in multiple environments. *Aust J Crop Sci* 7:1789–1800
- You FM, Duguid SD, Lam I, Cloutier S, Rashid KY et al (2016b) Pedigrees and genetic base of flax cultivars registered in Canada. *Can J Plant Sci* 96:837–852
- You FM, Jia G, Xiao J, Duguid SD, Rashid KY et al (2017a) Genetic variability of 27 traits in a core collection of flax (*Linum usitatissimum* L.). *Front Plant Sci* 8:1636
- You FM, Rashid K, Yao Z, Cloutier S (2017b) Genome-wide association study for Fusarium wilt resistance in flax (*Linum usitatissimum* L.) Joint Meeting of Canadian Phytopathological Society and Canadian Society of Agronomy, Winnipeg, Canada, 18–22 June, pp P23
- You FM, Xiao J, Li P, Yao Z, Gao J et al (2018a) Chromosome-scale pseudomolecules refined by optical, physical, and genetic maps in flax. *Plant J* 95:371–384
- You FM, Xiao J, Li P, Yao Z, Jia G et al (2018b) Genome-wide association study and selection signatures detect genomic regions associated with seed yield and oil quality in flax. *Int J Mol Sci* 19:2303
- Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol* 17:155–160
- Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* 178:539–551
- Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208
- Zhang J, Feng JY, Ni YL, Wen YJ, Niu Y et al (2017a) pLARM EB: integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. *Heredity (Edinb)* 118:517–524
- Zhang J, Yang J, Zhang L, Luo J, Zhao H et al (2020a) A new SNP genotyping technology target SNP-seq and its application in genetic analysis of cucumber varieties. *Sci Rep* 10:5623
- Zhang M, Liu Q, Yang X, Xu J, Liu G et al (2020b) CRISPR/Cas9-mediated mutagenesis of *Clpsk1* in watermelon to confer resistance to *Fusarium oxysporum* f.sp. *niveum*. *Plant Cell Rep* 39:589–595
- Zhang Y, Bai Y, Wu G, Zou S, Chen Y et al (2017b) Simultaneous modification of three homoeologs of *TaEDR1* by genome editing enhances powdery mildew resistance in wheat. *Plant J* 91:714–724
- Zhang YM, Jia Z, Dunwell JM (2019) Editorial: The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. *Front Plant Sci* 10:100
- Zhang YM, Tamba CL (2018) A fast mrMLM algorithm for multi-locus genome-wide association studies. *bioRxiv*. <https://doi.org/10.1101/341784>
- Zhang YW, Wen YJ, Dunwell JM, Zhang YM (2020c) QTL.gCIMapping.GUI v2.0: an R software for detecting small-effect and linked QTLs for quantitative traits in bi-parental segregation populations. *Comput Struct Biotechnol J* 18:59–65
- Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 44:821–824