Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Environment-dependent genetic architecture of seed vigor reveals antagonistic allelic effects under water deficit in rapeseed (Brassica napus)

Khadidiatou Demba
Institut Agro Rennes-Angers
Vanessa Clouet
Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement
Cécile Baron
Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement
Marie-Hélène Wagner
GEVES
Didier Demilly
GEVES
Sylvie Ducournau
GEVES
Anne Laperche
Institut Agro Rennes-Angers
Nathalie Nesi
nathalie.nesi@inrae.fr

Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement

Research Article

Keywords: rapeseed, seed germination, vigor, water stress, QTL stability

Posted Date: August 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10394115/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Environment-dependent genetic architecture of seed vigor reveals antagonistic allelic effects under water deficit in rapeseed (*Brassica napus*)

Khadidiatou Demba[1], Vanessa Clouet[1], Cécile Baron[1], Marie-Hélène Wagner[2], Didier Demilly[2], Sylvie Ducournau[2], Anne Laperche[1,*], Nathalie Nesi[1,*]

[1] Institute of Genetics, Environment and Plant Protection (IGEPP), INRAE, Institut Agro Rennes-Angers, Université de Rennes, 35650 Le Rheu, France

[2] Groupe d'Etude et de Contrôle des Variétés Et des Semences (GEVES), 49071 Beaucouzé, France

*Authors for correspondence : anne.laperche@institut-agro.fr; nathalie.nesi@inrae.fr

Background

Seed vigor is the ability of seeds to germinate quickly and uniformly and produce seedlings capable of emerging under a wide range of environmental conditions. It is a key determinant of crop establishment, particularly under water limited conditions. However, the genetic architecture underlying seed vigor and its stability across environment remain poorly understood. This study aims to investigate the genetic basis of seed vigor under water deficit. This was achieved by using a panel of 192 rapeseed accessions encompassing winter, spring, and semi-winter genotypes, and by phenotyping germination capacity and speed under osmotic stress induced by polyethylene glycol (-0.5 MPa) using a high-throughput imaging platform.

Results

Genetic analyses identified 22 QTL associated with seed vigor related-traits, explaining between 1.2 and 13.1% of phenotypic variance. Germination speed and germination capacity traits were controlled by distinct genomic regions, although several co-localizations suggested

shared genetic control of several components of seed germination. Evaluation of QTL stability across water regimes pointed out four significant QTL-by-environment interactions, revealing strong environmental dependency of allelic effects. Notably, favorable alleles under water deficit became unfavorable under optimal conditions, indicating antagonistic genetic effects across environments. Candidate gene analyses within environment-responsive QTL identified nine genes combining coding-sequence polymorphism and germination-associated expression pattern.

Conclusion

This study reveals a complex and environment-dependent genetic architecture of seed vigor and provides new targets for breeding oilseed rape varieties with improved germination under water-limited conditions.

Keyword: rapeseed, seed germination, vigor, water stress, QTL stability

Background

Crop establishment is a key factor in yield stability. Rapid, uniform and complete germination and as well as seedling emergence are essential for successful crop development. However, these processes are impacted by unpredictable weather conditions at the time of sowing, particularly a lack of rainfalls and the increasing frequency of heatwaves. This issue is of particular importance for winter oilseed rape (*Brassica napus*), a crop that is highly sensitive to early-stage environmental conditions, particularly under water limitation. Early sowings, traditionally recommended to secure crop establishment, are now often subject to combined stresses caused by water shortages and heat. These conditions severely impair germination and result in poor and uneven seedling emergence. Consequently, insufficient plant density results in inadequate canopy development and increased competition with weeds, leading to yield losses [1]. Conversely, delaying sowing to avoid drought stress often increases exposure to biotic stresses, particularly from the flea beetle (*Psylliodes chrysocephalus*). Therefore, farmers have to face critical challenges between abiotic and biotic constraints when setting sowing dates. In this regard, it is therefore essential to improve seed vigor in order to cope with environmental uncertainties during the early stages of development and to ensure optimal crop establishment in the context of climate change.

Seed vigor is defined as the ability of seeds to germinate rapidly, uniformly and robustly across a variety of environmental conditions [2,3]. It results from a complex set of traits encompassing the process of seed germination and seedling emergence, resulting from a series of physiological processes regarding both the seed and the developing plantlet [1,4]. Seed vigor is influenced by multiple factors, including the genetic background, environmental conditions and the maternal environment during seed development, as well as post-harvest handling conditions [3]. More broadly, seed vigor arises from complex interplays between the genotype, pedoclimatic conditions and crop practices, before, during and after sowing [1,3,5].

The genetic architecture of seed germination has been extensively investigated using a wide range of phenotypic traits. Many studies have focused on static traits such as final germination percentage, or indirect proxies of seedling vigor such as root or hypocotyl growth [6,7]. However, germination is inherently a dynamic process, involving multiple sequential mechanisms from imbibition to radicle emergence. To better characterize plant responses to environmental constraints while limiting phenotypic dimensionality, some studies have adopted integrative approaches. These approaches include stress tolerance indices or dimensionality reduction methods, which aggregate correlated traits into synthetic descriptors of stress response [6]. In contrast, other approaches decompose germination into multiple components, including germination speed, germination percentage at different time and uniformity [8,9]. This strategy allows a finer characterization of germination dynamic and highlighting the multi-dimensional nature germination. Together, these studies have identified numerous QTL associated with these traits, supporting a polygenic architecture dominated by loci of small to moderate effect.

Most studies investigating the genetic architecture of germination have been conducted under a single environmental condition [6-9]. Under optimal conditions, seeds from different genotypes often show similarly high levels of germination. Conversely, under stressful conditions, their ability to germinate can differ markedly due to variation in seed vigor [3]. In this context, genotype by environment (G×E) interactions play central role in shaping germination behavior across contrasting environmental conditions. Such interactions may lead to genotype re-ranking between environments, thereby reducing the stability and predictive value of genetic associations identified through genetic approaches. These observations highlight the importance of integrating environmental variability into the study of genetic bases of seed vigor. In particular, it remains unclear whether genetic loci controlling germination dynamics are stable across environments or largely environment-specific. Addressing this

question is essential for breeding climate-resilient oilseed rape varieties adapted to water-limited conditions.

To dissect the genetic architecture of seed vigor and its response to water availability, a Genome Wide Association Study (GWAS) was conducted using high-resolution phenotypic data collected under polyethylene glycol (PEG) induced osmotic stress. The stability of the identified QTL was subsequently evaluated using germination data obtained under optimal water conditions. A population including winter, spring and semi-winter accessions of *Brassica napus*, thereby capturing a broad genetic diversity within cultivated oilseed rape, was tested for germination. Germination dynamics were monitored using an automated high-throughput phenotyping platform enabling the precise characterization of temporal germination traits. Our results led to (i) identify genetic loci associated with seed vigor-related traits in oilseed rape under water stress, (ii) quantify the contribution of QTL×E interactions by distinguishing stable and environment-specific loci and (iii) prioritize candidate genes underlying these genomic regions. Overall, this work provides new insights into the environment-dependent genetic architecture of seed vigor and identifies potential targets for breeding oilseed rape varieties with improved capacity under water-limited conditions.

Materials and Methods

Seed materials and genotyping data

The diversity panel used in this study consisted of 192 genetically diverse *B. napus* inbred lines, including 100 winter oilseed rape (WOSR), 60 spring oilseed rape (SOSR), 12 winter fodder rape (WFR), 2 swedes and 18 accessions classified as exotics or others (**Additional Table 1**). This panel is fully included in the one described previously [9]. To minimize maternal and seed production environmental effects, all genotypes were grown in a common field nursery in Le

Rheu, France, during the 2020/2021 crop season. Seeds lots were produced under open pollination conditions and were harvested simultaneously. Seeds were subsequently stored under the same conditions (room temperature). All accessions were genotyped using the Brassica 60K Infinium® SNP array (Illumina, Inc., San Diego, CA) [10]. SNP quality was performed using thresholds of 5% for the minor allele frequency (MAF), 10% for missing data, and 10% for heterozygosity, resulting in a final dataset of 33,544 SNP. For genotype filtering, accessions with more than 10% missing data or more than 8% heterozygosity were excluded. The missing SNP data were imputed using BEAGLE software following the method described in [11]. SNP positions were anchored to the *B. napus* Darmor-*bzh* v10 reference genome [12]. Genotyping data arising from an exome capture assay were also used [13]. This dataset consists in 377 K SNP [14] available for 74 of our accessions (**Additional Table 1**).

Acquisition and analyses of germination data

Seed germination under water-limiting conditions was evaluated using the high-throughput germination phenotyping platform at the National Seed Testing Station (GEVES) in Angers, France [15]. For each genotype, four biological replicates of 25 seeds were sown on filter (Whatman™ Grade 3644 seed test papers 220 x 225 mm, GE Healthcare Life Sciences) placed into a Petri dish. Water shortage was imposed using a polyethylene glycol solution (PEG-8000, BP233-1, Fisher BioReagents ™) in order to induce an osmotic potential of -0.5 MPa (corresponding to a solution of PEG-8000 at 19 g/l). Germination tests were conducted in darkness at 20°C for 120 h. Surface temperature was recorded every 30 minutes using the WIT e@sy V9.2.0 system. Images were acquired every two hours, allowing automated monitoring of germination kinetics through image analysis as described by [16]. Germination-related traits were as following (**Table 1**): first germination time (FGT), time to reach 20% or 50% germination (T20 and T50 respectively), mean germination time (MGT), slope of the germination curve (SG), germination uniformity (Uni, calculated as the variance around MGT,

136 [17]), area under the germination curve (AUC), germination percentage at 36h or 96h after
137 sowing (GP36, GP96 respectively) and the final germination percentage (GP).

138 All statistical analyses were performed using R software (version 4.4.2) and the packing *lme4*
139 was used to fit mixed linear models. The package *betareg* was used to transform GP data using
140 beta regression, making these traits suitable for mixed linear model analyses. The package
141 *emmeans* was used to calculate the Best Linear Unbiased Estimators (BLUE). In addition, to
142 explore the phenotypic distribution into the panel of diversity, the package *dplyr* was performed
143 to calculate the minimum, maximum, mean, standard deviation (SD) and coefficient of
144 variation (CV) for each trait. The coefficient correlation between the different seed vigor traits
145 was obtained using the package *corrplot* and based on s*pearman* method due to nonlinear
146 relationship between traits.

147 BLUE values were calculated for each genotype and each trait using the flowing linear model
148 for the experimentation carried out under water limitation condition:

149 $$(1)\ Y_{ij} = \mu + G_i + R_j + \varepsilon_{ij}$$

150 where Y_{ij} is the observation corresponding to genotype i and repetition j, G_i corresponds to the
151 fixed genetic effect for genotype i, R_j is the fixed effect of repetition j and ε_{ij} is the residual
152 effect. We assumed that ε_{ij} was independent and to follow a normal distribution

153 The broad-sense heritability (H^2) was calculated for each trait by using the following linear
154 mixed model to extract the genotypic and error variance:

155 $$(2)\ Y_{ij} = \mu + G_i + R_j + \varepsilon_{ij}$$

156 where Y_{ij} is the observation corresponding to genotype i and repetition j, G_i corresponds to the
157 random genetic effect for genotype i, R_j is the fixed effect of repetition j and ε_{ij} is the residual
158 effect. H^2 was calculated as follows:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_e^2}{r}}$$

where σ_G^2 is the genotypic variance, σ_e^2 is the error variance, and r is the number of replicates.

Genetic analyses

The genetic structure of the diversity set was characterized using the first two axes of a principal coordinates analysis (PCoA, [18]) based on the SNP data. This analysis relies on the genetic distances among genotypes, calculated from a dissimilarity matrix defined as (1-K).

Genomic kinship (K) was estimated following the method of [19] algorithm such as:

$$K = \frac{WW^T}{M}$$

where W represents the allelic frequency-scaled matrix of size N × M, with N corresponding to the number of genotypes, M to the number of molecular markers and W^T to the transposed W matrix.

GWAS was performed using a mixed linear model implemented in FaST-LMM algorithm [20]. In addition, to reduce false associations (types I and II errors) due to population structure and kinship between genotypes, the kinship matrix was re-estimated for each chromosome using 'leave-one-chromosome-out' (LOCO) approach [21]. Thus, for each chromosome, the kinship matrix was calculated excluding markers from the chromosome being tested. The following mixed model that takes these two factors into account [22] was used:

$$(3)\ \mathrm{y} = \mu + x_{SNP}\beta_{SNP} + \mathrm{Qv} + \mathrm{Z}u + \varepsilon$$

where y is the vector of the phenotypic values under study (BLUEs obtained for each genotype), μ correspond to the intercept, x_{SNP} to the vectors of incidence (0, 2) that reports the alleles composition of each genotype for the SNP under tested genotype SNP, β_{SNP} to the SNP effects

vector, Q to the matrix describing the population structure obtained with PCoA, v to the population structure effects; Z the incidence matrix for the polygenic effect and u to the random polygenic effect $u \sim \mathcal{N}(0, K_{LOCO}\sigma_G^2)$; K_{LOCO} correspond to the kinship matrix calculated which LOCO approach and σ_G^2 to the genetic variance. $\varepsilon \sim \mathcal{N}(0, I\sigma_e^2)$ where σ_e^2 correspond to the residual variance.

To take account the correlation between markers due to linkage disequilibrium (LD), the genome-wide significance threshold was determined using the simpleM method [23]. Markers were considered as significantly associated with a trait if the −log10 (p-value) exceeded a 5% threshold (using our data, the 5% threshold was 3.57).

For each SNP detected by GWAS, the favorable allele corresponded to the allele associated with improved phenotypic performance e.g., increasing GP36, GP96, GP, AUC and SG values while decreasing FGT, MGT, T20, T50 and Uni values.

Definition of QTL and their confidence intervals

A QTL region was defined when at least one SNP exceeded the significance threshold (3.57) and was supported by at least one neighboring SNP displaying a similar or near significant association signal. Two complementary approaches were then used to define a QTL confidence interval. The first approach was based on local linkage disequilibrium (LD). For each QTL, the peak SNP was defined as the marker with the highest - log10 (p-value). The confidence interval was subsequently determined by including all neighboring SNPs in LD (r^2) with the peak SNP ($r^2 > 0.2$), a threshold commonly used in GWAS studies [24]. The second approach relied on haplotype block structure. Haplotype blocks were identified using Haploview (version 4.2) and the solid spine of LD algorithm. This method defines haploblocks as group of contiguous SNP connected by strong LD. Specially, SNPs located at the boundaries of the block must be in strong LD with the central SNP, although not all internal SNP pairs are required to exhibit

strong LD with each other [25]. This approach generates continuous haplotype blocks along chromosomes, while avoiding the definition of isolated SNP regions. Using this framework, a QTL confidence interval hereafter referred to as the "LD block interval", was defined as the genomic region encompassing all haplotype blocks containing at least one SNP whose -log10 (p-value) exceeded the significance threshold.

Analysis of QTL×E interactions

To assess the stability of seed vigor-associated loci across contrasting water regimes, phenotypic data generated under non-limiting water condition were obtained from [9]. Briefly, the same diversity set was evaluated using the same phenotyping platform under optimal water conditions for 96 h. In this experiment, seeds were directly sown on blotting paper directly placed on Jacobsen germination tables. Only traits common to both experiments were retained for comparative analyses, including FGT, MGT, T50, GP36, GP and AUC. Hereafter, non-limiting and water-limited conditions are referred to W+ and W-, respectively. To determine whether QTL effects were stable across water regimes or exhibited environment-dependent effects, loci identified under W- conditions were additionally tested under W+ conditions. Analyses were performed using genotype BLUE values estimated separately for each experiment, with genotypes considered as replicates. QTL×E interactions were assessed using the following linear model:

$$(4)\ Y = \mu + QTL + E + QTLxE$$

Where Y correspond to observed phenotypic trait, QTL to loci effect and E to experimental condition (W+, W-).

Genomic resources

Putative gene functions were inferred using *Arabidopsis thaliana* orthologs. Germination related expression profiles were explored using public databases, including TAIR

(https://www.arabidopsis.org/) and e-plant https://bar.utoronto.ca/eplant/. For comparative genomic analyses, multiple *Brassica napus* reference genomes were used, including *Darmor-bzh* v10 [12], *Quinta*, *Gangan*, *Shengli*, *ZS11* [26], *Express* [27] and *Da-Ae* [28]. Multiple sequence alignments across divers references genomes were performed to detected potential structural variations including missing or rearranged regions, which could affect marker positioning and QTL assignment.

Results

Seed vigor phenotypic structure is independent of genetic structure

Analysis of the genetic structure of the diversity set using SNP data and principal coordinate analysis (PcoA) revealed that the first two axes explained 47% of the total genetic variation (**Fig. 1A**). Three distinct genomic clusters were identified. Cluster 1 was predominantly composed of spring rapeseed genotypes (n=44), cluster 2 mainly included Asian rapeseed genotypes (n =23), whereas cluster 3 consisted primarily of winter rapeseed genotypes (n =125). These results indicate a strong genetic structuring of the diversity set. Principal component analysis (PCA) performed on the phenotypic traits showed that the first axes explained 84% of the total phenotypic variation (**Fig. 1B**). The analysis also identified three distinct phenotypic groups corresponding to genotypes exhibiting rapid and efficient germination (group 1), slow and poor germination performance (group 2), and intermediate germination behavior (group 3). When genotypes were projected onto the phenotypic PCA according to their genomic cluster membership, genotypes were broadly distributed across the phenotypic groups. Thus, genomic clusters did not drive the phenotypic groups identified by PCA. This lack of concordance suggests that variation in seed vigor-related traits is distributed throughout the genetic diversity of the panel and it is not primarily driven by its overall population structure.

252 Seed vigor traits exhibit contrasting levels of genetic control

253 The assessment of phenotypic variability in seed vigor traits, based on the coefficient of
254 variation (CV) and broad-sense heritability (H^2) estimates, revealed differences among traits
255 (**Table 2**). T20, T50, MGT, GP96, GP and AUC exhibited moderate levels of phenotypic
256 variation, with CV value ranging from 8.19% to 15.71%, together with high broad-sense
257 heritability estimates ($H^2 > 0.72$). These results indicate that genetic factors account for a large
258 proportion of the phenotypic variation observed for these traits. In contrast, Uni and FGT
259 displayed low to moderate phenotypic variation (10,43% and 17,11% respectively) and
260 relatively low heritability estimates ($H^2 \approx 0.34$), suggesting a stronger influence of
261 environmental factors on their expression. SG showed higher phenotypic variation (CV =
262 32.71%), while maintaining a high heritability estimate ($H^2 > 0.73$), indicating that most of the
263 observed variation is genetically determined. Conversely, GP36 exhibited the highest level of
264 phenotypic variability (CV = 54.81) together with a low heritability ($H^2 = 0.22$), suggesting a
265 limited genetic contribution under our experimental conditions. The relationships among seed
266 vigor traits were subsequently investigated through phenotypic correlation analysis (**Table 3**).

267 Distinct components of seed vigor showed correlations

268 Significant positive correlations were observed among FGT, T20, T50 and MGT ($0.28 < r <$
269 0.96, $p < 0.001$), as well as among GP96, GP and AUC ($0.69 < r < 0.94$, $p < 0.01$) (**Table 3**).
270 These associations are consistent with the biological meaning of these traits. A longer delay
271 before the first germination (FGT) was associated with an increase in mean germination time
272 (MGT) and in the time to reach 20% (T20) and 50% (T50) germination. Likewise, a higher
273 gemination percentage at 96h (GP96) was associated with higher final germination percentage
274 (GP) and greater overall germination performance (AUC). In contrast, traits related to
275 germination speed (FGT, T20, T50 and MGT) were negatively correlated with traits describing
276 germination capacity and overall germination performance (GP96, GP and AUC). These

277 relationships indicate that genotypes exhibiting rapid germination generally achieve higher
278 germination percentages and superior overall germination performance. Interestingly, the early
279 germination percentage measured at 36h (GP36) was not significantly correlated with later
280 germination traits, including GP96 and GP. However, GP36 remained positively correlated
281 with AUC, suggesting that early germination contributes to overall germination performance
282 without necessarily predicting the final germination capacity of a genotype. The slope of
283 germination curve (SG) and germination uniformity (Uni) were positively correlated with each
284 other and with GP36, GP96, GP and AUC. Conversely, both traits were negatively correlated
285 with FGT, T20, T50 and MGT. Together, these results indicate that genotypes characterized by
286 rapid and uniform germination tend to exhibit higher germination percentages and enhanced
287 overall germination performance. For the subsequent analysis traits were generally classified
288 into two categories according to the biological processes they described: (i) germination speed
289 traits, related to germination timing and kinetics (FGT, T20, T50, MGT, SG and Uni), and (ii)
290 germination capacity traits, related to the proportion of germination over the time and global
291 performance (GP36, GP96, GP and AUC).

292 Genetic analyses revealed 22 QTL controlling seed vigor under water stress
293 To unravel the genetic architecture of seed vigor under water limiting conditions, GWAS was
294 performed on the panel using the FASTLMM model. This analysis initially identified a total of
295 120 SNP peaks associated with seed vigor related traits that were consolidated and summarized
296 into 23 QTL regions using local linkage disequilibrium (LD) and haplotype block structure
297 (**Additional Tables 2, 3, 4**).

298 However, two of these loci, namely QTL.A06.3 and QTL.C07.1, exhibited a strong and
299 unexpected LD pattern ($r^2 \approx 1$) (**Fig. 2**). Given that the corresponding genomic regions were
300 known to be homoeologous, we hypothesized that one of the two QTL was a false positive.
301 This was confirmed by sequence alignments of the corresponding genomic regions in A06 and

C07, using seven reference genomes (**Additional Fig. 1**). Indeed, the *Darmor-bzh* genome was shown to contain a gap in the neighborhood of QTL.A06.3, which led to a biased mapping of the SNP associated with A06 onto the homologous region C07. Overall, these results showed that QTL.C07.1 was a false positive locus. Consequently, 22 unique QTL regions were retained for further analyses.

QTL were detected for all seed vigor traits at the exception of SG (**Table 4**). These loci individually explained between 1.2% and 13.1%. of the phenotypic variance (R^2), supporting the polygenic architecture of the genetic determinism. Among these QTL, 18 were located on the A sub-genome and 4 on the C sub-genome, illustrating an uneven genomic distribution of seed vigor associated loci. Six QTL were exclusively associated with germination speed traits. Four loci were trait-specific, QTL.A03.1 and QTL.A03.2 for FGT; QTL.A06.1 and QTL.A09.1 for T50. In addition, two loci were associated with multiple speed-related traits, QTL.A07.2 (MGT and T50) and QTL.A08.1 (T50, Uni and T20), suggesting either pleiotropic effects or closely linked loci controlling multiple components of germination dynamics. Fourteen QTL were exclusively associated with germination capacity traits. Co-localization was identified for QTL.C03.1 (AUC and GP96) and QTL.C03.2 (AUC, GP96 and GP), highlighting genomic regions potentially involved in the overall regulation of germination performance. Several trait-specific loci, were specially associated with GP36 (QTL.A02.1, QTL.A04.1, QTL.A05.1, QTL.A06.2, QTL.A07.1, QTL.A10.1 and QTL.C06.1) and GP (QTL.A06.3, QTL.A10.2, QTL.A10.3, QTL.A10.4, QTL.C02.1). Finally, two loci were associated with both germination speed and germination capacity traits. QTL.A07.3 (T50 and AUC) and QTL.A07.4 (T20, GP96, GP and AUC). All genomic regions controlling several traits (e.g., GP96, GP and AUC) was consistent with the high correlation between these traits ($0.69 < r < 0.94$, $p < 0.01$). For most QTL, favorable allele corresponded to the minor allele within the diversity set (Table 4), indicating that beneficial variants are generally present at low frequency in the population. In

most co-localizing regions, allelic effects were consistent across traits. However, QTL.A08.1 represented a notable, as the allele associated with reduced T50 and MGT values was simultaneously associated with increased Uni values, indicating a potential trade-off between rapid germination and germination synchrony.

QTL stability and QTL × environment interactions

To evaluate the stability of seed vigor associated loci across contrasting water conditions, the effects of identified QTL under limited water (W-) conditions were tested under non-limiting water (W+) conditions using the phenotypic dataset previously described by [9]. Phenotypic comparisons showed that water stress significantly increased the germination delay, mean germination time and the time to reach 50% of germination. Conversely, it decreased the early germination percentage and overall germination performance, while the final germination percentage remained unaffected (**Additional Fig. 2**). The analysis of variance of QTL by environment interactions (QTL×E) identified four loci exhibiting significant interactions with water availability (**Table 5**), indicating strong water-depending genetic effects. For these traits (FGT, GP36 and GP), the individual main effects were generally limited, whereas QTL×E interactions consistently contributed significantly to phenotypic variation (1.3% to 3%, $p<0.001$). Environmental conditions explained a substantial proportion of phenotypic variation for GP36 and FGT (33% to 47%, $p<0.001$) and not affected GP (0.1, $p>0.05$). These results support the hypothesis that the effects of these loci strongly depend on water availability.

To further characterize these interactions, allelic effects were estimated separately under each water condition using BLUE values (**Fig. 3**). Several loci displayed allelic effects across environment. For QTL.A03.1, associated with first germination time (FGT), the favorable allele (1) reduced germination delay from 38h to 35h under W- conditions, but increased it from 28.9h to 30.5h under W+ conditions. Similarly, the favorable alleles at QTL.A02.1 and QTL.A04.1, increased early germination (GP36), under W- (3% to 4%) but reduced GP36 under W+

conditions (18.3% to 13%). A similar antagonistic pattern was observed for QTL.A06.3, where the favorable allele slightly increased GP (93% to 93.9%) under W- but reduced germination under W+ conditions (93% to 89%).

Candidate genes

To identify candidate genes underlying the detected QTL, loci exhibiting significant QTL×E interactions were first prioritized. To further refine these genomic regions intervals and increase mapping resolution, an additional resource was used consisting of exome capture data, enabling to detection of polymorphisms within coding regions. Because the exome capture data were available for only 74 accessions, we first assessed the representativeness of this subset relative to the entire panel. To this end, SNP minor allele frequency (MAF) distribution from the 60K array was compared between the exome capture subset and the full diversity set (**Additional Table 5**). Coding region polymorphism within the QTL intervals were then identified by selecting SNP with MAF >0.05 in the exome capture subset. This approach enabled the identification of genes carrying potentially functional coding-sequence variation. Therefore, significant polymorphisms were detected exclusively within the two loci associated with early germination (GP36) namely QTL.A02.1 and QTL.A04.1. Whitin these regions, 8 genes in QTL.A02.1 and 16 genes within QTL.A04.1 showed significant exon level polymorphism (**Table 6**). To further refine the list of candidates, public available germination related expression datasets (e-plant and TAIR), were examined. Five candidate genes within QTL.A02.1 and four genes within QTL.A04 were expressed during germination. Together, these analyses identified high confidence candidate genes combining both coding region polymorphism and germination associated expression pattern.

Discussion

Potential false-positive QTL arising from SNP misalignment in homoeologous regions

In this study we performed a GWAS analysis to identify loci involved in seed vigor in oilseed rape. However, quality of the outputs of genetic analysis not only relies on the quality of phenotyping but also on the genotyping. Polyploid genomes contain highly similar homoeologous regions that can complicate marker assignment to unique genomic positions [29-33]. In our study, the apparent duplication of QTL signals between chromosome A06 and C07 likely resulted from the absence of the corresponding of A06 sequence in the Darmor-*bzh* reference genome, leading to the misalignment of markers onto the homeologous region on C07. This discovery is supported by multiple-genome alignment analyses, which indicate that A06 derived sequences preferentially align to the corresponding region on C07 when the A06 reference sequence is missing. More broadly, these results emphasize the limitations of relying on a single reference genome in GWAS analyses and support pangenome approaches [34]. Overall, our results demonstrate that homeologous regions can substantially influence marker positioning, LD patterns and QTL interpretation in *B. napus*. This highlights the importance of integrating comparative genomics approaches to improve the reliability of QTL identification and to avoid overestimating the number of independent loci. Despite these methodological challenges, a subset of QTL displayed consistent effects across traits and environmental conditions, supporting their biological relevance. Consequently, the following sections focus on the biological interpretation of these robust QTL in terms of seed vigor dynamics, environmental plasticity and underlying genes.

Seed vigor is a dynamic and environmentally responsive character

This study provides new insights into genetic architecture of seed vigor in *Brassica napus* by combining fine scale phenotyping, GWAS and cross-environment validation. Our results confirmed that seed vigor is a complex and dynamic character, structured around both

germination speed and germination capacity components. The strong correlations observed within these groups of traits, together with their contrasting relationships previously reported under optimal condition [9], suggest coordinated regulation of germination processes across successive developmental stages. Genotypes exhibiting faster germination generally also displayed higher germination capacity, indicating that rapid germination is closely associated with improved overall germination performance. Importantly, the use of dynamic traits, including germination time courses and AUC, enabled a more integrative characterization of germination performance compared to classic endpoint measurements. In contrast to final germination percentage, these dynamic traits capture both the temporal progression and the overall efficiency of germination [3]. Among them, AUC emerged as a robust synthetic trait, combining both temporal and quantitative aspects of germination, and may therefore represents a relevant target for genetic analyses and breeding. In addition, the relatively low heritability observed for FGT, GP36 and Uni highlights the strong influence of environmental conditions on the earliest stages of germination and on germination synchrony. This result supports the environmental conditions can influence several traits related to seed vigor.

Polygenic architecture and sub-genome contribution to seed vigor

The identification of 22 unique QTL with small to moderate effects supports a polygenic architecture of seed vigor, consistent with previous studies [8,9]. This result confirms that seed vigor is controlled by multiple loci contributing jointly to different components of seed vigor rather than few major effects genes. Most QTL were located on the A sub-genome suggesting an uneven contribution of the two sub-genomes to seed vigor related traits. This pattern may reflect differences in evolutionary history or historical selection between sub-genomes [35]. Several QTL regions showed co-localization across multiple traits (e.g., AUC, GP96 and GP). This pattern is consistent with high correlation between these traits ($0.69 < r < 0.94$) and suggests either pleiotropic effects or tight linkage between loci controlling different aspects of

germination. Notably, most co-localization regions displayed consistent allelic effects across traits, supporting the hypothesis of a shared genetic control. However, the antagonistic effect for QTL.A08.1 indicates that trade-offs may exist between germination speed and synchrony, which could have important effects for selection strategies. Interestingly, three QTL identified in this study co-localized with regions previously reported by [9], suggesting that some genomic regions are both implicated in seed vigor under water limiting and non-limiting conditions. In addition, one of these shared regions contains a candidate gene previously implicated in seed germination processes, further supporting its biological relevance. One additional QTL overlapped with a region identified by [8] for germination percentage at 72h after imbibition, reinforcing the robustness of this locus across independent studies. An important outcome of this study is the observation that favorable alleles associated with improved seed vigor were frequently carried by minor allele within the diversity panel. This suggests that beneficial variants for seed vigor, particularly under drought stress conditions may not have been widely selected in breeding programs inversely to germination senses strict.

Strong environmental dependence and antagonistic effects of seed vigor QTL

A major finding of this study is the limited stability of QTL effects across contrasting water conditions. By comparing QTL identified by [9], several QTL identified under drought showed no significant main effects under non-limiting condition but exhibited significant QTL×environment interactions. This indicates that their expression is highly context-dependent and may only be revealed under specific environmental constraints. More strikingly, some QTL exhibited antagonistic allelic effects depending on water availability, with allele conferring improved performance under stress having neutral or even deleterious effects under optimal conditions. Such crossover interactions are characteristic of strong QTL by environment (QTL×E) effects and illustrate the complex and plastic nature of the genetic regulation underlying seed vigor. The strong environment dependence observed in this study is consistent

with the well-established sensitivity of seed vigor to external conditions [3,1]. Similarly, environment dependent responses have been reported in other crops, including peanuts, where seed vigor varies according to temperature conditions [36] and soybean, where seed production environments significantly influence seed vigor [37]. Together, these studies support the view that seed vigor should be considered a highly plastic character strongly influenced by environment context. Our results therefore emphasize that QTL detected in a single environment may have limited predictive value when transferred to contrasting conditions. This finding has important implications for breeding strategies targeting seed vigor performance under drought stress. In particular, it suggests that selection for seed vigor under drought condition should explicitly account for environmental specificity rather than relying solely on performance under optimal conditions.

Candidate genes underlying environment responsive QTL

To identify candidate genes underlying early germination (GP36), we combined intergenic polymorphisms detected by GWAS with coding-sequence polymorphisms identified through exome capture sequencing. Among the identified candidates, AT5G14750 encoding the transcription factor MYB66/WER, is a special interest because transcriptional regulation likely plays central role in coordinating germination progression and stress responses. Previous studies showed that homozygous *myb66* mutants exhibit altered phyA mediated germination responses [38], supporting a potential role of this gene in the regulation of early germination. Additional candidate genes, including AT5G41520 (RPS10B), AT5G14720 (MAP4K4) and AT5G14850 (APTG1) have been associated with cell division, the primary root development and hypocotyl growth [39-41]. These functions are consistent with physiological processes required during the early stages of seed germination and seedling establishment.

It should be noted that the number of candidate genes identified within a QTL region strongly depends on how QTL confidence intervals are defined. Different studies have used distinct

strategies to delineate QTL intervals. For example, [9] defined intervals based on chromosome wide LD decay estimates resulting relatively large QTL regions, whereas [8] relied on local LD patterns surrounding significant markers resulting smaller regions. In the present study, QTL confidence intervals were defined using LD block analysis based one widely used r^2 threshold (0.2), allowing a more locally informed delimitation of associated genomic regions.

Conclusion

This study highlights that seed vigor in *Brassica napus* is governed by dynamic and environmentally dependent genetic architecture. Although multiple QTL contribute to seed vigor related traits, many exhibited limited stability or even antagonistic effects across contrasting water conditions, emphasizing the critical role of genotype by environment interactions. Our results further show that integrating fine-scale phenotyping with cross-environment validation is essential to identify biologically relevant and environmentally robust loci. In particular, the combination of dynamic traits, comparative genomics and exon level polymorphism analyses improved both QTL interpretation and candidate gene prioritization. Overall, this work provides new insights into genetic basis seed vigor plasticity under water stress and offers a framework for identifying of stable and environment specific loci that could support the breeding of rapeseed varieties better adapted to water limited conditions.

References

1. Nelson MN, Nesi N, Barrero JM, Fletcher AL, Greaves IK, Hughes T, et al. Chapter Two - Strategies to improve field establishment of canola: A review. In: Sparks DL, editor. Advances in Agronomy. Academic Press; 2022. p. 133–77. https://doi.org/10.1016/bs.agron.2022.05.001.

2. Perry DA. The Influence of Seed Vigour on Vegetable Seedling Establishment. Scientific Horticulture. 1982;33:67–75.

3. Finch-Savage WE, Bassel GW. Seed vigour and crop establishment: extending performance beyond adaptation. Journal of Experimental Botany. 2016;67:567–91. https://doi.org/10.1093/jxb/erv490

4. Gu J, Hou D, Li Y, Chao H, Zhang K, Wang H, et al. Integration of proteomic and genomic approaches to dissect seed germination vigor in Brassica napus seeds differing in oil content. BMC Plant Biol. 2019;19:21. https://doi.org/10.1186/s12870-018-1624-7

5. Sharma S. Seed Vigour Testing: Principles and Methods. 2018;17.

6. Luo T, Zhang Y, Zhang C, Nelson MN, Yuan J, Guo L, et al. Genome-Wide Association Mapping Unravels the Genetic Control of Seed Vigor under Low-Temperature Conditions in Rapeseed (Brassica napus L.). Plants (Basel). 2021;10:426. https://doi.org/10.3390/plants10030426

7. Ding Y, Hou D, Yin Y, Chen K, He J, Yan S, et al. Genetic dissection of Brassica napus seed vigor after aging. Theor Appl Genet. 2024;137:141. https://doi.org/10.1007/s00122-024-04648-3

8. Hatzig SV, Frisch M, Breuer F, Nesi N, Ducournau S, Wagner M-H, et al. Genome-wide association mapping unravels the genetic control of seed germination and vigor in Brassica napus. Front Plant Sci. 2015;6:221. https://doi.org/10.3389/fpls.2015.00221

9. Laurençon M, Legrix J, Wagner M-H, Demilly D, Baron C, Rolland S, et al. Genomic and phenomic predictions help capture low-effect alleles promoting seed germination in oilseed

517 rape in addition to QTL analyses. Theor Appl Genet. 2024;137:156.
518 https://doi.org/10.1007/s00122-024-04659-0
519 10. Clarke WE, Higgins EE, Plieske J, Wieseke R, Sidebottom C, Khedikar Y. A high-density
520 SNP genotyping array for Brassica napus and its ancestral diploid species based on optimised
521 selection of single-locus markers in the allotetraploid genome. Theor Appl Genet.
522 2016;129:1887–99. https://doi.org/10.1007/s00122-016-2746-7
523 11. Browning BL, Zhou Y, Browning SR. A One-Penny Imputed Genome from Next-
524 Generation Reference Panels. The American Journal of Human Genetics. 2018;103:338–48.
525 https://doi.org/10.1016/j.ajhg.2018.07.015
526 12. Rousseau-Gueutin M, Belser C, Da Silva C, Richard G, Istace B, Cruaud C, et al. Long-
527 read assembly of the Brassica napus reference genome Darmor-bzh. Gigascience.
528 2020;9:giaa137. https://doi.org/10.1093/gigascience/giaa137
529 13. Leveugle M, Duborjal H, West M, Pothier A, Clementi A, Nesi N, Dyrszka E, Pichon J,
530 Riviere N. Rapsodyn Whole Exome capture and genotyping in the polyploid Brassica napus,
531 in: JOBIM. 2015. Clermont-Ferrand.
532 14. Corlouer E, Sauvage C, Leveugle M. Envirotyping within a multi-environment trial allowed
533 identifying genetic determinants of winter oilseed rape yield stability. Theor Appl Genet. 2024;
534 137, 164. https://doi.org/10.1007/s00122-024-04664-3
535 15. Wagner M-H, Demilly D, Ducournau S, Durr C. Computer vision for monitoring seed
536 germination from dry state to youg seedlings. Seed Test Int. 2011;142:49–51.
537 https://www.seedtest.org/api/rm/6FV66QEXE2NDT3D/sti142oct-2011.pdf#page=51
538 16. Demilly D, Ducournau S, Wagner M-H. Digital imaging of seed germination. In: Gupta
539 SD. Ibaraki Y (eds) Plant image analysis: fundamentals and applications. CRC Press. Boca
540 Raton. 2014;pp 147–162

17. Ranal MA, Santana DG de. How and why to measure the germination process? Braz J Bot. 2006;29:1–11. https://doi.org/10.1590/S0100-84042006000100002

18. Gower JC. Multivariate Analysis and Multidimensional Geometry. Journal of the Royal Statistical Society Series D (The Statistician). 1967;17:13–28. https://doi.org/10.2307/2987199

19. Astle W, Balding DJ. Population Structure and Cryptic Relatedness in Genetic Association Studies. Statistical Science. 2009;24:451–71. https://doi.org/10.1214/09-STS307

20. Lippert C. Linear mixed models for genome-wide association studies. 2013.

21. Rincent R, Moreau L, Monod H, Kuhn E, Melchinger AE, Malvar RA, et al. Recovering Power in Association Mapping Panels with Variable Levels of Linkage Disequilibrium. Genetics. 2014;197:375–87. https://doi.org/10.1534/genetics.113.159731

22. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet. 2006;38:203–8. https://doi.org/10.1038/ng1702

23. Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genetic Epidemiology. 2008;32:361–9. https://doi.org/10.1002/gepi.20310

24. Alemu SK, Huluka AB, Tesfaye K, Haileselassie T, Uauy C. Genome-wide association mapping identifies yellow rust resistance loci in Ethiopian durum wheat germplasm. PLOS ONE. 2021;16:e0243675. https://doi.org/10.1371/journal.pone.0243675

25. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21:263–5. https://doi.org/10.1093/bioinformatics/bth457

26. Song J-M, Guan Z, Hu J, Guo C, Yang Z, Wang S, et al. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. Nat Plants. 2020;6:34–45. https://doi.org/10.1038/s41477-019-0577-7

27. Lee H, Chawla HS, Obermeier C, Dreyer F, Abbadi A, Snowdon R. Chromosome-Scale Assembly of Winter Oilseed Rape Brassica napus. Front Plant Sci. 2020;11. https://doi.org/10.3389/fpls.2020.00496

28. Davis JT, Li R, Kim S, Michelmore R, Kim S, Maloof JN. Whole-genome sequence of synthetically derived Brassica napus inbred cultivar Da-Ae. G3 Genes|Genomes|Genetics. 2023;13:jkad026. https://doi.org/10.1093/g3journal/jkad026

29. Edwards SV. Next-generation QTL mapping: crowdsourcing SNPs, without pedigrees. Molecular Ecology. 2013;22:3885–7. https://doi.org/10.1111/mec.12401

30. Fu D, Mason AS, Xiao M, Yan H. Effects of genome structure variation, homeologous genes and repetitive DNA on polyploid crop research in the age of genomics. Plant Science. 2016;242:37–46. https://doi.org/10.1016/j.plantsci.2015.09.017

31. Schiessl S-V, Katche E, Ihien E, Chawla HS, Mason AS. The role of genomic structural variation in the genetic improvement of polyploid crops. The Crop Journal. 2019;7:127–40. https://doi.org/10.1016/j.cj.2018.07.006

32. Hu D, Lu J, Li W, Yang Y, Xu J, Qin H, et al. The occurrence, inheritance, and segregation of complex genomic structural variation in synthetic *Brassica napus*. The Crop Journal. 2024;12:515–28. https://doi.org/10.1016/j.cj.2024.01.002

33. Wang T, van Dijk ADJ, Zhao R, Bonnema G, Wang X. Contribution of homoeologous exchange to domestication of polyploid Brassica. Genome Biol. 2024;25:231. https://doi.org/10.1186/s13059-024-03370-z

34. Hurgobin B, Edwards D. SNP Discovery Using a Pangenome: Has the Single Reference Approach Become Obsolete? Biology. 2017;6. https://doi.org/10.3390/biology6010021

35. Lu K, Wei L, Li X, Wang Y, Wu J, Liu M, et al. Whole-genome resequencing reveals Brassica napus origin and genetic loci involved in its improvement. Nat Commun. 2019;10:1154. https://doi.org/10.1038/s41467-019-09134-9

591 36. Grey TL, Branch WD, Tubbs RS, Snider JL, Webster TM, Arnold J, et al. The Impact of
592 Genotype × Environment Effects on Runner-Type Peanut Seed Vigor Response to
593 Temperature. Agronomy Journal. 2016;108:1424–33. https://doi.org/10.2134/agronj2015.0420
594 37. Szareski VJ, Carvalho IR, Demari GH, Rosa TC da, Souza VQ de, Villela FA, et al.
595 Multivariate index of soybean seed vigor: a new biometric approach applied to the effects of
596 genotypes and environments. J Seed Sci. 2018;40:396–406. https://doi.org/10.1590/2317-
597 1545v40n4198333
598 38. Ibarra SE, Auge G, Sánchez RA, Botto JF. Transcriptional Programs Related to
599 Phytochrome A Function in Arabidopsis Seed Germination. Molecular Plant. 2013;6:1261–73.
600 https://doi.org/10.1093/mp/sst001
601 39. Stirnberg P, Liu J-P, Ward S, Kendall SL, Leyser O. Mutation of the cytosolic ribosomal
602 protein-encoding RPS10Bgene affects shoot meristematic function in Arabidopsis. BMC Plant
603 Biol. 2012;12:160. https://doi.org/10.1186/1471-2229-12-160
604 40. Dai XR, Gao X-Q, Chen GH, Tang LL, Wang H, Zhang XS. Abnormal pollen tube
605 guidance1, an Endoplasmic Reticulum-Localized Mannosyltransferase Homolog of
606 glycosylphosphatidylinositol 10 in Yeast and phosphatidylinositol glycan anchor biosynthesis
607 B in Human, Is Required for Arabidopsis Pollen Tube Micropylar Guidance and Embryo
608 Development. Plant Physiol. 2014;165:1544–56. https://doi.org/10.1104/pp.114.236133
609 41. Pan L, Fonseca de Lima CF, Vu LD, van de Cotte B, De Winne N, Gevaert K, et al.
610 Heterodimerization domains in MAP4 KINASEs determine subcellular localization and
611 activity in Arabidopsis. Plant Physiol. 2024;195:1807–17.
612 https://doi.org/10.1093/plphys/kiae176
613

List of abbreviations

BLUE, best linear unbiased estimators; CV, coefficient of variation; GWAS, genome wide association study; H^2, broad-sense heritability; LD, linkage disequilibrium; LOCO, leave-one-chromosome-out; MAF, minor allele frequency; PEG, polyethylene glycol; PCA, principal component analysis ; PCoA, principal coordinates analysis ; QTL, quantitative trait loci ; SD, standard deviation; SNP, single nucleotide polymorphism; SOSR, spring oilseed rape; WFR, winter fodder rape; WOSR: winter oilseed rape.

Additional information

This manuscript contains additional tables and figures.

Ethics approval and consent to participate

Not applicable.

Consent for publication

All the authors read and approved the manuscript.

Availability of data and materials

The datasets generated and analyzed during the current study are available in the Recherche Data Gouv repository, through the following links: https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi:10.57745/PPKWJV and https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi:10.57745/JAAO8R.

Competing interests

The authors declare they have no competing interests.

Funding This work and the PhD grant of K.D. were supported by the program “Seeds for the future” funded by Institut Agro Foundation.

Author contribution A.L. and N.N. designed the study and obtained funding. K.D., M-H.W., D.D., C.B. and S.D. carried out the germination assays on the PHENOTIC platform at GEVES (Angers, France). K.D performed all the statistical data, developed the scripts, and ran the genetic analyses. V.C. and C.B. performed genomic comparative analyses and genes candidate research. K.D, A.L., and N.N. interpreted the results. K.D, A.L, and N.N wrote the manuscript. All the authors read and approved the manuscript.

Corresponding authors

Correspondence to: anne.laperche@institut-agro.fr; nathalie.nesi@inrae.fr.

Acknowledgements

The authors are grateful to Elise Alix, Bernard Moulin (IGEPP, France) and the "Domaine de la Motte" Experimental Unit (INRAE, France) for seed multiplication. We would like to thank Jean-Louis Queyreix and Angélique Juret (GEVES) for supporting the germination phenotyping effort, Mathieu Rousseau-Gueutin (IGEPP) for helpful discussions regarding genomic comparative analyses. We also thanks Béatrice Teulat (Institut Agro Renne-Angers, France) Laurent Leport (Université de Rennes, France) and Khadidiatou Ndoye (ENSA, Senegal) for the constructive discussions about germination and seed vigor. The GenOuest bioinformatics platform (https://www.genou est.org/) is acknowledged for providing resources in informatics.

Tables

Table 1. Germination traits scored in this study

Trait	Abbreviation (unit)	Formula
First germination time	FGT (h)	Notation of the first hour of germination
Time to reach 20% or 50% germination	T20, T50 (h)	Deduced from germination curve
Mean germination time	MGT (h)	$\frac{\text{germination time of each seed i}}{\text{total number of seeds sown}}$
Slope of the germination curve	SG (h)	$\frac{\%\ \text{Germ at T1} - \%\ \text{Germ at T0}}{T1 - T0}$
Uniformity	Uni (h)	$\sqrt{\frac{\Sigma(\text{time i} - \text{MGT})^2 \text{ x germinated seed at}}{\text{nb total}}}$
Germination percentage at 36h	GP36 (%)	$\frac{\text{nb germinated seeds at 36h}}{\text{total nb of seeds sown}} * 100$
Germination percentage at 96h	GP96 (%)	$\frac{\text{nb germinated seeds at 96h}}{\text{total nb of seeds sown}} * 100$
Final germination percentage	GP (%)	$\frac{\text{final germinated seed nb}}{\text{total nb of seeds sown}} * 100$
Area under curve	AUC (%.h^{-1})	$\int_{h2}^{h1} fh\ dn$

659 **Table 2. Phenotypic variations for 10 traits associated with seed vigor in oilseed rape**

660 **(*n*=192)**

Trait	Min	Max	Mean	SD	CV	H^2
FGT	19.00	57.50	36.78	6.29	17.11	0.34
T20	39.00	75.17	50.14	6.12	12.22	0.72
T50	43.73	88.77	60.28	8.09	13.42	0.73
MGT	46.23	82.79	60.67	6.4	10.57	0.73
SG	1.05	7.20	3.52	1.15	32.72	0.73
Uni	21.75	42.85	34.58	3.61	10.43	0.32
GP36	1.70	12.03	3.41	1.87	54.81	0.22
GP96	50.56	98.18	90.34	9.79	10.84	0.84
GP	56.90	98.47	93.29	7.65	8.19	0.84
AUC	2874.00	7111.00	5664.72	889.86	15.71	0.83

661 For details of trait abbreviations refer to Table 1 and "Material and methods". Minimum (Min),

662 maximum (Max), standard deviation (SD), coefficient of variation (CV) and broad-sense

663 heritability (H^2).

664

Table 3. Spearman correlation coefficient (r) for trait pairs associated with seed vigor under water stress

	FGT	T20	T50	MGT	SG	Uni	GP36	GP96	GP
T20	0.48***								
T50	0.28***	0.89***							
MGT	0.42***	0.92***	0.96***						
SG	0	-0.44***	-0.7***	-0.6***					
Uni	-0.45***	-0.91***	-0.95***	-1***	0.59***				
GP36	-0.86***	-0.37***	-0.15**	-0.28***	-0.15	0.3***			
GP96	-0.02	-0.49***	-0.62***	-0.56***	0.54***	0.56***	-0.03		
GP	0.04	-0.42***	-0.54***	-0.45***	0.5***	0.44***	-0.08	0.94***	
AUC	-0.29***	-0.87***	-0.96***	-0.93***	0.68***	0.92***	0.19**	0.76***	0.69***

Significant values are shown in boldface with significances calculated at levels of: *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. For details of trait abbreviations refer to Table 1 and "Material and methods". Green corresponds to positive correlation and red to negative correlation.

671 **Table 4. Summary of 22 unique putative QTL identified by GWAS for seed vigor traits**

QTL name	Chr	Trait	SNP peak (pb)	LD block interval (bp)	QTL size (kb)	Major/Minor allele	MAF	R^2
QTL.A02.1	A02	GP36	Bn_A02_p3635114 (**2.672.718**)	2.533.090-2.672.718	140	T/**C**	0.24	2.16
QTL.A03.1	A03	FGT	Bn_A03_p4193022 (**4.329.413**)	4.329.413-4.353.548	24	C/**T**	0.41	7.8
QTL.A03.2	A03	FGT	Bn_A03_p8038504 (**8.372.347**)	8.365.935-8.412.954	47	T/C	0.31	10.17
QTL.A04.1	A04	GP36	Bn_A04_p8310242 (**11.488.873**)	11.465.940-11.610.401	144	A/**G**	0.31	4.2
QTL.A05.1	A05	GP36	Bn_A05_p20852240 (**36.340.387**)	36.301910-36.340.387	38	A/**G**	0.23	2.3
QTL.A06.1	A06	T50	Bn_A06_p2756792 (**3.222.322**)	3.218.164-3.224.107	6	G/**A**	0.21	7.14
QTL.A06.2	A06	GP36	Bn_A06_p6555456 (**7.189.793**)	7.141.638-7.277.905	136	T/**C**	0.19	6.9
QTL.A06.3	A06	GP	Bn_A06_p26135283 (**45.080.482**)	45.079.971-45.122.416	42	A/**G**	0.14	6.3
QTL.A07.1	A07	GP36	Bn_A07_p3746131 (**8.659.517**)	8.517.414-8.724320	207	T/**C**	0.17	1.2
QTL.A07.2	A07	MGT T50	Bn_A07_p6410362 (**11.404.373**)	11.261.580-11.430.012	168	G/**A**	0.17	6.4 6.3
QTL.A07.3	A07	AUC T50	Bn_A07_p10847704 (**15.917.912**) Bn_A07_p10873126 (**15.958.334**)	15.917.912-15.962.246	44	T/C T/G	0.21 0.18	12.62 7.4
QTL.A07.4	A07	AUC GP GP96 T20	Bn_A02_p676719 (**17.195.682**) Bn_A02_p839171 (**17.362.341**)	17.173.959-17.371.734	198	C/T C/T	0.17 0.08	13.1 6.55 6.1 8.4
QTL.A08.1	A08	MGT T50 Uni T20	Bn_scaff_24108_1_p411736 (**5.244.405**) Bn_A08_p5697643 (**5.339.671**)	4.746.745-5.772.883	1026	**C**/**A** A/C	0.47 0.46	7.03 8.99 6.36 8.02
QTL.A09.1	A09	T50	Bn_A09_p1570558 (**3.739.243**)	3.739.243-3.923.574	184	T/**C**	0.48	7.3

QTL.A10.1	A10	GP36	Bn_A10_p9770913 (**13.954.270**)	13.941.982-13.981.679	40	G/**T**	0.38	1.98
QTL.A10.2	A10	GP	Bn_A10_p10866946 (**15.068.401**)	15.068.401-15.225.977	157	A/G	0.25	4.79
QTL.A10.3	A10	GP	Bn_A10_p13049564 (**16.027.945**)	16.027.945-16.084.725	57	C/T	0.14	5.15
QTL.A10.4	A10	GP	Bn_A10_p16737974 (**19.865.384**)	19.810.981-19.881.377	70	C/T	0.25	3.01
QTL.C02.1	C02	GP	Bn_scaff_16139_1_p259814 (**62.060.284**)	61.808.134-62.129.578	321	T/C	0.27	2.7
QTL.C03.1	C03	AUC GP96	Bn_scaff_18936_1_p327129 (**3.505.604**)	3.434.436-3.511.837	77	C/T	0.47	12.68 3.4
QTL.C03.2	C03	AUC GP96 GP	Bn_scaff_17298_1_p1055071 (**27.393.791**)	27.242.023-27.522.588	280	T/G	0.12	11.94 11.67 10.77
QTL.C06.1	C06	GP36	Bn_scaff_20294_1_p235008 (**46.428.288**)	46.412.614-46.428.288	15	T/**G**	0.15	2.6

672 QTL were named according to their position on the genome. For each QTL, information is
673 given about the peak SNP and flanking markers position (LD block interval), QTL size, as well
674 as major and minor allele, followed by the minor allele frequency (MAF). R^2 corresponds to
675 phenotypic variance explained by the SNP. In the column Major/minor allele, the characters in
676 bold indicate the favorable allele. Favorable allele corresponded to the allele associated with
677 improved phenotypic performance e.g., increasing GP36, GP96, GP, AUC and SG values while
678 decreasing FGT, MGT, T20, T50 and Uni values. Positions are relative to the reference genome
679 Darmor-*bzh* V10 [12].

680

Table 5. Analysis of variance of QTL × Environment

Source of variation	**FGT** QTL.A03.1	**GP36** QTL.A02.1	**GP36** QTL.A04.1	**GP** QTL.A06.3
Main QTL effect	0.3	0.9*	0.9*	0.6.
E	32.9***	47.1***	47.1***	0.1
QTL×E	3.1***	1.9***	2.17***	1.3*
Residuals	63.73	50.17	49.86	97.9123

E: environment, Values correspond to repartition of the sum of squares depending on the considered effect. For this analysis, BLUE obtained for each genotype in each environment condition (water limiting and non-limiting) were gathered and fitted in a mixed linear model (4). Only QTL showing significant QTL×E interactions are presented in this table. Significances were calculated at levels of: *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$

688 **Table 6. Candidate genes within QTL for germination percentage at 36 h**

QTL	Gene position on Darmor-*bzh* genome	Gene annotation	*A. thaliana* ortholog	Expression during germination
QTL.A02.1	2.599.377 -2.606.077	A02p02690.1_BnaDAR	AT5G14720	**Yes**
	2.617.897 - 2.619.292	A02p02720.1_BnaDAR	AT5G14750	**Yes**
	2.630.574 - 2.632.303	A02p02740.1_BnaDAR	AT5G14790	**Yes**
	2.632.940 - 2.636.182	A02p02750.1_BnaDAR	AT5G14850	**Yes**
	2.650.083 - 2.651.222	A02p02780.1_BnaDAR	AT5G15660	No
	2.651.596 - 2.663.976	A02p02790.1_BnaDAR	AT5G14950	**Yes**
	2.665.174 - 2.667.553	A02p02800.1_BnaDAR	AT5G14960	No
	2.667.843 - 2.669.700	A02p02810.1_BnaDAR	AT5G14970	No
QTL.A04.1	11466421 - 11468452	A04p14490.1_BnaDAR	AT5G41350	No
	11472329 - 11477147	A04p14500.1_BnaDAR	AT5G41370	No
	11486667 - 11488477	A04p14510.1_BnaDAR	AT5G41380	No
	11490392 - 11492253	A04p14520.1_BnaDAR	AT5G41390	No
	11496047 - 11497938	A04p14530.1_BnaDAR	NO HIT	No
	11519239 - 11519943	A04p14550.1_BnaDAR	AT3G46160	No
	11520775 - 11524030	A04p14560.1_BnaDAR	AT5G41460	No
	11530387 - 11534339	A04p14570.1_BnaDAR	AT5G41550	No
	11543639 - 11546355	A04p14580.1_BnaDAR	NO HIT	No
	11555364 - 11563081	A04p14590.1_BnaDAR	AT5G41540	**Yes**
	11563475 - 11564126	A04p14600.1_BnaDAR	AT1G16930	**Yes**
	11564127 - 11564905	A04p14610.1_BnaDAR	AT1G16930	**Yes**
	11565748 - 11571934	A04p14620.1_BnaDAR	AT5G40910	No
	11572062 - 11573724	A04p14630.1_BnaDAR	AT5G41520	**Yes**
	11603726 - 11604946	A04p14670.1_BnaDAR	AT5G41560	No
	11606738 - 11610230	A04p14680.1_BnaDAR	AT5G41570	No

689 **Figure legends**

690 **Figure 1. Genetic structure and phenotypic distribution of the diversity set**

691 **A. Genetic structure of the diversity**. Principal coordinates principal analysis (PcoA) revealed
692 three distinct clusters on the basis of genomic markers. Cluster 1 (*n*=44, in green) primarily
693 consists of spring oilseed rape; cluster 2 (*n*=23, in grey) includes most of the semi-winter type
694 accessions and cluster 3 (*n*=125, in purple) mainly comprises the winter oilseed rape. **B.**
695 **phenotypic distribution of the diversity.** Principal component analysis showed three clusters
696 using phenotypic data. Group 1 (black cercle) corresponds to genotypes with rapid germination
697 and high germination efficiencies. Group 2 (orange cercle) corresponds to genotypes with slow
698 germination and low germination efficiencies. Group 3 (blue cercle) corresponds to genotypes
699 with intermediate performances.

700 **Figure 2. Matrix of linkage disequilibrium (r^2) between the 23 genomic regions detected**
701 **by GWAS.**

702 QTL were named according to their position on the genome. The analysis showed LD (r^2)
703 patterns among SNP markers associated with QTL, in order to investigate potential genetic
704 relatedness and redundancy between regions.

705 **Figure 3. QTL × Environment effects**

706 Contrasting allelic effects of QTL on FGT, GP36 and GP under water-limited (solid orange
707 spine, W-) and non-limiting conditions (dashed light blue line, W+). On the x axis 0 represents
708 the unfavorable allele (associated with an increase in FGT, or a decrease in GP36 or GP), while
709 1 represents the favorable allele (associated with a decrease in FGT, or an increase in GP36 or
710 GP).

711

Additional Tables

Additional Table 1. List of the 192 genotypes used in this study

Additional Table 2. Selected QTL regions based on peak SNPs with -log10 (p-value) exceeding the significance threshold (3.57) and supported by at least one neighboring SNP displaying a similar or near significant association signal

Additional Table 3. QTL confidence interval based on local LD determined by including all neighboring SNP in LD (r^2) with the peak SNP ($r^2 > 0.2$)

Additional Table 4. QTL confidence interval using local linkage disequilibrium (LD) and haplotype block structure

Additional Table 5. MAF distribution in the exome capture subset and the complete diversity panel

Additional Figures

Additional Figure 1. Multi-alignment of the end of chromosome A06 from different reference genomes of *Brassica napus*.

Additional Figure 2. Phenotypic comparison between optimal and water limited conditions.

Figures

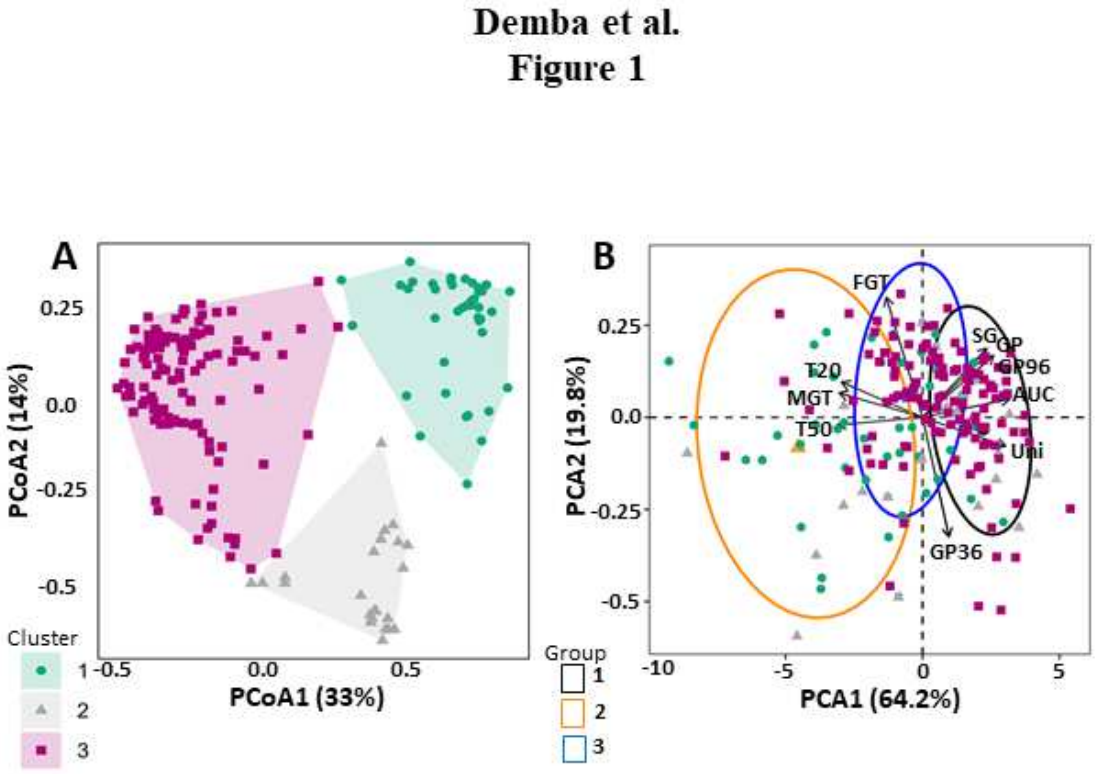


Figure 1

Genetic structure and phenotypic distribution of the diversity set

A. Genetic structure of the diversity. Principal coordinates principal analysis (PcoA) revealed three distinct clusters on the basis of genomic markers. Cluster 1 (*n*=44, in green) primarily consists of spring oilseed rape; cluster 2 (*n*=23, in grey) includes most of the semi-winter type accessions and cluster 3 (*n*=125, in purple) mainly comprises the winter oilseed rape. **B. phenotypic distribution of the diversity.** Principal component analysis showed three clusters using phenotypic data. Group 1 (black cercle) corresponds to genotypes with rapid germination and high germination efficiencies. Group 2 (orange cercle) corresponds to genotypes with slow germination and low germination efficiencies. Group 3 (blue cercle) corresponds to genotypes with intermediate performances.

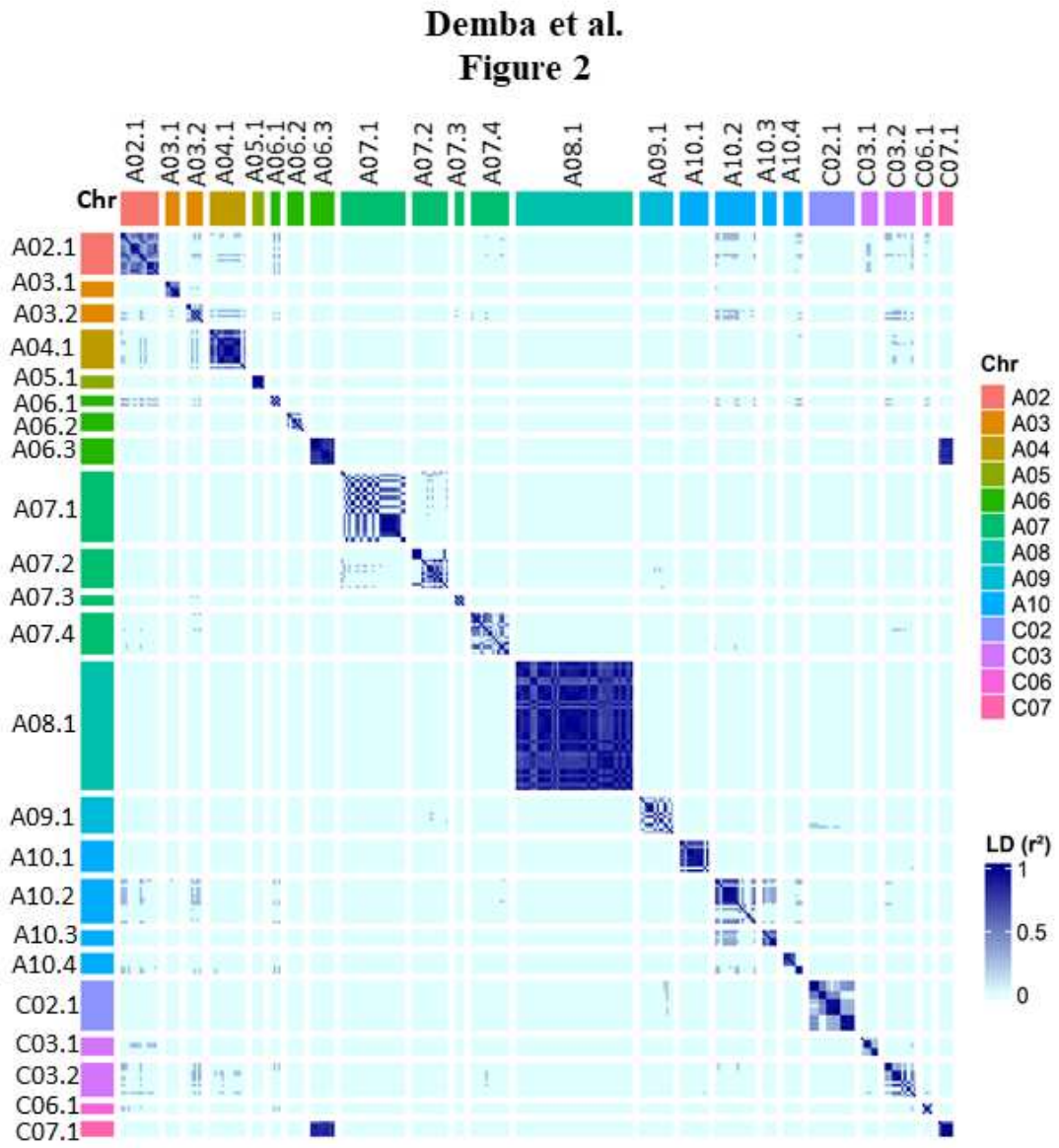


Figure 2

Matrix of linkage disequilibrium (r^2) between the 23 genomic regions detected by GWAS.

QTL were named according to their position on the genome. The analysis showed LD (r^2) patterns among SNP markers associated with QTL, in order to investigate potential genetic relatedness and redundancy between regions.

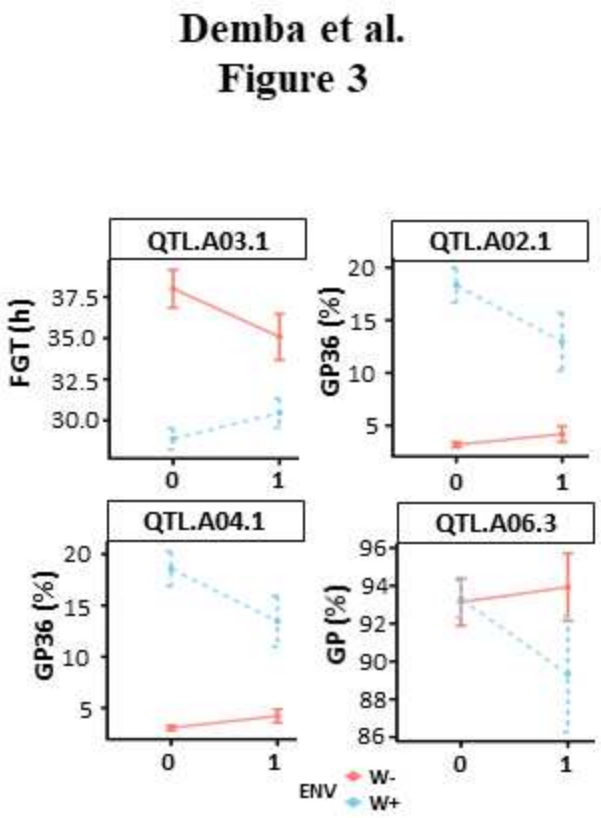


Figure 3

QTL × Environment effects

Contrasting allelic effects of QTL on FGT, GP36 and GP under water-limited (solid orange spine, W-) and non-limiting conditions (dashed light blue line, W+). On the x axis 0 represents the unfavorable allele (associated with an increase in FGT, or a decrease in GP36 or GP), while 1 represents the favorable allele (associated with a decrease in FGT, or an increase in GP36 or GP).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- DembaetalBMCGenomicsAdditionalFigures.pptx
- DembaetalBMCGenomicsAdditionalTables.xlsx