Genomic Prediction of Testcross Performance in Canola (*Brassica napus*)

Habib U. Jan[1], Amine Abbadi[2], Sophie Lücke[3], Richard A. Nichols[4], Rod J. Snowdon[1]*

1 Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany, 2 NPZ Innovation GmbH, Hohenlieth, 24363 Holtsee, Germany, 3 Norddeutsche Pflanzenzucht Hans-Georg Lembke KG, Hohenlieth, 24363 Holtsee, Germany, 4 School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom

* rod.snowdon@agrar.uni-giessen.de

Abstract

Genomic selection (GS) is a modern breeding approach where genome-wide single-nucleotide polymorphism (SNP) marker profiles are simultaneously used to estimate performance of untested genotypes. In this study, the potential of genomic selection methods to predict testcross performance for hybrid canola breeding was applied for various agronomic traits based on genome-wide marker profiles. A total of 475 genetically diverse spring-type canola pollinator lines were genotyped at 24,403 single-copy, genome-wide SNP loci. In parallel, the 950 F1 testcross combinations between the pollinators and two representative testers were evaluated for a number of important agronomic traits including seedling emergence, days to flowering, lodging, oil yield and seed yield along with essential seed quality characters including seed oil content and seed glucosinolate content. A ridge-regression best linear unbiased prediction (RR-BLUP) model was applied in combination with 500 cross-validations for each trait to predict testcross performance, both across the whole population as well as within individual subpopulations or clusters, based solely on SNP profiles. Subpopulations were determined using multidimensional scaling and K-means clustering. Genomic prediction accuracy across the whole population was highest for seed oil content (0.81) followed by oil yield (0.75) and lowest for seedling emergence (0.29). For seed yield, seed glucosinolate, lodging resistance and days to onset of flowering (DTF), prediction accuracies were 0.45, 0.61, 0.39 and 0.56, respectively. Prediction accuracies could be increased for some traits by treating subpopulations separately; a strategy which only led to moderate improvements for some traits with low heritability, like seedling emergence. No useful or consistent increase in accuracy was obtained by inclusion of a population substructure covariate in the model. Testcross performance prediction using genome-wide SNP markers shows considerable potential for pre-selection of promising hybrid combinations prior to resource-intensive field testing over multiple locations and years.

Introduction

Genomic selection (GS) is a modern breeding approach whereby genome-wide single-nucleotide polymorphisms (SNP) marker profiles are used to estimate individual breeding values of untested genotypes [1–3]. This novel biometrical approach was initially proposed in animal breeding [4] but is actively gaining currency for the improvement of various complex traits in plant breeding [5–8]. In GS, genome-wide markers are used that capture a distribution of genetic effects, from small to large, and thus, potentially accounts for a majority of the genetic variance for a given trait [9]. Thus, GS approach can be used without prior information on the effect of individual markers. Instead a 'black box' approach is adopted where combined marker effects are estimated [10].

GS in plant breeding is considered more challenging than in animal breeding due to the complex nature of genotype-by-environment interactions and their strong influence on plant reproduction [11]. On the other hand, genomic prediction can potentially shorten the breeding cycle by enabling early selection and increased selection intensity. In combination with potentially higher selection accuracy for traits with low heritability, this can ultimately boost genetic gain in comparison to conventional selection [12–13]. The implementation of cost-effective screening systems for high-density, genome-wide SNP markers makes genomic prediction approaches increasingly attractive [14–15].

The ridge regression best linear unbiased prediction (RR-BLUP) approach [4, 16] is becoming a method of choice in genome-wide prediction models due its computational ease combined with a high accuracy in predicting both polygenic and even non-polygenic traits. In crop plants RR-BLUP has been applied in various practical and experimental crop breeding scenarios [7, 17–20]. The development of GS models generally involves the use of a training population (TP) and a validation or prediction population (VP) [1, 21]. Both genotype data (molecular markers) and phenotype (field) data are collected from the individuals of the training population, while the validation population is used to test the performance (accuracy) of a statistical prediction model based solely on genotype data. Selection accuracy in GS can be affected by various factors including trait genetic architecture, linkage disequilibrium (LD) between the markers and quantitative trait loci (QTL), the number of markers, the size of the training population and the genetic relatedness of the training and validation populations [22–24]. A GS model for which the selection accuracy is equivalent or better in comparison to conventional selection, typically based on field performance in multi-environment evaluations, can potentially improve the selection gain by improving selection intensity or reducing the generation interval. In cases where a trait can only be accurately assessed using expensive phenotyping strategies, GS can also reduce costs for a given selection gain in a breeding programme.

Brassica napus L., commonly known as canola, oilseed rape or rapeseed, is one of the world's most important oilseed crops because it delivers a rich source of high-quality edible oil and animal feed as extracted seed meal [25–27]. In Europe, the oil from winter-sown oilseed rape is also widely used as a sustainable biofuel [28]. The allopolyploid *B. napus* formed only a few thousand years ago [29] from spontaneous inter-specific hybridisations between turnip rape (*Brassica rapa*; AA, 2n = 20) and cabbage (*Brassica oleracea* CC, 2n = 18), and the gene pool of modern breeding materials is narrow due to this restricted genetic background and strong selection for essential seed quality traits [30–32]. According to Cowling [33], the effective population size of spring-type canola grown in Australia is just $N_e \leq 11$, reflecting a huge loss of genetic diversity compared to the progenitor species *B. rapa* and *B. oleracea* [27].

Hybrid breeding has been instrumental in the exploitation of heterosis for yield gain and yield stability in plant breeding [34]. Due to its well-defined pollination control systems, *B. napus* can be used successfully for hybrid seed development [35–36]. On the other hand, the

relatively narrow genetic diversity in modern breeding pools restricts the heterotic potential [37]. In comparison to classical hybrid crops like maize, in which genetically distinct heterotic pools have been established over many decades of hybrid breeding, there are no such clear heterotic pools available within canola germplasm. Development of new heterotic pools within adapted germplasm types, particularly through marker-assisted introgressions of novel germplasm from the diploid progenitors or other exotic gene pools, is an important strategy to overcome this problem [38–42]. Nevertheless, even in the absence of heterotic pools, genomic prediction of testcross performance is a highly promising method in canola breeding to select promising germplasm for advancement into male-sterile maternal lines or fertility restorers [43]. Recently genomic prediction has been demonstrated for estimation of testcross performance in various crops, for example in maize [7–8, 44], sugar beet [45], and rye [46].

In a hybrid breeding programme, efficient selection of the most promising combinations between male and female parental lines is a vital step to avoid expensive field testing of poor performing hybrids [20]. This becomes particularly important in crops like canola where the absence of distinct genetic pools prohibits a *per se* assumption of heterotic potential between any two potential hybrid parents. Various studies have reported methods for optimum exploitation of heterosis in crop breeding using both morphological and molecular marker data [47–51]. Piepho [52] described how the performance of untested hybrids can also be predicted effectively using genomic selection methodology.

In hybrid breeding system, male inbred lines are crossed with genetically distant 'testers' and general combining ability (GCA) and specific combining ability (SCA) values are estimated. The variances of GCA and SCA depend on the kind of gene actions involved. GCA includes additive effects of the total variances which make up the major portion of variances whereas SCA refers to the non-additive gene actions mainly comprising dominance and epistatic deviations. Information on GCA plays an important role in a breeder's decision making to identify a viable hybrid [53]. GCA information has been used recently in various genome-wide prediction studies [20, 47]. In situations where GCA variance is predominant compared to SCA variance, prediction of hybrid performance based on parental GCA effects is an accurate approach [54]. Experimental studies in hybrid canola revealed that variance due to GCA is more pronounced and mainly additive effects contribute to hybrid performance compared to SCA effects [55].

Technical difficulties associated with the development of male-sterile lines in canola generally lead breeders to choose relatively small panels of maternal lines. On the other hand, some of the most widely used male-sterility systems have the benefit that all known *B. napus* accessions are restorers, so that testcross performance with available maternal lines is an important selection criterion for breeding of pollinators. The restorer lines have nuclear genes and are able to restore fertility in hybrid crosses. The goal of the present study was to investigate prediction of the best possible parental combination of pollinators crossed with the two testers lines in a testcross performance for a number of important traits in spring canola based on genome-wide SNP profiles. In particular, our objectives were (1) to examine strategies for genomic selection of suitable parental combination for the use in canola hybrid breeding, (2) to explore the effect of training population sample size on the prediction accuracy, and (3) to evaluate the potential for genomic prediction of GCA effects contributing to canola hybrid performance.

Materials and Methods

Genetic material

The experimental materials comprised a diverse population of spring-type *B. napus* with double-low seed quality (low erucic acid, low glucosinolate content) from a commercial canola

breeding programme. The materials were carrying introgressions from the diploid progenitors of *B. napus*. Two representative male sterile female testers (tester 1 and tester 2) from a pool of testers carrying the *Male Sterility Lembke* (MSL) sterility system (NPZ Lembke, Hohenlieth, Germany) were crossed with a total of 475 pollinators to generate seed from 950 F1 hybrids.

Phenotype data

The 950 testcrosses were evaluated at four different locations across Denmark, Germany, Poland and Estonia during the 2012 growing season. The locations for the field trials are commercial plant breeding testing sites. No specific permissions are required to grow conventional oilseed rape on these agricultural locations. Commercial crops like oilseed rape are not endangered or protected species. Un-replicated trials were performed in each of the four locations for various traits of commercial importance including seed yield (dt/ha), oil yield (dt/ha), seed oil content (% volume per seed dry weight), content of total seed glucosinolate (GSL; μmol/g seed), seedling emergence (visual observation ranging from a minimum value of 1 to maximum 9), lodging resistance (visual observation ranging from a minimum value of 1 to maximum 9) and days to onset of flowering (DTF; measured as number of days from sowing until 50% flowering plants per plot).

The restricted maximum likelihood (REML) method was used to estimate variance components. A best linear unbiased estimate (BLUE) was made for each trait (S1 File). All calculations were performed using the statistical software package SPSS Statistics for Windows Version 22.0 (IBM Corp., Armonk, NY, USA). Variance estimates were used to calculate broad sense heritability (H^2) following the method given in [56].

$$H^2(\%) = \left[\frac{\sigma^2_g}{(\sigma^2_g + \sigma^2_\varepsilon/n)} \right] \times 100$$

where σ^2_g is the genotypic variance, σ^2_ε is the estimated residual variance, and *n* is the number of locations (Table 1). Estimates of residual variance σ^2_ε were divided by the number of locations (in this case four).

In our genomic prediction analysis, we used GCA values as phenotype matrix on all the pollinator lines. Pearson's correlation coefficients (r) were calculated from the BLUE values between all the traits.

Genotype data

Each of the 475 pollinator lines was genotyped using the *Brassica* 60k SNP Infinium consortium array (Illumina Inc., San Diego, CA; USA). Genomic DNA was extracted from young leaf samples collected 20 days after sowing, shock frozen in liquid nitrogen and stored at -20°C until further processing. The DNA extractions were performed using a BioSprint 96 magnetic bead nucleic acid extraction robot (Qiagen, Hilden, Germany) according to the manufacturer's instructions. After fluorometric quantification of DNA concentrations using a Qubit 2.0 fluorometer (Life Technologies, Darmstadt, Germany), samples were diluted to 20ng/μl in sterile double distilled water, and quality checks of all DNA samples were carried out by gel electrophoresis on a 96 capillary Fragment Analyser (Advanced Analytical, Ames, IA, USA).

Genotyping on the 60k *Brassica* SNP array was outsourced to TraitGenetics GmbH (Gatersleben, Germany). All called SNPs were mapped to the *Brassica napus* cv. Darmor-bzh reference genome [29] using the basic local alignment search tool (BLAST) with no mismatches permitted in the flanking oligonucleotides. All SNPs showing multiple BLAST hits or a non-random distribution were removed and a total of 28,286 single-position SNPs remained.

Table 1. Summary statistics for seed yield (dt/ha), oil yield (dt/ha), seed oil content (%), seed glucosinolate content (GSL; μmol/g), seedling emergence (visual observation scale 1–9; good = 9), lodging resistance (visual observation scale 1–9; good = 9) and days to onset of flowering (DTF) in field trials with 950 spring canola F1 testcross phenotypes in 4 independent field locations throughout Europe. σ^2_g: genetic variance, σ^2_ε: estimated residual variance, H^2: broad sense heritability.

Traits	Mean	Min.	Max.	σ^2_g	σ^2_ε	H^2(%)
Seed yield (dt/ha)	31.17	23.94	38.38	1.56	7.95	44
Oil yield (dt/ha)	14.54	4.5	24.55	0.56	0.95	70
Oil content (%)	48.41	44.08	52.73	1.91	0.81	90
GSL (μmol/g)	9.22	6.91	11.57	1.35	3.12	63
Emergence (good = 9)	6.66	4.53	8.8	0.048	0.41	32
Lodging resistance (good = 9)	7.16	5.05	8.65	0.119	0.47	50
DTF	171.26	160.23	182.28	0.808	5.26	38

doi:10.1371/journal.pone.0147769.t001

Furthermore, markers with allelic frequencies smaller than 0.05 and markers having more than 20% missing values were removed (S2 File). A total of 24,403 unique, single-copy SNPs remaining after filtration were used in the subsequent genomic prediction of testcross performance.

Determination of population structure

Genetic relatedness between different genotypes can be estimated using molecular markers [57]. In the absence of clearly defined heterotic pools in *B. napus*, we analysed genetic relatedness and population substructure among the parental lines using the genome-wide SNP data. Roger's genetic distances [58] were calculated among pairs of inbred lines. A multidimensional scaling (MDS) analysis based on principal coordinate analysis (PCoA) was performed using the filtered panel of 24,403 unique, single-copy genome-wide SNPs. The software package *SelectionTools* [59; www.uni-giessen.de/population-genetics/downloads] was used in *R* [60] for PCoA and *K*-means clustering.

Clusters of genetically related individuals were identified using the *K*-means method, following the algorithm of [61]. A diagnosis of the optimal number of clusters in the dataset was performed using the method described by [62].

Scenarios for the genomic prediction of breeding values

Three independent scenarios based on the population structure were applied to estimate marker effects by genomic prediction. In scenario 1, the genomic prediction was performed across the whole population (WP). We further tested the prediction accuracy across the whole population using a model that included the population substructure as a covariate. To investigate genomic prediction accuracy separately in the different genetic backgrounds of cluster 1 (C1) and cluster 2 (C2), respectively, we developed scenario 2 (prediction within C1) and scenario 3 (prediction within C2) (Fig 1). However, we did not directly compare prediction accuracies among these three prescribed scenarios due to confounding caused by their different sizes, and rather reported them separately. Results from predictions within subpopulation C3 alone are not reported due to the very small size of the test and validation populations in this case. In addition, we tested two more scenarios where we sampled the training sets across three subpopulations and validated these within the two main subpopulations C1 and C2, respectively.

Genomic prediction using the RR-BLUP mixed model

Genomic prediction accuracies were estimated using the RR-BLUP model described by [4], [16], assuming the same distribution of marker effects across the whole-genome. The following
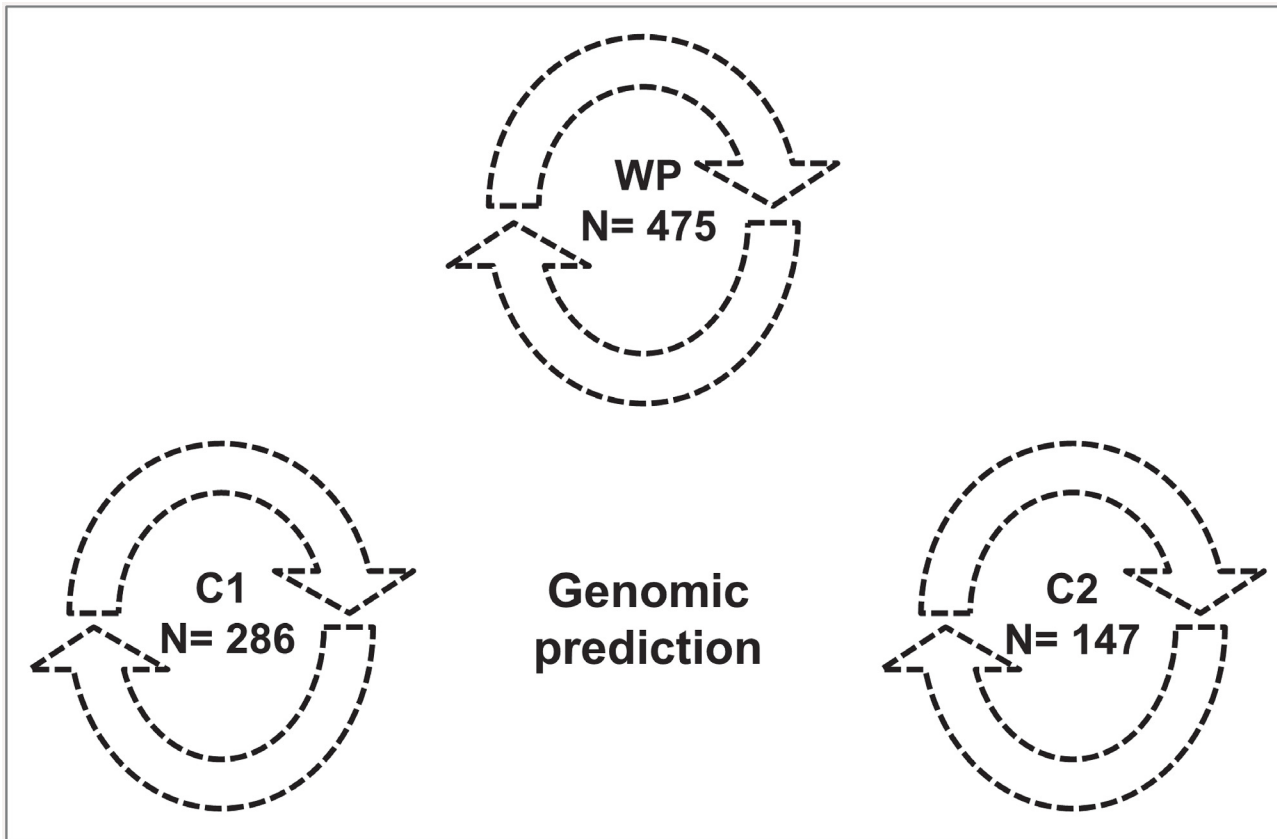
Fig 1. Genomic prediction across the whole population (WP) and genomic predictions within cluster 1 (C1) and cluster 2 (C2) separately are represented by dotted circular arrows. Fig 1 illustrates three independent genomic scenarios.

doi:10.1371/journal.pone.0147769.g001

model was used:

$$y = \mu + \sum_{i=1}^{N_m} X_i a_j + e,$$

where:

y is a $N \times 1$ vector of phenotype (vector of BLUEs across locations);

μ is the overall mean;

N_m is the number of SNPs;

a_j is the effect of the jth marker;

X_i is a $N \times 1$ vector of genotypes (coded as 0,-1,+1) of the lines for each marker j, and variance of a_j is assumed to be uniformly distributed and is σ^2_G / N_m [4].

Marker imputation and genomic prediction accuracy

Monomorphic SNPs and markers having more than 20% missing data were removed from the dataset. The rr-BLUP package in R [17] was used to estimate genomic predictions with the remaining missing data replaced using the default method (mean imputation). Genomic prediction accuracy, denoted as r_{GPA} was calculated for each trait. In some previous studies, prediction accuracies have been standardised by dividing the square root of the heritability to remove the corrected influence of heritability [54]. In our study, on the other hand, we report

prediction accuracy (r_{GPA}) as the Pearson correlation, $r(y, \hat{y})$, between the predicted values ($\hat{y}$) and observed BLUE values (y), using the rr-BLUP package [17]. Because the heritability is not considered for calculation of the BLUE values, it is not necessary or appropriate to correct for inferred heritability in the prediction model we applied.

Model cross validation

For determination of the optimum composition of training population size, we tested the prediction accuracies for each of the seven traits in the whole population under incremental increase of the training population from 10% up to 90% of the 475 lines. Based on the results of this test (see below), the training population for all further analyses and scenarios testing was set up at 70% of the total lines in the given dataset. Hence, in each run, the dataset was divided into a random 70 percent training population (TP) containing both genotyped and phenotyped data, and 30 percent validation (VP) or prediction population having only SNP data and SNP effects with no consideration of phenotype values. For each scenario the data for each trait was cross-validated for 500 rounds and a mean value was subsequently considered.

Results

Population structure among pollinators

The results of the principal coordinate analysis (PCoA) between the parental inbred lines are shown in Fig 2a. The genetic variance explained by the first four principal coordinates comprised 25.12%, 18.43%, 8.01% and 6.06% respectively, making a total of 57.62%.

The PCoA indicated the existence of subpopulations within the dataset. The *K*-means clustering revealed a tendency to two main clusters and one relatively smaller cluster. This assumption was supported by the results of the Caliński-Harabasz [62] clustering, which also suggested three optimum clusters, as shown in Fig 2b. These are subsequently referred to as cluster 1 (C1; n = 286), cluster 2 (C2; n = 147) and cluster 3 (C3; n = 42), respectively.

Broad sense heritability and variance components

Broad sense heritabilities, along with summary statistics for all the traits under consideration, are shown in Table 1. Heritability values ranged from 32% for seedling emergence to 90% for seed oil content. Best linear unbiased estimates (BLUE) of each trait followed approximately the normal distribution expected for quantitative traits (S1 Fig). This was further confirmed by Q-Q plots drawn individually for each trait using R. The highest genetic variance was observed for seed oil content, while seedling emergence had the lowest genetic variance. As expected, positive correlations were observed between oil yield and seed oil content (r = 0.66) and between seed yield and oil yield (r = 0.57). Similarly, the expected highly negative correlation was observed between seed oil content and seed GSL (r = -0.34) Fig 3.

Estimation of genomic prediction accuracy for different traits

1. Prediction across whole population: Fig 4 and Table 2 show the accuracies of genomic prediction for the respective traits based on GCA values, along with their respective standard errors for testcross performance using the whole population (WP) without consideration of population structure. For the seven traits considered, the highest prediction accuracy was recorded for seed oil content (r_{GPA} = 0.81) followed by oil yield (r_{GPA} = 0.75), seed glucosinolate content (r_{GPA} = 0.61), days to flowering (r_{GPA} = 0.56), seed yield (r_{GPA} = 0.45), lodging resistance (r_{GPA} = 0.39) and the least heritable trait, seedling emergence (r_{GPA} = 0.29).
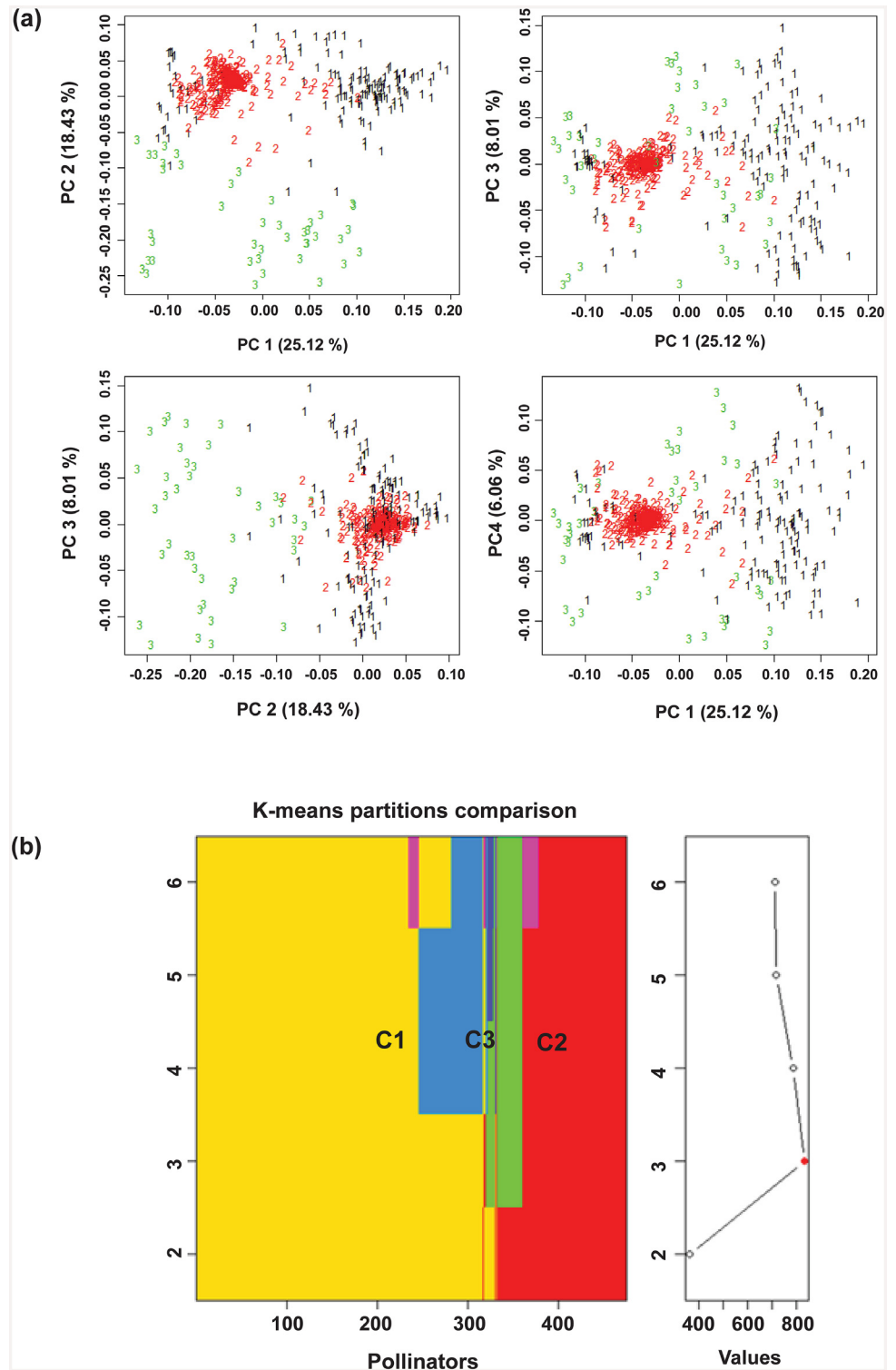
Fig 2. (a) Principal coordinates analysis (PCoA) among the population of 475 spring-type canola pollinators used for the testcross production. The PCoA is estimated using a panel of 24,403 filtered single nucleotide polymorphism (SNP) markers. Proportions of explained variance of principal coordinates 1, 2, 3 and 4 are given in parentheses. (b) K-means clustering of the 475 pollinator lines using the method of Caliński-Harabasz (1974) showing three clusters, i.e. cluster 1 (C1), cluster 2 (C2) and cluster 3 (C3) respectively. Fig 2 (a, b) shows sub-population structure in the dataset.

doi:10.1371/journal.pone.0147769.g002
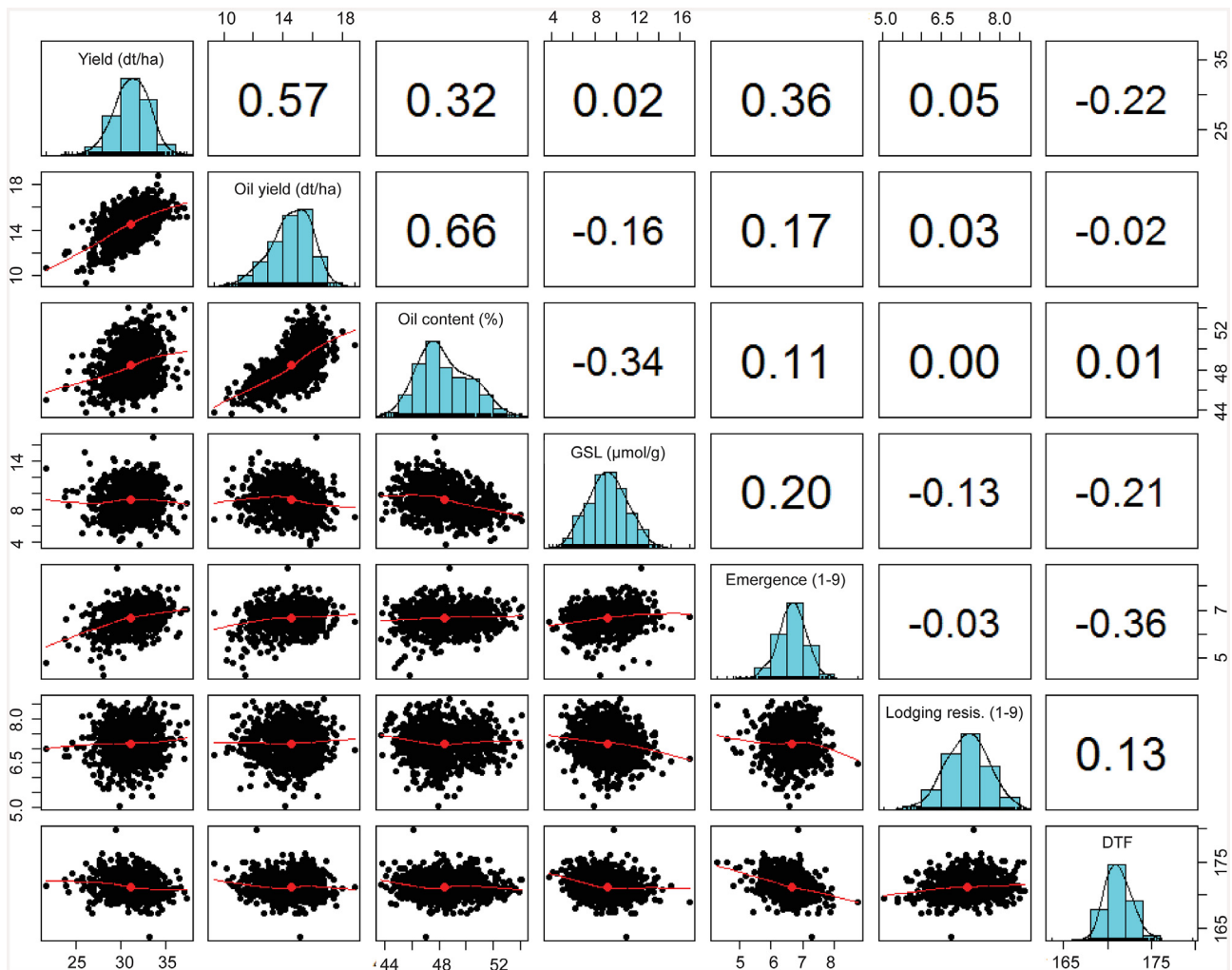
Fig 3. The Pearson's correlations between all the seven traits. Highest positive correlation recorded between oil yield and seed oil content and lowest negative correlation recorded between seed oil content and seed GSL.

doi:10.1371/journal.pone.0147769.g003

Scatter plots showing the correlations between true observed trait values and genomic predicted values for all the traits are shown in S2 Fig.

2. Predictions within subpopulations: Fig 5, S3 File show the independent prediction accuracies within subpopulations C1 and C2, respectively. Interestingly, an improved prediction accuracy (r_{GPA} = 0.39) was observed for the low-heritability trait seedling emergence within subpopulation C1, the largest subpopulation but also the narrowest in terms of genetic diversity. Predictions accuracies also improved for two other traits with low to moderate heritability, seed yield (r_{GPA} = 0.47) and DTF (r_{GPA} = 0.59) (Fig 5, S3 File). Similarly, within the second-largest subpopulation, C2, the prediction accuracies improved to r_{GPA} = 0.65 for GSL and r_{GPA} = 0.49 for lodging resistance, respectively (S3 File). For seed oil content and oil yield we observed no improvement in prediction accuracy within subpopulations compared to the whole population. Prediction accuracies in our additional two scenarios, where training set was taken across the three subpopulations and validated within C1 and C2, were either negative or close to zero for the majority of the traits (S4 File).
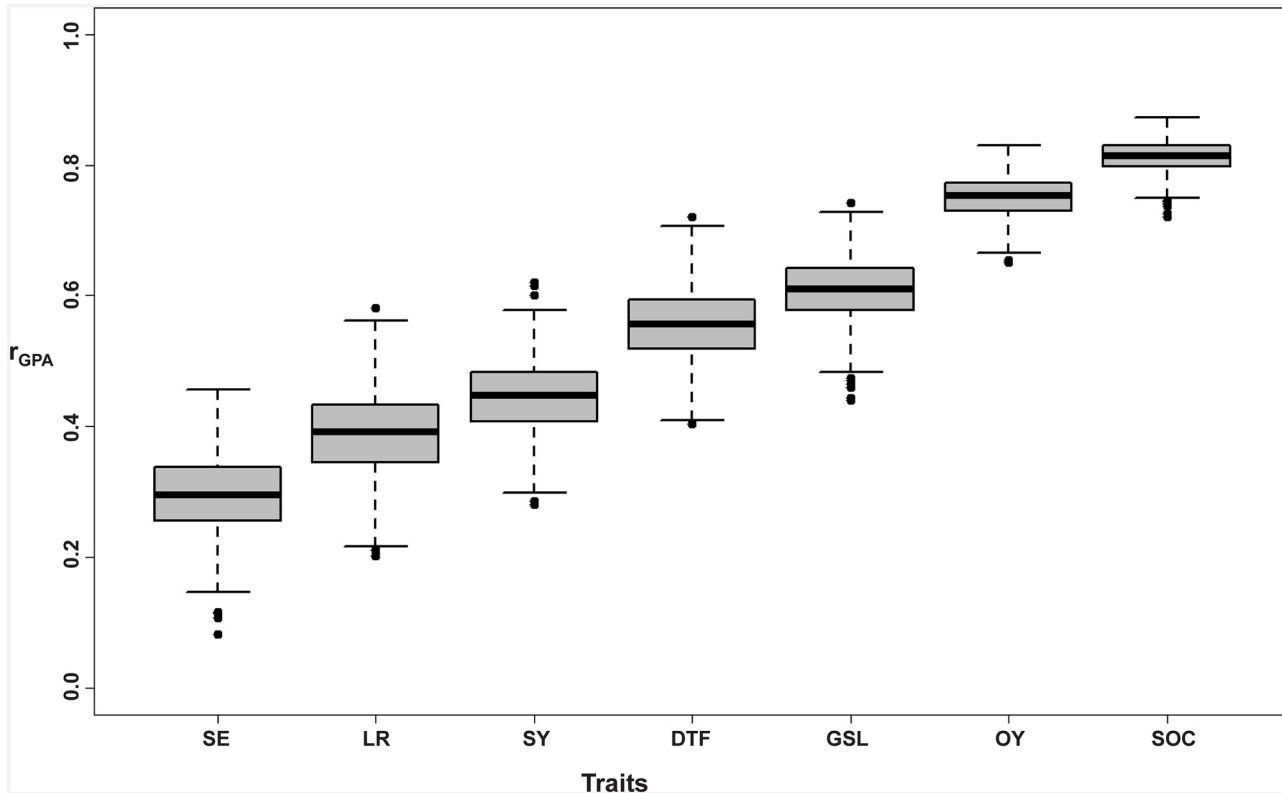
Fig 4. r$_{GPA}$ across the whole test population for seedling emergence; SE, lodging resistance; LR, seed yield; SY, days to flowering; DTF, seed glucosinolate content; GSL, oil yield; OY and seed oil content; SOC, respectively. Fig 4 shows genomic prediction accuracies.

doi:10.1371/journal.pone.0147769.g004

Prediction accuracy and training population (TP) size

As expected, increasing the size of the TP resulted in improvement of the genomic prediction accuracy (Fig 6). All the traits showed a plateau of prediction accuracy at a TP proportion of 80% except days to flowering, and only insignificant increases in accuracy as the TP size increased from 70% to 90%. We therefore, set an arbitrary TP size at 70% for all subsequent analyses and scenario testing.

Discussion

The first investigation of the potential of genomic selection in *B. napus* breeding [63] investigated a relatively narrow set of winter oilseed rape breeding lines derived from 9 elite parental lines that were genotyped with only 253 SNP markers. To our knowledge, our study is the first

Table 2. Average prediction accuracies (r$_{GPA}$) and standard errors (SE) for seed yield (dt/ha), oil yield (dt/ha), seed oil content (%), seed glucosinolate content (GSL; µmol/g), seedling emergence (visual observation scale 1–9; good = 9), lodging resistance (visual observation scale 1–9; good = 9) and days to onset of flowering (DTF) derived from 500 rounds of cross-validation across the whole-population.

Traits	Seed yield (dt/ha)	Oil yield (dt/ha)	Seed oil content (%)	GSL (µmol/g)	Seedling emergence (good = 9)	Lodging resistance (good = 9)	DTF
r$_{GPA}$±SE	0.45±0.002[a]	0.75±0.001	0.81±0.001	0.61±0.002	0.29±0.002	0.39±0.003	0.56 ±0.002

[a]Approximate standard errors (SE) attached.
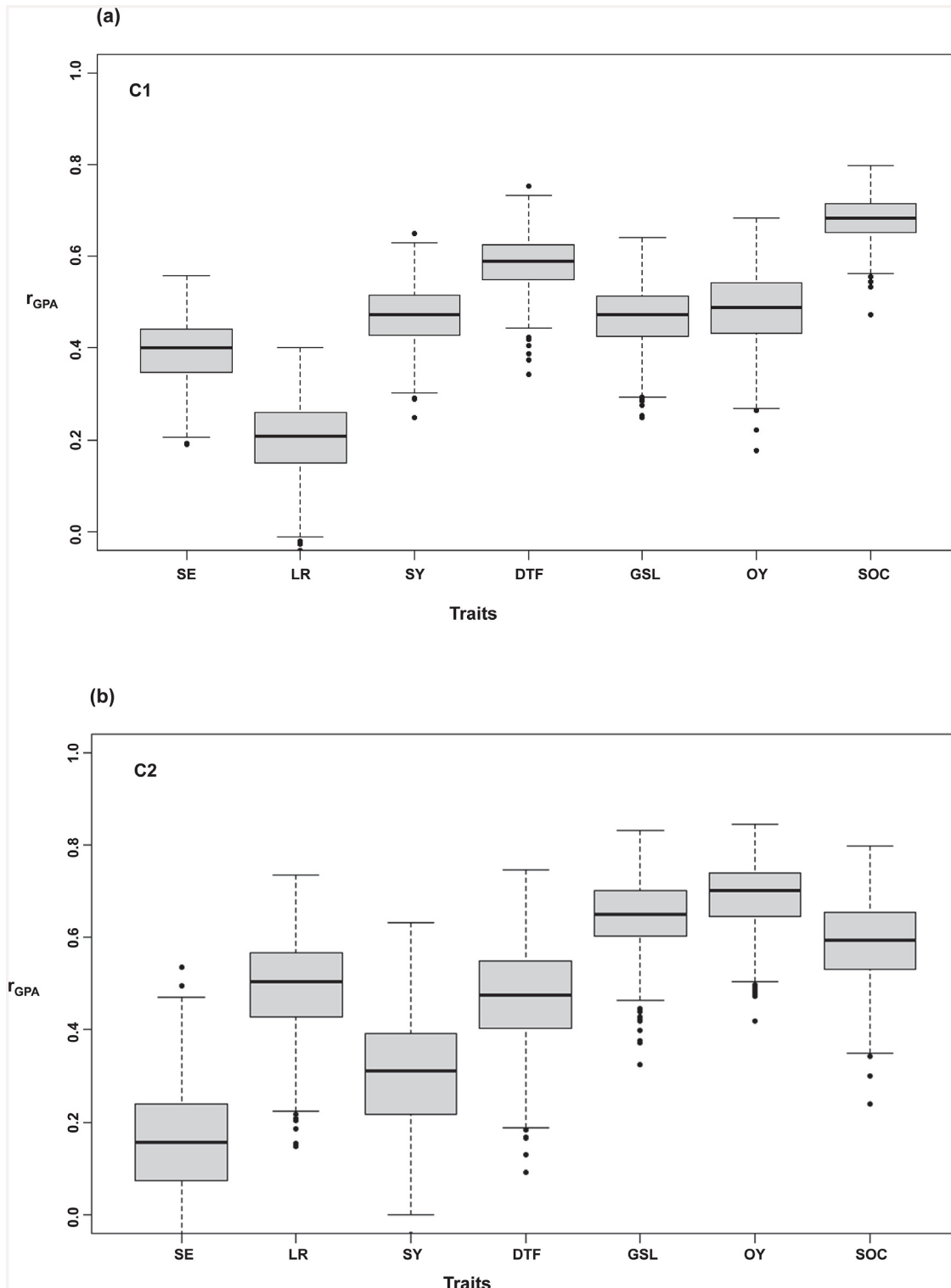
doi:10.1371/journal.pone.0147769.t002

Fig 5. (a) Genomic prediction accuracies (r_{GPA}) within cluster 1 (C1) for seedling emergence; SE, lodging resistance; LR, seed yield; SY, days to flowering; DTF, seed glucosinolate content; GSL, oil yield; OY and seed oil content; SOC, respectively. (b) Genomic prediction accuracies (r_{GPA}) within cluster 2 (C2) for seedling emergence; SE, lodging resistance; LR, seed yield; SY, days to flowering; DTF, seed glucosinolate content; GSL, oil yield; OY and seed oil content; SOC, respectively. Fig 5 (a, b) shows genomic prediction accuracies across the two sup-populations.

doi:10.1371/journal.pone.0147769.g005
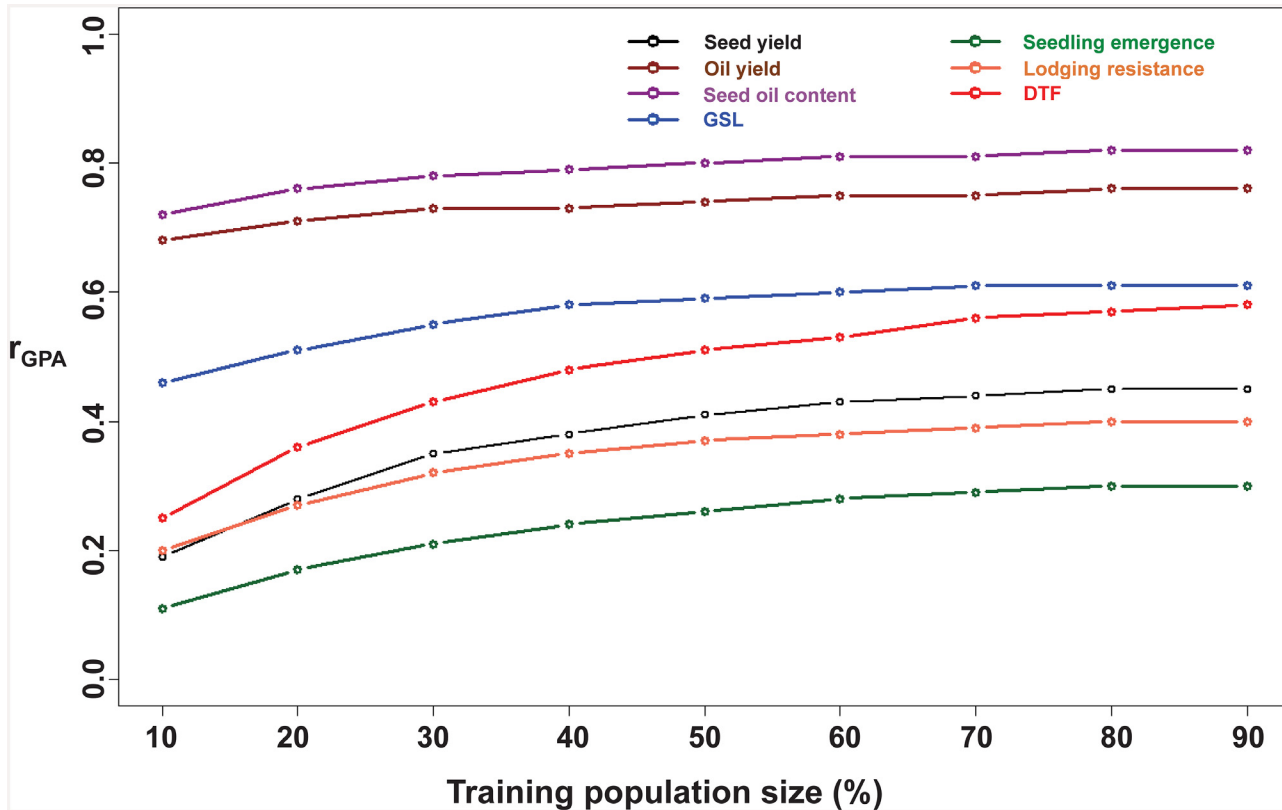
Fig 6. Influence of the size of the training population (TP; % of whole population size) on the genomic prediction accuracy (r_{GPA}) for the seven traits seedling emergence, lodging resistance, seed yield, days to flowering (DTF), seed glucosinolate content (GSL), oil yield and seed oil content. Fig 6 shows the effect of training population size on genomic prediction accuracies.

doi:10.1371/journal.pone.0147769.g006

report of testcross performance prediction in this important oil crop species. The population size, the represented genetic diversity and the number of SNP markers used for our analysis were all considerably larger than the previous study of [63].

We investigated genomic prediction accuracies for seven key agronomic traits including seed yield, oil content and quality related traits using a diverse population of spring-type canola. The RR-BLUP method used for the prediction modeling has been shown to be effective in accounting for both major and minor effect quantitative trait loci (QTL) in plant breeding [19–20, 63].

Independent genomic prediction across the whole population

First we investigated genomic prediction accuracy for each trait across the whole-population based on GCA values. Taking the whole population under consideration, the lowest genomic prediction accuracy was estimated for seedling emergence and highest for seed oil content. The low genomic prediction accuracy for seedling emergence under scenario 1 may be explained by the low heritability and genetic variance for this trait. One strategy to increase prediction accuracy in such traits could be to combine these with other correlated high heritable traits in a multi-trait genomic prediction model which has been shown in a previous study [64]. In the case of seed oil content, the prediction accuracy remained high across the whole population. This is presumably due to the high heritability and the comparatively simple genetic architecture

underlying this trait, where a few major QTL control maximum phenotypic variance [65, 66]. Oil yield showed the second highest prediction accuracy across the WP after seed oil content. This may be due to the strong positive correlation between these two traits. In our prediction analysis, genomic prediction accuracies based on additive genetic effects were higher in majority of the traits within the two tester pools.

Riedelsheimer et al. [47] and Saatchi et al. [67] reported that population substructure might affect genomic prediction accuracies. In our dataset, implementation of independent prediction within subpopulations increased prediction accuracies in specific subpopulations for low to moderate heritability traits like seed glucosinolate content, lodging resistance, DTF and seedling emergence. This is in line with the previous studies that reported higher prediction accuracies when genetically closely individuals were used in the TP and VP [22, 68]. The most straightforward explanation for such improvements might be that these traits are affected by variants at major-effect loci in some subpopulations that are rare or absent in the remainder of the materials. For some traits no improvement in accuracy were observed within subpopulations. This indicates that a large TP, in which the captured diversity strongly represents the diversity in the corresponding VP, may overcome the potential disadvantage caused by use of genetically distant individuals in the TP and VP. On the other hand, lowest prediction accuracies were obtained when a training set was derived from across all three subpopulations and validated within the two main subpopulations. This may indicate a lack of correspondence between the linkage phase of markers and QTL alleles across the different subpopulations. Adding a covariate to the prediction model which identified the clusters in the whole population did not improve the overall prediction accuracy for any trait. This scenario may be rather specific for canola, in which modern, adapted breeding pools have a particularly narrow genetic basis due to conscious selection for specific traits [31–33]. The situation is very different in maize or cattle, for example, where genetic differentiation among subpopulations or races are highly pronounced and population differences in gene and allele content are therefore often decisive [22, 68, 69]. We conclude that adjustment of prediction models on a case-by case basis in canola can potentially give small improvement in prediction of specific traits depending on the variance within a given breeding population.

For the high-value traits of seed oil content, oil yield and seed glucosinolate content, for which high heritabilities can be attributed to a good rank correlation among locations, we consistently obtained very high prediction accuracies in predictions across the entire population regardless of substructure. This may be further due to the modulating maternal influence of the two common male-sterile testers on embryo-related traits like seed size and oil content.

Effect of TP sample size on genomic prediction accuracy

In simulation studies [68] as well as real datasets [8, 15, 67], it has been shown earlier that an increase in the training population size has a positive impact on the overall genomic prediction accuracy. In predictions across the entire test population, a TP comprising 70% of the overall population size (333 lines from 475) was sufficient to accurately predict the performance of the remaining lines for testcross performance. With the exception of flowering time, where the prediction accuracy still did not achieve a plateau even with 90% TP, only small or insignificant increases in accuracy were achieved with a TP proportion greater than 70%. The failure to achieve a plateau for flowering time suggests the presence of some accessions with distinctly different genetic control of flowering time. From a breeder's viewpoint a smaller TP size is of course advantageous to reduce phenotyping costs. The most satisfying solution is the one in which adequate selection gains are achieved without surpassing current phenotyping costs.

Genomic selection prospects in hybrid rapeseed

At the dawn of canola hybrid breeding various authors reported considerable heterosis in F1 hybrids [70–72]. The use of molecular markers to accelerate the differentiation of hybrid pools and investigate the genetic basis of heterosis [73–76] further increased hybrid performance; however levels of yield improvement seen in more classical hybrid crops like maize are still not achieved in canola. The development of heterotic pools in canola has made only slow progress in comparison to maize due to the generally low diversity within the species. The highly complex allopolyploid genome of *B. napus*, with multiple interacting homoeologous copies of almost all genes [29], increases the difficulty in prediction of individual gene actions [31]. Hence, genomic prediction of testcross performance could be a promising avenue for improving important traits without consideration of detailed *a priori* knowledge of their underlying genetics.

Conclusions

The main purpose of genomic selection is the utilisation of large and inexpensive DNA marker datasets to bring an improvement to the mean performance of a certain population [77]. Seed yield, seed oil content and other polygenic traits are under the influence of complex genetic and biochemical interactions, and hundreds or thousands of small-effect QTL might be involved in their expression.

From a breeder's perspective the implementation of genomic prediction is only worthwhile if equivalent or greater selection gain can be achieved with equal or reduced time and/or cost than using conventional selection methods (generally multiple-year, multiple-location field evaluations). Depending on the selection intensity, the results presented in this paper clearly demonstrates the value of performance predictions based on high-density SNP markers in hybrid canola. Relatively higher genomic prediction accuracies in the majority of the traits, based on the additive genetic effects in our study, indicate a lack of distinct heterotic pools in our sample. Even where no improvement on phenotypic selection gain is achieved through genomic prediction, the method is still of considerable value for traits like seedling emergence, where the very low heritability seedlots generated in multiple maternal environments combined with multi-location field evaluations. In such cases an increase in genetic gain might still be achieved if the early pre-selection approach enables a shortening of the breeding cycle. The results of our study suggest that prediction of testcross performance in canola breeding, based on genome-wide SNP markers, can be a powerful, fast and low-cost method to pre-select promising pollinators for combinations with available male-sterile maternal lines. Genomic testcross performance prediction can hence allow breeders to optimise the allocation of breeding resources. Despite the absence of clearly defined heterotic pools in canola, predictions within large subpopulations can in some cases improve prediction accuracy for selected traits with low heritability.

Supporting Information

S1 Fig. Trait distribution: a-g) Histograms and h-n) Q-Q plots of best linear unbiased estimates (BLUEs) for a,h) seed yield (dt/ha), b,i) oil yield (dt/ha), c,j) seed oil content (%), d,k) seed glucosinolate content (GSL; μmol/g), e,l) emergence (visual observation scale 1–9; good = 9), f,m) lodging resistance (visual observation scale 1–9; good = 9) and g,n) days to onset of flowering (DTF) in field trials from 8 independent locations.
(TIF)

S2 Fig. Scatter plots showing correlations between observed mean trait values (observed) and genomic predicted (predicted) for the seven traits evaluated.
(TIF)

S1 File. BLUE values for each trait.
(XLSX)

S2 File. *Spring type Brassica napus* **(canola) genotype data (SNP) used in our analysis.**
(XLSX)

S3 File. Genomic prediction accuracies (r_{GPA}) for seven evaluated traits calculated independently within the two subpopulations C1 and C2.
(XLSX)

S4 File. Genomic prediction accuracies (r_{GPA}) for seven evaluated traits calculated with a training set sampled from the three subpopulation and independently validated within the two subpopulations C1 and C2.
(XLSX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: RS AA. Performed the experiments: SL HJ. Analyzed the data: AA SL HJ RN. Contributed reagents/materials/analysis tools: SL AA RS. Wrote the paper: HJ RS RN.

References

1. Heffner EL, Sorrells ME, Jannink J. Genomic Selection for Crop Improvement. Crop Science 49 (1): 1–12. doi: 10.2135/cropsci2008.08.0512

2. Jannink J, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Briefings in functional genomics. 2009; 9 (2): 166–177. doi: 10.1093/bfgp/elq001

3. Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, et al. Genomic Selection in Plant Breeding: Knowledge and Prospects. Advances in Agronomy. 2011; 110(C):77–123.

4. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001; 157(4): 1819–1829. PMID: 11290733

5. Lorenzana RE, Bernardo R Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theoretical and Applied Genetics. 2009; 120 (1): 151–161. doi: 10.1007/s00122-009-1166-3 PMID: 19841887

6. Crossa J, de Los Campos Gustavo, Pérez P, Gianola D, Burgueño J, Araus JL, et al. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics. 2010; 186 (2): 713–724. doi: 10.1534/genetics.110.118521 PMID: 20813882

7. Albrecht T, Wimmer V, Auinger H, Erbe M, Knaak C, Ouzunova M, et al. Genome-based prediction of testcross values in maize. Theoretical and Applied Genetics. 2011; 123 (2): 339–350. doi: 10.1007/s00122-011-1587-7 PMID: 21505832

8. Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin FH, et al. Accuracy of genomic selection in European maize elite breeding populations. Theoretical and Applied Genetics. 2012; 124 (4): 769–776. doi: 10.1007/s00122-011-1745-y PMID: 22075809

9. Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE. Reducing dimensionality for prediction of genome-wide breeding values. Genetics Selection Evolution. 2009; 41:29. doi: 10.1186/1297-9686-41-29

10. Jonas E, de Koning D. Does genomic selection have a future in plant breeding. Trends in biotechnology. 2013; 31 (9): 497–504. doi: 10.1016/j.tibtech.2013.06.003 PMID: 23870753

11. Riedelsheimer C, Endelman JB, Stange M, Sorrells ME, Jannink JL, Melchinge AE. Genomic predictability of interconnected biparental maize populations. Genetics. 2013; 194 (2): 493–503. doi: 10.1534/genetics.113.150227 PMID: 23535384

12. Schaeffer LR. Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet. 2006; 123(4):218–23. PMID: 16882088

13. Heffner EL, Lorenz AJ, Jannink J, Sorrells ME. Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. Crop Science. 2010; 50 (5): 1681. doi: 10.2135/cropsci2009.11.0662

14. Zhong S, Dekkers, Jack C M, Fernando RL, Jannink J. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a Barley case study. Genetics. 2009; 182 (1): 355–364. doi: 10.1534/genetics.108.098277 PMID: 19299342

15. Heffner EL, Jannink J, Sorrells ME. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. The Plant Genome. 2011; 4:65–7.

16. Whittaker JC, Thompson R, Denham MC. Marker-assisted selection using ridge regression. Genet Res Camb. 2000; 75:249–252. doi: 10.1111/j.1469-1809.1999.ahg634_0351_17.x

17. Endelman JB. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. The Plant Genome. 2011; 4 (3): 250. doi: 10.3835/plantgenome2011.08.0024

18. Resende MFR, Muñoz P, Resende M D V, Garrick DJ, Fernando RL, Davis JM, et al. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). Genetics. 2012; 190 (4): 1503–1510. doi: 10.1534/genetics.111.137026 PMID: 22271763

19. Würschum T, Langer SM, Longin C. Friedrich H., Korzun V, Akhunov E,Ebmeyer E, et al. Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. Theoretical and Applied Genetics. 2013; 126 (6): 1477–1486. doi: 10.1007/s00122-013-2065-1 PMID: 23429904

20. Reif JC, Zhao Y, Würschum T, Gowda M, Hahn V Genomic prediction of sunflower hybrid performance. Plant Breed. 2013; 132 (1): 107–114. doi: 10.1111/pbr.12007

21. Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink JL. Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats. The Plant Genome. 2011; 4 (2): 132. doi: 10.3835/plantgenome2011.02.0007

22. Hayes BJ, Visscher PM, Goddard ME. Increased accuracy of artificial selection by using the realized relationship matrix. Genetics research. 2009; 91 (1): 47–60. doi: 10.1017/S0016672308009981 PMID: 19220931

23. de Roos APW, Hayes BJ, Goddard ME. Reliability of genomic predictions across multiple populations. Genetics. 2009; 183 (4): 1545–1553. doi: 10.1534/genetics.109.104935 PMID: 19822733

24. Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA. The impact of genetic architecture on genome-wide evaluation methods. Genetics. 2010; 185 (3): 1021–1031. doi: 10.1534/genetics.110.116855 PMID: 20407128

25. Shahidi F. Rapeseed and canola: global production and distribution. In: Shahidi F (Ed.), Canola and Rapeseed: Production, Chemistry, Nutrition and Processing Technology, AVI Book, New York; 1990. pp. 3–14.

26. Iniguez-Luy FL, Federico ML. The genetics of *Brassica napus* L. In: Bancroft I., Schmidt R. Genetics and Genomics of *Brassicaceae*. Springer, New York; 2001. pp. 291–322. ISBN 978-1-4419-7118-0.

27. Delourme R, Falentin C, Fomeju BF, Boillot M, Lassalle G, André I, et al. High-density SNP-based genetic map development and linkage disequilibrium assessment in *Brassica napus* L. BMC Genomics. 2013; 14: 120. doi: 10.1186/1471-2164-14-120 PMID: 23432809

28. Snowdon R, Friedt W. Renewable energy: European biodiesel can be sustainable. Nature. 2012; 490 (7418): 37. doi: 10.1038/490037d

29. Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, et al. Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science. 2014; 345 (6199): 950–953. doi: 10.1126/science.1253435 PMID: 25146293

30. Becker HC, Engqvist GM, Karlsson B. Comparison of rapeseed cultivars and resynthesized lines based on allozyme and RFLP markers. Theoretical and Applied Genetics. 1995; 91 (1): 62–67. doi: 10.1007/BF00220859 PMID: 24169668

31. Hasan M, Seyis F, Badani AG, Pons-Kühnemann J, Friedt W, Lühs W, et al. Analysis of Genetic Diversity in the *Brassica napus* L. Gene Pool Using SSR Markers. Genet Resour Crop Evol. 2006; 53 (4): 793–802. doi: 10.1007/s10722-004-5541-2

32. Bus A, Körber N, Snowdon RJ, Stich B. Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. Theoretical and Applied Genetics. 2011; 123 (8): 1413–1423. doi: 10.1007/s00122-011-1676-7 PMID: 21847624

33. Cowling WA. Genetic diversity in Australian canola and implications for crop breeding for changing future environments. Field Crops Research. 2007; 104 (1–3): 103–111. doi: 10.1016/j.fcr.2006.12.014

34. Duvick J. Prospects for Reducing Fumonisin Contamination of Maize through Genetic Modification. Environ Health Prospect. 2001; 109 (s2): 337–342.

35. Buzza GC. Plant breeding. In: Kimber DS, McGregor DI. (Ed.) *Brassica* oilseeds: Production and utilization. CABI publishing, Wallingford, CT; 1995. pp. 153–175.

36. Renard M, Delourme R, Pierre J. Market introduction of rapeseed hybrid varieties: GCIRC Bulletin. 1997; 114–119.

37. Basunanda P, Radoev M, Ecke W, Friedt W, Becker HC, Snowdon RJ. Comparative mapping of quantitative trait loci involved in heterosis for seedling and yield traits in oilseed rape (*Brassica napus* L.). Theoretical and Applied Genetics. 2010; 120 (2): 271–281. doi: 10.1007/s00122-009-1133-z PMID: 19707740

38. Qian W, Chen X, Fu D, Zou J, Meng J. Intersubgenomic heterosis in seed yield potential observed in a new type of *Brassica napus* introgressed with partial *Brassica rapa* genome. Theoretical and Applied Genetics. 2005; 110 (7): 1187–1194. doi: 10.1007/s00122-005-1932-9 PMID: 15806350

39. Basunanda P, Spiller TH, Hasan M, Gehringer A, Schondelmaier J, Lühs W, et al. Marker-assisted increase of genetic diversity in a double-low seed quality winter oilseed rape genetic background. Plant Breed. 2007; 126 (6): 581–587. doi: 10.1111/j.1439-0523.2007.01404.x

40. Zou J, Zhu J, Huang S, Tian E, Xiao Y, Fu D, et al. Broadening the avenue of intersubgenomic heterosis in oilseed *Brassica*. Theoretical and Applied Genetics. 2010; 120 (2): 283–290. doi: 10.1007/s00122-009-1201-4 PMID: 19911158

41. Girke A, Schierholt A, Becker HC. Extending the rapeseed gene pool with resynthesized *Brassica napus* II: Heterosis. Theoretical and Applied Genetics. 2012; 124 (6): 1017–1026. doi: 10.1007/s00122-011-1765-7 PMID: 22159759

42. Snowdon RJ, Abbadi A, Kox T, Schmutzer T, Leckband G. Heterotic Haplotype Capture: precision breeding for hybrid performance. Trends in plant science. 2015; 20 (7): 410–413. doi: 10.1016/j.tplants.2015.04.013 PMID: 26027461

43. Snowdon RJ, Iniguez Luy, Federico L. Potential to improve oilseed rape and canola breeding in the genomics era. Plant Breed. 2012; 131 (3): 351–360. doi: 10.1111/j.1439-0523.2012.01976.x

44. Windhausen VS, Atlin GN, Hickey JM, Crossa J, Jannink JL, Sorrells ME, et al. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3. 2012; 2 (11): 1427–1436. PMID: 23173094

45. Hofheinz N, Borchardt D, Weissleder K, Frisch M Genome-based prediction of test cross performance in two subsequent breeding cycles. Theoretical and Applied Genetics. 2012; 125 (8): 1639–1645. doi: 10.1007/s00122-012-1940-5 PMID: 22814724

46. Wang Y, Mette MF, Miedaner T, Gottwald M, Wilde P, Reif JC, et al. The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. BMC Genomics. 2014; 15: 556. doi: 10.1186/1471-2164-15-556 PMID: 24997166

47. Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, et al. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nature Genetics. 2012; 44 (2): 217–220. doi: 10.1038/ng.1033 PMID: 22246502

48. Wang X, Wang H, Long Y, Li D, Yin Y, Chen L, et al. Identification of QTLs associated with oil content in a high-oil *Brassica napus* cultivar and construction of a high-density consensus map for QTLs comparison in *B. napus*. PloS one. 2013; 8 (12): e80569. doi: 10.1371/journal.pone.0080569 PMID: 24312482

49. Cho Y, Park C, Kwon S, Chin J, Ji H, Park KJ, et al. Key DNA Markers for Predicting Heterosis in F1 Hybrids of japonica Rice. Breed. Sci. 2004; 54 (4): 389–397. doi: 10.1270/jsbbs.54.389

50. Teklewold A, Becker HC. Comparison of phenotypic and molecular distances to predict heterosis and F1 performance in Ethiopian mustard (*Brassica carinata* A. Braun). Theoretical and Applied Genetics. 2006; 112 (4): 752–759. doi: 10.1007/s00122-005-0180-3 PMID: 16365759

51. Schrag TA, Melchinger AE, Sørensen AP, Frisch M. Prediction of single-cross hybrid performance for grain yield and grain dry matter content in maize using AFLP markers associated with QTL. Theoretical and Applied Genetics. 2006; 113 (6): 1037–1047. doi: 10.1007/s00122-006-0363-6 PMID: 16896712

52. Piepho HP. Ridge Regression and Extensions for Genomewide Selection in Maize. Crop Science. 2009; 49 (4): 1165. doi: 10.2135/cropsci2008.10.0595

53. Beck DL, Vaal SK and Crossa J. Heterosis and combining ability of CIMMYT's tropical early and intermediate maturity maize (*Zea mays*) germplasm. Maydica. 1990; (35:): 279–285.

54. Zhao Y, Zeng J, Fernando R, Reif C. Genomic prediction of hybrid wheat performance. Crop Science. 2013; 53:802–810. doi: 10.2135/cropsci2012.08.0463

55. Qian W, Sass O, Meng J, Frauen M, Jung C. Heterotic patterns in rapeseed (*Brassica napus* L.): I. Crosses between spring and Chinese semi-winter lines. Theoretical and Applied Genetics. 2007; 115: 27–34. doi: 10.1007/s00122-007-0537-x PMID: 17453172

56. Bekele WA, Fiedler K, Shiringani A, Schnaubelt D, Windpassinger S, Uptmoor R, et al. Unravelling the genetic complexity of sorghum seedling development under low-temperature conditions. Plant Cell & Environment. 2014; 37 (3): 707–723. doi: 10.1111/pce.12189

57. Melchinger AE, Gumber RK. Overview of heterosis and heterotic groups in agronomic crops. In: Lamkey KR, Staub JE, editors. Concepts and breeding of heterosis in crop plants. Madison, Wisconsin: CSSA. 1998. pp. 29–44.

58. Rogers JS. Measures of genetic similarity and genetic distances. Studies in Genetics. VII. Univ. Texas Publ. 1972; 7213, 145–153.

59. Hofheinz N, Frisch M. Heteroscedastic ridge regression approaches for genome-wide prediction with a focus on computational efficiency and accurate effect estimation. G3 (Bethesda, Md.). 2014; 4 (3): 539–546. doi: 10.1534/g3.113.010025

60. R 3.1.0 Development Core Team R. A language and environment for statistical computing. 2014; R Foundation for Statistical Computing, GWDG Gottingen, Germany, http://www.R-project.org (accessed 15 April, 2014).

61. Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. Applied Statistics. 1979; 28 (1): 100.

62. Caliński T & Harabasz J. A dendrite method for cluster analysis. Commun. Stat. 1974; 3, 1–27. doi: 10.1080/03610927408827101

63. Würschum T, Abel S, Zhao Y, Léon J. Potential of genomic selection in rapeseed (*Brassica napus* L.) breeding. Plant Breed. 2014; 133 (1): 45–51. doi: 10.1111/pbr.12137

64. Jia Y, Jannink JL. Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics. 2012; 192 (4): 1513–1522. doi: 10.1534/genetics.112.144246 PMID: 23086217

65. Wu J, Shi C, Zhang H. Partitioning genetic effects due to embryo, cytoplasm and maternal parent for oil content in oilseed rape (*Brassica napus* L.). Genet. Mol. Biol. 2006; 29 (3): 533–538. doi: 10.1590/S1415-47572006000300023

66. Delourme R, Falentin C, Huteau V, Clouet V, Horvais R, Gandon B, et al. Genetic control of oil content in oilseed rape (*Brassica napus* L.). Theoretical and Applied Genetics. 2006; 113 (7): 1331–1345. doi: 10.1007/s00122-006-0386-z PMID: 16960716

67. Saatchi M, McClure MC, McKay SD, Rolf MM, Kim J, Decker JE, et al. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genetics Selection Evolution. 2011; 43: 40. doi: 10.1186/1297-9686-43-40

68. Habier D, Fernando RL, Dekkers JCM. The impact of genetic relationship information on genome-assisted breeding values. Genetics. 2007; 177 (4): 2389–2397. doi: 10.1534/genetics.107.081190 PMID: 18073436

69. Technow F, Riedelsheimer C, Schrag TA, Melchinger AE. Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. Theoretical and Applied Genetics. 2012; 125 (6): 1181–1194. doi: 10.1007/s00122-012-1905-8 PMID: 22733443

70. Grant I, Beversdorf WD. Heterosis and combining ability estimates in spring-planted oilseed rape (*Brassica napus* L.). Can. J. Genet. Cytol. 1985; 27 (4): 472–478. doi: 10.1139/g85-069

71. Lefort-Buson M, Guillot-Lemoine B, Dattee Y. Heterosis and genetic distance in rapeseed (*Brassica napus* L.): crosses between European and Asiatic selfed lines. Genome. 1987; 29 (3): 413–418. doi: 10.1139/g87-072

72. Brandle JE, McVetty PBE. Heterosis and Combining Ability in Hybrids Derived from Oilseed Rape Cultivars and Inbred Lines. Crop Science. 1989; 29 (5): 1191. doi: 10.2135/cropsci1989.0011183X002900050020x

73. Li Y, Ma C, Fu T, Yang G, Tu J, Chen Q, et al. Construction of a molecular functional map of rapeseed (*Brassica napus* L.) using differentially expressed genes between hybrid and its parents. Euphytica. 2006; 152 (1): 25–39. doi: 10.1007/s10681-006-9173-9

74. Badani AG, Snowdon RJ, Wittkop B, Lipsa FD, Baetzel R, Horn R, et al. Colocalization of a partially dominant gene for yellow seed colour with a major QTL influencing acid detergent fibre (ADF) content in different crosses of oilseed rape (*Brassica napus*). Genome. 2006; 49, pp. 1499–1509. doi: 10.1139/g06-091 PMID: 17426765

75. Radoev M, Becker HC, Ecke W. Genetic analysis of heterosis for yield and yield components in rapeseed (*Brassica napus* L.) by quantitative trait locus mapping. Genetics. 2008; 179 (3): 1547–1558. doi: 10.1534/genetics.108.089680 PMID: 18562665

76. Mei J, Fu Y, Qian L, Xu X, Li J, Qian W. Effectively widening the gene pool of oilseed rape (*Brassica napus* L.) by using Chinese *B. rapa* in a 'virtual allopolyploid' approach. Plant Breed. 2011; 130 (3): 333–337.

77. Bernardo R, Yu J. Prospects for Genomewide Selection for Quantitative Traits in Maize. Crop Science. 2007; 47 (3): 1082. doi: 10.2135/cropsci2006.11.0690