



OPEN

## Genomic background selection to reduce the mutation load after random mutagenesis

Nirosha L. Karunarathna<sup>1,3</sup>, Dilan S. R. Patiranaige<sup>1</sup>, Hans-Joachim Harloff<sup>1</sup>, Niharika Sashidhar<sup>1,4</sup> & Christian Jung<sup>1,2</sup>✉

Random mutagenesis is a standard procedure to increase allelic variation in a crop species, especially in countries where the use of genetically modified crops is limited due to legal constraints. The chemical mutagen EMS is used in many species to induce random mutations throughout the genome with high mutation density. The major drawback for functional analysis is a high background mutation load in a single plant that must be eliminated by subsequent backcrossing, a time and resource-intensive activity. Here, we demonstrate that genomic background selection combined with marker-assisted selection is an efficient way to select individuals with reduced background mutations within a short period. We identified BC<sub>1</sub> plants with a significantly higher share of the recurrent parent genome, thus saving one backcross generation. Furthermore, spring rapeseed as the recurrent parent in a backcrossing program could accelerate breeding by reducing the generation cycle. Our study depicts the potential for reducing the background mutation load while accelerating the generation cycle in EMS-induced winter oilseed rape populations by integrating genomic background selection.

Oilseed rape (*Brassica napus*) is an important oil crop grown worldwide with a broad adaptation to different climates and daylength regimes. Winter oilseed rape is mainly grown in Europe. It only flowers after a period of cold exposure. In contrast, spring oilseed rape, grown primarily in Canada and Australia, does not need vernalization for flowering. Semi-winter oilseed rape is widely cultivated in China and requires moderate vernalization. The demand for oilseed rape has increased over the years, as oilseed rape is not only used for the production of edible oil but also as a protein source for feed, biofuel, and industrial raw material<sup>1</sup>. The generation cycle of winter rapeseed is much longer because it needs vernalization for flowering. Under field conditions, winter rapeseed requires one year to complete its life cycle, while in the greenhouse, six to seven months are needed for seed production. In contrast, early spring accessions need 3–4 months to produce seeds in the greenhouse yielding three to four generations per year<sup>2</sup>. Rapid generation cycling *B. napus* has been proposed, producing up to six generations per year (55 days generation cycle) under high light conditions in the greenhouse at 24 °C<sup>3</sup>.

Oilseed rape is a young crop with a short history of cultivation and breeding<sup>4</sup>. Genetic variation within this species is considerably small<sup>5</sup>. Mutagenesis is a powerful tool for creating new genetic variation. CRISPR-Cas mediated mutagenesis has been successfully applied to develop new allelic variation in oilseed rape<sup>6–8</sup>. However, its practical application is hampered by legal constraints in the European Union and other countries<sup>9</sup>. Therefore, chemical or irradiation-induced random mutagenesis is an important alternative. Numerous mutants have been produced with improved agronomical characters such as shatter resistance, flowering time, seed weight, and yield components<sup>10–13</sup>. However, the usage of mutants in breeding programs is limited due to their high mutation load. In our EMS rapeseed population, each plant has ~46,000 mutations<sup>6</sup>. While they can be grown in the greenhouse to study the effect of a mutation in a target gene, their application for field studies is limited. Plant development and flowering time are delayed, and they are suffering from low vigor and poor abiotic stress tolerance. In our field studies, they did not even survive the winter. The mutation load is even more increased after crossing different mutants to produce double mutants. This is the only way to get a new phenotype in polyploids because single mutants do not show a phenotypic effect<sup>6</sup>. Therefore, the EMS donor genome must be replaced by an elite genome, which requires at least 4–6 generations of backcrossing<sup>14</sup>.

<sup>1</sup>Plant Breeding Institute, Christian-Albrechts-University of Kiel, Olshausenstr. 40, 24098 Kiel, Germany. <sup>2</sup>Plant Breeding Institute, Christian-Albrechts-University of Kiel, Olshausenstrasse 40, 24118 Kiel, Germany. <sup>3</sup>Present address: KWS LOCHOW GMBH, Zuchtstation Wetze, Wetze 3, 37154 Northeim, Germany. <sup>4</sup>Present address: Max Planck Institute for Plant Breeding Research, Carl-von-Linne-Weg 10, 50829 Cologne, Germany. ✉email: c.jung@plantbreeding.uni-kiel.de

Gene name	Population (seed code)	Generation	Number of plants genotyped	Recurrent parent genome share		Plants with higher Peace genome share
				Per plant (%)	Average (%)	
<i>Bna.SFAR4.A06a</i>	A (180901)	BC <sub>1</sub>	85	68.3–83.7	75.4	50 (58.8%)
<i>Bna.SFAR4.C03a</i>	B (180902)	BC <sub>1</sub>	99	65.7–84.5	74.8	48 (48.5%)
<i>Bna.2-PGK2.A10</i>	C (183295)	BC <sub>1</sub>	33	69.6–85.7	79.0	26 (78.8%)
<i>Bna.MRP5.A10</i>	D (180858)	BC <sub>1</sub>	20	67.8–81.6	75.5	12 (60.0%)
<i>Bna.MRP5.C05</i>	E (180837)	BC <sub>1</sub>	20	71.5–80.6	75.8	11 (55.0%)
<i>Bn2-PGK2</i>	F (183297)	F <sub>1</sub> × F <sub>1</sub>	94	42.5–65.3	54.1	75 (79.8%)
<i>BnMRP5</i>	G (180859)	F <sub>1</sub> × F <sub>1</sub>	60	45.2–65.4	54.9	50 (83.3%)

**Table 1.** Recurrent parental genome contribution in seven BC<sub>1</sub> and F<sub>1</sub> oilseed rape generations. Genome share was calculated as % of the Peace genome. In the last column, plants with a higher Peace genome share were defined as plants with more than the expected average.

To accelerate breeding, breeders use different methods like single seed descent (SSD) in pedigree breeding and doubled haploid (DH) technology<sup>15</sup>. A recent study claimed to accelerate crop research and breeding by "speed breeding"<sup>2</sup>. Among other species, they grew canola (*B. napus*) under prolonged photoperiods with a day length of 22 h in a temperature-controlled glasshouse fitted with high-pressure sodium lamps. Under these controlled-environmental conditions, they reduced generation cycles for spring canola to 73 ± 9 days<sup>2</sup>. Thus, four generations per year could be achieved.

Recurrent backcrossing is commonly used to transfer genes responsible for favorable agronomic traits from a donor line to the recurrent parent<sup>14</sup>. Typically, the recurrent parent is an elite line with low relatedness to the donor parent. Marker-assisted selection enables efficient detection of the target gene. In contrast, genomic background selection retains the original characters of the recurrent parent, thus reducing the number of generation cycles needed to obtain an improved elite line<sup>16,17</sup>. High throughput marker genotyping technologies are used in oilseed rape breeding and research<sup>18,19</sup>. *B. napus* Illumina Infinium arrays ranging from 6 to 60 K<sup>20–23</sup> have been applied in germplasm genotyping<sup>24</sup>, genome-wide association studies<sup>25–28</sup>, and QTL mapping<sup>29,30</sup>.

Here, we propose genomic background selection in early backcross generations. Offspring from EMS treated rapeseed plants were backcrossed with an elite line. Different backcross populations were genotyped with a 15 K Illumina Infinium array. Plants with a high share of the backcross parent genome were identified in the first backcross generation. This study demonstrates that plants with substantially lower mutation load can be selected from the first backcross generation. The application for breeding with EMS mutagenized plants is discussed.

## Results

**Marker-assisted foreground selection of mutant plants in segregating populations.** We produced segregating populations with a varying share of the mutagenized donor genomes either by backcrossing or crossing two F<sub>1</sub> hybrids carrying different mutant alleles. We selected six different EMS-induced nonsense mutants for backcrossing with Peace, two *BnSFAR4-a* mutants (*Bna.SFAR4.A06a* and *Bna.SFAR4.C03a*)<sup>6</sup>, two *Bn2-PGK2* mutants (*Bna.2-PGK2.A02* and *Bna.2-PGK2.A10*)<sup>31</sup>, and two *BnMRP5* mutants (*Bna.MRP5.A10* and *Bna.MRP5.C05*)<sup>32</sup> (Supplementary Table 1). In the case of *BnSFAR4-a*, to produce BC<sub>1</sub> populations, we backcrossed F<sub>1</sub> hybrids (M<sub>3</sub> × Express 617 mutant) with Peace, while for *BnMRP5* and *Bn2-PGK2*, we backcrossed M<sub>3</sub> mutants with Peace (Supplementary Fig. 1). We also generated F<sub>1</sub> × F<sub>1</sub> double mutants for *BnMRP5* and *Bn2-PGK2* by combining F<sub>1</sub> hybrids after crossing the M<sub>3</sub> mutant with Peace. In total, we selected seven families for marker-assisted backcross selection. For easier understanding, we used a short letter code A through G (Table 1).

We questioned whether F<sub>1</sub> plants from a single cross with Peace would flower without vernalization. Therefore, *BnSFAR4-a* single mutant F<sub>1</sub> and BC<sub>1</sub> plants were grown in the greenhouse (16 h light/ 8 h dark, 23–24 °C) without vernalization. Four months after sowing, none of the F<sub>1</sub> plants had entered the generative phase (data not shown).

We selected BC<sub>1</sub> and F<sub>1</sub> × F<sub>1</sub> plants carrying the mutant alleles using allele-specific markers and Sanger sequencing (foreground selection) (Supplementary Table 2). Expected 1:1 and 1:1:1:1 segregation ratios for the mutant alleles were confirmed for BC<sub>1</sub> (populations A, C, D, and E) and F<sub>1</sub> × F<sub>1</sub> (populations F and G), respectively, except for population B. However, the number of heterozygous mutants in this population was sufficient for genomic background selection. In summary, we identified 256 heterozygous plants in the BC<sub>1</sub> populations and 154 heterozygous double mutants in the F<sub>1</sub> × F<sub>1</sub> populations.

**Population structure analysis.** We genotyped 256 BC<sub>1</sub> and 154 F<sub>1</sub> × F<sub>1</sub> plants carrying the mutant alleles along with M<sub>3</sub>/F<sub>1</sub> mutant parents and the non-mutagenized Express 617 and recurrent Peace parents by using the SNP array resulting in 13,416 functional SNPs (Table 2). From populations C through G, F<sub>1</sub> (M<sub>3</sub> × Peace) hybrids were taken as parental controls (Supplementary Fig. 1). Monomorphic markers are not informative as they cannot distinguish between parental genotypes. Therefore, we filtered SNPs for minor allele frequencies of 0.001 to exclude both genotyping artifacts and monomorphic markers, which resulted in 7,686 informative SNPs. Of these, 6,887 were polymorphic between Peace and Express 617 (Table 2).

	Number of SNPs
Markers on the SNP array	13,714
Functional markers for BC <sub>1</sub> and F <sub>1</sub> × F <sub>1</sub> genotyping	13,416
Markers after filtering	7686

**Table 2.** Results from genotyping BC<sub>1</sub> and F<sub>1</sub> × F<sub>1</sub> populations with the 15 K Illumina Infinium SNP array. Functional markers are the markers without missing SNP data. SNP information filtering was done for a minor allele frequency of 0.001.

Then, we analyzed the structure of seven populations. A principal component (PC) analysis was carried out separately for each population. The first two PCs (PC1 and PC2) explained 25.93% to 68.50% of the population's total genetic variation (Fig. 1). We expected to observe four and three clusters in the BC<sub>1</sub> and F<sub>1</sub> × F<sub>1</sub> populations, respectively. Based on the top two PCs, we detected in all BC<sub>1</sub> single mutant populations four main clusters, i.e., Express 617, Peace, parents, and BC<sub>1</sub> (Fig. 1). F<sub>1</sub> × F<sub>1</sub> double mutant populations displayed three main clusters, Express 617, Peace, and F<sub>1</sub> double mutants. In these populations, parents were clustered together with F<sub>1</sub> mutants. As expected, backcrossed plants and the recurrent genotype Peace were less distant from each other, indicating that backcrossed plants carried a higher Peace background percentage. In populations A and B, Express 617 and M<sub>3</sub> parents were clustered along with the second principal component, reflecting low genetic variation between these genotypes (Fig. 1).

**Genome composition in backcross generations.** We aimed to select the plants carrying the highest Peace genomic background. Therefore, we calculated the proportion of accumulated recurrent and donor parents' genomes for each plant. In a BC<sub>1</sub> generation, the expected share of the recurrent parent's genome is 75%. In the respective populations (A, B, C, D, and E), the proportion of the backcross parent genome ranged between 65.7% and 85.7% (Fig. 2, Table 1 and Supplementary data 1). The average share in four populations (A, B, D, and E) ranged from 74.8 to 75.8%, which is very close to the expected value. A slightly higher value was calculated for population C (79.0%). In the double-hybrid populations, the expected share of the Peace genome is 50%. The values calculated for individual plants of populations F and G ranged from 42.5% to 65.4% (Fig. 2 and Table 1). The average share of the recurrent parent genome in populations F and G was 54.1% and 54.9%, which is higher than the expected value (50%).

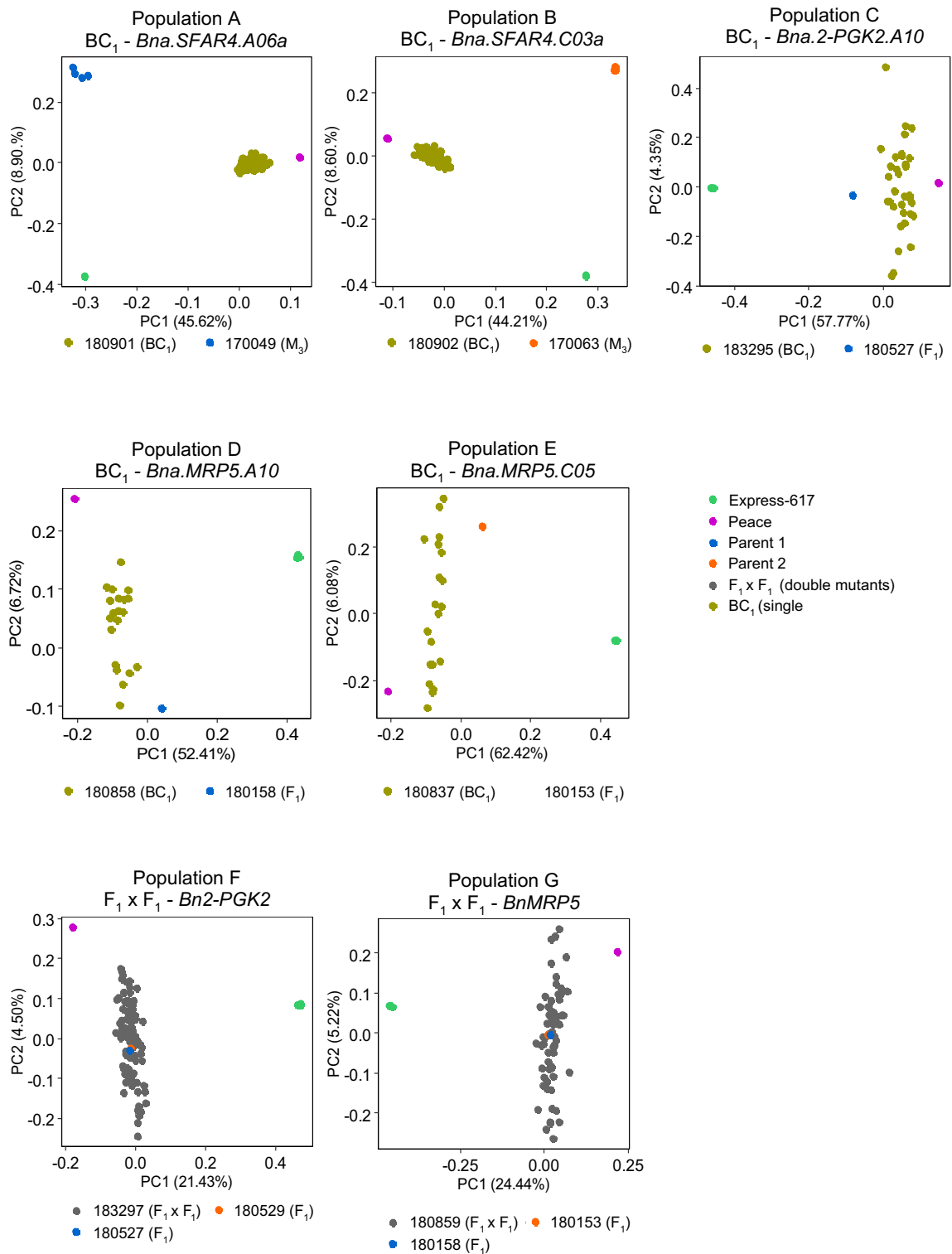
This study's primary aim was to select plants with a higher than average share of the backcross parent (Peace) genome. Across all BC<sub>1</sub> populations, the proportion of plants meeting these expectations ranged between 48.5% and 78.8% (Table 1). In the F<sub>1</sub> × F<sub>1</sub> double mutant populations F and G, the frequency of plants with a Peace genome share above 50% was 79.8% and 83.3%, respectively (Table 1). Based on these results, individual plants with substantially higher backcross genome share could be selected from the first segregating generations.

## Discussion

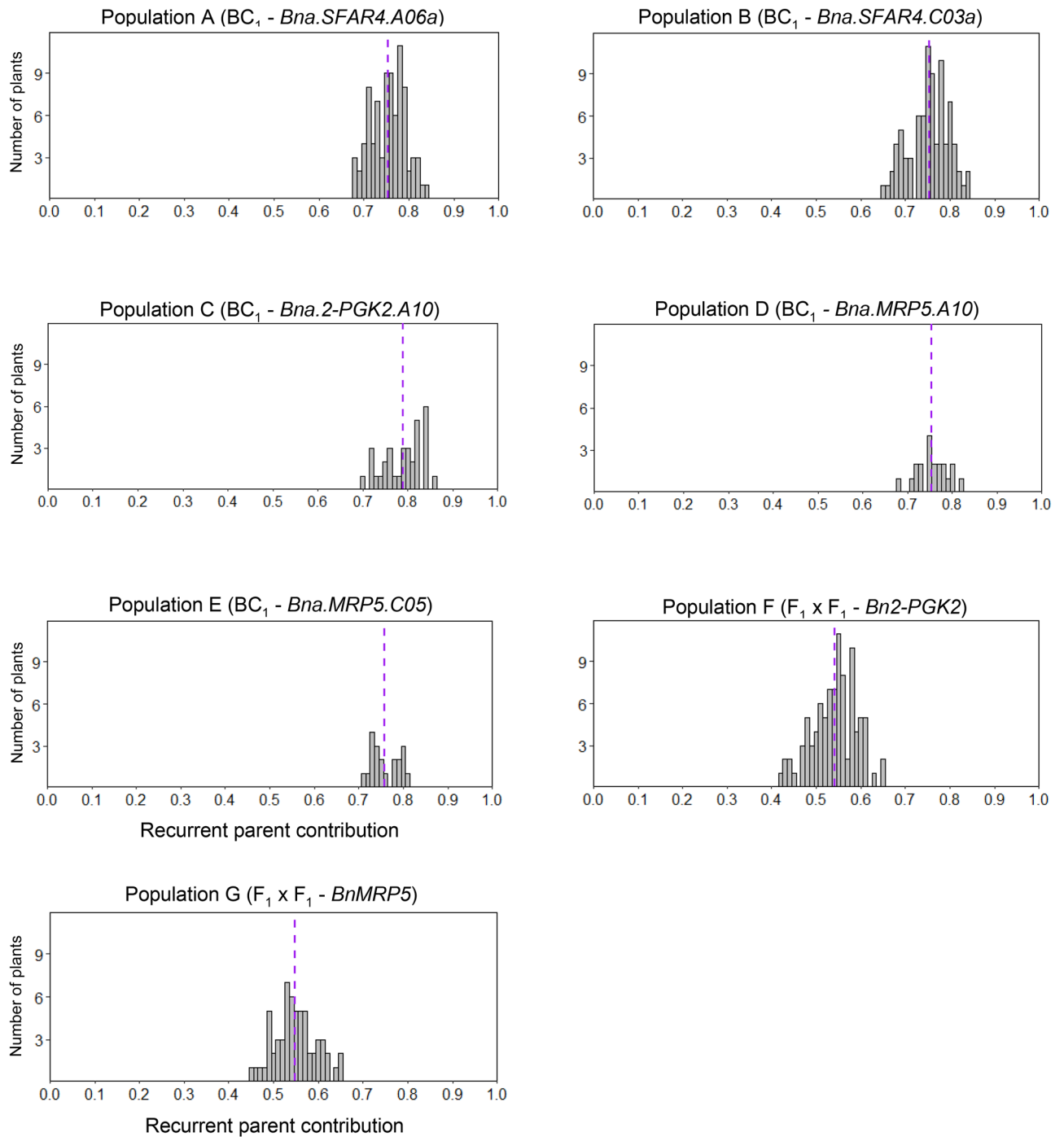
Repeated backcross breeding is a common practice that many breeders use to incorporate new alleles into elite plant material. However, conventional backcrossing is a time-consuming procedure, mainly in biennial species with long generation cycles. This study demonstrates the potential of genomic background selection for introducing EMS-induced alleles into oilseed rape breeding lines with a fast reduction of background mutation load.

A high mutation load is a major problem arising with EMS mutagenized plants<sup>6,33</sup>. Hence, background mutations should be eliminated by subsequent backcrossing with elite germplasm. The winter oilseed rape in this study has a strong vernalization requirement, which takes a minimum of six months to complete one generation under greenhouse conditions. Repeated backcrossing is required to produce a mutant line with a low mutation load. Here, we investigated two parameters to shorten the period for producing an improved elite line with a substantially reduced mutation load. First, we used a spring-type line as a recurrent parent, saving three months per generation compared to a cross with a winter-type. Applying the single seed descent (SSD) method under highly controlled growth conditions has reduced generation time in wheat and barley<sup>2</sup>. Harvesting and germination of the immature seeds, thereby shortening the generation cycle, have been proven for wheat, pigeon pea, and faba bean<sup>2,34,35</sup>. Hence, there is a potential for further reducing the generation cycle, at least by three to four weeks in our oilseed rape populations, by combining SSD with immature seed harvesting and germination under greenhouse conditions. This approach, however, is only applicable for research projects where the effect of mutant alleles is studied among a non-mutated genome background. In the European winter breeding program, backcrossing with spring type is avoided because of their non-adapted flowering time.

Then, we applied marker-assisted selection to save repeated cycles of backcrossing, which included two steps, foreground selection for the mutant allele and genomic background selection for the recurrent parent genome. We chose the 15 K Illumina Infinium SNP array for genotyping for two reasons. First, the information content is high because thousands of loci can be genotyped within a short time. Second, co-dominance enables identifying heterozygous and homozygous loci, which is essential in BC<sub>1</sub> populations. The recovery rate of the recipient genome by marker-assisted backcrossing relies on the number of backcross generations, marker density, and population size<sup>17</sup>. In general, higher recovery of the recipient genome can be obtained in early backcross generations using high marker densities and large populations<sup>36</sup>. The marker density in this study was around 6.09 markers/Mbp, which is far enough for background selection. In an earlier study, it has been shown that even a



**Figure 1.** Principal component (PC) analysis with five BC<sub>1</sub> single mutant and two F<sub>1</sub> × F<sub>1</sub> double mutant oilseed rape populations. PCA was performed for each population separately, and PC1 and PC2 were displayed using the ggplot2 package in R. Seed codes of populations and respective parents are given at the bottom of each plot. Size of each population as follows, population A = 85, population B = 99, population C = 33, population D = 20, population E = 20, population F = 94, and population G = 60. M<sub>3</sub> mutants were used as the parents for populations A and B, while F<sub>1</sub> (M<sub>3</sub> × Peace) hybrids were taken as parental controls for populations C to G.



**Figure 2.** Proportion of the recurrent parent in five  $BC_1$  single mutant and two  $F_1 \times F_1$  double mutant populations. The purple lines indicate the average of the recurrent parent's genome share.

lesser number of markers and plants are sufficient for high response to selection<sup>37</sup>. Low-cost multiplex marker systems such as AFLPs could offer an alternative; however, the workload is much higher than for SNP arrays.

On average, in the  $BC_1$  generation, a 75% recurrent parent genome is expected with one target gene introgression<sup>38</sup>. Our results met the expectations based on common genetic knowledge. Similar data were reported for rice, wheat, and maize<sup>39–41</sup>. As a side effect, our results demonstrate the accurateness of hand crosses because we found no self-pollination events.

It has been shown that marker-assisted background selection accelerates the recovery of the recurrent parent genome<sup>38,42,43</sup>. We could identify a  $BC_1$  plant with a recurrent genome share of 85.7%, the approximate average recipient genome share in a  $BC_2$  generation. Thus, one generation of backcrossing could be saved, leading to higher genetic gain. In polyploids, single mutations rarely have a phenotypic effect due to gene redundancy<sup>6,31,44</sup>. If gene families are targeted, a minimum of two knockout alleles must be combined in one genome, as has recently

been demonstrated for *BnSFAR4* and *Bn2-PGK2* gene families in rapeseed<sup>6,31</sup>. Although CRISPR-Cas mediated mutagenesis turned out to be superior, random mutagenesis like EMS is the only choice for European breeders to work with novel alleles due to legal constraints<sup>9</sup>. Therefore, the  $F_1 \times F_1$  double mutant hybrids of our study are of high practical relevance. We could select individual plants with 65% of the recovered recurrent parent's genome from these populations, far exceeding the theoretical average (50%). The reason could be that these populations are derived from two meiotic generations resulting in more recombination events during the meiosis.

In conclusion, we suggest using double hybrid populations for marker-assisted background selection to reduce the mutation load after random mutagenesis. This study was a proof of principle experiment with a considerably low number of plants. We expect that plants with higher recurrent genome share will be identified in practical breeding programs working with much larger populations.

## Materials and methods

**Plant materials and crossing scheme.** We used previously identified EMS-induced single mutants of winter oilseed rape (Supplementary Table 1) as the primary materials to perform different crosses. Seed fatty acid reducer (*SFAR*) genes encode lipases active during seed development. Gene knockout resulted in elevated seed oil contents<sup>6</sup>. The *multi-drug resistant protein 5* (*MRP5*) encodes a phytic acid transporter protein whereas the *2-phosphoglyceric acid kinase gene* (*2-PGK*) encodes an enzyme that is part of the phytic acid biosynthesis pathway. Knockout of both genes resulted in drastically reduced phytic acid seed contents<sup>31,32</sup>. We crossed *BnSFAR4*  $M_3$  mutants with Express 617, which is the EMS donor genotype. The resulting  $F_1$  plants were backcrossed twice with the Canadian spring variety 'Peace' to produce  $BC_1$  populations (Supplementary Fig. 1). We used 'Peace' for backcrossing because it is flowering very early in contrast to adapted German varieties which need a long time for vernalization. The mutant's agronomic value could only be assessed in backcross generations. *BnMRP5* and *Bn2-PGK2*  $M_3$  mutants were first crossed with Peace to produce  $F_1$  plants. The  $F_1$  plants were crossed with each other within a subfamily to produce double mutants ( $F_1 \times F_1$ ) or backcrossed with Peace to produce  $BC_1$  single mutants (Supplementary Fig. 1). Seeds were sown in  $3 \times 3$  cm 35-multi-well palettes in the greenhouse, and plants were grown in  $9 \times 9$  cm pots for seed production. Plants were grown under greenhouse conditions (16 h light/8 h dark, 23–24 °C). For vernalization, plants were transferred to a cold chamber (16 h light/ 8 h dark, 4 °C) for eight weeks and then returned to the greenhouse. Plastic selfing bags were mounted after bolting to control pollination.

**DNA isolation and selection of mutant plants.** Leaf genomic DNA was extracted from greenhouse-grown plants using the cetyltrimethylammonium bromide (CTAB) method<sup>45</sup>. We used allele-specific markers to select mutant genotypes for genotyping  $BC_1$  and  $F_1 \times F_1$  plants (Supplementary Table 2). PCR was performed under the following cycling conditions: 94 °C for 3 min, 40 cycles of 94 °C for 30 s, 60–66 °C for 30 s, and 72 °C for 1 min, followed by 72 °C for 5 min final elongation, and PCR products were separated on agarose gels. PCR products carrying the mutant alleles were Sanger sequenced (Institute of Clinical Molecular Biology, Kiel) to confirm the mutations.

**Genotyping with the 15 K Illumina Infinium SNP array.** The genomic DNA of mutant plants was normalized to a concentration of 50–200 ng/ $\mu$ l by using a NanoDrop2000 spectrophotometer (ThermoFisher Scientific, Waltham, United States). From the normalized DNA of  $BC_1$  and  $F_1 \times F_1$  plants, 20  $\mu$ l aliquots were sent for genotyping (TraitGenetics GmbH, Gatersleben, Germany) using the *Brassica* 15 K Illumina Infinium SNP array (Illumina, San Diego, CA).

Genotyping data were displayed in an MS Excel file format. First, we converted the Excel files to the Hapmap (Haplotype Map) format applying customized R scripts. Using TASSEL (Trait Analysis by aSSociation, Evolution and Linkage)<sup>46</sup>, the Hapmap file was converted into the VCF (Variant Call Format) format<sup>47</sup>. We filtered SNPs for minor allele frequencies lower than 0.001 using the VCFtools<sup>47</sup>. With this filtering, we attained the VCF file with high-quality SNPs.

We performed Principal Component (PC) analysis using the SNPRelate package<sup>48</sup> in R 3.6.1<sup>49</sup>. First, two PCs were plotted using the ggplot2 package in R (Wickham H (2016)<sup>50</sup>. Using the TASSEL software, we converted all alleles into the ABH file format. The Peace and Express 617 alleles are written as 'A' and 'B', respectively, and 'H' represents heterozygosity. Then, genotypes were exported to a CSV (comma-separated values) file. The resulting genotype file was used to calculate the share of the Peace and Express 617 genome (Supplementary data 1), following the formula.

$$\% \text{ of parent genome} = \frac{(\text{homozygous allele} \times 2) + (\text{heterozygous allele} \times 1)}{\text{total number of alleles}} \times 100$$

**Statement for the plant materials.** The authors confirmed that the collection of plant material or the collection of seeds in the study complies with relevant institutional, national, and international guidelines and legislation.

## Data availability

The authors declare that data supporting the finding of this study are available from this manuscript and its supplementary information files. Extra data, information, and plant materials used/produced in this study are available from the corresponding authors upon request.

Received: 26 March 2021; Accepted: 13 September 2021

Published online: 30 September 2021

## References

- Friedt, W., Tu, J. & Fu, T. In *The Brassica napus Genome* (eds Liu, S. et al.) 1–20 (Springer International Publishing, 2018).
- Watson, A. et al. Speed breeding is a powerful tool to accelerate crop research and breeding. *Nat Plants* **4**, 23–29. <https://doi.org/10.1038/s41477-017-0083-8> (2018).
- Williams, P. H. & Hill, C. B. Rapid-cycling populations of *Brassica*. *Science* **232**, 1385–1389. <https://doi.org/10.1126/science.232.4756.1385> (1986).
- Chalhoub, B. et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345**, 950–953. <https://doi.org/10.1126/science.1253435> (2014).
- Lu, K. et al. Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nat. Commun.* **10**, 1154. <https://doi.org/10.1038/s41467-019-09134-9> (2019).
- Karunaratna, N. L., Wang, H. Y., Harloff, H. J., Jiang, L. X. & Jung, C. Elevating seed oil content in a polyploid crop by induced mutations in *SEED FATTY ACID REDUCER* genes. *Plant Biotechnol. J.* <https://doi.org/10.1111/pbi.13381> (2020).
- Sashidhar, N., Harloff, H. J., Potgieter, L. & Jung, C. T. Gene editing of three *BnITPK* genes in tetraploid oilseed rape leads to significant reduction of phytic acid in seeds. *Plant Biotechnol. J.* **18**, 2241–2250. <https://doi.org/10.1111/pbi.13380> (2020).
- Braatz, J. et al. CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed rape (*Brassica napus*). *Plant Physiol.* **174**, 935–942. <https://doi.org/10.1104/pp.17.00426> (2017).
- Kupferschmidt, K. EU verdict on CRISPR crops dismays scientists. *Science* **361**, 435–436 (2018).
- Shah, S., Karunaratna, N. L., Jung, C. & Emrani, N. An APETALA1 ortholog affects plant architecture and seed yield component in oilseed rape (*Brassica napus* L.). *BMC Plant Biol.* <https://doi.org/10.1186/s12870-018-1606-9> (2018).
- Braatz, J., Harloff, H. J. & Jung, C. EMS-induced point mutations in *ALCATRAZ* homoeologs increase silique shatter resistance of oilseed rape (*Brassica napus*). *Euphytica* <https://doi.org/10.1007/s10681-018-2113-7> (2018).
- Simmonds, J. et al. A splice acceptor site mutation in TaGW2-A1 increases thousand grain weight in tetraploid and hexaploid wheat through wider and longer grains. *Theor. Appl. Genet.* **129**, 1099–1112. <https://doi.org/10.1007/s00122-016-2686-2> (2016).
- Guo, Y., Hans, H., Christian, J. & Molina, C. Mutations in single *FT*- and *TFL1*-paralogs of rapeseed (*Brassica napus* L.) and their impact on flowering time and yield components. *Front. Plant Sci.* <https://doi.org/10.3389/Fpls.2014.00282> (2014).
- Hasan, M. M. et al. Marker-assisted backcrossing: a useful method for rice improvement. *Biotechnol. Biotechnol. Equip.* **29**, 237–254. <https://doi.org/10.1080/13102818.2014.995920> (2015).
- Lenaerts, B., Collard, B. C. Y. & Demont, M. Review: Improving global food security through accelerated plant breeding. *Plant. Sci.* **287**, 1. <https://doi.org/10.1016/j.plantsci.2019.110207> (2019).
- Jamaluddin, M. et al. Marker Assisted Gene Pyramiding (MAGP) for bacterial blight and blast resistance into mega rice variety “Tellahamsa”. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0234088> (2020).
- Herzog, E. & Frisch, M. Selection strategies for marker-assisted backcrossing with high-throughput marker systems. *Theor. Appl. Genet.* **123**, 251–260. <https://doi.org/10.1007/s00122-011-1581-0> (2011).
- Snowdon, R. J. & Luy, F. L. I. Potential to improve oilseed rape and canola breeding in the genomics era. *Plant Breed.* **131**, 351–360. <https://doi.org/10.1111/j.1439-0523.2012.01976.x> (2012).
- Obermeier, C. & Friedt, W. 16 - Applied oilseed rape marker technology and genomics. In *Applied Plant Genomics and Biotechnology* (eds Poltronieri, P. & Hong, Y.) 253–295 (Woodhead Publishing, 2015). <https://doi.org/10.1016/B978-0-08-100068-7.00016-1>.
- Delourme, R. et al. High-density SNP-based genetic map development and linkage disequilibrium assessment in *Brassica napus* L. *BMC Genomics* **14**, 120. <https://doi.org/10.1186/1471-2164-14-120> (2013).
- Delourme, R. et al. Genes and Quantitative Trait Loci Mapping for Major Agronomic Traits in *Brassica napus* L. In *The Brassica napus Genome* (eds Liu, S. et al.) 41–85 (Springer International Publishing, 2018). [https://doi.org/10.1007/978-3-319-43694-4\\_3](https://doi.org/10.1007/978-3-319-43694-4_3).
- Dalton-Morgan, J. et al. A high-throughput SNP array in the amphidiploid species *Brassica napus* shows diversity in resistance genes. *Funct. Integr. Genomic* **14**, 643–655. <https://doi.org/10.1007/s10142-014-0391-2> (2014).
- Clarke, W. E. et al. A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theor. Appl. Genet.* **129**, 1887–1899. <https://doi.org/10.1007/s00122-016-2746-7> (2016).
- Mason, A. S. et al. High-throughput genotyping for species identification and diversity assessment in germplasm collections. *Mol. Ecol. Resour.* **15**, 1091–1101. <https://doi.org/10.1111/1755-0998.12379> (2015).
- Li, F. et al. Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). *DNA Res* **21**, 355–367. <https://doi.org/10.1093/dnares/dsu002> (2014).
- Hatzig, S. V. et al. Genome-wide association mapping unravels the genetic control of seed germination and vigor in *Brassica napus*. *Front. Plant Sci.* **6**, 221. <https://doi.org/10.3389/fpls.2015.00221> (2015).
- Xu, L. et al. Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (*Brassica napus* L.). *DNA Res* **23**, 43–52. <https://doi.org/10.1093/dnares/dsv035> (2016).
- Wei, L. J. et al. Genome-wide association analysis and differential expression analysis of resistance to Sclerotinia stem rot in *Brassica napus*. *Plant. Biotechnol. J.* **14**, 1368–1380. <https://doi.org/10.1111/pbi.12501> (2016).
- Qu, C. M. et al. Identification of Candidate Genes for Seed Glucosinolate Content Using Association Mapping in *Brassica napus* L. *Genes-Basel* **6**, 1215–1229. <https://doi.org/10.3390/genes6041215> (2015).
- Zhang, J. et al. Identification of putative candidate genes for water stress tolerance in canola (*Brassica napus*). *Front. Plant Sci.* <https://doi.org/10.3389/Fpls.2015.01058> (2015).
- Sashidhar, N., Harloff, H.-J. & Jung, C. Identification of phytic acid mutants in oilseed rape (*Brassica napus*) by large scale screening of mutant populations through amplicon sequencing. *New Phytol.* <https://doi.org/10.1111/nph.16281> (2019).
- Sashidhar, N., Harloff, H. J. & Jung, C. Knockout of *MULTI-DRUG RESISTANT PROTEIN 5* genes lead to low phytic acid contents in oilseed rape. *Front Plant Sci* **11**, 1. <https://doi.org/10.3389/fpls.2020.00603> (2020).
- Harloff, H. J. et al. A mutation screening platform for rapeseed (*Brassica napus* L.) and the detection of sinapine biosynthesis mutants. *Theor. Appl. Genet.* **124**, 957–969. <https://doi.org/10.1007/s00122-011-1760-z> (2012).
- Saxena, K. B., Saxena, R. K., Hickey, L. T. & Varshney, R. K. Can a speed breeding approach accelerate genetic gain in pigeonpea?. *Euphytica* **215**, 202. <https://doi.org/10.1007/s10681-019-2520-4> (2019).
- Mobini, S. H., Lulsdorf, M., Warkentin, T. D. & Vandenberg, A. Plant growth regulators improve in vitro flowering and rapid generation advancement in lentil and faba bean. *In Vitro Cell Dev-Pl* **51**, 71–79. <https://doi.org/10.1007/s11627-014-9647-8> (2015).
- Herzog, E. & Frisch, M. Efficient marker-assisted backcross conversion of seed-parent lines to cytoplasmic male sterility. *Plant Breed.* **132**, 33–41. <https://doi.org/10.1111/pbr.12021> (2013).
- Frisch, M. & Melchinger, A. E. Selection theory for marker-assisted backcrossing. *Genetics* **170**, 909–917. <https://doi.org/10.1534/genetics.104.035451> (2005).
- Bernardo, R., Murigneux, A., Maisonneuve, J. P., Johnsson, C. & Karaman, Z. RFLP-based estimates of parental contribution to F<sub>2</sub>- and BC<sub>1</sub>-derived maize inbreds. *Theor. Appl. Genet.* **94**, 652–656. <https://doi.org/10.1007/s001220050462> (1997).

39. Sagare, D. B., Shetti, P., Surender, M. & Reddy, S. S. Marker-assisted backcross breeding for enhancing  $\beta$ -carotene of QPM inbreds. *Mol Breed.* <https://doi.org/10.1007/s11032-019-0939-x> (2019).
40. Rai, N. *et al.* Marker-assisted backcross breeding for improvement of drought tolerance in bread wheat (*Triticum aestivum* L. em Thell). *Plant Breed.* **137**, 514–526. <https://doi.org/10.1111/pbr.12605> (2018).
41. Chukwu, S. C. *et al.* Genetic analysis of microsatellites associated with resistance against bacterial leaf blight and blast diseases of rice (*Oryza sativa* L.). *Biotechnol. Biotechnol. Equip.* **34**, 898–904. <https://doi.org/10.1080/13102818.2020.1809520> (2020).
42. Randhawa, H. S. *et al.* Rapid and targeted introgression of genes into popular wheat cultivars using marker-assisted background selection. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0005752> (2009).
43. Frisch, M. & Melchinger, A. E. Marker-assisted backcrossing for simultaneous introgression of two genes. *Crop Sci.* **41**, 1716–1725. <https://doi.org/10.2135/cropsci2001.1716> (2001).
44. Emrani, N., Harloff, H. J., Gudi, O., Kopisch-Obuch, F. & Jung, C. Reduction in sinapine content in rapeseed (*Brassica napus* L.) by induced mutations in sinapine biosynthesis genes. *Mol. Breed.* <https://doi.org/10.1007/s11032-015-0236-2> (2015).
45. Saghaimarouf, M. A., Soliman, K. M., Jorgensen, R. A. & Allard, R. W. Ribosomal DNA spacer-length polymorphisms in Barley–Mendelian inheritance, chromosomal location, and population-dynamics. *Proc. Natl. Acad. Sci. Biol.* **81**, 8014–8018. <https://doi.org/10.1073/pnas.81.24.8014> (1984).
46. Bradbury, P. J. *et al.* TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308> (2007).
47. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330> (2011).
48. Zheng, X. W. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328. <https://doi.org/10.1093/bioinformatics/bts606> (2012).
49. Team, R. C. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012. <http://www.R-project.org> (2019).
50. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>

## Acknowledgements

We thank Monika Bruisch for the greenhouse work. We acknowledge the Institute of Clinical Molecular Biology in Kiel for providing Sanger sequencing. We thank Prof. Habibur Rahman (University of Alberta, Canada) for providing seeds of the spring variety ‘Peace’. This work was financially supported by the Sino-German Center for Scientific Research (Grant No. GZ 1099) and the German Research Foundation (DFG) (Grant No.GZ: JU205/25-1 and Grant No. JU 205/26-1).

## Author contributions

N.L.K. and N.S. designed and performed experiments. N.L.K. and D.S. produced and analyzed data. N.L.K. drafted the manuscript. D.S., H.-J.H., N.S., and C.J. revised the manuscript. H.-J.H. and C.J. led the design of the study and supervised data analysis. All authors read and approved the final article.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-98934-5>.

**Correspondence** and requests for materials should be addressed to C.J.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



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

© The Author(s) 2021