PLANT GENETICS · ORIGINAL PAPER

Population structure and genetic diversity of a germplasm for hybrid breeding in rye (*Secale cereale* L.) using high-density DArTseq-based silicoDArT and SNP markers

Agnieszka Niedziela[1] · Piotr Tomasz Bednarek[1] ⓘ

Abstract

Investigating genetic structure and diversity is crucial for the rye hybrid breeding strategy, leading to improved plant productivity and adaptation. The present study elucidated the population structure and genetic diversity of 188 rye accessions, comprising 94 pollen fertility restoration lines (RF) and 94 cytoplasmic male-sterile (CMS) lines with Pampa sterilizing cytoplasm using SNP and silicoDArT markers from the diversity array technology (DArT)-based sequencing platform (DArTseq). Expected heterozygosity (H_e) and Shanon's diversity (I) indexes varied slightly between marker systems and groups of germplasms ($H_e = 0.34$, $I = 0.51$ for RF and CMS lines genotyped using SNPs; $H_e = 0.31$, $I = 0.48$, and $H_e = 0.35$, $I = 0.53$ for RF and CMS using silicoDArTs, respectively). ANOVA indicated moderate variation (7%) between RF and CMS breeding materials. The same parameter varied when chromosome-assigned markers were used and ranged from 5.8% for 5R to 7.4% for 4R. However, when silicoDArT markers were applied, the respective values varied from 6.4% (1R) to 8.2% (3R and 4R). The model-based (Bayesian) population structure analysis based on the total marker pool identified two major subpopulations for the studied rye germplasm. The first one (P1) encompasses 93 RF accessions, and the second one (P2) encompasses 94 CMS and one RF accession. However, a similar analysis related to markers assigned to selected chromosomes failed to put plant materials into any of the populations in the same way as the total marker pool. Furthermore, the differences in grouping depended on marker types used for analysis.

Keywords Rye · Genetic diversity · CMS Pampa · Hybrid breeding

Introduction

In rye breeding projects, it is crucial to make high-yielding cultivars resistant to diseases, lodging, and sprouting. Grain output and quality might vary greatly depending on cultivars (Arseniuk and Oleksiak 2003; Hansen et al. 2004; Bujak et al. 2006; Dynkowska et al. 2015; Alijošius et al. 2016; Iwańska et al. 2020). Hybrid rye cultivars have been shown to produce 20–25% more grain than the best population cultivars while maintaining grain quality (Arseniuk and Oleksiak 2003; Laidig et al. 2017; Linina et al. 2019). The heterosis phenomenon causes this beneficial impact.

✉ Piotr Tomasz Bednarek
p.bednarek@ihar.edu.pl

1   Plant Breeding and Acclimatization Institute - National Research Institute, 05-870 Błonie, Radzików, Poland

According to Fisher et al. (2010), heterosis is stronger when the parent forms come from different genetic populations. The populations are called heterotic groups. A group of related or unrelated genotypes from the same or other populations that exhibit comparable combining capacity and heterotic response when crossed with genotypes from other genetically dissimilar germplasm groups is referred to as a "heterotic group" (Melchinger and Gumber 1998). So, Reif et al. (2003) say that it depends on the available materials to determine the genetic makeup and/or genetic diversity of heterotic groupings. Regardless of the year and location of the experiment, some research showed that the larger the genetic distance between parental plants, the more significant the heterosis effect for most of the observed features (Reif et al. 2003; Liersch et al. 2010; Tomkowiak et al. 2019, 2020). For example, in studies of maize (Becker and Link, 2002; Betrán et al. 2003; Zhang et al. 2017a, b; Tomkowiak et al. 2019, 2020) and rapeseed (Becker and Engqvist 1995; Liersch et al. 2010), researchers found a link between

the heterosis effect on seed yield and the genetic distance between the parents.

However, the genetic distance coefficients employed for germplasm analysis affect how genetic pools form (Reif et al. 2003; Tomkowiak et al. 2020). Additionally, the nature of the marker system may be significant (Frisch et al. 2010; Tomkowiak et al. 2020). Tomkowiak et al. (2020) found that AFLP and RAPD markers were less valuable than SSR, SNP, and silicoDArTs when choosing parental components for crossing maize. Frisch et al. (2010) recommended a different solution based on transcriptome information. The author showed that general estimates of hybrid performance based on field data or older models with DNA markers are less accurate than distances based on the transcriptome.

It should be emphasized that predicting the consequences of heterosis is not always possible by relying solely on genetic groups derived using similarity indices based on molecular markers (Kaeppler 2012). Wheat or durum wheat, for instance, does not demonstrate a relationship between genetic distance and heterosis (Martin et al. 1995; Barbosa-Neto et al. 1996; Perenzin et al. 1998; Corbellini et al. 2002; Dreisigacker et al. 2005; Krystkowiak et al. 2009). That could be because the species lacks genetic diversity or because the studies only used a small amount of parental germplasm, making it hard to find links between genetic distance and heterosis. On the other hand, parental forms of sunflower hybrids (Gvozdenović et al. 2009), waxy and sweet corn (Dermail et al. 2020), and oilseed rape (Yu et al. 2005) showed an important link between genetic distance and heterosis impact. Also, Fischer (2010), Zhang et al. (2017a, b), and Boyaci et al. (2020) found that using heterotic pools to make hybrids by mixing the tester with each trait could lead to better results.

Rye hybrid breeding programs are a prime example of using heterotic gene pools for economic gain (Geiger and Miedaner 1996). In Germany, two heterotic genetic pools that represented the maternal ("Petkus" pool) and paternal ("Carsten" pool) genotypes were selected (Geiger and Miedaner 1996). Plants from both gene pools were selfed repeatedly to make the parental inbred rye lines (Fischer 2010). In the breeding process, inbred lines were primarily generated from crosses between elite inbreeds within heterotic pools, reducing genetic diversity (Duvick et al. 2004). Alternatives are being sought since using the breeding strategy to discover heterotic pools based on, for example, combining ability is relatively laborious. Furthermore, some species or materials might not have heterotic groups.

Numerous population studies have been carried out on the rye germplasm. DNA-based molecular marker techniques such as simple sequence repeats (SSR), amplification fragment length polymorphism (AFLP), random amplification polymorphic DNA (RAPD), diversity array technology (DArT), and single nucleotide polymorphism (SNP) were used to measure the genetic diversity (Ćwiklińska et al. 2010; Myśków et al. 2010; Chikmawati et al. 2012; Bolibok-Brągoszewska et al. 2014; Targońska et al. 2016; Sidhu et al. 2019; Vendelbo et al. 2020; Targońska-Karasek et al. 2020). The findings suggest that the genetic diversity of the *Secale* genus is greater than that of landraces or even cultivated accessions (Shang et al. 2006; Bolibok-Brągoszewska et al. 2014). Genetic differences (Phi_{PT}) based on DArT markers range from 0.15 to 0.20 between landraces, breeding materials, varieties from the PAS BG seed bank in Warsaw-Powsin (Poland), and accessions from professor A. Łukaszewski's collection (Bolibok-Brgoszewska et al. 2014). The same authors found no difference when comparing landraces, cultivated materials, and varieties from PAS BG (Phi_{PT} below 0.05). On the other hand, restorer lines (R) from the "Petkus" and "Carsten" gene pools were genetically distinct ($F_{ST} = 0.332$) from non-restorer germplasm (NRG) combined with cytoplasmic male-sterile (CMS) lines based on the "Gülzow" (G) type of cytoplasm, according to a study by Vendelbo et al. (2020) using SNP markers. Hierarchical clustering and principal component analysis showed that the seed parent (NRG&CMS) and pollen parent (R) populations were genetically different. Small populations of the identical seed and pollen parent pools from the "Petkus" and "Carsten" showed a significant difference ($F_{ST} = 0.229$) when Bauer et al. (2017) employed markers from a 600 K high-density SNP array. This trend was supported by a principal coordinate analysis (PCoA) that showed a clear difference between the parent populations. A similar study was not done on the maternal and restorer lines used by Polish breeding companies, which is unfortunate. However, rye hybrid breeding companies have extensive expertise in assessing plant material. Additionally, different companies preferred breeding maternal and parental lines. This makes it more likely that genetically different pools have formed. Knowing this could help assess heterogenic pools in the rye.

We hypothesize that the choice of a DNA-based marker system might be crucial for the identification of putative pools, and that due to breeding pressure, the markers assigned to the chromosomes under such pressure would be more effective in differentiating materials than the entire marker set.

Material and methods

Plant material and DNA extraction

For the investigation, a total of 188 elite hybrid breeding components of rye (*Secale cereale* L.) were chosen, including 94 restorer lines (RF) and 94 cytoplasmic male-sterile (CMS) lines with Pampa sterilizing cytoplasm (Table S1). For RF and CMS-based lines, the inbreeding stage ranged

from F3 to F11 and F2 to F3 generations, respectively. The plants were assisted by DANKO Plant Breeding Ltd., located in Choryń 27, 64–000 Kościan, Poland. All restorers and 83 CMS lines originated from the Breeding Department of Choryń and 11 CMS lines from the Breeding Department of Laski.

Total DNA samples were extracted from fresh leaf tissue of three plants (equal parts) representing the given line using a DNeasy Plant Mini Kit 250 according to the manufacturer's instructions. The quality of the isolated DNA was measured on a 1% agarose gel and with a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

Genotyping

DNA samples were processed for genotyping using DArTseq™ technology offered by Diversity Arrays Technology (Pty) Ltd. in Canberra, Australia. DArTseq is a genotyping-by-sequencing system that sequences the most informative representations of genomic DNA samples using next-generation sequencing platforms (HiSeq 2500 in our case). The protocol has been described previously in detail by Kilian et al. (2012) and Melville et al. (2017). The results consist of two marker types: SNP (codominant) and silicoDArT (dominant). SNP markers were coded in a binary matrix. Each locus was represented by two consecutive lines to preserve its codominant nature. The presence of an SNP relative to the reference sequence was denoted as 1, while the absence was 0. So, the array showed homozygotes as 1/1 or 0/0 and heterozygotes as 1/0. SilicoDArT markers were coded in a binary fashion according to their absence (0) or presence (1) in genomic representations. The markers were filtered for reproducibility ≥ 0.95 and CallRate ≥ 0.90. Furthermore, missing data (< 10%), and minor allele frequency < 5% were discarded for future analysis.

Marker position on the rye chromosomes was defined by diversity arrays technology (Pty) Ltd. according to the position(s) on contig(s) with the best alignment of marker/tag to the "Rye_v2" model genome. The values reflect distance as the number of the base pair from the beginning of the chromosome.

Genetic diversity and population structure

Primary genetic diversity indices were calculated using the GenAlEx v.6.5 Excel add-in (Peakall and Smouse 2012). These included the percentage of polymorphic markers (PPL), the number of different alleles (N_a), the number of effective alleles (N_e), the number of private alleles/bands, Shannon's information index (I), expected heterozygosity (H_e), unbiased expected heterozygosity (uH_e), observed heterozygosity (H_o), and the fixation index (F).

For the evaluation of polymorphic information content (PIC) of dominant bi-allelic silicoDArT markers, the following formula was used: $PIC = 1 - (p^2 + q^2)$, where "p" is the frequency of present alleles and "q" is the frequency of null alleles (Serrote et al. 2000). Botstein et al. (1980) came up with a formula to figure out the PIC for codominant (SNP) markers:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2$$

In this equation, n is the number of alleles, and p_i and p_j are the allele frequencies in populations i and j, respectively. PIC values range from 0 (monomorphic) to 0.5 (highly informative, where many alleles have the same frequency). Results were given for the total number of markers searched in the RF and CMS lines and for the markers assigned to each rye chromosome separately.

The GeneAlex Excel add-in software (Peakall and Smouse, 2012) used the analysis of molecular variance (AMOVA) to divide molecular variance into two groups: (1) between RF and CMS breeding materials and (2) within RF and CMS breeding materials. The pairwise F_{ST} (codominant SNP markers) or Phi_{PT} (dominant silicoDArT markers) coefficients and Nei's minimum genetic distance (Nei 1972) between the RF and CMS lines were calculated in GeneAlex. The Phi_{PT}/F_{ST} values were calculated to evaluate within-population variance and population differentiation. The estimated probability values from 1000 permutations were used to figure out if the way the variance components were split up was significant. The analysis was done using the total number of markers and the markers assigned to each chromosome.

The dartR package was used to look at a pairwise genetic distance (GD) matrix of codominant SNP markers. The dissimilarity between the two sets was measured using Jaccard's distance. GD matrices were calculated for each of the rye chromosomes. A pairwise genetic similarity (GS) matrix (Jaccard coefficients) (Jaccard 1908) based on molecular markers assigned to the given chromosome (and total marker pool) and utilizing dominant silicoDArT markers was calculated in PAST software, Version 4.06b (Hammer et al. 2001). For the analysis, ten hundred markers with the smallest amount of missing data per individual (less than 4.5%) were applied to fulfill PAST requirements. GS values were converted into GD by subtracting GS from 1. In PAST software, a violin plot was used to show how the GS values for all markers and chromosomes were spread out in terms of density. The genetic distance matrices were compared using the Mantel test (Mantel 1967) in the GeneAlex Excel add-in software. The comparison led to correlation and determination coefficients.

In the XlStat software (XlStat 2019), a one-way analysis of variance (ANOVA) with Tukey's pairwise comparisons was used to look for differences between the GD means of different chromosomes.

The putative structure among all 188 rye accessions genotyped with SNP and silicoDArT markers was tested using principal component analysis (PCA) based on the Jaccard distance matrices in PAST software, Version 4.06b (Hammer et al. 2001). Next, Bayesian analysis with Structure 2.2.3 (Pritchard et al. 2000b) was used to figure out how many populations (K) could capture the main structure in the data. The admixture model, a burn-in period of 50,000 MCMC iterations, and 100,000 run lengths were used in the analysis. There were eight separate runs for every simulated value of K, from 1 to 5. The most likely K-value was determined by structure harvester (Earl and Vonholdt 2012) using the log probability of the data [LnP(D)] and delta K (ΔK) based on the rate of change in [LnP(D)] between successive K-values (Evanno et al. 2005). Next, the average genetic structure was estimated using the CLUMPP software (Jakobsson and Rosenberg 2007). With the Distruct software (Rosenberg, 2004), a bar graph of the structure of the population was made. The analysis was done separately for the total number of SNP and silicoDArT markers and markers assigned to the given chromosomes.

Results

Genetic diversity

In total, 188,150 (62,580 SNP and 125,780 silicoDArT) signals were generated for 188 rye accessions using DArT-seq technology. After filtering, there were 57,550 markers (14,300 SNP and 43,220 silicoDArT).

Percentages of polymorphic markers (PPL) for investigated materials equaled 99% (SNP) and 100% (silicoDArTs). The genetic variation parameters calculated for SNP markers, such as observed heterozygosity (H_o), the diversity index (H_e), and fixation index (F) for the entire set of accessions,

had values of 0.074, 0.342, and 0.786 for RF and 0.125, 0.338, and 0.634 for CMS, respectively. Shannon's information index (I) was 0.51 for both the RF and CMS lines. There were 226 (RF) and 240 (CMS) private alleles (PrA) among SNP markers (Table 1).

The genetic variation parameters calculated for silicoDArTs were much lower for restorers (H_e = 0.313, I = 0.484) than for cytoplasmic male-sterile lines (H_e = 0.348, I = 0.525). In the RF and CMS materials, there were three and seventeen private bands (PrB) among the silicoDArTs.

There were 1498, 1901, 1559, 1450, 1787, 1477, and 1540 SNP markers on the chromosomes 1R, 2R, 3R, 4R, 5R, 6R, and 7R, respectively, in the RF and CMS materials that were studied (Fig. 1). The chromosomal location of the 3088 markers remains unknown. In the case of silicoDArTs, 2962 (1R), 3273 (2R), 2573 (3R), 3242 (4R), 3440 (5R), 2807 (6R), and 2794 (7R) were assigned to the particular rye chromosomes (Fig. 1), and 22,165 remained unassigned (NA). The markers (SNPs and silicoDArTs) were evenly distributed along the rye chromosomes. On average, when SNPs were used, markers were separated by a distance ranging from 26.1 to 35.1 bp in the case of the 2R and 1R chromosomes. When silicoDArTs were used, similar values
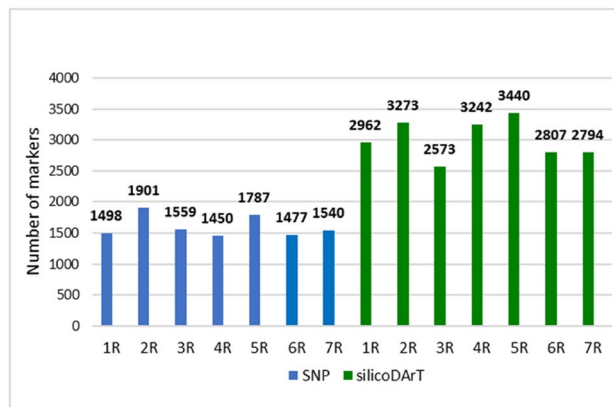

Fig. 1 Distribution of SNP and silicoDArT makers across all chromosomes in the rye genome

Table 1 The mean of different genetic parameters, including the number of different alleles (Na), number of effective alleles (Ne), Shannon's information index (I), number of private alleles (SSR)/bands (silicoDArT) (PrA/B), observed heterozygosity (H_o), expected heterozygosity (H_e), unbiased expected heterozygosity (uH_e), fixation index (F), and percentage of polymorphic loci (PPL) in RF and CMS parental components used for hybrid breeding in Poland

Type of the markers	lines	N_a	N_e	I	PrA/B	H_o*	H_e	uH_e	PPL	F*
SNP	RF	1.99	1.583	0.513	226	0.074	0.342	0.345	99.0	0.786
	CMS	1.99	1.576	0.509	240	0.125	0.338	0.341	99.1	0.634
silicoDArT	RF	2.00	1.452	0.484	3	-	0.313	0.315	99.9	-
	CMS	2.00	1.522	0.525	17	-	0.348	0.349	99.9	-

*For SNP only (Ho is available for co-dominant markers only)

The analysis omits a single restorer line which was classified with CMS lines according to the structure results

varied from 14.5 to 23.1 bp for the 5R and 6R chromosomes, respectively. It should be stressed, however, that the distal arms were poorly saturated with markers. In those regions, only 1 to 2% of markers were available. If those gaps were considered, the average marker distribution ranged from 33.7 bp (2R) to 56 bp (7R) for SNPs and from 21.5 bp (5R) to 43.9 bp (1R) for silicoDArTs. The most extended gaps were found on the 7R (8160 bp for SNPs) and 1R (29,631 bp for silicoDArTs) chromosomes.

The average PIC value for all analyzed SNP markers was 0.34 for both the RF and CMS lines. A very high PIC value (0.4–0.5) was observed for about 44.3% of the markers. As many as 6.2% (RF) and 7.6% (CMS) SNP markers with a PIC value of > 0.1 were present. When the markers were put on chromosomes, markers on 2R (CMS), 5R (RF), and 6R (RF) had a slightly higher average PIC (0.35 vs. 0.33–0.34) (Fig. 2).

In the silicoDArT markers, the PIC values varied for RF and CMS lines and reached 0.31 and 0.35, respectively. In total, 51.5% (RF) and 36.6% (CMS) of them were characterized by a PIC value between 0.4 and 0.5, whereas only 2.9% (RF) and 1.5% (CMS) had a PIC value ranging from 0 to 0.1. The average PIC values were comparable for all RF (0.34) and CMS (0.38) chromosomes. However, they varied significantly for not-assigned markers (0.25 for RF and 0.28 for CMS lines) (Fig. 2).

The average GD among rye genotypes calculated with a total number of SNP markers ranged from 0.030 ("L1909"– "L1961") to 0.619 ("SE58P/12"– "SO79R") and reached an average value of 0.535. The average GD value differentiated slightly but significantly ($p < 0.0001$) across the chromosomes and reached 0.562, 0.575, 0.567, 0.567, 0551, 0.565, and 0.566 for 1R, 2R, 3R, 4R, 5R, 6R, and 7R, respectively (Fig. 3a). When the plants were looked at with

markers that did not belong to any of the chromosomes, the mean GD value was 0.452.

The average genetic distance (GD) calculated using Jaccard's coefficient for rye breeding materials using a total number of silicoDArT markers was 0.799 and varied from the lowest of 0.098, which occurred between RF line "L1909" and "L1961," to the highest of 0.991, which occurred between RF "L1995" and CMS lines: "S195P/18" (0.897), "S246P/18" (0.897), "S235P/18" (0,896), "LS379P/18" (0.895), "S46P/18" (0,892), "S222P/18" (0.892), "LS483P/18" (0,891), "S29P/18" (0.891) and "S211P/18" (0.890). RF "L1995" reached the highest values of GD (> 0.850) for all CMS lines. The average GD values across the chromosomes were 0.633, 0.629, 0.635, 0.633, 0.630, 0.633, and 0.637 for 1R, 2R, 3R, 4R, 5R, 6R, and 7R, respectively (Fig. 3b). The highest mean value of GD (0.803) was determined when markers that did not correspond to any of the chromosomes were used.

The lowest genetic distance between RF lines "L1909" and "L1961" was noticed independently of the marker types and the chromosomes for which the matrices were calculated. The highest GD depends on marker types and chromosomes. Maximum SNP values range from 0.659 (5R) to 0.718 (3R), while maximum silicoDArT values range from 0.843 (4R) to 0.866 (7R). The RF and CMS lines with the highest GD values were different depending on the chromosome. However, in silicoDArTs, the same RF line, "L1995," always showed up in such a pair with the earlier-mentioned CMS lines, no matter the chromosome.

Comparing the genetic distance, matrices for all of the RF and CMS materials that were analyzed, evaluated using SNP or silicoDArTs markers, and assigned to the rye chromosomes revealed (Fig. 3) that the GD between plant materials was higher when calculated using SNPs as opposed

Fig. 2 Average polymorphic information content (PIC) values of SNP and silicoDArT markers calculated for the entire genome (Total) and particular chromosomes (1R-7R) of restorer (RF) and cytoplasmic male-sterile (CMS) lines. NA, markers not assigned to any chromosome
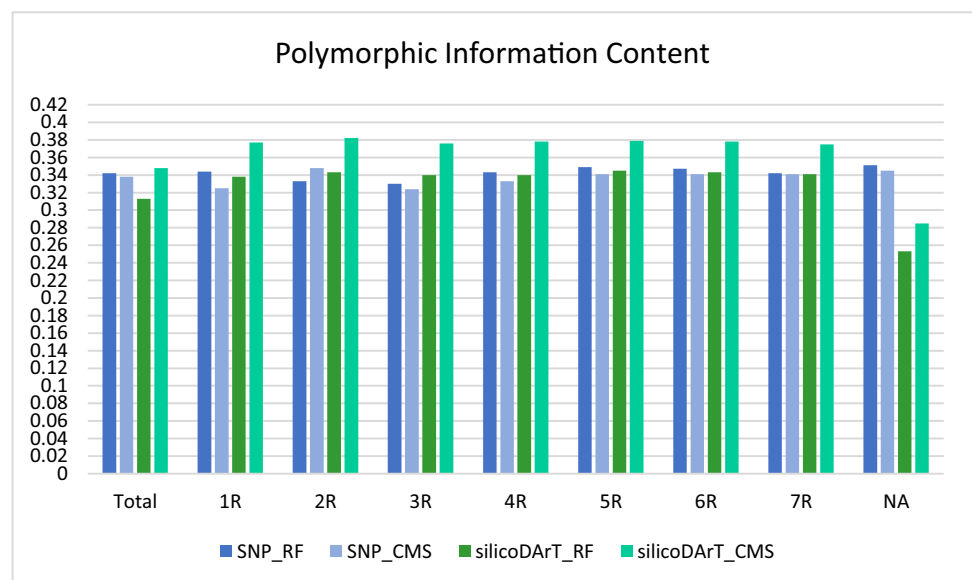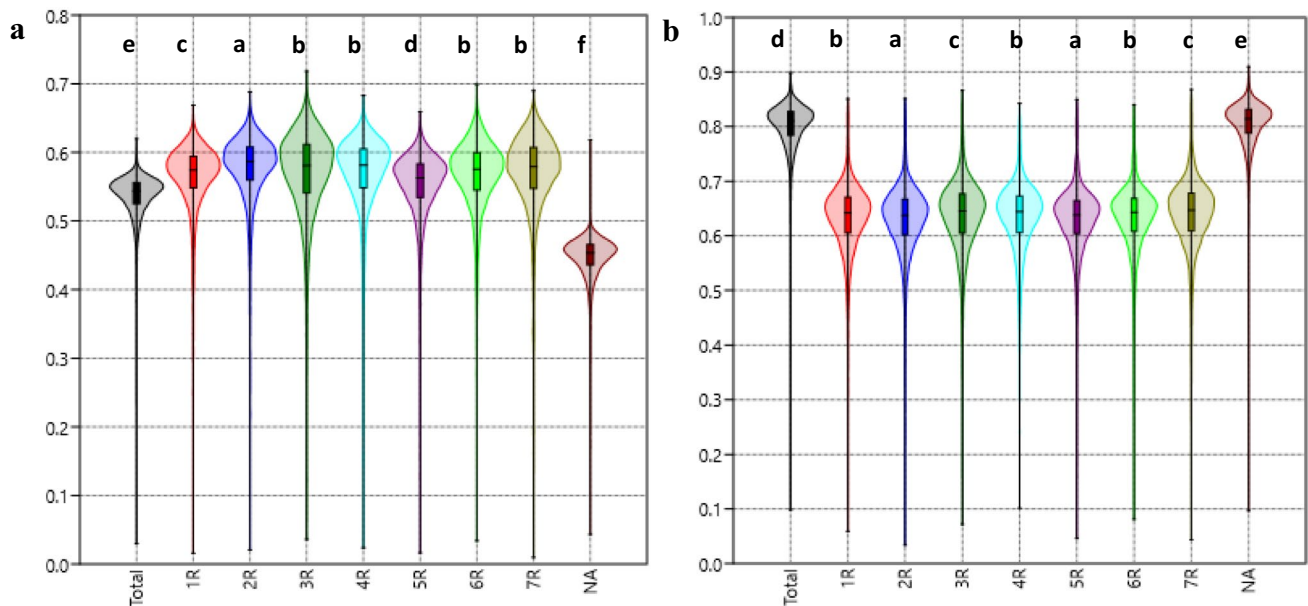
Fig. 3 Violin plots show the density distribution between plant materials evaluated based on genetic distance (GD) values calculated with SNP (**a**) and silicoDArT (**b**) markers. The abscisic axis displays data based on markers allocated to all rye chromosomes, markers assigned to individual chromosomes, and markers that have not yet been assigned to any chromosome. The ordinate axis displays genetic dis-tance. The violin plots illustrate the range of genetic distances across plant materials. Vertical bars reflect the genetic distances between studied materials lying within the most typical range, whereas hori-zontal lines inside the box shown mean values. The letters at the top of each plot show how the Tukey's HSD mean comparison test with a *p*-value of 0.0001 grouped the data

to silicoDArTs for all of the analyzed chromosomes. When GD matrices for the total number of markers and those for markers not assigned to any chromosomes were compared, the relative density distribution values for SNPs were lower than for silicoDArTs. Also, the data from the SNP markers assigned to a chromosome showed less separation between plant materials than the GD matrices made from the entire pool of markers and those not assigned to any chromosome. The silicoDArTs showed the opposite behavior. It should be noted that several materials had a significant GD distance. But the mean GD values for materials based on SNPs or silicoDArTs and assigned to chromosomes were closer. The violin charts illustrate the problem.

The analysis of molecular variance for all GD matrices using various marker types showed that the variations in mean GD values were significant (SNP: $F = 4969.5$, $p < 0.0001$; silicoDArT: $F = 22,952.9$, $p < 0.0001$). When SNP-based matrices were considered, the chromosomes 3R, 4R, 6R, and 7R formed one group, whereas the rest formed the other group. Additionally, the total and NA markers formed distinct groups. The silicoDArT-based matrices revealed that 1R, 4R, and 6R; 2R and 5R; and 3R and 7R formed separate groups. Again, the GD matrices based on total and NA markers formed separate groups.

An analysis of the Mantel test showed that the genetic distance matrices of SNPs and silicoDArTs evaluated for all the analyzed plant materials were significantly correlated

($r = 0.874$; $p < 0.001$). However, the correlation between GD matrices of rye chromosomes varied with the marker system used for the analysis and was relatively low for SNPs (from $r = 0.262$ between 3 and 7R, to $r = 0.383$ between 1 and 4R) and somewhat higher for silicoDArTs (from $r = 0.536$ between 3 and 7R, to $r = 0.676$ between 1 and 4R) (Table 2). No matter what marker system was used, all distance matrices from individual chromosomes were very similar to those from a whole set of markers and not just assigned ones. The 1R-based GD matrices exhibited the highest correlation values with most other chromosome-based matrices. Furthermore, the 4R GD matrix evaluated based on silicoDArTs was highly correlated with the 5R and 6R, whereas the 5R matrix was highly correlated with the 6R and 2R. The SNP-based markers also showed a similar pattern, but it was not as clear.

AMOVA analysis revealed that RF and CMS breeding materials had a much higher proportion of within-variation (93%) than RF and CMS breeding materials (7%). The F_{ST}/Phi_{PT} values between RF and CMS lines using the total number of SNP and silicoDArT markers were 0.074 ($p < 0.001$) and 0.077 ($p < 0.001$), respectively (Table 3). The same parameter varied when chromosome-assigned markers were used. Then the F_{ST} values ranged from 0.058 for 5R to 0.074 for 4R. But silicoDArT markers were used, the Phi_{PT} values were between 0.064 (1R) and 0.082 (3R and 4R). Nei's minimum distance (Table 3) showed a similar pattern of difference between RF and CMS lines.

Table 2 Correlations of genetic distance matrices obtained with total pool and chromosome-specific marker sets ($p = 0.001$)

		Total	1R	2R	3R	4R	5R	6R	7R
SNP	1R	0.648							
	2R	0.663	0.348						
	3R	0.603	0.300	0.281					
	4R	0.649	0.383	0.330	0.277				
	5R	0.663	0.363	0.361	0.307	0.361			
	6R	0.637	0.366	0.325	0.280	0.345	0.357		
	7R	0.652	0.329	0.316	0.262	0.318	0.346	0.340	
	NA	0.940	0.600	0.588	0.522	0.623	0.591	0.590	0.606
silicoDArT	1R	0.779							
	2R	0.729	0.631						
	3R	0.697	0.613	0.549					
	4R	0.780	0.676	0.608	0.583				
	5R	0.746	0.646	0.605	0.605	0.636			
	6R	0.765	0.671	0.606	0.587	0.645	0.645		
	7R	0.717	0.611	0.553	0.536	0.599	0.602	0.608	
	NA	0.992	0.765	0.718	0.689	0.786	0.738	0.766	0.719

Table 3 Fixation index (F_{ST}), pairwise values (Phi_{PT}) and Nei's minimum genetic distance between restorer and cytoplasmic male-sterile lines included in the Polish hybrid breeding program

	SNP		silicoDArT	
	F_{ST}*	Nei's genetic distance	Phi_{PT}*	Nei's genetic distance
Total	0.074	0.050	0.077	0.030
1R	0.072	0.052	0.074	0.040
2R	0.065	0.049	0.069	0.037
3R	0.072	0.052	0.082	0.043
4R	0.074	0.055	0.082	0.044
5R	0.058	0.044	0.064	0.034
6R	0.072	0.055	0.075	0.040
7R	0.070	0.052	0.073	0.040
NA	0.067	0.051	0.083	0.021

*F_{ST} is a statistic measure for comparison between codominant allelic data (SNP) and Phi_{PT} is an analogous measure for binary data (silicoDArT)

Population structure

Principal component analysis of 94 RF and 94 CMS rye genotypes provides an estimate of the proportion explained by individual principal components (PCs) and served in our study to get a first-hand visualization of the genetic diversity and architecture. The PCA based on the total number of SNP markers shows that the first and second PCs explained only 7.8% and 4.0% of the total genetic variability of the original data set, respectively (Fig. S1). The grouping of genotypes based on total and not assigned SNP markers is similar. It reveals that at least two separate subgroups exist. However,

17 of the RF lines are located in a subgroup represented by mostly CMS lines, and 15 of the CMS lines are located in the RF subgroup. Data on individual chromosomes need to be clarified to determine how to divide them into subgroups (Fig. S1); however, such division is not excluded in some cases.

The PCA done on all of the silicoDArT markers put the 188 rye accessions into two main groups, which were made up of the RF and CMS lines (Fig. S2). The first and second principal components explained 5.8% and 3.0% of genetic variability, respectively. When the pool of unassigned markers was used in the analysis, it showed the same grouping. The only exceptions are the restorer line ("L1409"), which shared ancestry with male-sterile genotypes, and the CMS line (S133P), which was grouped with the RF materials. When markers assigned to individual chromosomes were applied for the PCA, the division into two groups was still present (Fig. S2). However, the differentiation of the RF and CMS materials was somewhat unclear, with apparent regions where representatives of the CMS and RF pools are mixed. Furthermore, variation in RF and CMS populations was observed depending on the chromosome-assigned markers used. For example, when 1R data is used, CMS line variation is much higher than RF; when 4R data is used, it is the other way around.

Based on the total number of SNPs or silicoDArTs, an analysis of the population structure showed that the data was split into two groups (P1 and P2). The highest ΔK value was observed at $K = 2$ (Fig. 4a and b). SNP markers revealed more vital structuring than silicoDArTs ($\Delta K = 1961.5$ for SNPs; $\Delta K = 959.3$ for silicoDArTs). Among all the accessions tested using SNP and silicoDArT markers, 93 restorer lines formed one population (Table 4, Table S1). In contrast,
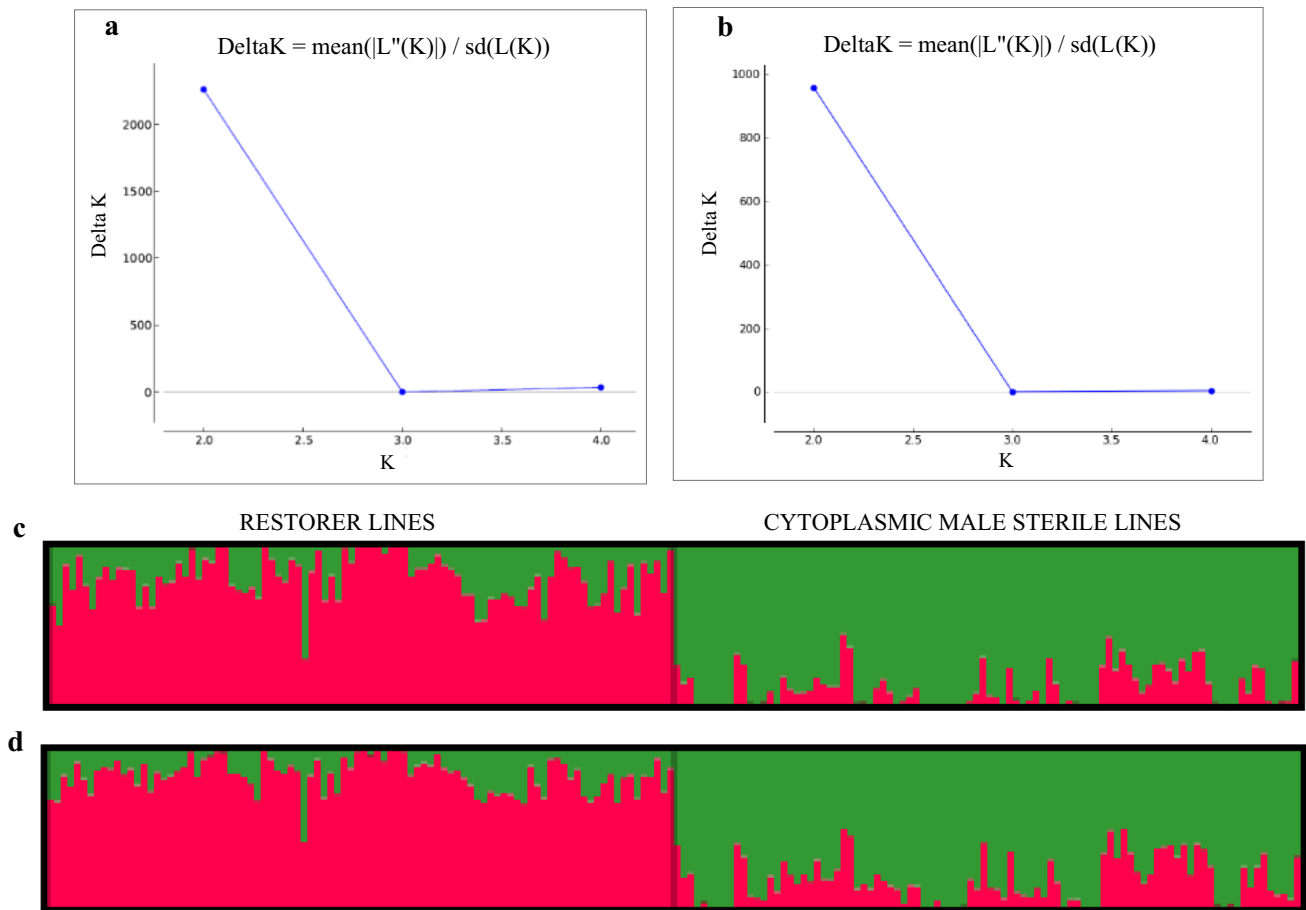
Fig. 4 Determination of the optimal value of K = 2 and population structure analysis of 94 restorer lines (RF) and cytoplasmic male-sterile (CMS) lines using silicoDArT (**a**) and SNP (**b**) data. Illustration of population structures evaluated on the total pool of SNP (**c**) and silicoDArT markers (**d**)

the remaining lines belonging to the CMS pool formed the second population (Fig. 4c and d). It should be stressed, however, that the CMS pool encompasses a single restorer line (L1409). Due to only minor deviations, the results of AMOVA for a structured population based on total pools of markers were the same as for RF and CMS pools (not shown).

The markers associated with the rye chromosomes generally assessed a similar number of populations (Table 4, Table S1). In the case of the 1R-6R chromosomes, *K* equals two. However, for the 7R chromosome, *K* equals three regardless of whether SNPs or silicoDArTs were employed. Chromosome-assigned silicoDArT markers (in contrast to the entire marker pool) revealed more robust plant material structuring than SNPs (Table 4). Based on chromosome-specific SNPs or silicoDArTs, the analyzed plant samples were not precisely put into one of the populations in the same way as was done with the whole marker pool. Using SNP markers, the correct classification of RF lines to P1 ranged from 52.1% for 7R to 98.9% for 6R, and silico-DArTs from 42.5% (7R) to 100% (4R). A

similar analysis shows that most CMS lines are classified as P2. The percentage values varied from 88.3% (5R) to 100% (2R, 4R, and 7R) for SNPs and from 82.9% (3R) to 100% (6R and 7R) for silicoDArTs. Regarding 7R-located markers, SNPs and silicoDarTs group the CMS-based lines into one population and divide the restorers into three populations.

Discussion

Genetic differentiation of breeding populations originating from distinct sources may indicate the formation of heterogenic pools (Vogt et al. 2020). Combined, they may result in extra heterozygosity, increasing yield, tolerance to abiotic stresses (Anioł and Gustafson 1984; Matos 2005), and confer extra resistance to pathogens (Miedaner et al. 2002). In the Polish breeding populations of rye, such pools were hardly recognized, mainly due to common breeding programs leading to the exchange of materials (Bolibok-Brągoszewska et al. 2014). However, with the development of the CMS

Table 4 The arrangement of structure results and "misassignment" of RF and CMS lines depending on marker type and their chromosomal location

		1R	2R	3R	4R	5R	6R	7R
SNP	K	2	2	2	2	2	2	3*
	ΔK value	318.5	261.6	149.1	216.8	399.7	204.6	477.2
	Number of **RF** lines in **P1**	66 (70.2%)	51 (54.3%)	72 (76.6%)	70 (74.5%)	70 (74.5%)	93 (98.9%)	49 (52.1%)
	Number of **CMS** lines in **P2**	93 (98.9%)	94 (100%)	90 (95.7%)	94 (100%)	83 (88.3%)	89 (94.7%)	94 (100%)
	Number of RF lines in P2	28	43	22	14	14	1	23
	Number of CMS lines in P1	1	0	4	0	11	5	0
silicoDArT	K	2	2	2	2	2	2	3*
	ΔK value	468.6	490.7	428.2	1329.9	323.3	354.6	501.8
	Number of **RF** lines in **P1**	89 (94.7%)	92 (97.9%)	86 (91.5%)	94 (100%)	92 (97.9%)	92 (97.9%)	40 (42.5%)
	Number of **CMS** lines in **P2**	92 (97.9%)	90 (95.7%)	78 (82.9%)	88 (93.6%)	86 (91.5%)	94 (100%)	94 (100%)
	Number of RF lines in P2	5	2	8	0	2	2	30
	Number of CMS lines in P1	2	4	16	6	8	2	0

*The rest of the RF lines represented P3

The percentage of RF and CMS lines in a given population (P1 or P2) is placed next to the numerical values

Pampa hybrid approach (Geiger and Schnell 1970) and its introduction by Lucjan Madej (Madej 1975), Polish companies started separating their materials. Some specialized in maternal, whereas the others in parental forms used as components of new cultivars. A good example is the "Gulden" cultivar (DANKO Plant Breeding Co. Ltd) released recently (COBORU 2022a) or those made available to farmers years ago (Arseniuk and Oleksiak 2003). Still, Polish rye hybrids' yield is lower than those developed in Germany (COBORU 2022b). The reason for that is the preference for 1R rather than 4R restorer QTL in Polish materials. The frequency of the 4R QTL is relatively minute. However, in some cases, such materials were identified (Niedziela et al. 2021a).

The selection of hybrid materials via a breeding approach is time-consuming. Thus, alternative methods based on molecular markers are needed. While the genomic selection approach is of the highest importance and gave acceptable results in wheat (Michel et al. 2017; Zhao et al. 2021), corn (Zhang et al. 2017a, b), and many other kinds of cereal (Cui et al. 2020; Wolfe et al. 2017), it was not tested for rye. Still, the limited pool of available materials for such selection makes the approach unjustified. The other option is to make markers for pollen fertility restoration QTLs and use them for selection. While many markers linked to or associated with the QTLs were evaluated (Niedziela et al. 2021b; Stojałowski et al. 2011), few were tested on a broad range of materials. Alternatively, molecular markers combined with clustering methods could be used. But the method was questioned because the results depended on the species and the marker system (Tomkowiak et al. 2020; Dziurdziak et al. 2021). The other option is to identify the respective genetic pools and assess their suitability for hybrid breeding.

Based on PCA with a full pool of SNPs or silicoDArTs, the current study shows that plant materials from two breeding companies that run maternal or parental components of the hybrids are different. Based on the results of AMOVA, the difference between the two pools explained 7.4% (SNPs) and 7.7% (silicoDArTs) of the variance. A similar comparison of the German hybrid breeding pool showed that the F_{ST} values were at least two times higher in the cases of a broad range of rye materials (Bauer et al. 2017) and about five times higher for the RF and CMS-based materials (Vendelbo et al. 2020). The fact was interpreted in terms of genetic pool formation. The PCA analysis confirmed the difference. It shows apparent differences between the RF and CMS materials, where only a small number of lines from each group were put in the wrong group regarding the breeding company. A study of the structure of the population confirmed the result, which showed that two different genetic pools of rye materials have grown up that are used only for hybrid breeding. But the differences may not be due to the creation of pools that can be used to choose materials for hybrid breeding. Instead, they may be the result of breeding preferences. Instead of using all the markers, looking at the ones unique to each rye chromosome might be more helpful. This would help us tell where the plants came from and whether they belonged to the RF or CMS groups. As expected, AMOVA on the markers put on each rye chromosome showed that RF and CMS materials differed. According to AMOVA, the difference between RF and CMS materials depends on the chromosome and marker type.

Furthermore, the highest values of explained variance were evaluated for the 4R (SNPs) and the 4R and 5R (silicoDArTs) chromosomes. The presented result is in line with our hypothesis that breeding materials for RF and

CMS pools should force selection at the level of chromosomes carrying pollen fertility restoration traits and that the differentiation of plant materials based on the markers assigned to those chromosomes should be stronger than the total number of markers of the given type. The outcome is consistent with genetic distance, PCA, and structure analyses. More exposed results were found for silicoDArTs than SNPs, showing that they are better at separating rye lines. The difference between marker systems may reflect their dominant and codominant natures. Codominant markers may identify heterozygotic loci, so they should be less distinctive than dominant markers. Alternatively, the number of SNPs and silicoDArTs may also be substantial. In general, more silicoDArTs were available for the analyses of those SNPs. However, independently of the marker type, they were evenly distributed along each rye chromosome without significant gaps. Thus, their number and distribution should not affect the results (Table S2). The differences reflect the available RF and CMS plant material differentiation.

The notion seems to be confirmed by the Tukey's HSD test comparing genetic distances evaluated based on silicoDArTs. It was shown that 1R, 4R, and 6R formed a single group. The other chromosomes, 2R and 5R, and 3R and 7R, formed another one, suggesting that the chromosomes belonging to the same group might be under comparable pressure. In the case of the first group, the grouping may reflect pollen fertility restoration as the chromosomes are responsible for the expression of the trait. A comparable analysis conducted on SNPs failed to identify similar grouping (3R, 4R, 6R, and 7R vs 1R, 2R, and 5R). This is in agreement with the fact that distinct information could be gained depending on the marker system and type. Therefore, selecting the marker system and type should be carefully considered for hybrid breeding purposes.

The results of the GD analysis show that the lowest averages based on SNPs for 1R (0.562) and 5R (0.551) might show less polymorphism in some genomic regions where there are QTLs that control traits that are important for agriculture. The 1RS is known to carry a cluster of genes encoding resistance to stem, leaf, and yellow rust (Mago et al. 2005), as well as one of the most influential and widely used Pm genes to control powdery mildew (Wricke et al. 1996; Hsam et al. 2000; Simkova et al. 2008). In European rye resources, Meidaner et al. (2000) found that the short arm of the 1R chromosome is home to the QTL that helps CMS Pampa grow back the best. The other essential traits, including plant height, spike length, the number of florets (Plaschke et al. 1993; Börner et al. 1999), and alpha-amylase activity (Masojć and Milczarski 2009), are localized on 5R. At the same time, GD shows that despite high within-population variance (93%), some materials were very similar. The result may suggest that the RF and CMS gene pools may be limited and need to be enriched with new genotypes.

The other question we are trying to answer with our work is whether or not any grouping analysis can be used to predict materials that would fit into the maternal or paternal restoration pool. The most logical approach would be to use chromosomal markers that reflect pollen fertility and/or sterility. For example, in rye pollen fertility restoration, QTLs are present on 1R, 3R, 4R, 5R, and 6R (Miedaner et al. 2000), with the most significant on 4R originating from Iranian and Argentinian germplasm (Miedaner et al. 2000).

Focusing attention on the classification of RF and CMS lines (based on structure analysis), it could be shown that CMS lines are usually completely separated from the pool of RF materials. The highest misclassification was about 17% for SNPs (5R) and silicoDArTs (3R) chromosomes. Generally, the RF lines were inappropriately classified, even in 46% of cases utilizing SNPs (2R) and 56.5% for silicoDArTs (7R). However, misclassification is less pronounced for the 1R and 4R chromosomes, which carry the most significant pollen fertility restoration QTLs. Only 30% of the RF and about 1% of CMS materials were misclassified based on SNPs. When silicoDArTs were applied, the level of misclassification equaled 5% and 2% for the RF and CMS materials, respectively. In the case of the 4R data, up to 74.5% of the RF and 100% of the CMS lines were assigned according to the known phenotypes based on SNPs. The silicoDArTs were even more efficient in the case of RF materials (100% correct assignment), and 93.6% of the CMS lines were grouped according to expectations. Based on the markers that were assigned to the 1R and 4R chromosomes, the data could be used to separate the hybrid breeding materials into two groups. So, the result may show that materials from different companies show signs of forming heterotic pools. The idea is backed up by the fact that when markers for the other chromosomes were used, there were more wrong classifications. To confirm the notion, though, more research needs to be done on combining the abilities of the two groups.

Conclusions

In conclusion, the choice of marker system is crucial when trying to find the genetic pools of RF and CMS materials. Dominant markers are more effective than codominant ones in separating RF and CMS-based materials. It is best to find out which chromosomes have the most to do with the trait since the markers on those chromosomes would be better at separating materials than the total number of markers. The most differentiating RF and CMS materials markers were those mapped to 1R and 4R chromosomes, which carry the most important QTLs for the trait. The two pools of plant materials derived from two companies and reflecting RF and CMS hybrid components have signs of differentiation, suggesting the opportunity for independent genetic pool

formation. The presence of genetically closely related materials in the RF and CMS materials suggests the necessity of extending the genetic pool.

Author contribution AN conducted the molecular experiments, collected and analyzed data, and co-wrote the manuscript; PB analyzed data and co-wrote the manuscript.

Data availability All supporting data are included within the article and its additional files.

Declarations

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest The authors declare no competing interests.

References

Alijošius S, Švirmickas GJ, Kliseviciute V et al (2016) The chemical composition of different barley varieties grown in Lithuania. Vet Med Zoot 73(95):9–13

Anioł A, Gustafson JP (1984) Chromosome location of genes controlling aluminium tolerance in wheat, rye, and triticale. Can J Genet Cytol 26:701–705. https://doi.org/10.1139/g84-111

Arseniuk E, Oleksiak T (2003) Rye production and breeding in Poland. Plant Breed Seed Sci 47:7–16. https://ojs.ihar.edu.pl/index.php/pbss/article/view/749. Accessed 20 Dec 2022

Barbosa-Neto JF, Sorrells ME, Cisar G (1996) Prediction of heterosis in wheat using coefficient of parentage and RFLP-based estimates of genetic relationship. Genome 39(6):1142–1149. https://doi.org/10.1139/g96-144. (PMID: 18469962)

Bauer E, Schmutzer T, Barilar I et al (2017) Towards a whole-genome sequence for rye (Secale cereale L). Plant J 89:853–869. https://doi.org/10.1111/tpj.13436

Becker H, Engqvist GM (1995) The potential of re-synthesized rapeseed for hybrid breeding. Proc 9th Int Rapeseed Congr, GCIRC, Cambridge, England, pp 113–115

Becker H, Link W. Heterosis in hybrid breeding. In: Mendel Centenary Congress MCC-2000. Brno, Czech Republic: Mendel Centenary Congress MCC-2000; 2000:319–327

Betrán FJ, Ribaut JM, Beck D, de León DG (2003) Genetic diversity, specific combining ability, and heterosis in tropical maize under stress and nonstress environments. Crop Sci 43(3):797. https://doi.org/10.2135/cropsci2003.7970

Bolibok-Brągoszewska H, Targonska M, Bolibok L, Kilian A, Rakoczy-Trojanowska M (2014) Genome-wide characterization of genetic diversity and population structure in Secale. BMC Plant Biol 14:184. https://doi.org/10.1186/1471-2229-14-184

Börner A, Korzun V, Voylokov AV, Weber WE (1999) Detection of quantitative trait loci on chromosome 5R of rye (Secale cereale L). Theor Appl Genet 98:1087–1090. https://doi.org/10.1007/s001220051171

Botstein D, White RL, Skalnick MH, Davies RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am J Hum Genet 32(3):314–31

Boyaci HF, Prohens J, Unlu A, Gumrukcu E, Oten M, Plazas M (2020) Association of heterotic groups with morphological relationships and general combining ability in eggplant. Agriculture 10(6):203. https://doi.org/10.3390/agriculture10060203

Bujak H, Dopierała A, Dopierała P, Nowosad K (2006) Analysis of genotype-environment interaction for the yield of rye cultivars (In Polish, with English abstract). J Bull Plant Breed Acclimatization Inst 240:151–160

Chikmawati T, Miftahudin M, Skovmand B, Gustafson JP (2012) Amplified fragment length polymorphism-based genetic diversity among cultivated and weedy rye (Secale cereale L.) accessions. Genet Resour Crop Evol 59:1743–1752. https://doi.org/10.1007/s10722-012-9796-8

COBORU (2022a) https://www.coboru.gov.pl/pl/szczegoly_odmiany?nrodm=47493. Accessed 20 Dec 2022

COBORU (2022b) COBO 93/2022b n.300+100

Corbellini M, Perenzin M, Accerbi M, Vaccio P, Borghi B (2002) Genetic diversity in bread wheat, as revealed by coefficient of parentage and molecular markers, and its relationship to hybrid performance. Euphytica 123:273–285. https://doi.org/10.1023/A:1014946018765

Cui Y, Li R, Li G, Zhang F, Zhu T, Zhang Q, Ali J, Li Z, Xu S (2020) Hybrid breeding of rice via genomic selection. Plant Biotechnol J 18(1):57–67. https://doi.org/10.1111/pbi.13170

Ćwiklińska A, Broda Z, Bocianowski J, Dobrzycka A (2010) The usefulness of RAPD and AFLP markers for deter-mining genetic similarity in rye (Secale L.) species and subspecies. Acta Biol Crac Ser Bot 52(1):19–25

Dermail A, Suriharn B, Chankaew S, Sanitchon J, Lertrat K (2020) Hybrid prediction based on SSR-genetic distance, heterosis and combining ability on agronomic traits and yields in sweet and waxy corn. Sci Hortic 259:108817. https://doi.org/10.1016/j.scienta.2019.108817

Dreisigacker S, Melchinger AE, Zhang P, Ammar K, Flachenecker C, Hoisington D, Warburton ML (2005) Hybrid performance and heterosis in spring bread wheat, and their relations to SSR-based genetic distances and coefficients of parentage. Euphytica 144:51–59. https://doi.org/10.1007/s10681-005-4053-2

Duvick DN, Smith JSC, Cooper M (2004) Long-term selection in a commercial hybrid maize breeding program. Plant Breed Rev 24:109–152. https://doi.org/10.1002/9780470650288.ch4

Dynkowska W, Cyran M, Ceglińska A (2015). Soluble and cell wall-bound phenolic acids and ferulic acid dehydrodimers in rye flour and five bread model systems: insight into mechanisms of improved availability. J Sci Food Agric 95. https://doi.org/10.1002/jsfa.7007

Dziurdziak J, Gryziak G, Groszyk J, Podyma W, Boczkowska M (2021) DArTseq genotypic and phenotypic diversity of barley landraces originating from different countries. Agronomy 11:2330. https://doi.org/10.3390/agronomy11112330

Earl DA, von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and

implementing the Evanno method. Conserv Genet Resour 4:359–361. https://doi.org/10.1007/s12686-011-9548-7

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fischer S, Maurer HP, Würschum T et al (2010) Development of heterotic groups in triticale. Crop Sci 50(2):584–590. https://doi.org/10.2135/cropsci2009.04.0225

Frisch M, Thiemann A, Fu J, Schrag TA, Scholten S, Melchinger AE (2010) Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize. Theor Appl Genet 120(2):441–450. https://doi.org/10.1007/s00122-009-1204-1

Geiger HH, Miedaner T (1996) Genetic basis and phenotypic stability of male-fertility restoration in rye. Vortr Pflanzenzüchtg (Germany) 35:27–38

Geiger HH, Schnell FW (1970) Cytoplasmic male sterility in rye (Secale cereale L.). Crop Sci 10:590–593. https://doi.org/10.2135/cropsci1970.0011183X001000050043x

Gruber B, Unmack PJ, Berry OF, Georges A (2018) dartr: an R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Mol Ecol Resour 18(3):691–699. https://doi.org/10.1111/1755-0998.12745

Gvozdenović S, SaftićPanković D, Jocić S, Radić V (2009) Correlation between heterosis and genetic distance based on SSR markers in sunflower (Helianthus annuus L.). J Agric Sci 54(1):1–10. https://doi.org/10.2298/JAS0901001G

Hammer O, Harper DAT, Ryan PD (2001) PAST: Paleontological statistics software package for education and data analysis. Palaeontol Electron 4:e9. http://palaeo-electronica.org/2001_1/past/issue1_01.htm. Accessed 20 Dec 2022

Hansen HB, Moller B, Andersen SB, Jorgensen JR, Hansen A (2004) Grain characteristics, chemical composition, and functional properties of rye (Secale cereale L.) as influenced by genotype and harvest year. J Agric Food Chem 52:2282–3229. https://doi.org/10.1021/jf0307191

Hsam S, Mohler V, Hartl L, Wenzel G, Zeller F (2000) Mapping of powdery mildew and leaf rust resistance genes on the wheat-rye translocated chromosome T1BL.1RS using molecular and biochemical markers. Plant Breed 119:87–89. https://doi.org/10.1046/j.1439-0523.2000.00444.x

Iwańska M, Paderewski J, Stępień M, Rodrigues PC (2020) Adaptation of winter wheat cultivars to different environments: a case study in Poland. Agronomy 10:632. https://doi.org/10.3390/agronomy10050632

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaudoise Sci Nat 44:223–70

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29(4):E25. https://doi.org/10.1093/nar/29.4.e25

Jakobsson M, Rosenberg N (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics (oxford, England) 23:1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Kaeppler S (2012) Heterosis: many genes, many mechanisms - end the search for an undiscovered unifying theory. Int Sch Res Not 2012(682824):12

Kilian A, Wenzl P, Huttner E et al (2012) Diversity arrays technology: a generic genome profiling technology on open platforms. Methods Mol Biol 888:67–89. https://doi.org/10.1007/978-1-61779-870-2_5

Krystkowiak K, Adamski T, Surma M, Kaczmarek Z (2009) Relationship between phenotypic and genetic diversity of parental

genotypes and the specific combining ability and heterosis effects in wheat (Triticum aestivum). Euphytica 165:419434. https://doi.org/10.1007/s10681-008-9761-y

Laidig F, Piepho HP, Rentel D, Drobek T, Meyer U, Huesken A (2017) Breeding progress, variation, and correlation of grain and quality traits in winter rye hybrid and population varieties and national on-farm progress in Germany over 26 years. Theor Appl Genet 130(5):981–998. https://doi.org/10.1007/s00122-017-2865-9

Liersch A, Popławska W, Ogrodowczyk M, Krótka K, Bartkowiak-Broda I, Bocianowski J (2010) Assessment of genetic distance of parental lines of F1 hybrids of winter oilseed rape and its relationship with phenotypic distance and heterosis effect (In Polish, with English abstract). Oilseed Crops 31(2):229–241

Linina A, Kunkulberga D, Kronberga A, Locmele I (2019) Winter rye grain quality of hybrid and population cultivars. Agron Res 17(2):1380–1389. https://doi.org/10.15159/AR.19.058

Madej L (1975) Research on male sterility in rye. Plant Breed Seed Sci 19(5/6):421–422

Mago R, Miah H, Lawrence GJ et al (2005) High-resolution mapping and mutation analysis separate the rust resistance genes Sr31, Lr26 and Yr9 on the short arm of rye chromosome 1. Theor Appl Genet 112:41–50. https://doi.org/10.1007/s00122-005-0098-9

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Martin JM, Talbert LE, Lanning SP, Blake NK (1995) Hybrid performance in wheat as related to parental diversity. Crop Sci 35:104–108. https://doi.org/10.2135/cropsci1995.0011183X003500010019x

Masojć P, Milczarski P (2009) Relationship between QTLs for pre-harvest sprouting and α-amylase activity in rye grain. Mol Breeding 23(1):75–84. https://doi.org/10.1007/s11032-008-9215-1

Matos M, Camacho MV, Pérez-Flores V, Pernaute B, Pinto-Carnide O, Benito C (2005) A new aluminum tolerance gene located on rye chromosome arm 7RS. Theor Appl Genet 111(2):360–369. https://doi.org/10.1007/s00122-005-2029-1

Melchinger AE, Gumber RK (1998) Overview of heterosis and heterotic groups in agronomic crops. In: Larnkey KR and Staub JE (eds) Concepts and breeding of heterosis in crop plants, Vol. 25, Madison, WI: Crop Science Society of America, Inc, pp 29–44. https://doi.org/10.2135/cssaspecpub25.c3

Melville J, Haines ML, Boysen K, Hodkinson L, Kilian A, Smith Date KL, Potvin DA, Parris KM (2017) Identifying hybridization and admixture using SNPs: application of the DArTseq platform in phylogeographic research on vertebrates. R Soc Open Sci 4(7):161061. https://doi.org/10.1098/rsos.161061

Michel S, Ametz C, Gungor H et al (2016) Genomic selection across multiple breeding cycles in applied bread wheat breeding. Theor Appl Genet 129(6):1179–1189. https://doi.org/10.1007/s00122-016-2694-2

Miedaner T, Glass C, Dreyer F, Wilde P, Wortmann H, Geiger HH (2000) Mapping of genes for male-fertility restoration in 'Pampa' CMS winter rye (Secale cereale L.). Theor Appl Genet 101:1226–1233. https://doi.org/10.1007/s001220051601

Miedaner T, Gey A-KM, Sperling U, Geiger HH (2002) Quantitative-genetic analysis of leaf-rust resistance in seedling and adult-plant stages of inbred lines and their testcrosses in winter rye. Plant Breed 121:475–479. https://doi.org/10.1046/j.1439-0523.2002.00765.x

Myśków B, Milczarski P, Masojć P (2010) Comparison of RAPD, ISSR and SSR markers in assessing genetic diversity among rye (Secale cereale L.) inbred lines. Plant Breed Seed Sci 62:107–115. https://doi.org/10.2478/v10129-011-0009-y

Nei M (1972) Genetic Distance between Populations. Am Nat 106:283–292

Niedziela A, Brukwiński W, Bednarek PT (2021a) Genetic mapping of pollen fertility restoration QTLs in rye (Secale cereale L.) with

CMS Pampa. J Appl Genetics 62:185–198. https://doi.org/10.1007/s13353-020-00599-9

Niedziela A, Wojciechowska M, Bednarek PT (2021) New PCR-specific markers for pollen fertility restoration QRfp-4R in rye (Secale cereale L.) with pampa sterilizing cytoplasm. J Appl Genet. https://doi.org/10.1007/s13353-021-00646-z

Peakall R, Smouse PE (2012) GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research - an update. Bioinformatics 28:2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Perenzin M, Corbellini M, Accerbi M, Vaccino P, Borghi B (1998) Bread wheat: F 1 hybrid performance and parental diversity estimates using molecular markers. Euphytica 100(1):273–279. https://doi.org/10.1023/A:1018324811038

Plaschke J, Börner A, Xie DX, Koebner RM, Schlegel R, Gale MD (1993) RFLP mapping of genes affecting plant height and growth habit in rye. Theor Appl Genet 85(8):1049–1054. https://doi.org/10.1007/BF00215046

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Genetics 155:574–578. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Reif JC, Melchinger AE, Xia XC et al (2003) Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. Crop Sci 43:1275–1282. https://doi.org/10.2135/cropsci2003.1275

Rosenberg N (2004) DISTRUCT: A program for the graphical display of population structure. Mol Ecol Notes 4:137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

<>14417510.1016/j.gene.2019.144175Serrote CML, Reiniger LRS, Silva KB, Rabaiolli SMDS, Stefanel CM (2000) Determining the Polymorphism Information Content of a molecular marker. Gene 726:144175. https://doi.org/10.1016/j.gene.2019.144175

Shang H-Y, Wei Y-M, Wang X-R, Zheng Y-L (2006) Genetic diversity and phylogenetic relationships in the rye genus Secale L. (rye) based on Secale cereale microsatellite markers. Genet Mol Biol 29(4):685–691. https://doi.org/10.1590/S1415-47572006000400018

Sidhu JS, Ramakrishnan SM, Ali S, Bernardo A, Bai G, Abdullah S, Ayana G, Sehgal SK (2019) Assessing the genetic diversity and characterizing genomic regions conferring Tan Spot resistance in cultivated rye. Plos One 14(3):e0214519. https://doi.org/10.1371/journal.pone.0214519

Simkova H, Safar J, Suchankova P et al (2008) A novel resource for genomics of Triticeae: BAC library specific for the short arm of rye (Secale cereale L.) chromosome 1R (1RS). BMC Genomics 9:237. https://doi.org/10.1186/1471-2164-9-237

Stojałowski SA, Milczarski P, Hanek M, Bolibok-Brągoszewska H, Myśków B, Kilian A, Rakoczy-Trojanowska M (2011) DArT markers tightly linked with the Rfc1 gene controlling restoration of male fertility in the CMS-C system in cultivated rye (Secale cereale L.). J Appl Genetics 52:313–318. https://doi.org/10.1007/s13353-011-0049-x

Targońska M, Bolibok-Brągoszewska H, Rakoczy-Trojanowska M (2016) Assessment of genetic diversity in Secale cereale based on SSR markers. Plant Mol Biol Report 34:37–51. https://doi.org/10.1007/s11105-015-0896-4

Targońska-Karasek M, Boczkowska M, Podyma W et al (2020) Investigation of obsolete diversity of rye (Secale cereale L.) using multiplexed SSR fingerprinting and evaluation of agronomic traits. J Appl Genet 61(4):513–529. https://doi.org/10.1007/s13353-020-00579-z

Tomkowiak A, Bocianowski J, Radzikowska D, Kowalczewski PŁ (2019) Selection of parental material to maximize heterosis using SNP and SilicoDarT markers in maize. Plants (basel) 8(9):349. https://doi.org/10.3390/plants8090349

Tomkowiak A, Bocianowski J, Kwiatek M, Kowalczewski P (2020) Dependence of the heterosis effect on genetic distance, determined using various molecular markers. Open Life Sciences 15(1):1–11. https://doi.org/10.1515/biol-2020-0001

Vendelbo NM, Sarup P, Orabi J, Kristensen PS, Jahoor A (2020) Genetic structure of a germplasm for hybrid breeding in rye (Secale cereale L.). Plos One 15(10):e0239541. https://doi.org/10.1371/journal.pone.0239541

Vogt M, Yates S, Sykes T, Luesink W, Koch M, Studer B (2020) Developing heterotic groups for successful hybrid breeding in perennial ryegrass. Agronomy 10(9):1410. https://doi.org/10.3390/agronomy10091410

Wolfe MD, Del Carpio DP, Alabi O et al (2017) Prospects for genomic selection in cassava breeding. The Plant Genome 10(3). https://doi.org/10.3835/plantgenome2017.03.0015. Accessed 20 Dec 2022

Wricke G, Dill P, Senft P (1996) Linkage between a major gene for powdery mildew resistance and an RFLP marker on chromosome 1R of rye. Plant Breeding 115:71–73. https://doi.org/10.1111/j.1439-0523.1996.tb00874.x

XlStat (2019) https://www.xlstat.com/en/solutions/premium. Accessed 20 Dec 2022

Yu CY, Hu SW, Zhao HX, Guo AG, Sun GL (2005) Genetic distances revealed by morphological characters, isozymes, proteins and RAPD markers and their relationships with hybrid performance in oilseed rape (Brassica napus L.). Theor Appl Genet 110(3):511–518. https://doi.org/10.1007/s00122-004-1858-7

Zhang X, Pérez-Rodríguez P, Burgueño J et al (2017a) Rapid cycling genomic selection in a multiparental tropical maize population. G3: Genes. Genomes, Genetics 7:2315–2326. https://doi.org/10.1534/g3.117.043141

Zhang X, Yong H, Zhou Z et al (2017b) Heterosis and combining ability of seven maize germplasm populations. Euphytica 213:45. https://doi.org/10.1007/s10681-016-1831-y

Zhao Y, Thorwarth P, Jiang Y et al (2021) Unlocking big data doubled the accuracy in predicting the grain yield in hybrid wheat. Sci Adv 7(24):eabf9106. https://doi.org/10.1126/sciadv.abf9106