

RESEARCH ARTICLE

Open Access

# Genome-wide characterization of genetic diversity and population structure in *Secale*

Hanna Bolibok-Brągoszewska<sup>1\*</sup>, Małgorzata Targońska<sup>1</sup>, Leszek Bolibok<sup>2</sup>, Andrzej Kilian<sup>3</sup>  
and Monika Rakoczy-Trojanowska<sup>1</sup>

## Abstract

**Background:** Numerous rye accessions are stored in *ex situ* genebanks worldwide. Little is known about the extent of genetic diversity contained in any of them and its relation to contemporary varieties, since to date rye genetic diversity studies had a very limited scope, analyzing few loci and/ or few accessions. Development of high throughput genotyping methods for rye opened the possibility for genome wide characterizations of large accessions sets. In this study we used 1054 Diversity Array Technology (DArT) markers with defined chromosomal location to characterize genetic diversity and population structure in a collection of 379 rye accessions including wild species, landraces, cultivated materials, historical and contemporary rye varieties.

**Results:** Average genetic similarity (GS) coefficients and average polymorphic information content (PIC) values varied among chromosomes. Comparison of chromosome specific average GS within and between germplasm sub-groups indicated regions of chromosomes 1R and 4R as being targeted by selection in current breeding programs. Bayesian clustering, principal coordinate analysis and Neighbor Joining clustering demonstrated that source and improvement status contributed significantly to the structure observed in the analyzed set of *Secale* germplasm. We revealed a relatively limited diversity in improved rye accessions, both historical and contemporary, as well as lack of correlation between clustering of improved accessions and geographic origin, suggesting common genetic background of rye accessions from diverse geographic regions and extensive germplasm exchange. Moreover, contemporary varieties were distinct from the remaining accessions.

**Conclusions:** Our results point to an influence of reproduction methods on the observed diversity patterns and indicate potential of *ex situ* collections for broadening the genetic diversity in rye breeding programs. Obtained data show that DArT markers provide a realistic picture of the genetic diversity and population structure present in the collection of 379 rye accessions and are an effective platform for rye germplasm characterization and association mapping studies.

## Background

Rye (*Secale cereale* L.;  $2n = 14$ , RR) is an out-crossing, wind-pollinated temperate zone cereal with low water and soil fertility requirements and good tolerance for biotic and abiotic stresses. It is an important crop in several Eastern, Central and Northern European countries with cultivation area of approximately 5 Million hectares worldwide in 2011 [<http://faostat.fao.org>]. The primary uses of rye include bread making, alcohol production,

and animal feed. Recently it is also gaining attention as a biomass crop. Rye products are a valuable diet component due to high dietary fiber content and rye bread was shown to have beneficial influence on human health [1,2]. The crop is also of interest to triticale (x *Triticosecale* Wittmack) and wheat (*Triticum* ssp.) geneticists and breeders as a source of genetic variation, since rye is a donor of the R genome of the triticale and the 1RS chromosome is one of the most frequently used sources of alien chromatin in wheat varieties [3,4].

Apart from *Secale cereale* L., three other species are currently recognized in the genus *Secale*: *S. sylvestre* Host, *S. vavilovii* Grossh. (both annual and self-pollinating), and *S. strictum* (C. Presl.) C. Presl. – perennial and

\* Correspondence: hanna\_bolibok\_bragoszewska@sggw.pl

<sup>1</sup>Department of Plant Genetics, Breeding and Biotechnology, Faculty of Horticulture, Biotechnology and Landscape Architecture, Warsaw University of Life Sciences, Warsaw, Poland

Full list of author information is available at the end of the article

open-pollinating. There are eight subspecies in *S. cereale* and five in *S. strictum* with *Secale cereale* ssp. *cereale* L. being the only cultivated rye [5].

Cultivated rye is believed to have originated from south-western Asia from where it was introduced via Russia to Poland, Germany and subsequently distributed throughout most of Europe. It is also hypothesized that there was a second route of migration of the species into Europe – via Turkey and across the Balkan Peninsula [6].

Traditional rye varieties are panmictic populations, characterized by high levels of heterozygosity and heterogeneity [7]. Over 20 years ago hybrid varieties were introduced and quickly gained popularity due to considerable increase in grain yield. Hybrid rye breeding became possible by using an Argentinean landrace as source of a cytoplasmic male sterility [8], while the most effective restorer genes originated from Iranian and South American collections [9].

Early landrace varieties and wild ancestors provide a broad representation of the natural variation that is present in a species and it was demonstrated that they contain genomic segments that can enhance the performance of some of the world's most productive crop varieties [10]. Increasing the extent of genetic variation available in rye breeding programs by utilization of exotic/primitive accessions is a particularly challenging task because, apart from being agriculturally poor-adapted, the materials in question are also heterozygous and self-incompatible. Nevertheless, it was shown that successful exploitation of these genetic resources is possible in rye even for quantitative traits – promising results were obtained through implementation of marker assisted backcrossing in a study aimed at improving baking quality, where a heterozygous Iranian primitive population was used as a donor [9].

Over eighty rye germplasm collections are maintained world-wide, with the total number of accessions estimated to be approximately 21 000 [11]. Characterization of genetic variation contained in germplasm collections is essential for efficient genebank management [12]. It is also crucial for effective utilization of the genetic resources available in breeding [13]. Over the years several molecular studies were undertaken to assess genetic diversity and relationships in *Secale* species, varieties and inbred lines. Various methods such as, Amplified Fragment Length Polymorphism (AFLP), Inter Simple Sequence Repeat (ISSR), Random Amplified Polymorphic DNA (RAPD), Selective Amplification of Microsatellite Polymorphic Loci (SAMPL), Sequence Specific Amplification Polymorphism (SSAP) and Simple Sequence Repeats (SSR) have been used for this purpose [5,6,14-17]. Analyses of chloroplast and mitochondrial genomes were also done [18]. However, the number of genotypes and the number of marker loci assessed in these studies were usually low. Additionally, in the majority of the studies

the markers used were of anonymous nature, with no information on their chromosomal location available. To our knowledge no extensive assessment of genetic diversity represented by any of the *Secale* genebanks was done up to date.

Recently, Diversity Arrays Technology (DArT) markers were developed for rye and their efficacy in detecting genetic diversity in this species was demonstrated [19]. Subsequently a DArT marker based consensus genetic map was constructed [20]. In consequence the chromosomal location of over four thousand DArT markers was determined and genome-wide genomic analyses became available for rye. The DArT method itself was developed in early 2000. It allows for simultaneous detection of several thousand DNA polymorphisms arising from single base changes and indels by utilizing selective hybridization to DNA fragments immobilized on solid-phase slides. Contrary to the other existing SNP genotyping platforms DArT does not rely on sequence information [21]. DArT markers have been successfully used for a variety of genetic studies including construction of high density linkage maps, association mapping and genomic selection [20,22-24]. Recently DArT markers find increasing use in extensive analyses of genetic diversity and population structure of various crops [12,25-29].

The present study was undertaken to analyze genetic diversity and population structure in a set of 379 diverse rye accessions using high density, genome-wide distributed DArT markers. In particular the objectives of the study were: a) to assess the genetic diversity represented by rye accessions from the collection of Polish Academy of Sciences Botanical Garden-Center for Biological Diversity Conservation in Powsin (PAS BG, Warsaw, Poland) including wild, primitive and historic cultivated germplasm from diverse geographic regions, b) to assess the level of genetic diversity represented by the rye varieties currently marketed in Central Europe, also in relation to diversity contained in the germplasm from an *ex situ* collection (c) to compare the distribution of DNA polymorphisms among rye chromosomes.

## Methods

### Plant material

Plant material consisted of 379 rye accessions: 153 landraces, 46 cultivated materials, 137 varieties, 26 breeding strains, and 17 wild accessions. A total of 306 accessions originated from the collection of the PAS BG, the remaining 73 forms were kindly supplied by breeding companies and by Prof A. Lukaszewski (University of California, Riverside). The samples of *Secale cereale* ssp. *cereale* from PAS BG were selected for maximum diversity and to represent a broad range of historic rye germplasm and geographical regions. Breeding companies supplied rye varieties, which are currently registered

and marketed in Europe, including population and hybrid varieties. Detailed information on rye accessions used, including source, improvement status and region of origin is in Additional file 1: Table S1.

#### DNA isolation and genotyping

DNA was isolated from ca. 2 weeks old greenhouse grown plants using Mag-Bind Plant DNA 96 kit (Omega Bio-Tek). Each rye accession was represented by 96 plants. DNA isolates were pooled into one representative DNA sample for each accession. DNA samples were genotyped with DArT markers using the procedure and the rye genotyping array described in [20] at Diversity Arrays Technology Pty. Ltd., Yarralumla ACT, Australia (Additional file 1: Table S1). For quality control 33% of DNA samples were genotyped in full technical replication. The reproducibility parameter threshold was set at 95%. Chromosomal locations of DArT markers were obtained from the rye consensus map [20] and the results of analysis of wheat-rye addition lines [19].

#### Genetic diversity and population structure

Polymorphic information content (PIC) was calculated for each DArT marker according to Alheit *et al.* [25]. Since DArTs are biallelic dominant markers, PIC values range from 0 (in case of fixation of one allele) to 0.5 (when the frequencies of both alleles are equal). Genetic similarity (GS) matrices based on the Jaccard's similarity coefficient [30] were constructed in NTSYS-pc 2.1 [31] based on combined information from all polymorphic markers with defined chromosomal location and also separately based on markers from individual chromosomes. Mantel's test was performed in NTSYS-pc 2.1 to estimate the correlation between the matrices of genetic similarity obtained using marker data from individual chromosomes. The distance (1-Jaccard) matrix based on combined data was used to construct a dendrogram in MEGA 5.1 [32] software with the Neighbor Joining method. Additionally, a dendrogram including only wild *Secale* accessions was constructed to verify their genetic relationships. Two varieties - Petkus from PAS BG and Dankowskie Nowe from Danko were included in this analysis as the representative accessions of *S. cereale* ssp. *cereale*.

The STRUCTURE [33] software was used to identify the number of populations (K) capturing the major structure in data. We used the admixture model, a burn-in period of 10,000 MCMC iterations and 100,000 run length. Five independent runs were performed for each simulated value of K ranging from 1 to 20. The most likely number of K was then determined using the DeltaK method [34] with the help of the Structure Harvester software [35]. Permutations of the output of the STRUCTURE analysis were done with CLUMPP software [36]

using independent runs to obtain a consensus matrix. A bar graph of the population structure results was generated with Distruct software [37].

GenALEx v.6.501 [38,39] was used to assess the amount of variation among the inferred populations by AMOVA and to calculate the pairwise  $\Phi_{iPT}$  values.  $\Phi_{iPT}$ , an analogue of  $F_{ST}$ , is the estimate of population genetic differentiation provided by GenALEx when binary data are analyzed. GenALEx was also used to investigate graphically the genetic relationships amongst the rye accessions via principal coordinates analysis (PCoA).

## Results

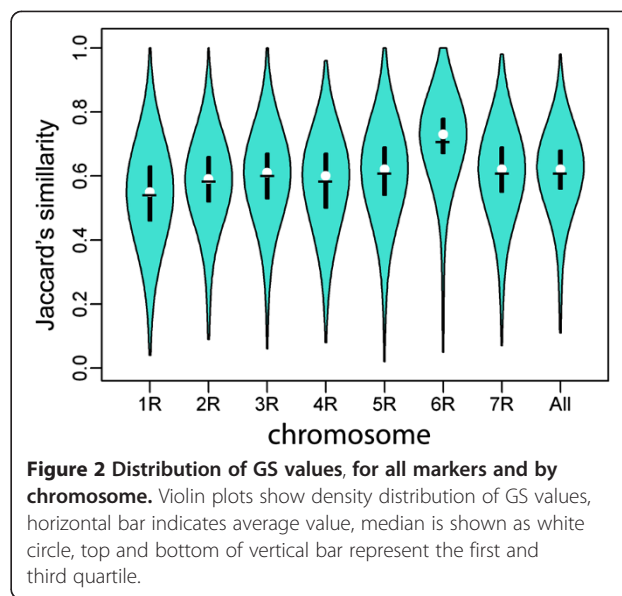
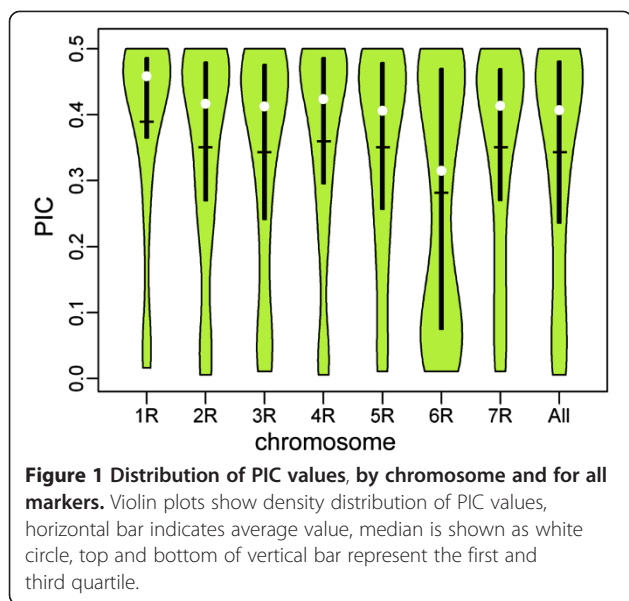
#### Genetic diversity patterns

In total 1054 polymorphic DArT markers with defined chromosomal location were found using the scoring reproducibility threshold of 95%. The average scoring reproducibility was 97.7%. The number of markers per chromosome ranged from 112 for 1R to 231 for 4R. The distribution of markers available by chromosome is given in Table 1.

The mean PIC value for all markers with chromosome location was 0.34. PIC values observed in our study ranged from 0.01 to 0.50 with a high proportion of markers with PIC values above 0.41 (51.7%). PIC values varied among chromosomes, from 0.28 for 6R to 0.39 for 1R. Markers from individual chromosomes exhibited similar PIC value distributions with exception of chromosome 6R, where a large proportion of markers with PIC value below 0.1 (ca. 29%) was observed. Chromosome 1R was characterized by the highest proportion of markers with PIC values above 0.41 (ca. 68%) and the lowest proportion of markers with intermediate PIC values. Violin plots showing distribution of PIC values by chromosome and chromosome specific averages are shown in Figure 1. Mean PIC values for all markers with chromosome location were also calculated separately for sub-groups of accessions created according to source and type of germplasm and varied from 0.26 for varieties from breeding companies to 0.39 for wild accessions. (Additional file 2: Table S2).

**Table 1 Distribution of DArT markers available by chromosome**

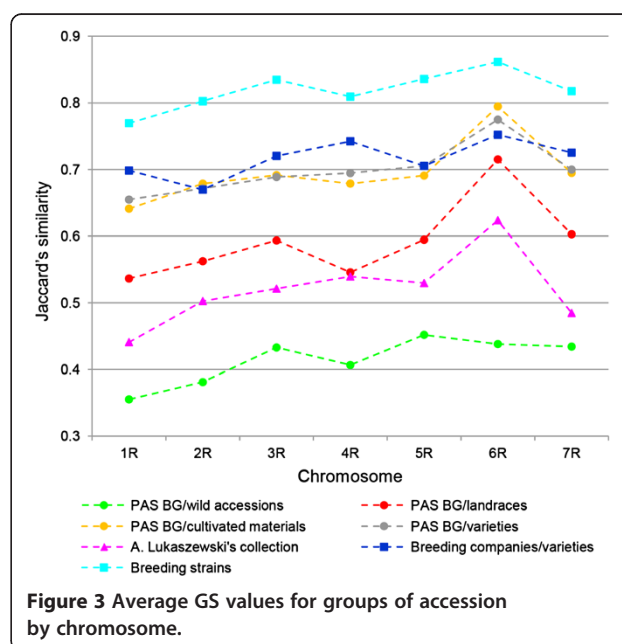
Chromosome	No. of DArT markers
1R	112
2R	125
3R	132
4R	231
5R	122
6R	204
7R	128
total	1054



GS values calculated based on all markers varied from 0.11 to 0.98 with the average 0.61. In the case of the individual chromosomes the highest average GS (0.71) was observed for chromosome 6R, the lowest (0.54) for chromosome 1R. While the combined markers were able to differentiate all accessions, markers in chromosome specific sets (with exception of markers from 4R and 7R) were unable to distinguish two accessions of *S. sylvestre* included in the study. Noteworthy is the relatively low GS value for two samples of Dankowskie Nowe variety, one from PAS BG and the other one from the breeding company Danko, equal 0.82, obtained using combined marker data. The average GS values for individual chromosomes were significantly different ( $p = 0.01$ ), except for the average GS for chromosomes 5R and 7R. Violin plots showing distribution of GS values by chromosome and chromosome specific averages are shown in Figure 2. The average GS values were also calculated for 7 sub-groups of accessions created according to the source and type of germplasm (Figure 3). The pattern of differences in chromosome specific average GS was moderately consistent in different germplasm sub-groups. The highest average GS values were observed for chromosome 6R for all sub-groups, with exception of wild accessions, where the highest GS value was obtained for 5R. The lowest average GS for accession groups occurred in the case of chromosome 1R, except for varieties supplied by breeding companies. In this sub-group the lowest average GS was recorded for chromosome 2R. Overall the lowest average GS was observed for wild accessions, followed by accessions from A. Lukaszewski's collection. The highest average GS was obtained for breeding strains from Danko. Compared to the pattern of chromosome specific average

GS values observed in germplasm sub-groups a markedly higher average GS value was observed in 4R for varieties from breeding companies. In general, the average GSs values for varieties and cultivated materials from PAS BG and for varieties from breeding companies were similar, especially in the case of chromosomes 2R and 5R.

The results of the Mantel's test for correlation between the matrices of genetic similarity obtained for individual chromosomes were significant ( $p = 0.001$ ) and positive with the correlation coefficient  $r$  values ranging from 0.64 for 1R and 7R to 0.77 for 2R and 6R (Table 2).





**Table 2 Correlations of GS matrices obtained with chromosome specific marker sets**

	1R	2R	3R	4R	5R	6R
2R	0.71					
3R	0.72	0.75				
4R	0.72	0.75	0.72			
5R	0.66	0.72	0.72	0.69		
6R	0.68	0.77	0.75	0.75	0.71	
7R	0.64	0.71	0.72	0.67	0.72	0.73

### Model-based population structure

It was estimated through the method of Evanno *et al.* [34] that there are 3 groups contributing significant genetic information in the analyzed *Secale* collection (Additional file 3: Figure S1). STRUCTURE results, grouped and graphed according to accession source, type and geographic region are shown in Figure 4. The classification of accessions into populations by the model-based method is given in Additional file 1: Table S1. In total 226 accessions (59.6%) were assigned to one of the three populations, where at least 70% of their inferred ancestry was derived from one of the three model-based populations. Populations 1, 2, and 3 (P1, P2, and P3) consisted of 140, 10 and 76 accessions, respectively. The remaining accessions were categorized as having admixed ancestry, including 33 admixtures between P1 and P2 (P1P2), 114 between P1 and P3 (P1P3), and two between P2 and P3 (P2P3). Four accessions had similar percentages of their inferred ancestry derived from each of the three model based population and were classified as heterogeneous (H).

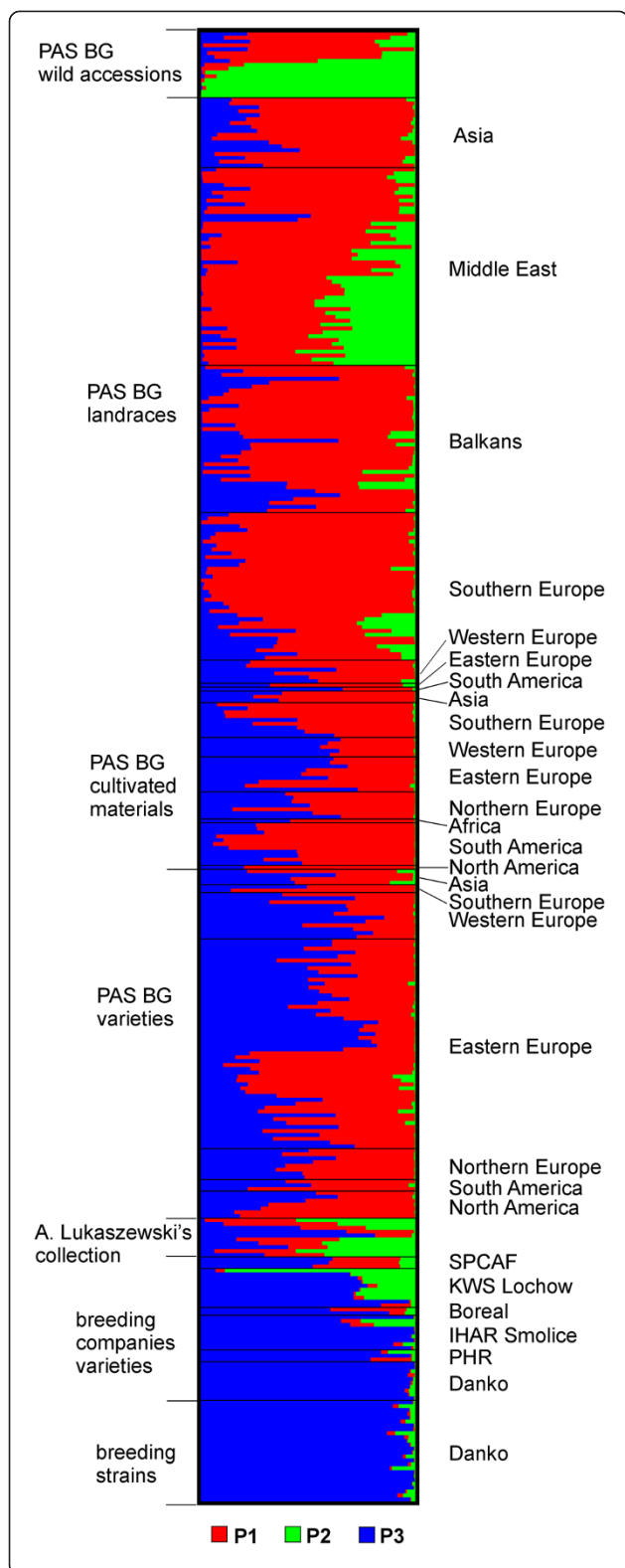
Population assignments of accessions from different germplasm sub-groups were as follows: wild *S. cereale* accessions (with exception of *S. cereale* ssp. *ancestrale*, categorized as heterogeneous) and *S. vavilovii* accessions were assigned to P1, whereas accessions of *S. strictum* and *S. sylvestre* were assigned to P2. Landraces were mostly categorized as P1 (ca. 60% landraces from different regions) The remaining landraces were classified as P1P3 (ca. 18% landraces from various regions) and P1P2. (also ca. 18%). Landraces constituted the majority of the later subpopulation (ca. 85%), and most of them (23 accessions – ca. 66% of all Turkish landraces in the collection) originated from Turkey. Cultivated materials had mostly admixed ancestry – almost 59% of accessions from this sub-group were classified as P1P3, the rest was assigned to P1. Similarly, the majority of varieties from PAS BG (61%) were assigned to P1P3. About 20% of varieties from PAS BG was categorized as P1, and ca. 19% – as P3. The assignment of cultivated materials and varieties from PAS BG to different subpopulations appeared to be rather uncorrelated with the geographical origin of accessions. All breeding strains and the vast majority of the varieties from breeding companies (84%)

were categorized as P3. All three varieties from Belarus and one of the varieties from Boreal – Rihii – were classified as admixtures P1P3. Only one variety from breeding companies –KWS Magnifico F<sub>1</sub> from KWS Lochow was classified as admixture P2P3. Noteworthy is that F<sub>1</sub> varieties included in the set and assigned to P3 had a relatively high proportions of alleles from P2 – from 18 to 30% (Figure 4). The population varieties from breeding companies, that were assigned to P3, had the percentage of their inferred ancestry derived from that population close to 100%. A single variety – Gonello F<sub>1</sub> from KWS Lochow – was assigned to P2. Accessions from A. Lukaszewski's collections had mostly admixed ancestry and myriad subpopulation assignments, with three accessions classified as heterogeneous. In general the classification of populations appeared rather uncorrelated with the geographical origin of rye accessions, but rather reflected the source of the accessions, and, to a degree, the improvement status of the accessions.

### Relationships among accessions based on PCoA and cluster analysis

The PCoA was largely consistent with the STRUCTURE results (Figure 5A). The percentages of genetic diversity explained by the first and the second coordinate were 17.85 and 8.44, respectively. The three model-based populations were well separated with admixtures and heterogeneous accessions located between populations. However, accessions classified as P3 based on STRUCTURE results were divided into two groups in the PCoA plot. One of the resulting groups contained all the materials from breeding companies assigned to P3, and was separated from all remaining accessions from the study. The second group consisted of all the varieties from PAS BG that were assigned to P3 and was located adjacent to the group of accessions from PAS BG classified as admixtures P1P3.

We also used PCoA to determine the extent to which the accessions from different sources and of different improvement status represent distinct areas of diversity space (Figure 5B). It was revealed that, indeed, most materials supplied by breeding companies (which included also several older, but still marketed varieties, such as Dankowskie Nowe and Amilo - release year 1976 and 1987, respectively), occupied a distinct area of diversity space, well distinguished from the remaining accessions. Varieties and cultivated materials from PAS BG overlapped in the PCoA plot, and occupied a relatively narrow space. Landraces represented much greater diversity, with some overlapping with PAS BG varieties and cultivated materials and with accessions from A. Lukaszewski's collection, which in turn occupied a space between materials from other sub-groups. Accessions of *S. sylvestre*, and *S. strictum* were separated from the rest of the rye accessions,



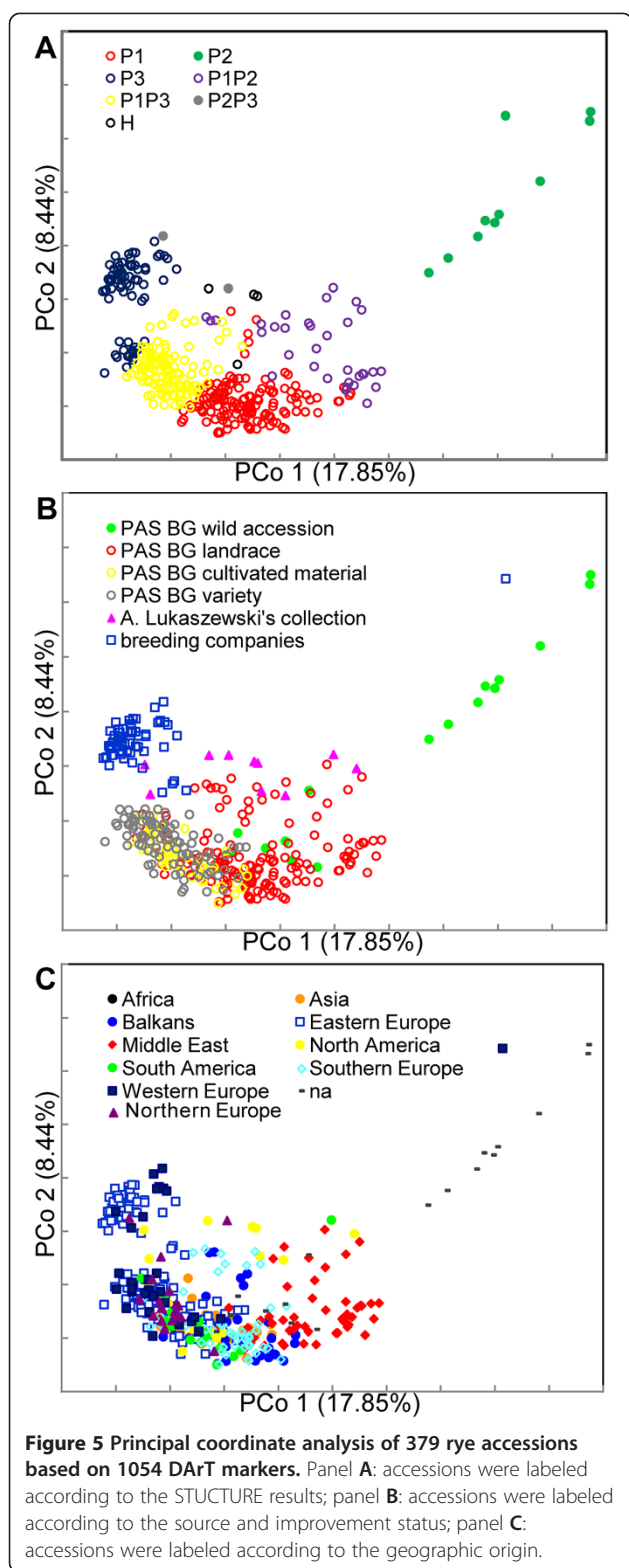
**Figure 4** STRUCTURE plot of the 379 rye accessions with  $K = 3$  clusters based on 1054 DArT markers. The plot is sorted according to accession source and type. The order of accessions in the plot corresponds to the order of accessions in Additional file 1: Table S1. Each accession's genome is represented by a single row, which is partitioned into colored segments in proportion to the estimated membership in the three subpopulations. Black line separates accessions of predefined groups.

whereas wild *S. cereale* and *S. vavilovii* accessions, were intermixed with landraces. Thus, it can be concluded that source and improvement status contributed significantly to the structure observed in the analyzed set of *Secale* germplasm.

On the other hand the distribution of the rye accessions in the 2D space appeared largely unrelated to geography (Figure 5C). It is particularly evident in varieties and cultivated materials from PAS BG, with accessions of different origins dispersed and intermixed in the space occupied by these sub-groups (Additional file 4: Figure S2a). Nevertheless in the case of landraces a certain separation between European and Middle Eastern accessions can be observed. Additionally, Asian landraces clustered relatively close together and occupied an area, where the European and Middle Eastern clusters overlapped (Additional file 4: Figure S2b). Based on PCoA landraces from South Europe, Balkans and Middle East were the most diverse subset of *S. cereale* ssp. *cereale* accessions from PAS BG analyzed in this study, with a subset of Middle Eastern landraces (corresponding to the sub-group P1P2 inferred by STRUCTURE analysis) representing a distinct diversity space, not overlapping with other accessions.

Clustering analysis based on Neighbor Joining allowed the detection of three major clusters: I, II and III (Figure 6). Cluster I, containing the majority of the accessions, could be further subdivided into three clusters a, b and c. The clustering of accessions in the unrooted Neighbour Joining tree was generally in agreement with the model-based population structure of the collection (Figure 6A). The inferred sub-populations were relatively well but not completely separated. Accessions from sub-population P1 were located mostly in clusters Ia and Ib, P2 in cluster III, P3 mostly in cluster II with a subset of accessions in cluster Ic. This separation of P3 into two sub-groups was consistent with PCoA results. Admixtures P1P3 were located mostly in cluster Ic, while the admixtures P1P2 grouped on the verge of cluster Ia, and also in cluster III adjacent to P2 accessions.

The Neighbour Joining tree topology reflected the source and the improvement status of the accessions (Figure 6B), with the accessions from breeding companies placed almost exclusively in cluster II, the varieties from PAS BG mostly in cluster Ic and the accessions



from A. Lukaszewski's collection in cluster III. Cultivated materials were dispersed in clusters Ia, Ib and Ic. Landraces constituted the majority of the accessions in clusters Ia

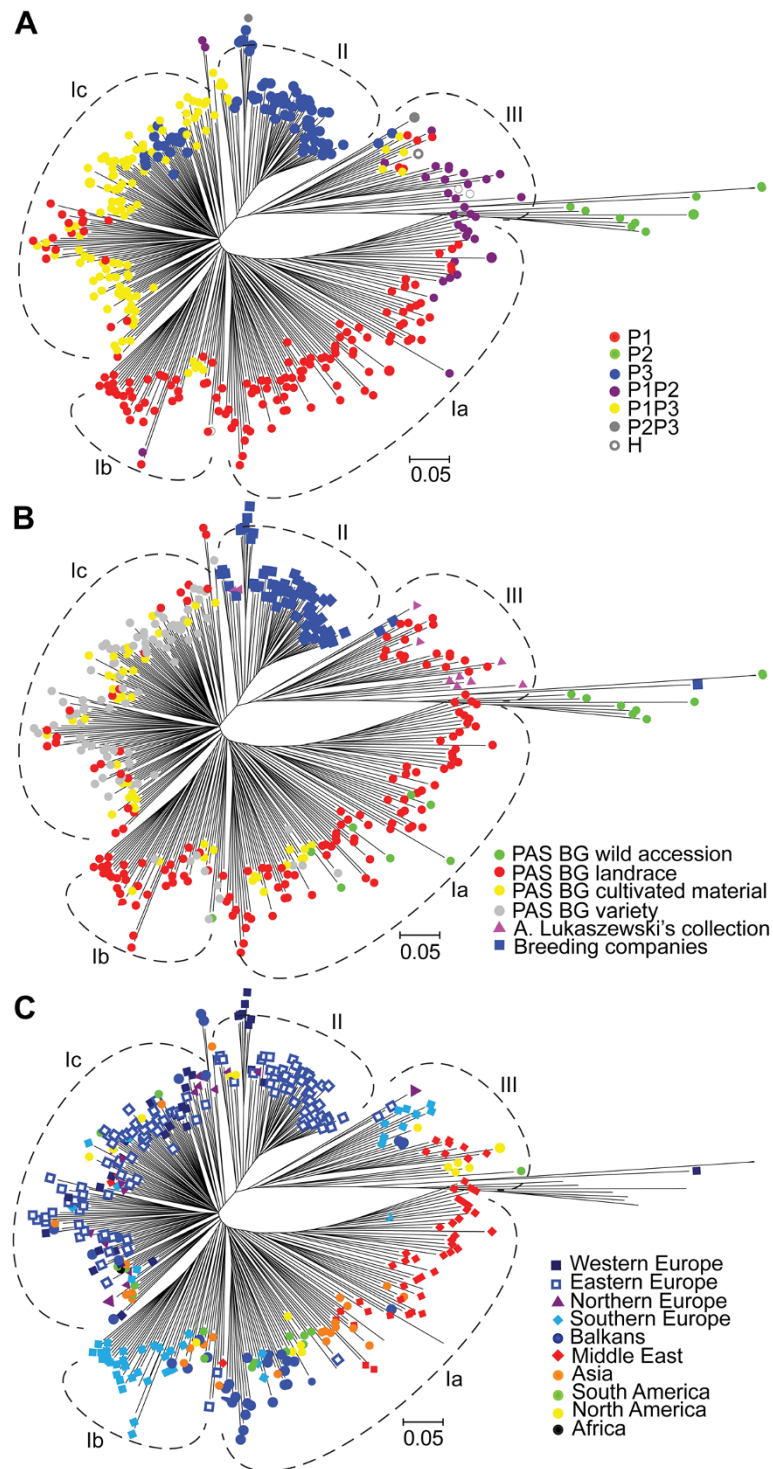
and Ib, although several landraces occurred also in clusters Ic and III.

Correlation of clustering with geographic region of origin was rather weak (Figure 6C). It was slightly more pronounced than in the case of PCoA, but also mostly restricted to landraces. A separation of Middle Eastern and Balkan accessions was visible in cluster Ia. Cluster Ib contained mostly Southern European accessions.

The DArT markers based hierarchical clustering of wild *Secale* accessions (Figure 7) revealed that *S. sylvestre* samples were very divergent from the rest and formed a separate group. The remaining accessions formed two clusters. One of them comprised only *S. strictum* accessions, the second cluster consisted of *S. cereale* subspecies, both *S. vavilovii* samples and *S. strictum ssp. ciliatoglume*.

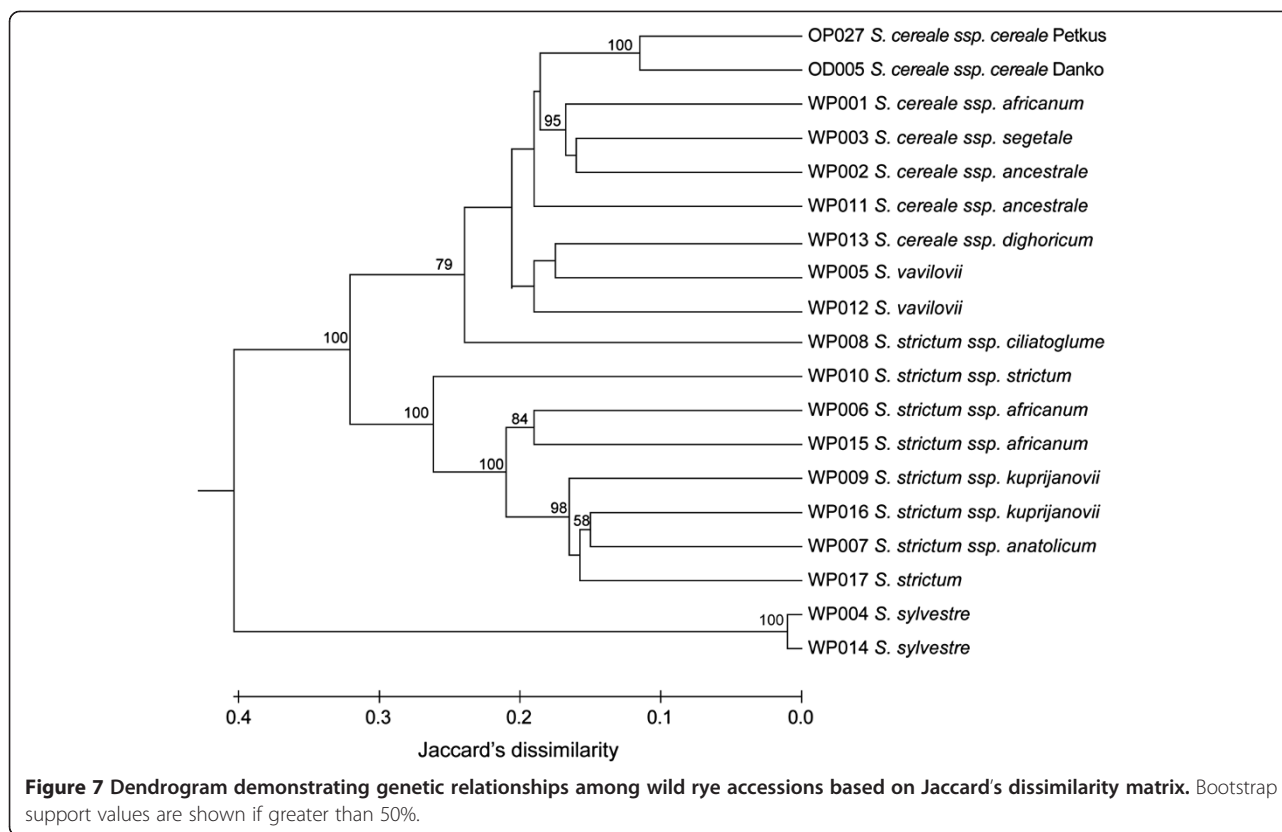
#### Genetic differentiation among accession sub-groups

AMOVA analysis of the model-based populations P1, P2 and P3 showed that there was a much greater proportion of variation within populations (75%) than among populations (25%,  $P < 0.001$ ). Pairwise  $\Phi_{PT}$  values indicated a high degree of differentiation between the populations P1 and P3 (0.21) and a very high genetic differentiation between population P2 and populations P1 and P3 (0.36 and 0.53, respectively,  $P < 0.001$ ). AMOVA analyses were also done to assess genetic differentiation among accessions grouped according to geographic region of origin and germplasm source, and improvement status. Again, a much greater proportion of variation within population than among populations was found: 90% and 10%, respectively, ( $P < 0.001$ ), when accessions were grouped based on region of origin, and 87% and 13%, ( $P < 0.001$ ), when source and improvement status of accessions were used as criterion for grouping. Based on pairwise  $\Phi_{PT}$  values, there was a substantial differentiation between wild accessions and the remaining accession sub-groups (Additional file 5: Table S3). Within cultivated ryes high genetic differentiation was observed between accessions from breeding companies and remaining four germplasm groups: landraces, cultivated materials, varieties from PAS BG and accessions from A. Lukaszewski's collection ( $\Phi_{PT}$  values ranging from 0.15 to 0.20). By contrast little genetic differentiation ( $\Phi_{PT}$  values below 0.05) was detected between landraces and cultivated materials, and between cultivated materials and varieties from PAS BG. When the accessions were grouped according to region of origin, the highest pairwise  $\Phi_{PT}$  values indicating high genetic differentiation were found between accessions from Middle East and accessions from tree European regions: Eastern Europe, Western Europe and Northern Europe (0.21, 0.18 and 0.16, respectively). Little or moderate genetic variation was found between accessions from remaining regions (Additional file 5: Table S3).



**Figure 6** Dendrogram demonstrating the genetic relationships among 379 rye accessions based on 1054 DArT markers, obtained using Neighbor Joining clustering from Jaccard's dissimilarity matrix. Panel **A**: accessions were labeled according to the STRUCTURE results; panel **B**: accessions were labeled according to the source and improvement status; panel **C**: accessions were labeled according to the geographic origin.





## Discussion

DNA markers rapidly became an indispensable tool of assessing genetic diversity contained in germplasm collections that supplements morphological evaluations. While in the early studies the laborious and time consuming procedures of detecting DNA variation allowed for sampling of only relatively limited numbers of accessions and loci, the development of high throughput genotyping methods, (fluorescence-based SSR detection on automated sequencers, and, in particular, highly parallel SNP genotyping assays) enabled a thorough characterization of whole germplasm collections [12,13,25,27,40-44].

The development of high throughput genotyping methods in rye has lagged behind that of other cereals and so far no large scale genetic diversity studies of *Secale* were conducted. The development of DArT genotyping array for rye opened the possibility for genome-wide genetic analyses in this crop [19]. In this study we applied 1054 genome-wide distributed DArT markers with defined chromosome location to assess genetic diversity and population structure in a collection of 379 rye accessions. To our best knowledge it is the most comprehensive study of genetic diversity in rye done until now.

In our study we analyzed bulked DNA samples, originating from 96 plants pro accession. This strategy allows to capture in one sample the genetic variability of a heterogenous accession, which will be of advantage in

future analyses of the assembled collection involving multiallelic markers. At the same time, however, this strategy equilibrates a part of genetic diversity of an accession, when, like in our study, dominant markers are used for genotyping. Nevertheless DArT markers performed well in providing the picture of genetic diversity in a large collection of rye germplasm. The combined data allowed us to distinguish all accessions. The average PIC (0.34) was found to be intermediate to that observed in studies of genetic diversity done using DArT markers in other species, such as cassava (0.42) [45], common wheat (0.40) [26], triticale (0.40) [25], barley (0.38) [46], *T. monococcum* (0.31) [28], *Lesquerella* (0.21) [12], sugar beet (0.28) [47] and *Asplenium* (0.21) [48]. The average PIC of rye DArT markers was thus lower than in autogamous crops, whereas open-pollinating species are generally expected to exhibit a higher level of polymorphism that self-pollinating ones [49]. This value was also slightly lower than mean PIC for the R genome (0.38) reported for DArT markers in triticale by Alheit et al. [25]. This lower than expected mean PIC value can result from certain limitations of DArT markers in analyses of heterozygous and heterogeneous samples [19], that will be discussed in more detail later, and from germplasm choice [45]. Whereas the accessions chosen for this study were selected to represent maximum diversity, most of them were not previously subjected to extensive molecular or

morphometric analyses. Hence the choice of accessions was made mostly based on the available pedigree information and region of origin. Genotyping with genome-wide DArT markers revealed a limited diversity in certain germplasm subgroups, such as breeding strains. This lower diversity, manifested by prevalence of one marker score (1 or 0) in the respective sub-groups, contributed to lower overall PIC values for a subset of DArT markers and, in effect, to a lower mean PIC value. By contrast, mean PIC value calculated based only on markers scores obtained in landraces, the most diverse germplasm subgroup of *S. cereale* ssp. *cer-eale* analyzed in our study, was higher (0.38). Similarly, in cassava (which is also an allogamous and highly heterozygous crop), a high mean PIC value of 0.42 was obtained when 35% of analyzed samples were wild relatives of cassava, while an approximately 27% lower mean PIC value (0.27) was achieved in an experiment, where wild relatives constituted only 7.9% of the accessions [45]. On the other hand, as noted by Badea *et al.* [50], the lower PIC values of DArT markers lead to a more defined genetic structure of accessions from distant geographic regions or genetic origin and thus can be of advantage in cluster analyses.

The range and distribution of PIC values was similar to those observed in other studies [28,45]. However, in general, a relatively higher proportion of markers with PIC values above 0,4 was observed. Average PIC values varied between chromosomes. Considerable differences in chromosome specific average PIC values were also reported by Alheit *et al.*, who performed genome-wide evaluation of genetic diversity in triticale using DArT markers [25]. Similarly to our results the highest average PIC occurred in winter triticale for 1R, but the lowest chromosome specific average PIC value, was observed for 3R. However, due to limited number of rye founder lines used for the establishment of triticale its R genome may not reflect accurately the genetic variation of the rye genome.

The GS values obtained in our study showed a greater range and a lower average than in previous studies on genetic diversity in rye [5,6], which was probably caused by inclusion of diverse wild and primitive accessions in the collection analyzed in our study. Ma *et al.* [6] assessed genetic diversity in 42 spring and winter rye varieties using RAPD markers. The obtained GS values ranged from 0.435 to 0.964, the average GS value was not reported. Shang *et al.* [5] analyzed separately 30 wild *Secale* accessions and 47 cultivated ryes with 24 SSR markers. GS values obtained in their study ranged from 0.326 to 0.932 (0.633 on average) and from 0.622 to 0.921 (0.773 on average) in wild and cultivated accessions, respectively. This difference in average GS between wild and cultivated *Secale* accessions (with the lower average GS in wild ryes) is in agreement with our results on average GS values in different germplasm subgroups. Large scale population

genomic analyses have the potential to provide insight into evolution of crop plants and their genomes, since the genomic regions that have been targeted by selection during crop evolution are expected to exhibit characteristic changes in levels of polymorphism [41]. The high average GS value on 6R that occurred in all sub-groups of cultivated germplasm, but was not observed in wild accessions might thus indicate that this chromosome contains regions that were subjected to strong selection pressure during domestication. Similarly, in the case of varieties supplied by breeding companies, the relatively high GS averages on 1R and 4R, that deviated from the general pattern of differences in chromosome specific average GS values observed within germplasm groups might reflect the presence of genomic regions with limited polymorphism, possibly resulting from selection for QTLs located therein [25] and controlling adaptive traits and quality characters relevant for cultivation in Central and Northern Europe. Chromosome 1R is known to contain genes controlling resistance to diseases and insects, improving adaptation and increasing yield [51], while chromosome 4R was found to harbour QTLs for alpha-amylase activity, preharvest sprouting, kernel thickness, heading time, chlorophyll content in leaves, and flag leaf length [52]. The relatively high GS averages on 1R and 4R that occurred in varieties supplied by breeding companies could also result from the locations of self fertility mutations and fertility restoration genes deployed in hybrid rye breeding. According to Lundqvist self incompatibility in rye is controlled by two multiallelic loci – S and Z. and mutations at these loci lead to self fertility [53,54]. The two loci S and Z were subsequently mapped on chromosomes 1R and 2R, respectively, and additional self fertility genes on chromosomes 4R, 5R and 6R were also described [55]. It can be hypothesized that self-fertility genes that are captured in commercial hybrid rye breeding programs and allow for development of inbred parental lines are located in chromosomes 1R and/or 4R. Effective fertility restorer genes, originating from Iranian and Argentinean germplasm and currently used in hybrid rye breeding, are located on chromosome 4RL [56,57].

In this work we used simultaneously three methods - Bayesian clustering, PCoA and Neighbor Joining clustering to obtain a picture of genetic relationship in the collection of 379 rye accessions. Despite minor differences, the results were largely consistent. First of all we found that genetic diversity within germplasm sub-groups consisting of improved rye accessions - obtained from breeding companies, as well as historic varieties and cultivated materials from PAS BG - is relatively narrow. It is possible that the genotyping of bulked DNA samples with dominant DArT markers resulted in lower estimates of genetic diversity and further cost- and labor-intensive analyses based on single plants within accessions with a method, which

enables detection of heterozygosity would be needed to unequivocally address the question of the relatively high genetic similarity of rye varieties. It is, however, noteworthy that similar results regarding rye varieties were obtained earlier using RAPD markers [6,58], AFLP markers [17], SSR markers [5], in analyses of organellar genome diversity [18] and in a preliminary small scale study on utility of DArT markers for assessment of genetic diversity in rye [19]. It was postulated that this relatively low genetic diversity could indicate common genetic background of improved rye accessions from diverse regions and can be attributed to an extensive germplasm exchange [6,18]. Interestingly, recently this supposition was partially confirmed by SSR analyses of single  $S_0$  plants from five Eastern European open-pollinated varieties. A close relationship between varieties was revealed, pointing to common ancestors and/or gene flow between the varieties [59]. The relatively high similarity of rye varieties observed in our study, and hypothesized to be indicative of common genetic background and germplasm exchange, is also in agreement with the available information on the rye breeding history and the ancestry of rye varieties. One of the leading rye varieties in the twentieth century was Petkus and many of the open pollinated rye varieties worldwide are selections from Petkus or include Petkus in their ancestry [59-61]. Eastern European population varieties were crossed with each other in the past, and Petkus has a large contribution in all of them [59], while in Germany, before the establishment of hybrid rye breeding, all leading population varieties belonged either to the Petkus or to the Carsten gene pool [56]. Moreover introgressions of foreign material were common in Eastern European breeding programs [62].

On the other hand accessions from breeding companies turned out to be distinct from the remaining accessions, both primitive and improved. This could reflect the adaptation to local conditions of Central and Northern Europe and to requirements of modern agriculture. However, in our opinion, it is at least partly attributable to different reproduction methods and to the accuracy of the procedures used by breeders and during genebank maintenance. We are inclined to this supposition based on the results obtained for the two samples of Dankowskie Nowe variety originating from different sources - the relatively low  $GS$  value and assignment to different clusters in PCoA and Neighbor Joining analysis. While accessions reproduced by breeders are subjected to conservative breeding, the reproduction of rye accessions in genebanks is carried out on 1 m<sup>2</sup> field plots and before flowering the whole plot area is covered with a 2-m-high metal frame covered with pollen-proof tissue [63]. Since in our study only one variety was represented by two independent samples of different origin no definitive conclusions can be drawn and additional comparative studies are needed. Nevertheless, changes in allele frequencies after reproduction cycles

were discovered by Chebotar *et al.* [7] in rye accessions from the IPK Gatersleben genebank. They also stressed the importance of making extended efforts in order to sustain the genetic identity of open pollinating rye accessions during *ex situ* maintenance, such as using plots large enough for growing a number of plants sufficient to cover the whole diversity of the populations, omitting harvest when a large proportion of plants is lost during a regeneration cycle and dividing the resources into base and active collections.

Another finding concerning accessions supplied by breeding companies is a relatively high proportion of the inferred ancestry derived from P2, that occurred in hybrid varieties. The unique features of hybrid varieties, differentiating them from open pollinated rye varieties and originating from exotic accessions are cytoplasmic male sterility and fertility restoration genes. In our study we used only DArT markers with defined chromosomal location. Therefore only nuclear DNA was analyzed. Fertility restoration genes can account for some proportion of P2 alleles, however, since after only four rounds of backcrossing, the percentage of the donor genome is reduced to 3.125%, it is unlikely that restorer genes alone account for the whole proportion of P2 alleles observed in hybrid varieties. Hence, further analyses would be required to identify the cause of the observed allelic composition.

Another interesting issue is the relation of the obtained results to the available information on the genetic background of hybrid rye varieties. In Germany, based on diallel variety crosses analyzed by Hepting [61], Petkus and Carsten gene pools were chosen for the development of seed- and pollinator lines, respectively, for hybrid rye breeding [56]. In our study, based on STRUCTURE analysis, Petkus together with hybrid varieties, was assigned to population P3, and Carsten's Roggen was classified as P1P3. While this results indicate close relationship of hybrid varieties to Petkus and Carsten, they do not mirror the expected high genetic divergence between these two varieties, which causes high hybrid performance. Moreover, the Kustro variety, which is a selection from Petkus, contrary to expectations, was not assigned to population P3 alongside Petkus, but, similarly to Carsten, was classified as admixture P1P3. These results should be, however, treated with a certain caution, since, as mentioned before, allelic compositions of these accessions could have changed during gene bank maintenance, and thus can be not equivalent to allelic composition of Carsten and Petkus populations maintained and exploited by breeders. On the other hand our results seem to be in agreement with the information concerning the existence of Eastern European gene pools, which proved to be heterotic counterparts to both Petkus and Carsten gene pools [56], since several of Eastern European varieties analyzed in our study were assigned to population P1.

A very striking result was obtained for the Gonello F<sub>1</sub> variety. This variety is advertised as having the highest frost tolerance. Thus it could be hypothesized that its divergence from other currently registered varieties included in our study is a result of using a very distinct genepool as a source of cold resistance genes. Unfortunately such supposition is not supported by the available information on the genetic background of the hybrid varieties from KWS Lochow: Gonello F<sub>1</sub> ((Lo115-P × Lo133-N) × LSR83 (Lo310 × Lo312)) is closely related to Guttino F<sub>1</sub> ((Lo115-P × Lo133-N) × LSR81 (Lo298 × Lo312)), and the male line Lo310 is included in the pedigree of Palazzo F<sub>1</sub>. Both Guttino F<sub>1</sub> and Palazzo F<sub>1</sub> were analyzed in our study and clustered closely with other varieties from breeding companies. Hence, additional studies involving Gonello F<sub>1</sub> are needed to verify the obtained results and to resolve this issue.

Our research showed also that rye landraces represented fairly large diversity. This result is consistent with expectations [10] and also in agreement with the result of a small scale (12 rye accessions) study done using RAPD markers [58]. The fact that the landraces were distant from material obtained from breeders and a subset of landraces occupied a distinct diversity space, not overlapping with other accessions (as shown in PCoA) demonstrates that there is a great genetic potential for detection of unexplored alleles to broaden the genetic diversity in current breeding programs. Distinctness of landraces from contemporary varieties was also observed in tomato [64].

AMOVA results obtained in our study indicate that there is a higher amount of genetic diversity within populations than among population, which is in agreement both with expectations for an out-crossing species [56], and with previous results obtained in genetic diversity analyses in rye using both codominant isoenzyme markers [60], and dominant RAPD markers [58].

In the case of wild *Secale* accessions, the outcome of cluster analysis separating the *Secale* species into three groups – *sylvestre*, *strictum*, and *cereale/vavilovii* is in agreement with the results of AFLP and SSR based analyses of phylogenetic relationships in the genus. It is also consistent with the revision of taxonomical classification of *Secale* done by Frederiksen and Petersen [65], who recognized only three species within the genus *Secale*: *S. sylvestre*, *S. strictum* and *S. cereale*, and included *S. vavilovii* in *S. cereale*. Similarly, Kobyljanskij [66] classified *S. vavilovii* as a subspecies of *S. cereale*. Furthermore, we observed that wild accessions clustered together at almost the same similarity level (about 40%) like in a previous SSR-based study of phylogenetic relationships in the genus *Secale*. It might suggest that the set of biallelic DArT markers used here has a similar discriminatory power as multiallelic SSRs. The fact that *S. strictum* spp.

*ciliatoglume* did not cluster with other *strictum* accessions is consistent with the AFLP based results of Chikmawati *et al.* [14], with results of SSCP and SSR analyses done in our laboratory (Targonska *et al.* in preparation), and with the observations of Frederiksen and Petersen, who based on morphometrical analyses suggested that *S. ciliatoglume* should be given an intra-specific rank [67].

Overall we observed rather poor correlation between clustering of rye accessions and geographic origin and lack thereof in rye varieties and cultivated materials. Similar results were obtained earlier in rye using organellar genome analyses [18], SSR markers [5] and AFLP markers [17] and can be explained by common genetic background of the accessions in question resulting from germplasm exchange. Conversely, Ma *et al.* [6] found a good correspondence of clustering to geographical location while analyzing 40 rye varieties mostly of European, North American and Chinese origin. However, they observed that temporal isolation has influenced the genetic diversity of rye more than geographical isolation, since winter and spring varieties formed separate clusters. The growth habit was found to be the primary determinant of population structure also in triticale [25] and barley [68]. This issue is however beyond the scope of our study, since the vast majority of the accession were winter types. Our results indicate that the population structure of the assembled rye collection was mostly determined by the accessions source and improvement status and that the significant influence of accessions source seems to be at least partly attributable to the influence of reproduction methods on the genetic identity of rye accessions. This is an interesting and rather unexpected finding and it would be worth while to verify it in a follow up study, involving the same set of markers and comparative analyses of fairly numerous accessions sets from several rye genebanks, as well as modern rye varieties bred for cultivation in geographic regions other than Northern/Central Europe. Such analyses would also allow to assess, if the rye collection assembled for this research is representative of the worldwide diversity.

Finally, a question might be posed to what extent the obtained results reflect the factual genetic relationships in the assembled rye collection. DArT markers are biallelic, which imposes certain limitations with the respect to the range of analyses that can be performed [29]. In addition, DArT markers are scored as dominant markers, in general, and it was done so also in this study. These features of DArT markers are of particular importance in the presented research, since most of the rye accessions analyzed are likely to be highly heterozygous and heterogeneous and to vary in allelic frequencies for a number of markers. Therefore it could be expected that the accuracy of the obtained results is compromised to some extent. The



unquestionable advantage of DArT markers, however, that in our opinion counterbalances the potential drawbacks of assessing genetic variation in an outcrossing crop using this method, is the genome coverage achieved – in this study, more than one hundred DArT markers were available for each chromosome. Moreover, it was demonstrated by Laido *et al.* [27] that in tetraploid wheat the estimates of genetic diversity between genotypes obtained independently using SSR and DArT were highly similar and that dendrograms obtained with the two types of marker data were highly congruent. A comparison of SSR and DArT markers in assessing genetic diversity of 436 heterozygous cassava accessions also revealed that both markers systems generated similar clustering patterns [69], while in an empirical evaluation of DArT, SNP and SSR markers for genotyping, clustering and assigning sugar beet hybrid varieties into populations it was found that the same accuracy of the results can be achieved when 4.9-13.3 times more DArT markers than SSR markers are used [47]. Therefore, in reverse, it could be supposed, that our data is comparable with results from almost 80 SSR ( $1054 / 13.3 = 79.25$ ). By contrast, so far only much smaller numbers of SSR markers were used in genetic diversity analyses of rye: 24 [5], 30 [59], 38 [15], 10 [7]. Consistence of DArT based clustering with taxonomical classification was reported in a study of *Physaria* and *Paysonia* accessions [12] and in tetraploid wheats [27]. Our results obtained in cluster analysis of wild *Secale* germplasm also indicate that DArTs markers reflect phylogenetic relationships faithfully, even when applied to analyses of heterozygous and heterogenous samples, such as out-crossing rye accessions. Therefore we think that DArT based analyses provide a realistic picture of the genetic diversity and population structure present in the 379 rye accessions. Apart from high discriminatory power, faithful reflection of phylogenetic relationships, extensive genome coverage and the availability of information on chromosomal location, the use of DArT markers offers advantages also in terms of transferability, allowing for cross studies comparisons, and much lower costs compared to the other SNP genotyping platform currently available for rye – Illumina iSelect HD Custom BeadChip [70]. Apart from lower cost-effectiveness, diversity studies based on Illumina iSelect platform (especially studies involving highly diverged germplasm pools) are likely to be affected by ascertainment bias caused by the SNP discovery process in which limited number of individuals from selected populations are used. It was shown that this ascertainment bias can distort measures of population divergence [71]. Additionally, the ongoing sequencing of DArT markers from rye genotyping panel will soon enable more in depth analyses, such as candidate gene identification. DArTs can be thus recommended as a method of choice for rye germplasm characterization, as well as for association mapping studies, where the

knowledge of population structure and genetic relationships is of critical importance [41,42].

## Conclusions

We conducted the largest and most comprehensive study of genetic diversity in rye to date and revealed a relatively limited diversity in improved rye accessions, both historical and contemporary, as well as lack of correlation between clustering of improved accessions and geographic origin, suggesting common genetic background of rye accessions from diverse geographic regions that probably results from extensive germplasm exchange. Genetic distinctness of contemporary varieties from the rest of accession that was also observed indicates a great genetic potential of *ex situ* collections for broadening the genetic diversity in current rye breeding programs.

One of the more interesting findings of our study is the identification of accession source as a major determinant of population structure. Our result point to possible influence of reproduction methods on the observed diversity pattern. Further studies involving accessions from different genebanks would be desirable to verify our findings concerning the source of population structure in rye.

We also revealed differences in distribution of genetic polymorphism among rye chromosomes that could be indicative of genome regions targeted by selection during domestication and improvement programs.

Obtained data indicate that DArT markers provide a realistic picture of the genetic diversity and population structure present in the collection of 379 rye accessions and are currently a method of choice for rye germplasm characterization and association mapping studies.

## Additional files

**Additional file 1: Table S1.** Excel spreadsheet containing detailed information on rye used in the study and their genotypic DArT marker data.

**Additional file 2: Table S2.** Excel spreadsheet containing mean PIC values for groups of accessions created according to source and type of germplasm.

**Additional file 3: Figure S1.** (Portable Document File). Delta K values for K values (number of populations assumed) ranging from 1 to 20.

**Additional file 4: Figure S2.** (Portable Document File). Principal coordinate analysis of 379 rye accessions based on 1054 DArT markers with defined chromosomal location. Accessions were labeled according to the geographic origin of accessions: panel a: only cultivated materials and varieties from PAS BG are shown; panel b: only landraces are shown.

**Additional file 5: Table S3.** Excel spreadsheet containing results of AMOVA and pairwise population  $\Phi_{PT}$  analyses for accession source and improvement status, and for region of origin.

## Competing interests

AK is an employee of Diversity Arrays Technology Pty. Ltd., which offers genome profiling service using the technology described in this report. This fact, however, has not interfered whatsoever with the full, objective, transparent and unbiased presentation of the research results described in the manuscript nor alters the authors' adherence to all the BMC policies on data and material release.

#### Authors' contributions

HBB conceived and designed the experiments, wrote the manuscript. HBB, MT performed the experiments. HBB, AK, LB analyzed the data. HBB, AK, LB, MT, MRT contributed reagents/materials/analysis tools. HBB, MT, MRT contributed to the selection and preparation of germplasm. All authors read and approved the final version of the manuscript.

#### Acknowledgments

We would like to thank Prof J. Puchalski (Polish Academy of Sciences Botanical Garden-Center for Biological Diversity Conservation in Powsin, Poland), Prof. A. Lukaszewski (University of California, Riverside, USA), Dr. Z. Banaszak and Mr. Z. Brukwiński (Danko Plant Breeders Ltd) and Dr. P. Dopierała (KWS Lochow Poland Ltd) for supplying the seeds of accessions used in the study. Additionally, we would like to thank Dr. P. Dopierała for providing information on genetic background of Gonello F1, Palazzo F1 and Guttino F1 varieties. We also thank Diversity Arrays Technology Pty. Ltd team members for their help and friendly support during execution of this project. This study was funded by the Ministry of Science and Higher Education grant N N310 080139.

#### Author details

<sup>1</sup>Department of Plant Genetics, Breeding and Biotechnology, Faculty of Horticulture, Biotechnology and Landscape Architecture, Warsaw University of Life Sciences, Warsaw, Poland. <sup>2</sup>Department of Silviculture, Faculty of Forestry, Warsaw University of Life Sciences, Warsaw, Poland. <sup>3</sup>Diversity Arrays Technology Pty. Ltd, Yarralumla ACT, Australia.

Received: 13 March 2014 Accepted: 27 June 2014

Published: 1 August 2014

#### References

- Holma R, Hongisto S, Saxelin M, Korpela R: Constipation is relieved more by rye bread than wheat bread or laxatives without increased adverse gastrointestinal effects. *J Nutr* 2010, **140**:534–541.
- Gråsten SM, Juntunen KS, Poutanen KS, Gylling HK, Miettinen TA, Mykka HM: Rye bread improves bowel function and decreases the concentrations of some compounds that are putative colon cancer risk markers in middle-aged women and men. *J Nutr* 2000, **130**:2215–2221.
- Baum M, Appels R: The cytogenetic and molecular architecture of chromosome 1R—one of the most widely utilized sources of alien chromatin in wheat varieties. *Chromosoma* 1991, **101**:1–10.
- Lukaszewski AJ: Frequency of 1RS.1AL and 1RS.1BL translocations in United States wheats. *Crop Sci* 1990, **30**:1151.
- Shang H, Wei Y, Wang X, Zheng Y: Genetic diversity and phylogenetic relationships in the rye genus *Secale* L. ( rye ) based on *Secale cereale* microsatellite markers. *Genet Mol Biol* 2006, **29**:685–691.
- Ma R, Yli-Mattila T, Pulli S: Phylogenetic relationships among genotypes of worldwide collection of spring and winter ryes (*Secale cereale* L.) determined by RAPD-PCR markers. *Hereditas* 2004, **140**:210–221.
- Chebotař S, Röder MS, Korzun V, Saal B, Weber WE, Börner A: Molecular studies on genetic integrity of open-pollinating species rye (*Secale cereale* L.) after long-term genebank maintenance. *Theor Appl Genet* 2003, **107**:1469–1476.
- Stojalowski S, Jaciubek M, Masojć P: Rye SCAR markers for male fertility restoration in the P cytoplasm are also applicable to marker-assisted selection in the C cytoplasm. *J Appl Genet* 2005, **46**:371–373.
- Falke KC, Susić Z, Hackauf B, Korzun V, Schondelmaier J, Wilde P, Wehling P, Wortmann H, Mank R, Rouppe Van Der Voort J, Maurer HP, Miedaner T, Geiger HH: Establishment of introgression libraries in hybrid rye (*Secale cereale* L.) from an Iranian primitive accession as a new tool for rye breeding and genomics. *Theor Appl Genet* 2008, **117**:641–652.
- McCouch S: Diversifying selection in plant breeding. *PLoS Biol* 2004, **2**:e347.
- FAO: *The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture*. Rome, Italy: FAO; 2010.
- Cruz VMV, Kilian A, Dierig DA: Development of DArT marker platforms and genetic diversity assessment of the U.S. collection of the new oilseed crop *lesquerella* and related species. *PLoS One* 2013, **8**:e64062.
- Lv J, Qi J, Shi Q, Shen D, Zhang S, Shao G, Li H, Sun Z, Weng Y, Shang Y, Gu X, Li X, Zhu X, Zhang J, van Treuren R, van Dooijeweert W, Zhang Z, Huang S: Genetic diversity and population structure of cucumber (*Cucumis sativus* L.). *PLoS One* 2012, **7**:e46919.
- Chikmawati T, Skovmand B, Gustafson JP: Phylogenetic relationships among *Secale* species revealed by amplified fragment length polymorphisms. *Genome* 2005, **48**:792–801.
- Bolibok H, Rakoczy-Trojanowska M, Hromada A, Pietrzykowski R: Efficiency of different PCR-based marker systems in assessing genetic diversity among winter rye (*Secale cereale* L.) inbred lines. *Euphytica* 2005, **146**:109–116.
- Bolibok-Bragoszewska H, Zabierzewska N, Hromada-Judycka A, Krzewska L: SSAP markers based on a novel Ty1-copia like element are a powerful tool for the assessment of genetic diversity in rye (*Secale cereale* L.) inbred lines. *Cereal Res Commun* 2012, **40**:204–209.
- Chikmawati T, Miftahudin M, Skovmand B, Gustafson JP: Amplified fragment length polymorphism-based genetic diversity among cultivated and weedy rye (*Secale cereale* L.) accessions. *Genet Resour Crop Evol* 2012, **59**:1743–1752.
- Isik Z, Parmaksiz I, Coruh C, Geylan-Su YS, Cebeci O, Beecher B, Budak H: Organellar genome analysis of rye (*Secale cereale*) representing diverse geographic regions. *Genome* 2007, **50**:724–734.
- Bolibok-Bragoszewska H, Heller-Uszyńska K, Wenzl P, Uszyński G, Kilian A, Rakoczy-Trojanowska M: DArT markers for the rye genome - genetic diversity and mapping. *BMC Genomics* 2009, **10**:578.
- Milczarski P, Bolibok-Bragoszewska H, Myśków B, Stojalowski S, Heller-Uszyńska K, Górala M, Bragoszewska P, Uszyński G, Kilian A, Rakoczy-Trojanowska M: A high density consensus map of rye (*Secale cereale* L.) based on DArT markers. *PLoS One* 2011, **6**:e28495.
- Jaccoud D, Peng K, Feinstein D, Kilian A: Diversity Arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res* 2001, **29**:e25.
- Resende MDV, Resende MFR, Sansaloni CP, Petrollo CD, Missaggi AA, Aguiar AM, Abad JM, Takahashi EK, Rosado AM, Faria DA, Pappas GJ, Kilian A, Grattapaglia D: Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol* 2012, **194**:116–128.
- Mace ES, Xia L, Jordan DR, Halloran K, Parh DK, Huttner E, Wenzl P, Kilian A: DArT markers: diversity analyses and mapping in *Sorghum bicolor*. *BMC Genomics* 2008, **9**:26.
- Cappa EP, El-Kassaby YA, Garcia MN, Acuña C, Borralho NMG, Grattapaglia D: Marcucci Poltri SN: Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in *Eucalyptus globulus*. *PLoS One* 2013, **8**:e81267.
- Alheit KV, Maurer HP, Reif JC, Tucker MR, Hahn V, Weissmann EA, Würschum T: Genome-wide evaluation of genetic diversity and linkage disequilibrium in winter and spring triticale ( $\times$  *Triticosecale* Wittmack). *BMC Genomics* 2012, **13**:235.
- Zhang L, Liu D, Guo X, Yang W, Sun J, Wang D, Sourdille P, Zhang A: Investigation of genetic diversity and population structure of common wheat cultivars in northern China using DArT markers. *BMC Genet* 2011, **12**:42.
- Laidò G, Mangini G, Taranto F, Gadaleta A, Blanco A, Cattivelli L, Marone D, Mastrangelo AM, Papa R, De Vita P: Genetic diversity and population structure of tetraploid wheats (*Triticum turgidum* L.) estimated by SSR, DArT and pedigree data. *PLoS One* 2013, **8**:e67280.
- Jing H-C, Bayon C, Kanyuka K, Berry S, Wenzl P, Huttner E, Kilian A, Hammond-Kosack KE: DArT markers: diversity analyses, genomes comparison, mapping and integration with SSR markers in *Triticum monococcum*. *BMC Genomics* 2009, **10**:458.
- White J, Law JR, MacKay I, Chalmers KJ, Smith JSC, Kilian A, Powell W: The genetic diversity of UK, US and Australian cultivars of *Triticum aestivum* measured by DArT markers and considered by genome. *Theor Appl Genet* 2008, **116**:439–453.
- Jaccard P: Nouvelles recherches sur la distribution florale. *Bull Soc Vaudoise Sc Nat* 1908, **44**:223–270.
- Rohlf FJ: *NTSYS-pc: numerical taxonomy and multivariate analysis system*; 2000.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011, **28**:2731–2739.
- Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. *Genetics* 2000, **155**:945–959.

34. Evanno G, Regnaut S, Goudet J: Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 2005, **14**:2611–2620.
35. Earl DA, von Holdt BM: STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 2011, **4**:359–361.
36. Jakobsson M, Rosenberg NA: CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 2007, **23**:1801–1806.
37. Rosenberg NA: Distruct: a program for the graphical display of population structure. *Mol Ecol Notes* 2003, **4**:137–138.
38. Peakall R, Smouse PE: Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 2006, **6**:288–295.
39. Peakall R, Smouse PE: GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 2012, **28**:2537–2539.
40. Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troglio M, Myles S, Martinez-Zapater JM, Zyprian E, Moreira FM, Grandi MS: Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biol* 2013, **13**:39.
41. Mandel JR, Nambesani S, Bowers JE, Marek LF, Ebert D, Rieseberg LH, Knapp SJ, Burke JM: Association mapping and the genomic consequences of selection in sunflower. *PLoS Genet* 2013, **9**:e1003378.
42. Soto-Cerda BJ, Diederichsen A, Ragupathy R, Cloutier S: Genetic characterization of a core collection of flax (*Linum usitatissimum* L.) suitable for association mapping studies and evidence of divergent selection between fiber and linseed types. *BMC Plant Biol* 2013, **13**:78.
43. De Jesus ON, Silva SDOE, Amorim EP, Ferreira CF, de Campos JMS, Silva GDG, Figueira A: Genetic diversity and population structure of *Musa* accessions in ex situ conservation. *BMC Plant Biol* 2013, **13**:41.
44. Blair MW, Soler A, Cortés AJ: Diversification and population structure in common beans (*Phaseolus vulgaris* L.). *PLoS One* 2012, **7**:e49488.
45. Xia L, Peng K, Yang S, Wenzl P, de Vicente MC, Fregene M, Kilian A: DArT for high-throughput genotyping of Cassava (*Manihot esculenta*) and its wild relatives. *Theor Appl Genet* 2005, **110**:1092–1098.
46. Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A: Diversity Arrays Technology ( DArT ) for whole-genome profiling of barley. *Proc Natl Acad Sci U S A* 2004, **101**:9915–9920.
47. Van Hintum TJL, Eujayl I, Simko I: Empirical evaluation of DArT, SNP, and SSR marker-systems for genotyping, clustering, and assigning sugar beet hybrid varieties into populations. *Plant Sci* 2012, **184**:54–62.
48. James KE, Schneider H, Ansell SW, Evers M, Robba L, Uszynski G, Pedersen N, Newton AE, Russell SJ, Vogel JC, Kilian A: Diversity arrays technology (DArT) for pan-genomic evolutionary studies of non-model organisms. *PLoS One* 2008, **3**:e1682.
49. Rafalski A: Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol* 2002, **5**:94–100.
50. Badea A, Eudes F, Salmon D, Tuvevsson S, Vrolijk A, Larsson C-T, Caig V, Huttner E, Kilian A, Laroche A: Development and assessment of DArT markers in triticale. *Theor Appl Genet* 2011, **122**:1547–1560.
51. Simkova H, Safar J, Suchankova P, Kovarova P, Bartos J, Kubalaková M, Janda J, Cihalikova J, Mago R, Lelley T, Dolezel J: A novel resource for genomics of Triticeae: BAC library specific for the short arm of rye (*Secale cereale* L.) chromosome 1R (1RS). *BMC Genomics* 2008, **9**:237.
52. Schlegel R, Korzun V: Genes, markers and linkage data of rye (*Secale cereale* L.). 7th updated inventory, 09.13. [http://www.rye-gene-map.de]
53. Lundqvist A: Self-incompatibility rye I. Genetic control in the diploid. *Hereditas* 1956, **42**:293–348.
54. Lundqvist A: The origin of self-compatibility in rye. *Hereditas* 1960, **46**:1–19.
55. Voytkov AV, Korzun V, Borner A: Mapping of three self-fertility mutations in rye (*Secale cereale* L.) using RFLP, isozyme and morphological markers. *Theor Appl Genet* 1997, **97**:147–153.
56. Geiger HH, Miedaner T: Rye breeding. In *Cereals (handbook of plant breeding, Vol 3)*. 1st edition. Edited by Carena MJ. New York, NY: Springer US; 2009:157–181.
57. Hackauf B, Korzun V, Wortmann H, Wilde P, Wehling P: Development of conserved ortholog set markers linked to the restorer gene Rfp1 in rye. *Mol Breed* 2012, **30**:1507–1518.
58. Persson K, Díaz O, von Bothmer R: Extent and patterns of RAPD variation in landraces and cultivars of rye (*Secale cereale* L.) from northern Europe. *Hereditas* 2001, **134**:237–243.
59. Fischer S, Melchinger AE, Korzun V, Wilde P, Schmiedchen B, Möhring J, Piepho H-P, Dhillon BS, Würschum T, Reif JC: Molecular marker assisted broadening of the Central European heterotic groups in rye with Eastern European germplasm. *Theor Appl Genet* 2010, **120**:291–299.
60. Persson K, Von Bothmer R: Genetic diversity amongst landraces of rye (*Secale cereale* L.) from northern Europe. *Hereditas* 2002, **136**:29–38.
61. Hepting L: Analyse eines 7x7-Sortendiallels zur Ermittlung geeigneten Ausgangsmaterials für die Hybridzüchtung bei Roggen. *Zeitschrift für Pflanzenzüchtung* 1978, **80**:188–197.
62. Li Y, Haseneyer G, Schön C-C, Ankerst D, Korzun V, Wilde P, Bauer E: High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. *BMC Plant Biol* 2011, **11**:6.
63. Schlegel RH: *Rye: Genetics, Breeding, and Cultivation*. Boca Raton, FL: CRC Press; 2013:387.
64. Corrado G, Piffanelli P, Caramante M, Coppola M, Rao R: SNP genotyping reveals genetic diversity between cultivated landraces and contemporary varieties of tomato. *BMC Genomics* 2013, **14**:835.
65. Frederiksen S, Petersen G: A taxonomic revision of *Secale* (Triticeae, Poaceae). *Nord J Bot* 1998, **18**:399–420.
66. Kobyljanskij VD: The system of the genus *Secale* L. *Trudy Prikl Bot* 1983, **79**:24–38.
67. Petersen SG: Morphometrical analyses of *Secale* (Triticeae, Poaceae). *Nord J Bot* 1997, **17**:185–198.
68. Rostoks N, Mudie S, Cardle L, Ramsay L, Booth A, Svensson JT, Wanamaker SI, Walia H, Rodriguez EM, Hedley PE, Liu H, Morris J, Close TJ, Marshall DF, Waugh R: Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. *Mol Genet Genomics* 2005, **274**:515–527.
69. Hurtado P, Olsen KM, Buitrago C, Ospina C, Marin J, Duque M, de Vicente C, Wongtiem P, Wenzl P, Killian A, Adeleke M, Fregene M: Comparison of simple sequence repeat (SSR) and diversity array technology (DArT) markers for assessing genetic diversity in cassava (*Manihot esculenta* Crantz). *Plant Genet Resour Charact Util* 2008, **6**:208–214.
70. Haseneyer G, Schmutzer T, Seidel M, Zhou R, Mascher M, Schön C-C, Taudien S, Scholz U, Stein N, Mayer KF, Bauer E: From RNA-seq to large-scale genotyping - genomics resources for rye (*Secale cereale* L.). *BMC Plant Biol* 2011, **11**:131.
71. Albrechtsen A, Nielsen FC, Nielsen R: Ascertainment biases in SNP chips affect measures of population divergence. *Mol Biol Evol* 2010, **27**:2534–2547.

doi:10.1186/1471-2229-14-184

Cite this article as: Bolibok-Brągoszewska et al.: Genome-wide characterization of genetic diversity and population structure in *Secale*. *BMC Plant Biology* 2014 **14**:184.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

