



Genetic Distinctiveness of Rye *In situ* Accessions from Portugal Unveils a New Hotspot of Unexplored Genetic Resources

Filipa Monteiro^{1*}, Patrícia Vidigal¹, André B. Barros², Ana Monteiro¹, Hugo R. Oliveira^{3,4} and Wanda Viegas¹

¹ Linking Landscape, Environment, Agriculture and Food, Instituto Superior de Agronomia, Universidade de Lisboa, Lisboa, Portugal, ² Colégio F3 Food, Farming and Forestry, Universidade de Lisboa, Lisboa, Portugal, ³ Plant Biology/Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Porto, Portugal, ⁴ Faculty of Life Sciences, Manchester Institute of Biotechnology, University of Manchester, Manchester, UK

OPEN ACCESS

Edited by:

Anna Maria Mastrangelo,
Centro di Ricerca per l'Orticoltura
(CRA), Italy

Reviewed by:

Manuela Matos,
Universidade de Trás-os-Montes e
Alto Douro, Portugal
Thomas Miedaner,
University of Hohenheim, Germany

*Correspondence:

Filipa Monteiro
fmonteiro@isa.ulisboa.pt

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 18 May 2016

Accepted: 18 August 2016

Published: 31 August 2016

Citation:

Monteiro F, Vidigal P, Barros AB,
Monteiro A, Oliveira HR and Viegas W
(2016) Genetic Distinctiveness of Rye
In situ Accessions from Portugal
Unveils a New Hotspot of Unexplored
Genetic Resources.
Front. Plant Sci. 7:1334.
doi: 10.3389/fpls.2016.01334

Rye (*Secale cereale* L.) is a cereal crop of major importance in many parts of Europe and rye breeders are presently very concerned with the restrict pool of rye genetic resources available. Such narrowing of rye genetic diversity results from the presence of “Petkus” pool in most modern rye varieties as well as “Petkus” × “Carsten” heterotic pool in hybrid rye breeding programs. Previous studies on rye’s genetic diversity revealed moreover a common genetic background on landraces (*ex situ*) and cultivars, regardless of breeding level or geographical origin. Thus evaluation of *in situ* populations is of utmost importance to unveil “on farm” diversity, which is largely undervalued. Here, we perform the first comprehensive assessment of rye’s genetic diversity and population structuring using cultivars, *ex situ* landraces along a comprehensive sampling of *in situ* accessions from Portugal, through a molecular-directed analysis using SSRs markers. Rye genetic diversity and population structure analysis does not present any geographical trend but disclosed marked differences between genetic backgrounds of *in situ* accessions and those of cultivars/*ex situ* collections. Such genetic distinctiveness of *in situ* accessions highlights their unexplored potential as new genetic resources, which can be used to boost rye breeding strategies and the production of new varieties. Overall, our study successfully demonstrates the high prospective impact of comparing genetic diversity and structure of cultivars, *ex situ*, and *in situ* samples in ascertaining the status of plant genetic resources (PGR).

Keywords: *Secale cereale*, microsatellite, *in situ* conservation, population structure, genetic pool

INTRODUCTION

Rye (*Secale cereale* L.) belongs to the Triticeae tribe, along with other economically important cereals such as wheat and barley. Much controversy about the taxonomy of the *Secale* genus remains, despite the large number of studies performed (e.g., Roshevitz, 1947; Delipavlov, 1962; Kobyljanskij, 1975; Sencer and Hawkes, 1980; Frederiksen and Petersen, 1998; Chikmawati et al., 2005). According to taxonomic system adopted by the American Germplasm Resources

Information Network (GRIN, <http://www.ars-grin.gov>), *Secale* comprises four species, namely *S. cereale* L., *S. sylvestre* Host, *S. vavilovii* Grossh., and *S. strictum* (Presl.) Presl. (syn. *S. montanum* Guss). Within *S. cereale*, eight subspecies are recognized: *afghanicum* (Vavilov) K. Hammer, *ancestrale* Zhuk., *dighoricum* Vavilov, forma unranked *rigidum* Antropov and Antropova, *segetale* Zhuk., *tetraploidum* Kobyl, *tsitsinii* Kobyl and *cereale* L. (the only cultivated). Within *S. strictum*, there are five subspecies: *africanum* (Stapf) K. Hammer, *anatolicum* (Boiss.) K. Hammer, *ciliatoglume* (Boiss.) K. Hammer, *kuprijanovii* (Grossh.) K. Hammer and *strictum* (syn. *S. montanum* Guss.; GRIN, <http://www.ars-grin.gov/cgi-bin/npgs/html/splist.pl?11022>, 20th July 2016).

Rye is commonly grown in Eastern and Northern Europe, mainly for the production of bread, alcohol, and animal feed (Evans, 1995). In contrast to most grain crops that are self-pollinating, rye is a cross-pollinating cereal, and such outbreeding nature results in a high intraspecific diversity (Schlegel, 2014). Additionally, rye harbors a broad tolerance to biotic and abiotic stress, absent in other temperate cereals (Rizvi and Scoles, 2014). Therefore, this crop also had a major importance on plant breeding strategies both through the production of the synthetic hybrid Triticale (x *Triticosecale* Wittmack) as well as through the introgression of rye chromatin in wheat varieties, particularly by the short arm of chromosome 1R (1RS), as a source of genes for agronomic and resistant improvement (Baum and Appels, 1991). Due to its higher ability to grow in poor soils and under greater adverse conditions than other cereals, rye is an economically important cover crop in Northern Europe and other rye-growing countries (Vaughan and Geissler, 2009). Rye culture is of marked importance in the northern region of Portugal where local farmers cultivate the same rye population for several centuries under a subsistence agricultural system in small areas for both food and feed. Traditional rye bread baking is an important share of both diet and cultural heritage not only in Portugal but also in other rye-producing countries. Until the middle of last century there was genetic exchanges as a result of transhumance linking the territory to remote and dispersed regions, mainly by pastoralism that have worked as ecological corridors from valleys to the mountainous areas. Considering, that 80% of Portuguese soils are acidic (Almeida, 1955), it has been shown that Northern rye populations display not only high genetic diversity on storage proteins (Ribeiro et al., 2012) but also on aluminum tolerance (Matos et al., 2001), probably responsible for rye maintenance in the regional agricultural system. Portugal has a wealth of rye germplasm with about 769 accessions conserved in several institutions (Bettencourt and Carnide, 1998); with many other local accessions yet to be preserved/identified. According to FAO (Food and Agricultural Organization of the United Nations, 2016), in Portugal, there was a 33% decline of total rye harvest area from 2004 to 2014 as a result of traditional agricultural abandonment, which poses a huge threat toward local rye landraces. There is therefore some urgency to characterize and evaluate landraces maintained on “on farm” conditions, in order to develop proper measures for *in situ* conservation and made available for utilization.

While much of the world’s rye harvest is based on modern high-yield varieties, traditional varieties grown locally have great importance as a resource for future crop improvement. Such local landraces may represent an intermediate stage of domestication between a wild ancestor and modern varieties, being important reservoirs of agronomically important genes. Landrace can be defined following (Camacho-Villa et al., 2005), as: “a dynamic population of a cultivated plant that has historic origin, distinct identity and lacks formal crop improvement, as well as often being genetically diverse, locally adapted and associated with traditional farming systems.” Furthermore, landraces can be separated in *ex situ* and *in situ* collections: the former being those detained in gene banks or botanical gardens which represent a comprehensive snapshot of the genetic diversity at a given time and place (Greene et al., 2014); while *in situ* allows adaptive evolutionary processes to continue shaping genetic diversity under farmer management. In plant genetic resources (PGR) conservation, it has long been recognized that effective strategies need to integrate *in situ* and *ex situ* approaches (Greene et al., 2014). Studies reported the complementary source of genetic variation between *in situ* and *ex situ* collections, as some of the alleles may have been lost *in situ* (e.g., Jensen et al., 2012) or *ex situ* (e.g., Li et al., 2005). Indeed, simple sequence repeat (SSR) data from bean (*Phaseolus vulgare* L.) landraces conserved *ex situ* and *in situ* indicated significant genetic differentiation in *ex situ* subpopulations as well as loss of alleles, gain of new alleles, and reduction of rare alleles with an increase of common alleles (Negri and Tiranti, 2010). In most crops, landraces usually display higher genetic diversity than breeding cultivars, due to the genetic bottleneck and selective effect associated to its improvement (Meyer and Purugganan, 2013). In fact, the narrowing of the genetic pool of modern crop varieties has become an increasing concern also for rye breeders (Fischer et al., 2010). “Petkus” was one of the leading cultivars in the twentieth century from which many of the open pollinated varieties (OPVs) were selected or include “Petkus” in their ancestry (Hepting, 1978; Miedaner, 1997; Fischer et al., 2010). The two genetic pools “Carsten” and “Petkus” were previously identified as the most promising heterotic pattern (Hepting, 1978), and from then onwards, hybrid rye breeding, a breeding system used for cross-fertilized crops, was and still is based on the “Petkus” × “Carsten” heterotic pattern (Geiger and Miedaner, 2009). Recent evidences point out for a genetic narrowing of “Carsten” pool (Fischer et al., 2010). Also, considering that the steadily improved “Petkus” was the parental ancestor of many OPVs, the probability of finding genetically diverse populations from “Petkus” pool is significantly reduced. As such, assessment of new genetically distinct populations is urgent for supplementing rye’s heterotic pool. Considering that rye cultivars are panmictic populations, characterized by high levels of heterozygosity and heterogeneity, they usually display similar genetic diversity levels as landraces, namely *ex situ* collections (Persson and von Bothmer, 2000, 2002; Persson et al., 2001; Parat et al., 2016). However, conflicting results reported higher levels of genetic diversity on Portuguese rye landraces than varieties (Ribeiro et al., 2012), as expected for most crops. Previous studies on the genetic diversity of rye accessions share common features, namely: lack of correlation

between accessions and geographic origin and similar genetic diversity between landraces (i.e., *ex situ*) and cultivars, which is indicative of a common genetic background, regardless of breeding level or geographical origin (Bolibok-Bragoszewska et al., 2014; Hagenblad et al., 2016; Parat et al., 2016; Targońska et al., 2016). Indeed, it was proposed that ecological and temporal isolation are key for shaping rye's genetic diversity rather than spatial or geographic isolation (Ma et al., 2004). A recent study has shown that, rather than distinction between landraces and cultivars, diversity patterns on rye seem to be related to the end use over time (Parat et al., 2016), uncovering a clear separation of rye for forage in the Mediterranean area and for grain in Northeast Europe.

SSRs have proven to be a marker of excellence for examining rye's genetic diversity (Shang et al., 2006; Akhavan et al., 2010; Gailite et al., 2013; Parat et al., 2016; Targońska et al., 2016). Several studies using a wide range of molecular markers systems, have contributed to a better knowledge on rye's genetic diversity (e.g., Persson and von Bothmer, 2000, 2002; Chikmawati et al., 2005, 2012; Bolibok-Bragoszewska et al., 2012; Ribeiro et al., 2012; Al-Beyroutiová et al., 2016; Parat et al., 2016; Santos et al., 2016; Targońska et al., 2016), with studies using rye cultivars and landraces, particularly from *ex situ* collections (e.g., Matos et al., 2001; Hagenblad et al., 2016; Parat et al., 2016; Targońska et al., 2016). As genetic diversity of *in situ* populations is dynamically maintained and it evolves along changing environments, evaluation of rye accessions will open insights into a new potential genetic diversity linked to the genetic adjustment on many traits during adaptation to local conditions and agriculture practices.

Our study aims at performing the first comprehensive assessment of the genetic diversity and population structuring on rye using cultivars, *ex situ* and *in situ* collections following a worldwide sampling scheme, by performing a molecular-directed analysis using SSRs markers. Data obtained exposes and delivers novel *in situ* genetic resources with potential for broadening the genetic diversity within the rye heterotic pool, thus opening a new venue for rye breeders.

MATERIALS AND METHODS

Plant Material

For this study, 28 *Secale cereale* L. subs. *cereale* accessions and its crop wild relative (*S. strictum* subs. *strictum*, referred as *S. strictum* hereforth) were selected from different geographic regions. The panel consist of eight *ex situ* accessions from gene banks, nine cultivars ("Imperial," "Kungs II," "Petkus," "Dankowskie Zlote," "Ailé," "Voima," "Alvão," "Pulawskie," and "Antoninskie"), along with 11 *in situ* accessions collected "on farm" in 2014 from Northeast Portugal (Table 1). *Ex situ* accessions were provided by the following gene banks with acronym, accession prefix and country: Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK, R, Germany) and Nordic Genetic Resource Center (NordGen, NGB, Sweden). Rye accessions were grouped and mentioned hereforth as cultivars, *ex situ*, and *in situ/on farm* accessions. Cultivars are those resulting from modern rye breeding, *in*

situ are collections held in farmers' fields and *ex situ* are those collections detained in gene banks or botanical gardens. Both *ex situ* and *in situ* accessions collectively will be referred as landraces. The 11 *in situ* Portuguese accessions were collected from local farms on Northern region (Figure 1A) where a great importance of rye culture still exists at regional level. The region sampled covers part of the Serra da Estrela Natural Park, the largest natural conservation area, and biggest mountain range in Portugal, with several valleys along the mountain assortment. Such region is characterized by harsh winters and samples selected are within 400–1100 m elevation (Figure 1B). Each accession represents a mixed sampling from small plots (maximum 0.5 ha) maintained by local farmers under a subsistence regime. The samples collected were sowed in September/October 2013 and harvested in mid-July 2014, as it is commonly performed. About 200 kg from each *in situ* population were obtained directly from each farmer and further used on genetic diversity studies and stored at the Instituto Superior de Agronomia, University of Lisbon for future characterizations of agronomical and morphological traits. All *in situ* accessions included in this study are available upon author request.

DNA Extraction

DNA from 5 to 10 individuals of each rye germplasm accessions (cultivars and *ex situ*), while for Portuguese *in situ* collections 16 individuals were used, for accounting within-population diversity. Genomic DNA (gDNA) was extracted from young leaves using the cetyltrimethylammonium bromide (CTAB) procedure adapted from Thomas et al. (1993). Briefly, young leaves (100 mg) from individual plants grown from seeds were directly ground by an Eppendorf-pestle in a 1.5 mL tube, thawed and resuspended in 300 μ l of extraction buffer [0.35 M Sorbitol, 0.1 M Tris pH 8.0 NaCl, 5 mM EDTA, 1 % (w/v) PVP-40] by adding 3.8 g/L of sodium bisulphite and 1 % (v/v) 2-mercaptoethanol (Sigma Aldrich) upon use. After a 5-min incubation on ice, 300 μ l of lysis buffer [0.2 M Tris pH 8.0, 2 M NaCl, 50 mM EDTA, 2% CTAB] was added along with 120 μ l of 5% sarkozyl (w/v) and RNase A (10 mg/mL). After incubation at 65°C for 15 min, an equal volume of chloroform: isoamyl alcohol (24:1) was then mixed by a brief vortex, and the aqueous phase was recovered by centrifugation at 16,000 g for 15 min. gDNA was precipitated with 0.6 volume of isopropanol and recovered by a 15-min centrifugation at 16,000 g, followed by a washing step with 70% ethanol. Following a centrifugation at 16,000 g for 10 min, the pellet was dried and resuspended in 40 μ l deionized water. DNA purity and concentration were measured at 260/280 nm using a spectrophotometer (NanoDrop-1000, Thermo Scientific) while DNA integrity was verified by agarose gel electrophoresis.

Microsatellite Genotyping

A set of 16 microsatellite markers (Supplementary Table S1) was used for screening rye accessions, consisting of nine genomic SSRs (gSSRs, Saal and Wricke, 1999) and seven Expressed Sequence Tags-SSRs (ESTs-SSRs, Hackauf and Wehling, 2002). Before multiplexing, each SSR marker was validated in singleplex polymerase chain reactions (PCR) using a three-primer PCR

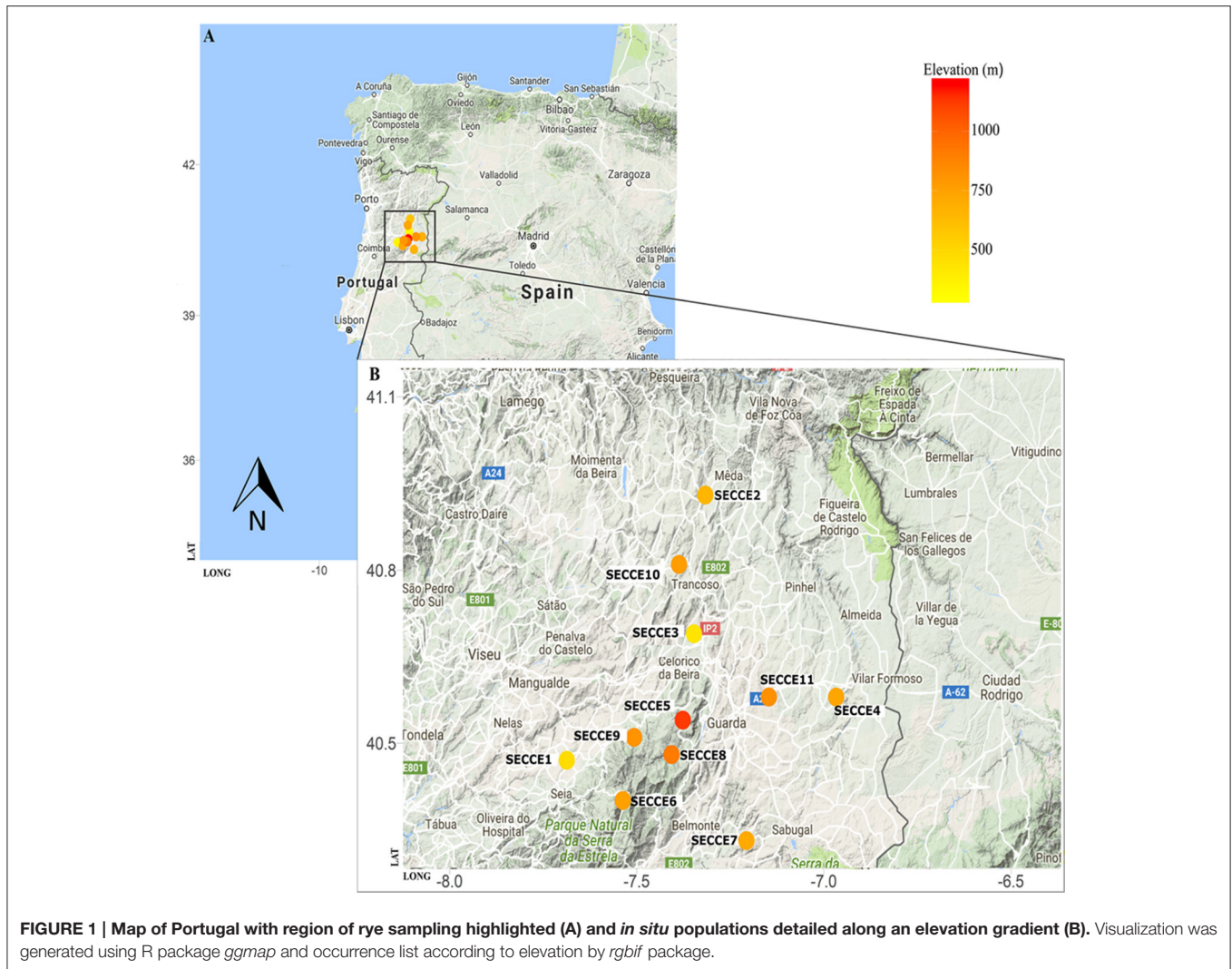
TABLE 1 | Rye sampled accessions.

Accession number	Taxon	Accession name	Location/country	Breeding level	Latitude	Longitude	Seed bank	N
R2119	<i>Secale cereale</i> subsp. <i>cereale</i>	Ailé	Spain	Cultivar	38.95	-5.13	IPK	6
R 2204	<i>S. cereale</i> subsp. <i>cereale</i>	Antoninskie	Poland	Cultivar	54.37	18.64	IPK	6
R 1633	<i>S. cereale</i> subsp. <i>cereale</i>	Dankowskie Zlote	Poland	Cultivar	54.37	18.64	IPK	6
R1150	<i>S. cereale</i> subsp. <i>cereale</i>	Imperial	Canada	Cultivar	47.51	-72.11	IPK	6
R 1265	<i>S. cereale</i> subsp. <i>cereale</i>	Kungs II	Sweden	Cultivar	62	15	IPK	6
R 1667	<i>S. cereale</i> subsp. <i>cereale</i>	Petkus	German Democratic Republic	Cultivar	50.87	12.08	IPK	6
R1454	<i>S. cereale</i> subsp. <i>cereale</i>	Voima	Finland	Cultivar	64	26	IPK	6
R 751	<i>S. cereale</i> subsp. <i>cereale</i>	Pulawskie	Bazanowka/Poland	Cultivar	49.60	22.05	IPK	6
70c	<i>S. cereale</i> subsp. <i>cereale</i>	Alvão	Portugal	Cultivar	40.28	-7.74	UTAD	6
NGB14283	<i>S. cereale</i> subsp. <i>cereale</i>	Svedjeråg 66-S (Sved)	Sweden	<i>ex situ</i> accession	59	13	NORDGEN	8
18/2015	<i>S. cereale</i> subsp. <i>cereale</i>	Riodeva	Spain	<i>ex situ</i> accession	38.95	-5.13	ISA/UL	6
R 2136	<i>S. cereale</i> subsp. <i>cereale</i>	R2136Russ	Leningrad/Russia	<i>ex situ</i> accession	59.95	30.32	IPK	6
R 780	<i>S. cereale</i> subsp. <i>cereale</i>	R780Spain	Badajoz/Spain	<i>ex situ</i> accession	38.95	-5.13	IPK	6
R 2694	<i>S. cereale</i> subsp. <i>cereale</i>	R2694West	Westfalen/Germany	<i>ex situ</i> accession	52	8	IPK	6
R 1148	<i>S. cereale</i> subsp. <i>cereale</i>	R1148Turkey	Van/Turkey	<i>ex situ</i> accession	38.48	43.68	IPK	6
R 1138	<i>S. cereale</i> subsp. <i>cereale</i>	R1138Italy	Bruzolo/Italy	<i>ex situ</i> accession	45.13	7.20	IPK	6
R 1133	<i>S. cereale</i> subsp. <i>cereale</i>	R1133PT	Trás-os-Montes (PT)	<i>ex situ</i> accession	41.70	-7.13	IPK	6
02/2015	<i>S. cereale</i> subsp. <i>cereale</i>	SECCE1	Seia (PT)	<i>in situ</i> accession	40.47	-7.69	ISA/UL	16
03/2015	<i>S. cereale</i> subsp. <i>cereale</i>	SECCE2	Aveloso (PT)	<i>in situ</i> accession	40.93	-7.32	ISA/UL	16
04/2015	<i>S. cereale</i> subsp. <i>cereale</i>	SECCE3	Celorico da Beira (PT)	<i>in situ</i> accession	40.69	-7.35	ISA/UL	16
05/2015	<i>S. cereale</i> subsp. <i>cereale</i>	SECCE4	Mesquitela (PT)	<i>in situ</i> accession	40.58	-6.97	ISA/UL	16
06/2015	<i>S. cereale</i> subsp. <i>cereale</i>	SECCE5	Videmonte (PT)	<i>in situ</i> accession	40.54	-7.38	ISA/UL	16
07/2015	<i>S. cereale</i> subsp. <i>cereale</i>	SECCE6	Manteigas (PT)	<i>in situ</i> accession	40.40	-7.54	ISA/UL	16
08/2015	<i>S. cereale</i> subsp. <i>cereale</i>	SECCE7	Trancoso (PT)	<i>in situ</i> accession	40.81	-7.39	ISA/UL	16
09/2015	<i>S. cereale</i> subsp. <i>cereale</i>	SECCE8	Guarda António (PT)	<i>in situ</i> accession	40.48	-7.41	ISA/UL	16
10/2015	<i>S. cereale</i> subsp. <i>cereale</i>	SECCE9	Gouveia (PT)	<i>in situ</i> accession	40.51	-7.51	ISA/UL	16
11/2015	<i>S. cereale</i> subsp. <i>cereale</i>	SECCE10	Sabugal (PT)	<i>in situ</i> accession	40.33	-7.21	ISA/UL	16
12/2015	<i>S. cereale</i> subsp. <i>cereale</i>	SECCE11	Guarda (PT)	<i>in situ</i> accession	40.58	-7.15	ISA/UL	16
XX-0-MG-19-44650	<i>S. strictum</i> (=S. <i>montanum</i>)	<i>S. strictum</i>	Unknown origin*	Crop wild relative	-	-	BGJG/UM	5

Accession numbers and names are provided, along taxonomic classification, breeding level, number of individuals screened (N), geographical provenance, and the seedbank from which seeds were requested (for cultivars and *ex situ*) or stored (*in situ*). IPK, Leibniz, Institut für Pflanzengenetik und Kulturpflanzenforschung (Germany); NGB, Nordic Genetic Resource Center (NordGen, Sweden), UTAD, Universidade de Trás-os-Montes e Alto Douro; BGJG/UM, Botanic Garden Johannes Gutenberg-Universität Mainz; ISA/UL, Instituto Superior de Agronomia, Universidade de Lisboa; PT, Portugal. *Distribution; Mediterranean, West Asia.

approach (*sensu* Schuelke, 2000) for reaction reproducibility and presence of PCR artifacts. Each SSR was PCR amplified in a 25 μ l volume reaction following cycling conditions previously described (Saal and Wricke, 1999; Hackauf and Wehling, 2002), using HotStarTaq DNA Polymerase kit (Qiagen), as per manufacturer's instructions. After, SSRs amplified fragments were run in an ABI 3130XL sequencer (Applied Biosystems) with the internal size standard GS500 LIZ (Applied Biosystems), while allele calling was performed in GeneMapper v 3.7 (Applied Biosystems). Stringent selection of markers to ensure the success of co-amplification loci using Multiplex Manager software (Holleley and Geerts, 2009), allowed building four SSRs panels assembled in 4-plex PCR reactions (Multiplex A, B, C, and D; Table 2), using four universal forward fluorescently labeled primers following Culley et al. (2013). To increase genotyping accuracy, a "Pig-tail" sequence was added at the 5' end of each

of the reverse primer (Brownstein et al., 1996). PCR multiplex amplifications were carried out using the QIAGEN Multiplex PCR kit (Qiagen), following the manufacturer's protocol, in a total volume of 25 μ L with 1 μ L of gDNA (50–100 η g) and 2.5 μ mol of each primer Forward and Reverse and 0.15 μ mol of each of the tailed fluorescently labeled primers (D1–D4). Reactions were done in 96 well-plates and on each plate one sample was repeated per run thus working as positive control for scoring. Negative PCR controls were included. Initially, a hot-start step at 95°C for 15 min was performed, followed by a touchdown cycling protocol adapted from Hackauf and Wehling (2002): 5 cycles of denaturation at 95°C for 45 s, primer annealing at 68°C for 5 min with -2°C/cycle; a sequence extension at 72°C for 1 min; 5 cycles of denaturation at 95°C for 45 s, primer annealing at 58°C for 2 min with -2°C/cycle and an extension step for 1 min at 72°C; 27 cycles at 95°C for 45 s, 47°C for 75 s, and 72°C for 1 min;



followed by a final extension step at 72°C for 10 min. Multiplex PCR products were run as described earlier and SSR allele sizes were aligned with the internal size standard and scored using the binning function in GeneMapper v3.7. (Applied Biosystems). To improve the SSR marker data quality, allele assignments were checked manually, and ambiguous results were set as “missing data.”

Genetic Diversity—Based Analysis

Genotyping errors were assessed using MICRO-CHECKER v2.2.3. (Van Oosterhout et al., 2004), and estimation of null alleles frequency was done with the EM algorithm of Dempster et al. (1977) as implemented in FreeNA (<http://www.montpellier.inra.fr/URLB/>). These values were computed, as described in Chapuis and Estoup (2007), with 10,000 bootstrap iterations, alternatively using and not using the Excluding Null Alleles (ENA) method, after assessment of null allele frequencies. Polymorphism information content (PIC) and genetic diversity indices were calculated with Microsatellite Toolkit v.3.1.1 (Park, 2001) and GenALEX 6.5 (Peakall and Smouse, 2006, 2012).

These included the total allele number and mean alleles per locus (N_a), private alleles, inbreeding coefficient (fixation index, F), observed (H_O), and expected (H_E) heterozygosity. Deviations from Hardy–Weinberg equilibrium (HWE) were assessed for each locus–population combination and linkage disequilibrium (LD) to determine the extent of distortion from independent segregation of loci using GenePop v4.5 (Rousset, 2008). Statistical significance for both HWE and LD was tested by running a Monte Carlo Markov Chain (MCMC) consisting of 10,000 iterations each, and p -values were corrected for multiple comparisons [$p < 0.00012$, (0.05/406)] by applying a sequential Bonferroni correction (Rice, 1989).

To detect isolation-by-distance (IBD) effects, $F_{ST}/(1 - F_{ST})$ and $F_{ST}^{ENA}/(1 - F_{ST}^{ENA})$ matrixes were done, with a geographic distance matrix defined as pairwise distances generated from geographical coordinates expressed in Km. Pairwise unbiased F_{ST} -values using the ENA method (F_{ST}^{ENA}) for each population comparison were calculated with FreeNA software, while F_{ST} -values were generated in GenALEX 6.5. Both Slatkin’s linearized F_{ST} ($F_{ST}/(1 - F_{ST})$) matrixes were obtained in GenoDive 2.0b27

TABLE 2 | Markers diversity measurements.

Locus	Sample analyzed	Allele number	PIC	He	H _o	F	Frequency null alleles (>0.20)
ESTs-SSRs (Hackauf and Wehling, 2002)							
SCM113	284	1	–	–	–	–	–
SCM166	285	4	0.45	0.48	0.51	–0.28	–
SCM63	255	12	0.73	0.76	0.54	0.02	0.26 <u>SECCE10</u> , 0.21 <u>SECCE2</u> , 0.25 <u>SECCE8</u>
SCM152	285	11	0.83	0.85	0.57	0.06	0.25 <u>Dankow</u>
SCM98	285	3	0.47	0.55	0.61	–0.41	–
SCM164	268	11	0.67	0.70	0.68	–0.16	0.25 <u>R1138Italy</u>
SCM66	285	3	0.39	0.51	0.80	–0.73	–
Mean		7.33	0.59	0.64	0.62	–0.25	–
gSSRs (Saal and Wricke, 1999)							
SCM39	281	10	0.63	0.66	0.54	0.06	0.21 <u>Alvão</u> , 0.25 <u>R1138Italy</u> , 0.25 <u>R1148Turkey</u>
SCM2	284	5	0.52	0.58	0.62	–0.34	–
SCM28	278	13	0.81	0.83	0.71	–0.05	–
SCM9	282	7	0.65	0.70	0.74	–0.25	–
SCM75	261	11	0.82	0.84	0.78	–0.15	0.22 <u>R1148Turk</u>
SCM43	284	9	0.79	0.81	0.79	–0.11	–
SCM138	285	10	0.85	0.86	0.80	–0.12	0.21 <u>R1133PT</u>
SCM86	285	12	0.77	0.80	0.85	–0.28	–
Mean		9.6	0.73	0.76	0.73	–0.15	–
Total Mean		8.64	0.67	0.71	0.68	–0.20	
Total		122					

The level of genetic diversity of each SSR marker was described with the parameters number of alleles, Polymorphism Information Content (PIC), gene diversity (expected heterozygosity, He), observed heterozygosity (H_o), inbreeding/fixation coefficient (F), and frequency of null alleles above 0.20 as calculated by FreeNa are presented with additional information on the identified population.

(Meirmans and Van Tienderen, 2004). The correlation between the two data matrices was assessed using a Mantel test and its significance estimated by *p*-values, the regression coefficient (R^2), and the mean correlation coefficient (R_{XY}) over 999 random permutations as implemented in GenALEx 6.5.

Population Structure

Population structure was addressed using three approaches: (i) estimating relations among populations using genetic distances; (ii) hierarchical genetic analysis by AMOVA; and (iii) individual-based clustering.

(i) Estimating Relations among Populations Using Genetic Distances

Relationships among populations were estimated with Cavalli-Sforza and Edward's chord genetic distances (DC , Cavalli-Sforza and Edwards, 1967) using the INA method computed in FreeNA (DC^{INA}), and Nei's distance (D , Nei, 1972) calculated in GenALEx 6.5. Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Neighbor-Joining (NJ) trees were produced using package *ape* v3.4. (Paradis et al., 2004) for R v3.2.3 (R Developmental Core Team, 2015) based on 10,000 bootstraps values assessed by *aboot* function from *poppr* v2.1.0. package (Kamvar et al., 2014, 2015). Trees were further edited in FigTree v1.4.2 (Rambaut, 2014).

(ii) Hierarchical Genetic Analysis (AMOVA)

The hierarchical distribution of genetic variation on the 28 populations (excluding the outgroup *Secale strictum*) was characterized using Analysis of Molecular Variance (AMOVA, Weir and Cockerham, 1984; Excoffier et al., 1992; Hill, 1996) with ARLEQUIN v3.5.1.3 (Excoffier and Lischer, 2010) and significance was assessed after 1000 permutations. Two 3-level AMOVAs were pursued: one using cultivars and landraces (i.e., *ex situ* and *in situ* accessions) as groups, and the second, narrowed to landraces using *ex situ* and *in situ* accessions alone. In each AMOVA, the total variance was partitioned into components to account for differences between two defined groups [V_a , (1) cultivars and landraces; (2) *ex situ* vs. *in situ* accessions], differences among populations within those groups (V_b), differences among individuals within populations (V_c). Variance components (V_a , V_b , and V_c) were used to calculate the fixation indices (F -statistics; F_{CT} , F_{SC} , F_{ST}) according to Weir and Cockerham (1984).

(iii) Individual-Based Clustering

To identify genetically distinct clusters, two individual-based assignment approaches were pursued: a bayesian clustering analysis using STRUCTURE (Pritchard et al., 2000) and a multivariate analysis by Discriminant Analysis of Principal Components (DAPC, Jombart et al., 2010). While STRUCTURE uses allele frequency and LD information from the dataset

directly; the latter is a multivariate method which attempts to summarize the genetic differentiation between groups, while overlooking within-group variation and not relying on a particular population genetics model and free of HWE assumptions (Jombart et al., 2010).

Bayesian model-based clustering algorithm implemented in STRUCTURE v.2.3.4 was used to identify genetic clusters under a model assuming admixture and correlated allele frequencies without using population information. An exploratory run was performed setting *K*-values from 1 to 30 with a 50,000 burn-in period followed by 100,000 MCMC iterations. Subsequent runs were set for a burn-in period length to 100,000 followed by 1,000,000 MCMC iterations with *K*-values narrowed from 1 to 10 with 10 runs computed for each *K*. StructureHarvester v0.6.94 (Earl and VonHoldt, 2012) was then used to calculate ΔK *ad hoc* statistics from Evanno et al. (2005) for estimating the most likely *K*-value, which is based on the rate of change of the “estimated likelihood” between successive *K*-values. CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007) was used to average replicate runs for the selected *K*-value, for accounting problems with multimodality and label switching between iterations of STRUCTURE runs. CLUMPP results were then plotted with DISTRUCT v1.1 (Rosenberg, 2004).

DAPC was implemented in R using *adegenet* v1.3.1 package (Jombart, 2008). The function *find.clusters* was used to find the ideal *K*-value, based on the computation of Bayesian Information Criterion (BIC) scores, maintaining default parameters and retaining all principal components (PCs). Cross validation using the *xvalDapc* function was pursued to determine the optimal number of PCs to retain in the Discriminant Analysis (DA).

RESULTS

SSRs Genotyping and Statistics

All 16 SSRs were tested in singleplex reactions at the estimated optimal annealing temperature, and only after this initial quality assessment, SSRs markers were grouped into 4-plex reactions (Supplementary Table S1). Upon multiplex reactions, SCM180 (Multiplex C) displayed a difficult allele scoring performance, not depicted in singleplex reactions. This fact may be ascribed to PCR dynamics under a multiplex reaction, where concentrations of different primers are equimolar, requiring in some cases, relative balanced primers concentrations (Sint et al., 2012). For the remaining 15 SSRs loci, allele profiles were clear and easy to score. No errors in the genotypic data matrix were detected, indicating the absence of potential errors associated with stuttering bands or large allele dropout in SSRs screened. In only 10 of the 406 locus-populations comparisons, the frequencies of null alleles were higher than 0.20 (Table 2). Deviations from Hardy-Weinberg Equilibrium (HWE) were observed in most loci except for SCM166, with 70 locus-population combinations statistically significance ($p < 0.05$); while after sequential Bonferroni correction only four loci (SCM64, SCM39, SCM63, SCM75) displayed significant deviations, matching 11 of the 406 locus-population combinations (Supplementary Table S2). All 15 loci were in linkage equilibrium after Bonferroni correction, thus being non-correlated, and alleles independently segregated and

inherited (data not shown). Negative fixation index (*F*) estimates were observed across several loci, exceptions for the EST-SSRs SCM152 (0.060), and SCM63 (0.017) and in the gSSR SCM39 (0.057, Table 2), which can reflect more heterozygotes than expected or other population structure complexities.

Genetic Diversity Estimates

Overall, a total of 122 alleles were detected in the 285 individuals analyzed (Table 2). All loci screened were polymorphic except SCM113, which revealed to be monomorphic (194 bp, Supplementary Table S1) being not used for further analysis. The total number alleles per locus ranged from 3 (SCM66 and SCM98) to 13 (SCM28) with an average of 8 alleles per locus (Table 2). When comparing gSSRs with EST-SSRs, the first revealed a higher number (5–13) of alleles per locus than EST-SSRs (3–12), with an average of 9.6 and 7.3 alleles, respectively (Table 2). Once the dataset was separated in cultivars and landraces (including *in situ* and *ex situ* accessions) accessions, results showed that in landraces the mean allele number is higher with 4.4 in all SSRs, 4.8 with gSSRs, and 4.4 in EST-SSRs, against the values in cultivars (3.3 alleles per locus in all SSRs, 3.7 in gSSRs, and 3.3 for EST-SSRs, Table 3). When analyzing landrace dataset, *ex situ* collections (2.97 all SSRs, 3.3 for gSSRs, 2.97 for EST-SSRs, Table 3) showed a substantially lower mean of alleles

TABLE 3 | Genetic diversity analysis by cultivars and landraces (i.e., *ex situ* and *in situ*).

	<i>N</i>	Loci	<i>H_e</i>	<i>H_e SD</i>	<i>H_o</i>	<i>H_o SD</i>	<i>N_a</i>	<i>N_a SD</i>	<i>F</i>
ALL SSRs									
Cultivars	54	14	0.61	0.04	0.70	0.05	3.28	1.20	-0.26
<i>ex situ</i>	50		0.56	0.05	0.67	0.05	2.97	1.06	-0.33
<i>in situ</i>	176		0.67	0.04	0.68	0.03	5.47	2.20	-0.07
Landraces	226		0.63	0.05	0.67	0.04	4.42	1.72	-0.18
Total	285		0.61	0.05	0.68	0.04	3.98	1.53	-0.23
gSSRs									
Cultivars	54	8	0.68	0.04	0.73	0.06	3.8	1.17	-0.18
<i>ex situ</i>	50		0.62	0.06	0.69	0.06	3.3	1.05	-0.24
<i>in situ</i>	176		0.70	0.04	0.74	0.04	6.0	1.95	-0.10
Landraces	226		0.67	0.05	0.72	0.05	4.8	1.57	-0.16
Total	285		0.67	0.05	0.72	0.05	4.4	1.44	-0.19
EST-SSRs									
Cultivars	54	6	0.61	0.04	0.70	0.05	3.28	1.20	-0.14
<i>ex situ</i>	50		0.56	0.05	0.67	0.05	2.97	1.06	-0.20
<i>in situ</i>	176		0.67	0.04	0.68	0.03	5.47	2.20	-0.12
Landraces	226		0.63	0.05	0.67	0.04	4.42	1.72	-0.16
Total	285		0.60	0.04	0.67	0.04	3.91	1.50	-0.16

Data are provided by total SSRs, gSSR, and EST-SSRs, following by the grouping scheme adopted, with sample size (*N*): cultivars, *ex situ*, and *in situ* collections, which collectively are referred as landraces and total sampling (i.e., sampled accessions including *S. strictum*). Genetic diversity indices for each group was assessed by expected heterozygosity (*H_e*) and observed heterozygosity (*H_o*) with corresponding standard deviation (*SD*) values, inbreeding/fixation coefficient (*F*), and mean alleles per locus (*N_a*).

per locus when compared to *in situ* accessions (5.5 all SSRs, 6 for gSSRs, 5.5 for EST-SSRs), but similar values to those obtained for cultivars.

Overall, Polymorphic Information Content (PIC) values ranged from 0.39 (SCM66) to 0.85 (SCM138) with a mean value of 0.67 (Table 2). Average PIC-values were higher in gSSRs ($PIC_{gSSRs} = 0.73$) than in EST-SSRs ($PIC_{EST-SSRs} = 0.59$). Similar PIC-values were observed in cultivars, ranging from 0.366 (SCM63) to 0.695 (SCM138), and in landrace accessions (SCM166: $PIC = 0.346$; SCM138: $PIC = 0.677$). In both cultivars and landraces, gSSRs revealed to be more polymorphic than EST-SSRs. In our 14-loci dataset, observed heterozygosity (H_O) varied from 0.51 (SCM166) to 0.85 (SCM86) with a mean of 0.68 (Table 2); and expected heterozygosity (H_e) varied between 0.480 (SCM166) and 0.86 (SCM138). When narrowing analysis to gSSRs and EST-SSRs, both mean H_O and H_e were higher in genomic ($H_O = 0.73$; $H_e = 0.76$) compared to the expressed loci ($H_O = 0.62$; $H_e = 0.64$).

The Fixation Index F (also called the Inbreeding Coefficient) exhibits values from -1 to $+1$. Values close to zero are expected under random mating, while substantial positive values indicate inbreeding or undetected null alleles. Negative values denote excess of heterozygosity, due to negative assortative mating, or selection for heterozygotes. Overall, negative F -values were observed across most accessions (Table 3), thus revealing an outbreeding scenario with increased number of heterozygotes. Positive F was encountered in six Portuguese *in situ* accessions (SECCE1–SECCE6), ranging from 0.002 (SECCE5) to 0.049 (SECCE1; Supplementary Table S3). Despite positive, their relatively small F -values indicate that these populations are at or near Hardy–Weinberg equilibrium, further supported by the lower observed heterozygosity values against the expected under HWE (Supplementary Table S3).

Pairwise Wright's F -statistics (F_{ST}) was used as a measure of the extent of genetic differentiation among subpopulations, with values ranging from 0 (no differentiation) to 1 (high differentiation). Mean pairwise F_{ST} was 0.136 (min = 0.017, "Pulawskie" and "Dankowskie"; max = 0.418, "Riodeva" and *S. strictum*) indicating an overall low level of population's differentiation (Supplementary Table S4). Most pairwise populations (749 out of 784) showed low to moderate genetic differentiation (pairwise $F_{ST} < 0.250$), while the remaining 35 pairwise populations displayed high genetic differentiation (pairwise F_{ST} ranged from 0.251 to 0.418). The presence of null alleles has not caused a significant overestimation of the level of population differentiation, as low population differentiation was also depicted with F_{ST}^{ENA} (mean = 0.138; min = 0.015, "Pulawskie" and "Dankowskie"; max = 0.417, "Riodeva" and *S. strictum*), with 747 pairwise populations showing low to moderate genetic differentiation (Supplementary Table S4). The remaining 37 pairwise comparisons exhibited high genetic differentiation (pairwise $F_{ST}^{ENA} > 0.250$). Overall, cultivars (mean $F_{ST} = 0.136$; $F_{ST}^{ENA} = 0.138$) and *ex situ* collections (mean $F_{ST} = 0.163$; $F_{ST}^{ENA} = 0.165$) presented a moderate population differentiation, whereas *in situ* accessions almost no genetic differentiation were depicted (mean $F_{ST} = 0.057$; $F_{ST}^{ENA} = 0.059$). An overestimation of F_{ST} -values due to

null alleles is widely known especially in cases of significant population differentiation, which is not the case in our study where weak population differentiation was detected.

In order to understand whether genetic variation is correlated with geographical gradients, Isolation-by-distance (IBD) effects were addressed. F_{ST} -values (Figure 2A either excluding, F_{ST}^{ENA} , or not null alleles F_{ST} Figure 2B) confirmed a small, yet significant, explanation of genetic diversity variation across a geographic range.

Population Structure

(i) Estimating Relations among Populations Using Genetic Distances

UPGMA and NJ trees were built using Nei's D and DC^{INA} genetic distances (Supplementary Table S5) across accessions screened. Regarding UPGMA trees, similar structure was observed with both D and DC^{INA} matrices, thus indicating a reliable topology regardless of the different genetic distances algorithms used. As such, only Nei's D distances matrices—derived trees are presented in Figure 3. In UPGMA-derived tree, two clusters are depicted (Figure 3A): one comprising all cultivars with most of *ex situ* accessions and another clade comprising *in situ* accessions, "Sved" and "Riodeva" *ex situ* samples and *S. strictum*. No clade seems to cluster accessions on the basis of a particular geographic origin. NJ dendrogram derived from Nei's distance matrix, grouped Portuguese landraces into different clusters (Figure 3B), displaying a different population structuring of *in situ* accessions compared to UPGMA trees: one group (i.e., SECCE1–SECCE5) without any link with other rye accessions and the other Portuguese accessions placed within the same clade as cultivars and *ex situ* accessions.

Conversely to UPGMA-derived dendrograms, the two genetic distance algorithms produced very dissimilar NJ-generated trees (Nei's D distance, Figure 3; DC^{INA} Supplementary Figure S1), which may be attributed to different assumptions adopted in each clustering methods, with a strict (UPGMA) or relaxed (NJ) molecular clock shown previously to have implications when inferring phylogenies considering that rates of evolution may vary among microsatellite loci (Putman and Carbone, 2014).

(ii) Analysis of Molecular Variance

When grouping cultivars vs. landraces (*ex situ* and *in situ* accessions), AMOVA results showed that molecular variation was mainly (86.84%) found within accessions, whereas variation among accessions within groups explained 8.86% and variance among groups represents only 4.30% of the total genetic variability (Table 4). Regarding *ex situ* vs. *in situ* accessions, a similar scenario was depicted, with genetic variation being higher within accessions (89.64%), rather than among groups (2.86%) or within groups (7.49%). In both cases, a high molecular variation was found within accessions, as expected for a cross-pollinated species, as previously detected in other rye studies (Gailite et al., 2013; Hagenblad et al., 2016; Parat et al., 2016; Targońska et al., 2016).

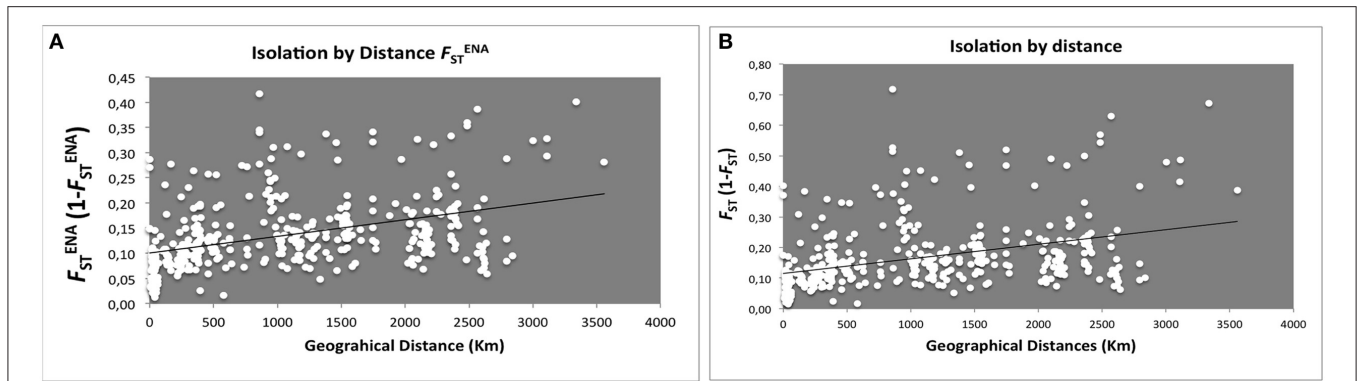


FIGURE 2 | Scatter plots of genetic distance vs. geographical distance for pairwise population comparisons. Each point represents one population pairwise using Slatink’s linearized (A) $F_{ST}^{ENA}/(1 - F_{ST}^{ENA})$; $R^2 = 0.1612$, $R_{xy} = 0.402$, $P = 0.001$; 999 permutations) and (B) $F_{ST}/(1 - F_{ST})$; $R^2 = 0.14067$, $R_{xy} = 0.375$, $P = 0.002$; 999 permutations) plotted against geographic distance (Km).

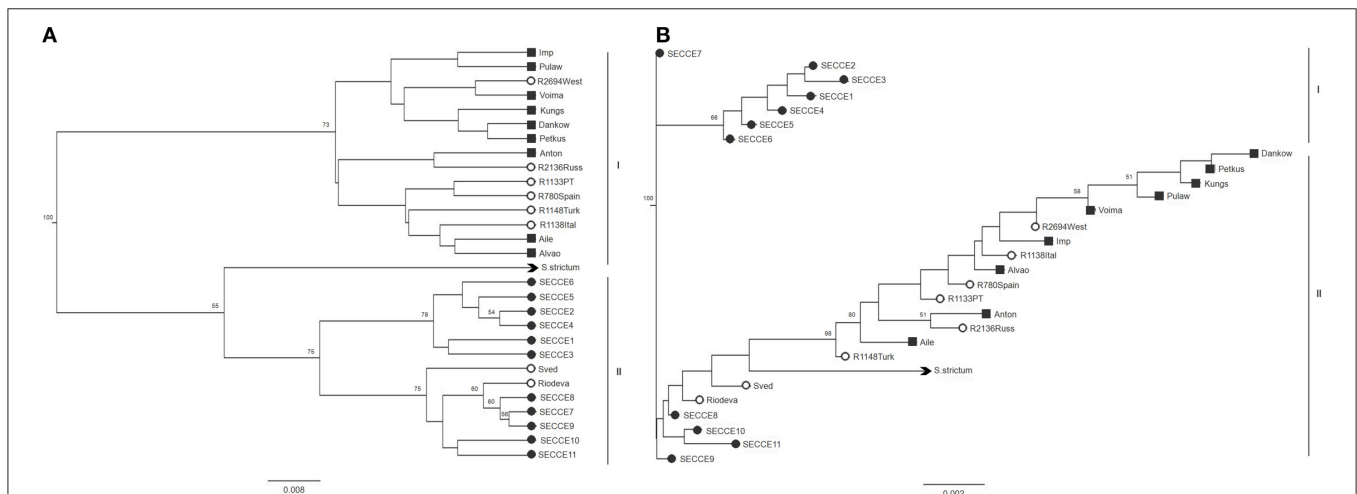


FIGURE 3 | UPGMA (A) and NJ (B) trees generated from Nei’s D distance matrix. Accessions are indicated by symbols reflecting grouping assignment: ■ cultivars, ● *ex situ* accessions, ● *in situ* accessions. *S. strictum* was used as a species outgroup (▶). Only bootstraps values above 50 are indicated.

(iii) Individual Based-Clustering Using Bayesian and a Multivariate Discriminant Analysis to Uncover Population Structure

Exploratory STRUCTURE run considering the biggest range of clusters conceivable ($K = 1-30$), determined $K = 2$ as the most likely model, following Evanno et al. (2005) ΔK method, with no clear plateau observed in the $\ln P(D) = L(K)$ for each value of K (data not shown). Particularly, K -models above 10 revealed high standard deviations of log-likelihood along low values following ΔK *ad hoc* statistics, thus showing that these cluster assignment are not reliable to describe our dataset. Though, this exploratory STRUCTURE results prompted us to subsequently constraint runs to 10 possible clusters ($K = 1-10$). This analysis assigned $K = 2$ as the optimal number of groups based on ΔK , with $K = 3$ also displaying high ΔK -values (Supplementary Figure S2). In $K = 2$, cultivars were grouped in a single cluster (blue cluster), along with the *ex situ* collections; whereas *in situ* are

grouped essentially in a second cluster (pink cluster) together with *S. strictum* (Figure 4A). It is worth mentioning that the *in situ* accession SECCE11 seems to be genetically clustered within blue cluster, along with cultivars and *ex situ* accessions, with some admixed individuals. The result of $K = 3$ was also analyzed (Figure 4C) as the next most likely model. In this assignment, cultivars along most *ex situ* accessions are grouped into a single cluster (pink cluster, C1), excluding “Sved” and “Riodeva” grouped in a different cluster (green cluster, C2), as in $K = 2$, with few or even no admixture. *In situ* populations SECCE7 to SECCE11 are assembled in the green cluster (C2), together with “Sved” and “Riodeva” *ex situ* accessions, with the occurrence of other admixed *in situ* accessions (SECCE1–SECCE6), which belong to the blue cluster (C3, Figure 5). Interestingly, *S. strictum* was assigned to the blue cluster along with some Portuguese admixed accessions (SECCE1–SECCE6), thus reflecting a common genetic diversity with the rye crop wild relative that might be linked to preservation of an ancient

TABLE 4 | AMOVA results including fixation indices F_{CT} , F_{SC} , and F_{ST} .

Source of variation	df	Sum of squares	Variance components	Variation (%)	Fixation indices
CULTIVARS vs. LANDRACES					
Among groups	1	40.572	$V_a = 0.211$	4.30	$F_{CT} = 0.132^*$
Among accessions within groups	26	337.191	$V_b = 0.435$	8.86	$F_{SC} = 0.093^*$
Within accessions	532	2265.938	$V_c = 4.260$	86.84	$F_{ST} = 0.043^*$
EX SITU vs. IN SITU ACCESSIONS					
Among groups	1	35.693	$V_a = 0.137$	2.86	$F_{CT} = 0.029^*$
Among accessions within groups	19	235.926	$V_b = 0.358$	7.49	$F_{SC} = 0.077^*$
Within accessions	455	1949.188	$V_c = 4.284$	89.64	$F_{ST} = 0.104^*$

The genetic differentiation among cultivars vs. landraces (i.e., ex situ/in situ accessions) and ex situ vs. in situ accessions groups is denoted as F_{CT} , among accessions within groups as F_{SC} and within accessions as F_{ST} . * $p < 0.001$.

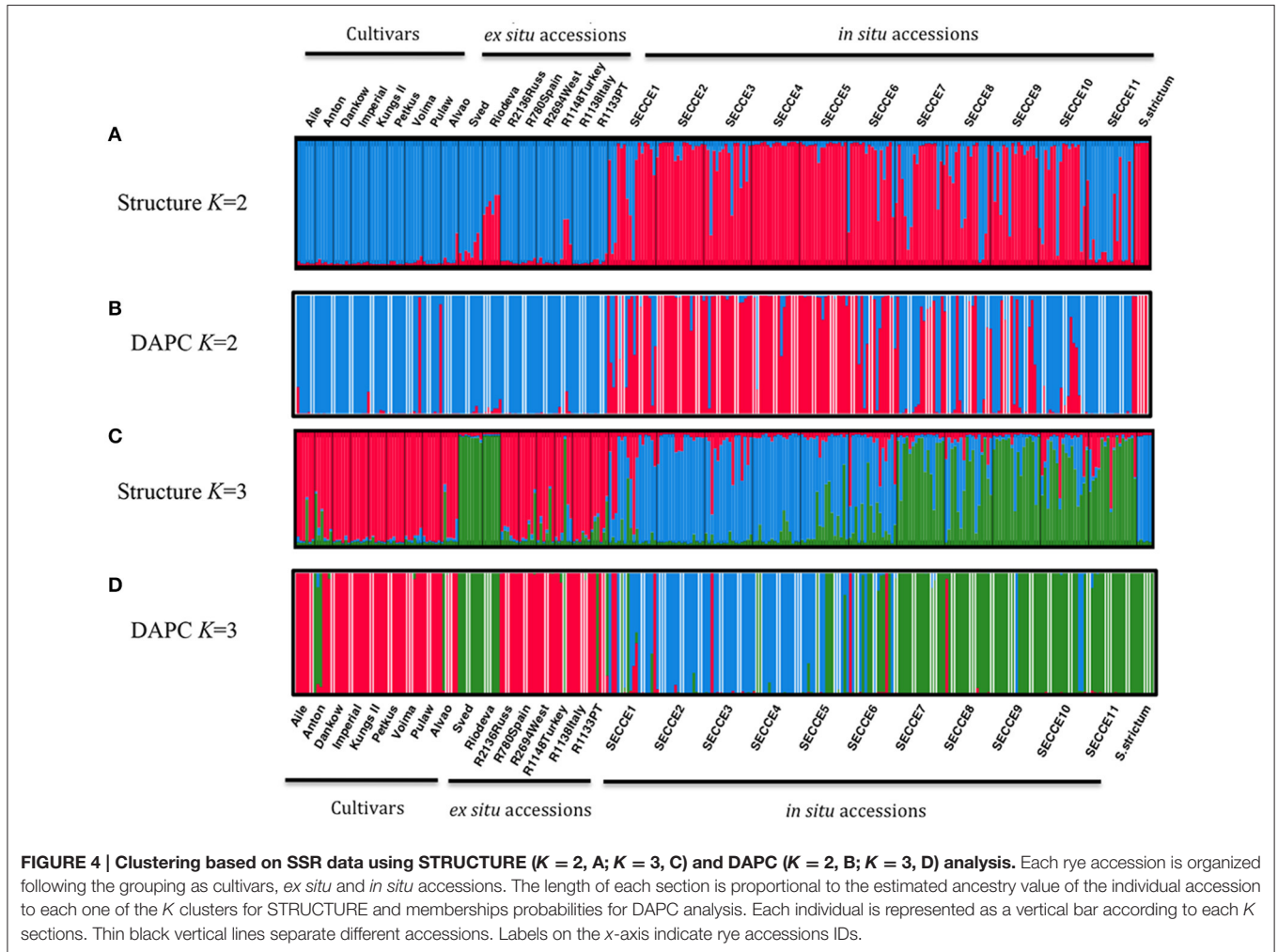
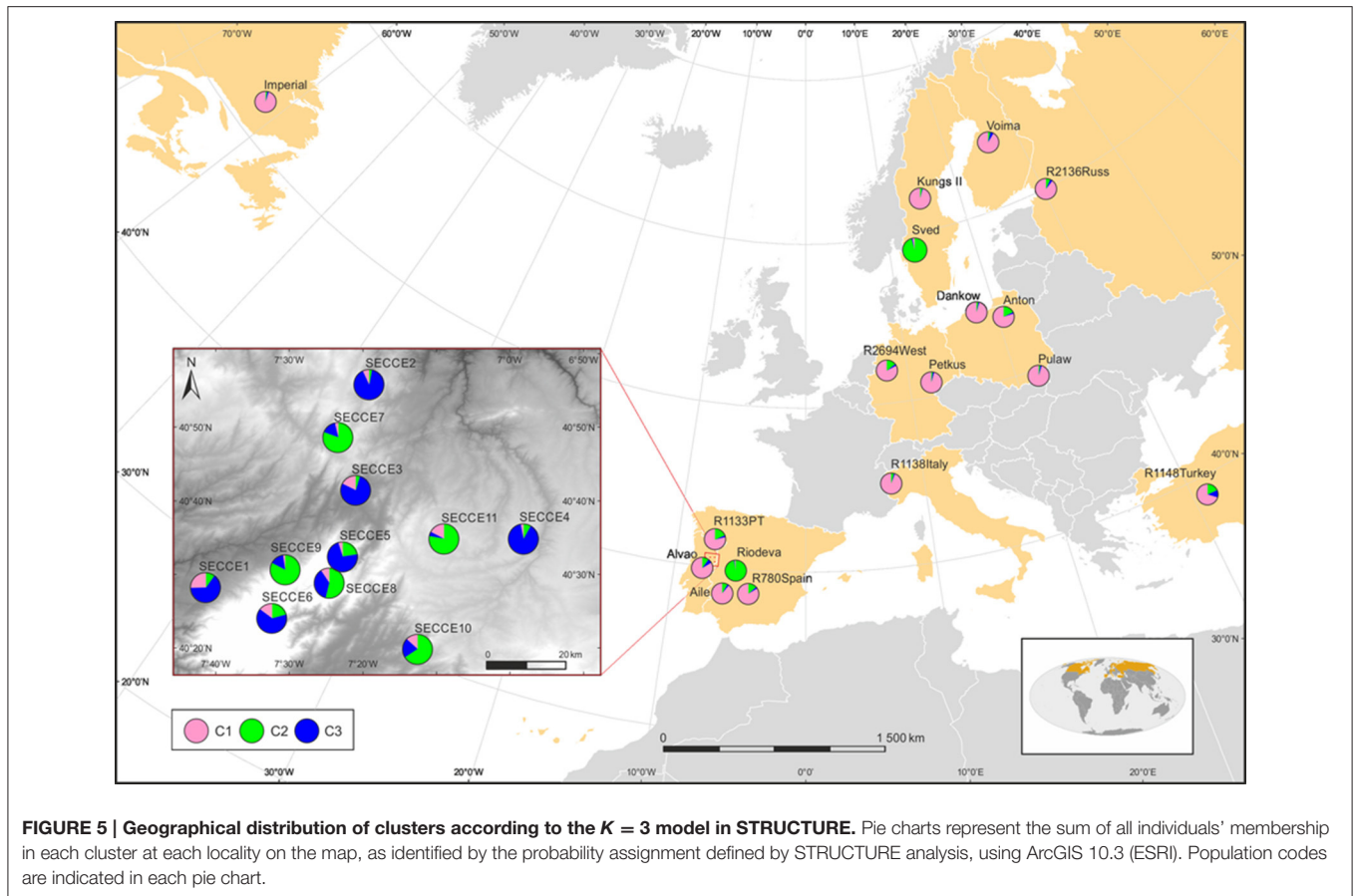


FIGURE 4 | Clustering based on SSR data using STRUCTURE ($K = 2$, A; $K = 3$, C) and DAPC ($K = 2$, B; $K = 3$, D) analysis. Each rye accession is organized following the grouping as cultivars, *ex situ* and *in situ* accessions. The length of each section is proportional to the estimated ancestry value of the individual accession to each one of the K clusters for STRUCTURE and memberships probabilities for DAPC analysis. Each individual is represented as a vertical bar according to each K sections. Thin black vertical lines separate different accessions. Labels on the x-axis indicate rye accessions IDs.

diversity resulting from the low diversification rate of Portuguese accessions.

DAPC analysis was made without any *a priori* group assignment. To infer the appropriate number of genetic clusters, the lowest Bayesian Information Criteria (BIC) score was selected, predicting a $K = 3$ (Supplementary Figure S3). Cross validation using the *xvaldapc* function outcome the number of

PCA axes retained against the proportion of successful outcome prediction, which allowed retaining 60 PCA axes (considering the highest successful assignment- 93.41%, with the lowest mean squared error, MSE- 7.76%) and 2 Discriminant Functions (explaining 92.9% of cumulative variance), for inferring the 3 genetic clusters. When displaying loading plots from both discriminant functions, one can determine which variables



(i.e., alleles/loci) contributed the most for the three-clustering assembly. As such, alleles 177 (SCM138), 369 (SCM152), 245 (SCM166) from DF1, and alleles 147 and 150 from SCM164 together with 182 (SCM75), 171 (SCM138), and 194 (SCM166) from DF2 are responsible for most of the genetic variation explaining the three genetic cluster assignment (Supplementary Figure S3). A scatterplot allows an overview of the 3 genetic groups clustering (Supplementary Figure S4), and when performing a DAPC membership probability plot as in STRUCTURE (Figure 4D), one can depict a similar clustering assignment as determined with $K = 3$ in STRUCTURE: pink cluster with cultivars and most *ex situ* populations, green cluster with SECCE7–11 *in situ* populations along *ex situ* “Riodeva” and “Sved” populations and *S. strictum*, and blue cluster comprising the remaining *in situ* populations (SECCE1–SECCE6). Analysis of $K = 2$ from DAPC (Figure 4B) was also performed in order to obtain a comparison with the ideal K inferred from STRUCTURE analysis. For this K clustering, cross-validation analysis following the Occam's razor principle determined the retention of 20 PCA axes (94.56% of successful assignment with 7.67% of MSE), capturing 60% of cumulative variance, and with only 1 Discriminant Function for describing the 2 genetic clusters (data not shown). Following this cluster assignment, two different genetic backgrounds of Portuguese *in situ* accessions can be depicted: one group (SECCE7–SECCE11) that shares allelic

diversity with cultivars and *ex situ* material in the blue cluster, and other group (SECCE1–SECCE6) grouped with *S. strictum* in the pink cluster, thus showing a common genetic diversity background with wild rye (Figure 4B).

Considering the overall pattern of genetic clustering and observed intraspecific variation, STRUCTURE, and DAPC produced similar results, showing optimal clustering of individuals that separates most cultivars and *ex situ* from Portuguese *in situ* accessions.

DISCUSSION

In this study, 28 rye accessions screened with SSRs included nine international rye cultivars, eight worldwide *ex situ* accessions and 11 *in situ* accessions from Northeast Portugal, one of the regions of excellence for rye cultivation in this country. Also, wild rye *S. strictum* was included in order to track any shared allelic diversity existing with the screened accessions. The observed genetic diversity and population structure of a global collection of rye cultivars and *ex situ* accessions along an exhaustive sampling of *in situ* Portuguese accessions indicate that (i) rye's genetic diversity do not follow a geographical/spatial trend (ii) *in situ* accessions display similar genetic diversity than cultivars/*ex situ* collections yet with a different genetic background and (iii) exists

an unexplored genetic diversity on *in situ* accessions which can represent an effective alternative to increment rye heterotic pool in future breeding programs.

SSRs Performance and Informativeness

Twenty-nine rye accessions were first genotyped with 16 SSRs, and after SSRs quality assessment 14 markers were further used for subsequent genetic diversity analysis (Supplementary Table S6). One SSR was discarded due to low reproducibility in a multiplex PCR amplification (SCM180) and other revealed to be monomorphic (SCM113) across the germplasm analyzed and thus not being an informative marker for posterior diversity analysis. Unexpectedly, no private alleles were detected, contrasting with other SSRs studies in rye populations (Parat et al., 2016; Targońska et al., 2016). In this context, by not detecting private alleles, one can speculate that alleles are present in one or more rye populations regardless of being cultivars or landraces thus representing significant shared allelic diversity. Observed heterozygosity (H_O) obtained in our study (0.51, SCM166–0.85, SCM86) with a mean of 0.68, is higher than those reported for 9 Latvian rye accessions using 9 genomic loci (Gailite et al., 2013) with a mean H_O of 0.58, ranging from 0.21 (SCM2) to 0.84 (SCM9). Furthermore, a recent study using 32 SSRs in 14 rye accessions (Parat et al., 2016), reported values of H_O per SSR of 0.44 ± 0.17 and 0.67 ± 0.14 for H_e ; while in our study, despite using only 14 SSRs both H_O and H_e were similar (0.64 ± 0.21 and 0.66 ± 0.22 , respectively) likely due to a higher number of rye populations screened ($n = 29$).

PIC-values obtained for SSRs used were high (average PIC = 0.67) which indicates their high informativeness, which was predictable since the SSRs markers were not selected randomly, but based on the previous performance analyses (Shang et al., 2006). Our PIC-values were higher when compared with a recent study using 22 SSRs (including EST-SSRs and gSSRs) to screen 367 Polish rye accessions, which displayed a 0.57 average PIC-value for all loci used (Targońska et al., 2016); but similar to those obtained in 14 rye accessions using 32 SSRs (PIC-values ranged from 0 to 0.92) with an average of 0.62 (Parat et al., 2016). We detected higher PIC-values in gSSRs (0.73) than in EST-SSRs (0.59), which is in agreement with a previous study using 8 gSSRs (PIC = 0.63) and 14 EST-SSRs (PIC = 0.54; Targońska et al., 2016). Inversely, another analysis of rye cultivars using both gSSRs and EST-SSRs, showed a lower polymorphism content in 13 genomic SSRs (0.38) than in 11 EST-SSRs (0.58; Shang et al., 2006). Deviations on PIC-values depicted in our study might be due to differences in plant material sources compared to former reports, which may influence the number of alleles detected at each SSR locus, though a potential influence of the lower number of SSRs loci used should not be discarded. When analysing null alleles presence and their effect on population structure, only 10 of the 406 loci harbored null alleles with a frequency higher than 0.20, and as such, no overestimation of F_{ST} due to null alleles was observed (mean $F_{ST}^{ENA} = 0.138$ vs. mean $F_{ST} = 0.136$).

Overall, considering the above analysis, one can predict that the SSRs loci selected to screen the rye accessions under study are suitable for downstream genetic diversity analysis.

Genetic Diversity between Cultivars and Landraces

Rye samples analyzed displayed different levels of diversity (number of alleles and heterozygosity). Overall, cultivars displayed lower mean allele number (3) than landraces (4), but similar genetic diversity was observed as denoted by levels of observed heterozygosity (average H_O Cultivars = 0.70, H_O Landraces = 0.67, **Table 3**). Accordingly, Matos et al. (2001) obtained a similar genetic diversity between Portuguese rye landraces and cultivars maintained in a Portuguese germplasm. Additionally, previous studies using RAPDs (Persson et al., 2001) and allozyme markers (Persson and von Bothmer, 2002, 2000) on Northern Europe rye showed that landraces and cultivars maintain roughly the same levels of genetic diversity. Overall, rye cultivars analyzed displayed similar genetic diversity as landraces, thus showing an unexpected absence of reduction in genetic diversity with increased improvement level, as recently reported for rye (Parat et al., 2016).

When narrowing analysis to landraces alone, *ex situ* collections revealed less allele per locus (3) than *in situ* (5) although showing similar heterozygosity (H_O *ex situ* = 0.67 vs. H_O *in situ* = 0.68, **Table 3**). This result is rather surprising, since generally higher genetic diversity would be expected within *in situ* compared to *ex situ* collections (e.g., Hou et al., 2012; Andrianasolo et al., 2013). In this study, *ex situ* and *in situ* accessions exhibit analogous genetic diversity levels, yet with different allelic diversity which may be indicative of a distinct subset of core alleles since *ex situ* are maintained under a steady environment without any selective pressure while *in situ* accessions are selected and grown by farmers in a subsistence agriculture context, being adapted under specific environments.

Population Structure

AMOVA showed that the majority of the genetic diversity lies within rye accessions with only little additional diversity present among groups or within groups. The large proportion of diversity found within accessions for the two types of groupings (cultivars vs. landraces; *ex situ* vs. *in situ*) suggests a high gene flow between accessions, which may be attributed to the wind-pollinated reproduction of rye allied with its outcrossing habit. One can therefore depict that rye-breeding system plays an important role in driving the generally observed high diversity within-accessions. Outcrossing plant species tend to have higher genetic variation within-populations, whereas selfing species or species with a mixed mating system are often genetically less variable (Nybom, 2004). Since rye is an outbreeder, negative to low inbreeding coefficients (F) were expected, which is in agreement with a recent rye study using SNPs (Hagenblad et al., 2016) but contrariwise to a SSRs study (Parat et al., 2016), where most populations displayed positive F -values which can be recorded in outcrossing populations with a strong population substructure. In the present study, low to moderate population differentiation among cultivars/*ex situ* was observed confirming the assumption of considerable population structure within cultivars and *ex situ* collection, which display steady population configuration as they are sequentially conserved in gene banks. Instead, *in*

situ collections showed no population differentiation (low F_{ST}), which may be suggestive of high gene flow between *in situ* populations from Northeast Portugal and/or an evidence of a single recent genetic source.

Generally, wind-pollinated species require a great isolation distance, since its windborne pollen may travel reasonably large distances, while in insect-pollinated species distances are related to insect activity (Richards, 1986). The low correlation between genetic and geographic distances obtained, is in agreement with earlier analysis (Hagenblad et al., 2016), and can be explained considering that pollen transfer between cultivated rye fields may easily occur as a result of its wind-pollination mode thus requiring relatively high isolation distances as established for other cross-pollinated crops (e.g., 1000–1600 m for cabbage, cauliflower). This lack of correlation between genetic and geographic distances is also in accordance with previous assumptions that attribute both temporal and ecological isolation for shaping rye's genetic diversity (Ma et al., 2004) in deterrence to spatial or geographic isolation. Moreover, individual-based clustering methods (STRUCTURE and DAPC) applied to our data highlights that genetic diversity scattering does not follow a geographic trend, regardless of being cultivars or landraces. Previous studies support the lack of clear structuring of the distribution of genetic diversity in different rye accessions depicted from geography, by using classical (i.e., allozymes, Persson and von Bothmer, 2000, 2002; RAPDs, Persson et al., 2001; AFLPs, Chikmawati et al., 2005, 2012; SSRs, Akhavan et al., 2010; Targońska et al., 2016) and modern (DaRT, Bolibok-Bragoszewska et al., 2014) molecular marker systems as well as organellar genome diversity analysis (Isik et al., 2007). However, a recent study using SNPs revealed a clustering of European landraces according to its geographic origins (Hagenblad et al., 2016). Likewise, Parat et al. (2016) managed to obtain two main subgroups indicating a differentiation according to both geography and end use, which can be described as “southern European forage rye” vs. “northern European grain rye.” In our data, despite no structuring was depicted along a geographical array, STRUCTURE analysis revealed two main subgroups indicating a differentiation between cultivars/*ex situ* with *in situ*/rye's crop wild relative. Genetic structuring between rye cultivars and landraces has been reported earlier (Persson and von Bothmer, 2000; Bolibok-Bragoszewska et al., 2014; Targońska et al., 2016), evidencing a genetic diversity assortment according to breeding status. In our data, population structuring obtained highlights a similar allelic diversity between *in situ* collections and *S. strictum*, while cultivars and *ex situ* collections do not seem to share alleles with CWR. Likewise, a multivariate approach using DAPC show that cultivars exhibit shared allelic diversity with *ex situ* accessions while *in situ* accessions might retain allelic diversity similar with the rye CWR ($K = 2$) or not ($K = 3$). Population structuring of cultivars with *ex situ* collections is further reinforced by the UPGMA analysis, which also highlights a highly supported clade consisting of *in situ* collections with the rye CWR. Altogether, our data provides evidences of cultivars and *ex situ* collections harboring a different genetic diversity in contrast to *in situ* accessions. As *ex situ* accessions offers a static genetic snapshot, reflecting

a population's adaptation to environmental conditions where they were collected, and considering that cultivars are a result of a controlled breeding process, a common gene pool can be depicted between cultivars and *ex situ* accessions here analyzed from eight European regions. Particularly, *in situ* populations screened in our study are originated from the region of excellence for rye production in Portugal, with diversity being preserved under farmer management. A recent study using dominant markers in three regional populations from the Northern Portugal shows the clustering of Portuguese populations in a different set than other rye cultivars, i.e., “Imperial,” “Dankowskie Zlote,” and the Portuguese cultivar “Alvão” (Santos et al., 2016), which is in accordance to our results. Considering the dissimilar population structuring of these Portuguese *in situ* populations with cultivars and *ex situ* collections, it cannot be ruled out that the region studied may be an hotspot of rye genetic diversity yet to be explored, and thus can provide valuable knowledge about genetic diversity resulting as part of the selection process adopted by local farmers through their agricultural practices.

***In situ* Collections As a Venue to a New Genetic Diversity: Portuguese Accessions As a Case Study**

Northeast Portuguese landraces have always been of great importance for local farmers, yet few studies have been performed for addressing its genetic diversity and population structure (Matos et al., 2001; Ribeiro et al., 2012; Santos et al., 2016). Adding other accessions with different geographic origins allows performing a comprehensive assessment of the genetic diversity of Portuguese landraces. Parat et al. (2016) by studying weedy, forage and grain ryes, which included a Portuguese forage landrace, determined a high fragmentation of membership coefficients in STRUCTURE analysis, which reflects a high diversity within Portuguese accessions. Indeed, our results support this former finding, with *in situ* Portuguese accessions displaying a high fragmentation with cultivars and *ex situ* collections. Interestingly, two *ex situ* accessions, “Sved” from Sweden and “Riodeva” from Spain, grouped along *in situ* accessions, as depicted by both UPGMA and model-based clustering (STRUCTURE and DAPC, $K = 3$) analysis. Interestingly, one cultivar (“Kungs II”) and one *ex situ* accession (“R780Spain”) originated from similar geographical regions as “Sved” and “Riodeva,” respectively, displayed a distinct genetic diversity from the former accessions and with the Portuguese *in situ* populations, thus highlighting the genetic distinctiveness of such rye accessions.

“Riodeva” is a Spanish rye inbred line resulting from a selection of a local landrace of Riodeva region (Lacadena et al., 1969), further bred over 30 generations of selfing (Gallego and Benito, 1997). In our study, admixture of “Riodeva” with Portuguese *in situ* populations was disclosed, in contrary with recent data using dominant markers (Santos et al., 2016), which showed no clustering of “Riodeva” with three regional Northern Portuguese populations. In our study, notwithstanding “Riodeva” displayed a lower mean number of alleles (Supplementary Table S3) comparing to the other rye accessions screened, a

higher genetic diversity was obtained with the codominant markers used. “Riodeva” inbred-line has been used as a control for aluminum (Al) susceptibility (Gallego and Benito, 1997) and studies on rye aluminum tolerance have included this accession (e.g., De Sousa et al., 2016). Therefore, admixture between Portuguese *in situ* accessions with “Riodeva” can only be related to an ancient genetic diversity that remained even after the selfing process or, to some extent, to a potential relation with Al sensitivity/tolerance in acidic soils, since rye is one of the most tolerant cereals to Al-stress, with Portuguese *in situ* populations could hold different Al-tolerance behavior yet to be uncovered. Nevertheless, further genomic studies will be needed to disclose genetic background shared between Portuguese *in situ* populations and “Riodeva,” along with the characterization of Al-tolerance behavior.

“Sved” accession is a Swedish rye landrace that clustered with another accession (also from Finnmarken, on the border between Norway and Sweden) in a way distinct from all other Scandinavian and European rye landraces (Hagenblad et al., 2016), demonstrating to be a distinct genotype not found earlier in other rye landraces, including those from the same geographical provenance. Hagenblad et al. (2016) linked its distinctiveness with historical human migrations, since Finnish farmers settled after leaving their native country in the sixteenth century (Ahokas, 2008). Considering Portuguese historic trading markets it cannot be ruled out a scenario of multiple rye introductions into Portuguese territory, especially from Northern Europe (i.e., Sweden). Indeed, historical records evidence rye grain being conveyed from Sweden to Portugal in the late eighteenth century (Ojal and Karvonen, 2012). Thus, the observed genetic similarity of a Swedish landrace with Portuguese accessions could be ascribed to a historical context. Overall, *in situ* accessions displayed a genetic kinship with a distinct landrace genotype (“Sved”) along with rye CWR, highlighting a hidden diversity on Portuguese rye gene pool yet to be uncovered.

Rye As a Rediscovered Crop: Implications to Genetic Diversity

In cross-fertilized species like rye, open-pollinated varieties (OPVs) constitute panmictic populations harboring high levels of genetic variation in their genetic build-up (Geiger and Miedaner, 2009). Moreover, improved varieties are grown in relatively uniform agricultural environments, which tend to narrow its genetic pool. Therefore, a high phenotypic variation exhibited by improved varieties might not always be a good predictor of the extent of their genetic variation (McCouch, 2004), and to surpass this concern both *in situ* and *ex situ* approaches are used to conserve the genetic diversity (Gepts, 2006). Our study shows unequivocally that *ex situ* collections display a similar genetic architecture with cultivars, sharing genetic material in a great extent with “Petkus,” one of rye’s heterotic pool (Hepting, 1978). Considering that not all parental lines could be uncovered from the cultivars used in our study, and that, as far as we know, none of the used cultivars have in its pedigree “Carsten” as a parental line, we can only infer about “Petkus” genetic pool. As such, *ex situ* collections screened do not present an effective alternative for supplementing “Petkus” pool. In contrast, *in situ* Portuguese

collections displayed a significant different genetic diversity than cultivars, including “Petkus,” thus being surprisingly distinctive genotypes from both cultivars and *ex situ* collections. Thus, Portuguese rye gene pool will be important for identifying new useful alleles that are linked to local adaptive processes and to its ends use, either forage or grain. This is an important finding as it sheds light onto new rye genotypes that remain to be uncovered and that could be useful for incrementing “Petkus” genetic pool. A recent study in wheat genetic diversity uncovers a pool of regional divergence, and highlights the need to increase regional breeding programs for the maintenance of crop diversity, rather than consolidation of commercial breeding alone (Novoselović et al., 2016). It is unquestionable that conservation of agricultural *in situ* genetic resources provide the genetic building blocks to improve plant varieties, and our findings unfold new *in situ* resources that will boost the improvement of new rye varieties delivering innovative information to rye breeders.

On Farm Conservation: A Growing Importance for Crop Diversity

Conservation of PGR through *ex situ* and *in situ* strategies have been implemented worldwide, yet in the last decades, there has been a growing interest in on farm conservation of landraces highlighted in the Convention of Biological Diversity (CBD), Agenda 21, and the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA), emphasizing the importance of on farm conservation as an essential component of sustainable agriculture. By assessing *in situ* genetic diversity one can determine which landraces may hold “new” genetic variation that could be useful to supplement crop cultivars diversity, in traits with agronomical importance (i.e., abiotic and biotic stress). Such genetic diversity is generally concentrated in centers of diversity as well as on farm conditions since landraces structure and dynamics result from both natural and human selection (Gepts, 2006). Such rich agro-historical heritage requires conservation policies to preserve the management of landraces in farmers’ fields where they originated, with the aim of maintaining the evolutive processes. The outcome of on farm conservation can be conceptualized as “evolutionary service” to agricultural and food systems, and to function, it depends on farmers’ preferences, knowledge, management, practices, and social organization.

On farm conservation of local landraces, particularly those found in our study, reflects a specific case, since we studied landraces not listed at both national and international gene banks. As such, we believe that the first intervention toward a conservation protocol is to implement a new national seed policy on landraces, which usually favors only varieties that are distinct, uniform and stable, discouraging the use of more heterogeneous, variable landraces. In agricultural systems as the one practiced in Northern Portugal rye fields, farmers typically save seed from one season to the next and may share seed with other farmers, being seed sourcing embedded in well-structured traditional systems with rules and expectations based on family and local social networks (Veteläinen et al., 2009). As such, on farm conservation protocols should be compatible with improved livelihoods and well-being among farmers who conserve such landraces, by incrementing ecosystem services at regional and national level

and by giving public benefits as a stimulus to promote specialized or novel marketing niches based on landraces and on local cultural heritage. Such landraces should be maintained as *in situ* genetic reserves and, besides *ex situ* conservation at national and international gene banks, an inventory periodically updated should be pursued to monitor on farm maintenance of landrace diversity.

Overall, our study successfully illustrates the significance of comparing the genetic diversity and structure of *ex situ* and *in situ* samples, along rye cultivars thus highlighting *in situ* collections from Northeast Portugal as new genetic resources being distinct genotypes to those reported for rye *ex situ* and cultivars. Identification of alleles/genes underlying such distinctive diversity would be of utmost importance for determining their usefulness for incorporating future rye breeding programs and to additionally propose on farm conservation policies at national level.

AUTHOR CONTRIBUTIONS

FM, AM, HO, and WV designed the research. FM carried out molecular work. FM, PV, and AB analyzed the data. FM wrote the manuscript and all authors improved upon versions. All authors read and approved the final manuscript.

REFERENCES

- Ahokas, H. (2008). "The ethnobotany of rye cultivation in Finland and its influence abroad," in *Plants that Tell Stories, Nordgen Seminar* (Hillerød).
- Akhavan, A., Saeidi, H., and Rahiminejad, M. R. (2010). Genetic diversity of *Secale cereale* L. in Iran as measured using microsatellites. *Genet. Resour. Crop Evol.* 57, 415–422. doi: 10.1007/s10722-009-9480-9
- Al-Beyroutiová, M., Sabo, M., Slezia, P., Dušinský, R., Bičrák, E., Hauptvogel, E., et al. (2016). Evolutionary relationships in the genus *Secale* revealed by DArTseq DNA polymorphism. *Plant Syst. Evol.* doi: 10.1007/s00606-016-1318-2. [Epub ahead of print].
- Almeida, L. A. V. (1955). A materia orgânica e a calagem na fertilização da Terra. *Bol. Ordem Eng.* 4, 1–16.
- Andrianasolo, D. N., Davis, A. P., Razafinarivo, N. J., Hamon, S., Rakotomalala, J. J., Sabatier, S. A., et al. (2013). High genetic diversity of *in situ* and *ex situ* populations of Madagascar coffee species: further implications for the management of coffee genetic resources. *Tree Genet. Genomes* 9, 1295–1312. doi: 10.1007/s11295-013-0638-4
- Baum, M., and Appels, R. (1991). The cytogenetic and molecular architecture of chromosome 1R - one of the most widely utilized sources of alien chromatin in wheat varieties. *Chromosoma* 101, 1–10. doi: 10.1007/BF00360680
- Bettencourt, E., and Carnide, V. (1998). "Status of *Secale* collections in Portugal-conservation, characterization, evaluation and documentation," in *Challenges in Rye Germplasm Conservation*, eds T. Gass, W. Podyma, J. Puchalski, and S. A. Eberhart (Rome: Proceedings of an international conference on crop germplasm conservation with special emphasis on rye, International Plant Genetic Resources Institute), 134–137.
- Bolibok-Bragoszewska, H., Targońska, M., Bolibok, L., Kilian, A., and Rakoczy-Trojanowska, M. (2014). Genome-wide characterization of genetic diversity and population structure in *Secale*. *BMC Plant Biol.* 14:184. doi: 10.1186/1471-2229-14-184
- Bolibok-Bragoszewska, H., Zabierzewska, N., Hromada-Judycka, A., and Krzewska, L. (2012). 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.* 40, 204–209. doi: 10.1556/CRC.40.2012.2.4

ACKNOWLEDGMENTS

The authors would like to acknowledge the following institutions and individuals for providing seeds: Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK, R, Germany) and Nordic Genetic Resource Center (NordGen, NGB, Sweden), Professor Dr. César Benito of Complutense University of Madrid (Spain) for kindly providing "Riodeva" accession, and Professor José Eduardo Lima Brito from Universidade de Trás-os-Montes e Alto Douro (UTAD, Portugal) for providing "Alvão" accession. The authors would like to thank the Associate Editor Anna Maria Mastrangelo and the two Reviewers, for their thoughtful comments, which greatly improved our paper. The authors would also like to acknowledge Ezequiel Coelho for the help on the ArcGIS graphical representation. The work was funded by the Portuguese Ministry of Agriculture Forests and Rural Development under the PRODER Measure 4.1 Project PA 55570, and by Fundação para a Ciência e a Tecnologia (FCT) LEAF Unit (UID/AGR/04129/2013).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01334>

- Brownstein, M. J., Carpten, J. D., and Smith, J. R. (1996). Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* 20, 1004–1006, 1008–1010.
- Camacho-Villa, T., Maxted, N., Scholten, M., and Ford-Lloyd, B. (2005). Defining and identifying crop landraces. *Plant Genet. Resour. Charact. Util.* 3, 373–384. doi: 10.1079/PGR200591
- Cavalli-Sforza, L. L., and Edwards, A. W. (1967). Phylogenetic analysis. Models and estimation procedures. *Am. J. Hum. Genet.* 19, 233–257. doi: 10.2307/2406616
- Chapuis, M.-P., and Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol.* 24, 621–631. doi: 10.1093/molbev/msl191
- Chikmawati, T., Miftahudin, M., Skovmand, B., and Gustafson, J. P. (2012). Amplified fragment length polymorphism-based genetic diversity among cultivated and weedy rye (*Secale cereale* L.) accessions. *Genet. Resour. Crop Evol.* 59, 1743–1752. doi: 10.1007/s10722-012-9796-8
- Chikmawati, T., Skovmand, B., and Gustafson, J. P. (2005). Phylogenetic relationships among *Secale* species revealed by amplified fragment length polymorphisms. *Genome* 48, 792–801. doi: 10.1139/g05-043
- Culley, T. M., Stamper, T. I., Stokes, R. L., Brzyski, J. R., Hardiman, N. A., Klooster, M. R., et al. (2013). An efficient technique for primer development and application that integrates fluorescent labeling and multiplex PCR. *Appl. Plant Sci.* 1:1300027. doi: 10.3732/apps.1300027
- Delipavlov, D. (1962). *Secale rhodopaenum* Delipavlov - A new species of rye from the Rhodope Mountains. *Dokl. Bolg. Akad. Nauk* 15, 407–411.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* 39, 1–38.
- De Sousa, A., AbdElgawad, H., Han, A., Teixeira, J., Matos, M., and Fidalgo, F. (2016). Oxidative metabolism of rye (*Secale cereale* L.) after short term exposure to aluminum: uncovering the glutathione-ascorbate redox network. *Front. Plant Sci.* 7:685. doi: 10.3389/fpls.2016.00685
- Earl, D. A., and VonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Evans, G. M. (1995). “Rye” in *Evolution of Crop Plants*, eds J. Smartt and N. Simmonds (Harlow: Longman Scientific & Technical), 166–170.
- Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x
- Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491.
- Food and Agricultural Organization of the United Nations (2016). *FAO Statistics Division: Food and Agricultural Commodities Production*. Available online at: <http://faostat3.fao.org/download/Q/QC/E> [Accessed July 21, 2016].
- Fischer, S., Melchinger, A. E., Korzun, V., Wilde, P., Schmiedchen, B., Möhring, J., et al. (2010). Molecular marker assisted broadening of the Central European heterotic groups in rye with Eastern European germplasm. *Theor. Appl. Genet.* 120, 291–299. doi: 10.1007/s00122-009-1124-0
- Frederiksen, S., and Petersen, G. (1998). A taxonomic revision of *Secale* (Triticeae, Poaceae). *Nord. J. Bot.* 18, 399–420. doi: 10.1111/j.1756-1051.1998.tb01517.x
- Gailite, A., Gaile, A., Gaile, I., Voronova, A., Veinberga, I., Kokare, A., et al. (2013). Genotypic assessment of the Latvian rye (*Secale Cereale* L.) collection. *Proc. Latv. Acad. Sci. Sect. B Nat. Exact Appl. Sci.* 3, 264–267. doi: 10.2478/prolas-2013-0046
- Gallego, F. J., and Benito, C. (1997). Genetic control of aluminium tolerance in rye (*Secale cereale* L.). *Theor. Appl. Genet.* 95, 393–399. doi: 10.1007/s001220050575
- Geiger, H., and Miedaner, T. (2009). “Rye Breeding,” in *Cereals*, ed M. J. Carena (New York, NY: Springer US), 157–181.
- Gepts, P. (2006). Plant genetic resources conservation and utilization. *Crop Sci.* 46, 2278. doi: 10.2135/cropsci2006.03.0169gas
- Greene, S. L., Kisha, T. J., Yu, L.-X., and Parra-Quijano, M. (2014). Conserving plants in gene banks and nature: investigating complementarity with *Trifolium thompsonii* Morton. *PLoS ONE* 9:e105145. doi: 10.1371/journal.pone.0105145
- Hackauf, B., and Wehling, P. (2002). Identification of microsatellite polymorphisms in an expressed portion of the rye genome. *Plant Breed.* 121, 17–25. doi: 10.1046/j.1439-0523.2002.00649.x
- Hagenblad, J., Oliveira, H. R., Forsberg, N. E. G., and Leino, M. W. (2016). Geographical distribution of genetic diversity in *Secale* landrace and wild accessions. *BMC Plant Biol.* 16:23. doi: 10.1186/s12870-016-0710-y
- Hepting, L. (1978). Analyse eines 7x7-Sortendiales zur Ermittlung geeigneten Ausgangsmaterials für die Hybridzüchtung bei Roggen. *Z. Pflanzenzüchtung* 80, 188–197.
- Hill, W. G. (1996). *Genetic Data Analysis II*. By Bruce S. Weir, Sunderland, Massachusetts. Sinauer Associates, Inc. 445 pages. ISBN 0-87893-902-4. *Genet. Res.* 68, 187.
- Holleley, C. E., and Geerts, P. G. (2009). Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. *Biotechniques* 46, 511–517. doi: 10.2144/000113156
- Hou, B., Tian, M., Luo, J., Ji, Y., Xue, Q., and Ding, X. (2012). Genetic diversity assessment and *ex situ* conservation strategy of the endangered *Dendrobium officinale* (Orchidaceae) using new trinucleotide microsatellite markers. *Plant Syst. Evol.* 298, 1483–1491. doi: 10.1007/s00606-012-0651-3
- Isik, Z., Parmaksiz, I., Coruh, C., Geylan-Su, Y. S., Cebeci, O., Beecher, B., et al. (2007). Organellar genome analysis of rye (*Secale cereale*) representing diverse geographic regions. *Genome* 50, 724–734. doi: 10.1139/G07-052
- Jakobsson, M., and Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23, 1801–1806. doi: 10.1093/bioinformatics/btm233
- Jensen, H. R., Dreiseitl, A., Sadiki, M., and Schoen, D. J. (2012). The Red Queen and the seed bank: pathogen resistance of *ex situ* and *in situ* conserved barley. *Evol. Appl.* 5, 353–367. doi: 10.1111/j.1752-4571.2011.00227.x
- Jombart, T. (2008). Adegnet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi: 10.1093/bioinformatics/btn129
- Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11:94. doi: 10.1186/1471-2156-11-94
- Kamvar, Z. N., Brooks, J. C., and Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* 6:208. doi: 10.3389/fgene.2015.00208
- Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281. doi: 10.7717/peerj.281
- Kobylyanskij, V. (1975). K sistematičke i filogenii roda *Secale* L. - Bjull. *VIR* 48, 64–71.
- Lacadena, J., Sánchez-Monge, E., and Villena, L. (1969). Selection for self-fertility in rye inbred lines. *An. Ia Estac. Exp. Aula Dei* 10, 846–855.
- Li, Q., He, T., and Xu, Z. (2005). Genetic evaluation of the efficacy of *in situ* and *ex situ* conservation of *Parashorea chinensis* (Dipterocarpaceae) in Southwestern China. *Biochem. Genet.* 43, 387–406. doi: 10.1007/s10528-005-6778-y
- Ma, R., Yli-Mattila, T., and Pulli, S. (2004). Phylogenetic relationships among genotypes of worldwide collection of spring and winter ryes (*Secale cereale* L.) determined by RAPD-PCR markers. *Hereditas* 140, 210–221. doi: 10.1111/j.1601-5223.2004.01844.x
- Matos, M., Pinto-Carnide, O., and Benito, C. (2001). Phylogenetic relationships among Portuguese rye based on isozyme, RAPD and ISSR markers. *Hereditas* 134, 229–236. doi: 10.1111/j.1601-5223.2001.00229.x
- McCouch, S. (2004). Diversifying selection in plant breeding. *PLoS Biol.* 2:e347. doi: 10.1371/journal.pbio.0020347
- Meirmans, P. G., and Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Mol. Ecol.* 4, 792–794. doi: 10.1111/j.1471-8286.2004.00770.x
- Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852. doi: 10.1038/nrg3605
- Miedaner, T. (1997). Breeding wheat and rye for resistance to *Fusarium* diseases. *Plant Breed.* 116, 201–220. doi: 10.1111/j.1439-0523.1997.tb00985.x
- Negri, V., and Tiranti, B. (2010). Effectiveness of *in situ* and *ex situ* conservation of crop diversity. What a *Phaseolus vulgaris* L. landrace case study can tell us. *Genetica* 138, 985–998. doi: 10.1007/s10709-010-9485-5
- Nei, M. (1972). Genetic distance between populations. *Am. Nat.* 106, 283–292. doi: 10.1086/282771
- Novoselović, D., Bentley, A. R., Šimek, R., Dvojković, K., and Sorrells, M. E., Gosman, N., et al. (2016). Characterizing Croatian wheat germplasm diversity and structure in a European context by DArT markers. *Front. Plant Sci.* 7:184. doi: 10.3389/fpls.2016.00184
- Nybom, H. (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol. Ecol.* 13, 1143–1155. doi: 10.1111/j.1365-294X.2004.02141.x
- Ojal, J., and Karvonen, L. (2012). Assessing the reliability of the sound toll accounts: comparing the data to the Swedish and Portuguese sources. *Early Mod. Econ. Trade Work*. Available online at: https://www.jyu.fi/hum/laitokset/hie/en/research/economy_and_trade/emet_abstracts/ojala_karvonen
- Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412
- Parat, F., Schwertfurm, G., Rudolph, U., Miedaner, T., Korzun, V., Bauer, E., et al. (2016). Geography and end use drive the diversification of worldwide winter rye populations. *Mol. Ecol.* 25, 500–514. doi: 10.1111/mec.13495
- Park, S. D. E. (2001). *Trypanotolerance in West African cattle and the Population Genetic Effects of Selection*. Ph.D. thesis, University of Dublin, Ireland.
- Peakall, R., and Smouse, P. E. (2006). GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol.* 6, 288–295. doi: 10.1111/j.1471-8286.2005.01155.x
- Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460
- Persson, K., Díaz, O., and von Bothmer, R. (2001). Extent and patterns of RAPD variation in landraces and cultivars of rye (*Secale cereale* L.) from northern Europe. *Hereditas* 134, 237–243. doi: 10.1111/j.1601-5223.2001.00237.x
- Persson, K., and von Bothmer, R. (2000). Assessing the allozyme variation in cultivars and Swedish landraces of rye (*Secale cereale* L.). *Hereditas* 132, 7–17. doi: 10.1111/j.1601-5223.2000.00007.x

- Persson, K., and von Bothmer, R. (2002). Genetic diversity amongst landraces of rye (*Secale cereale* L.) from northern Europe. *Hereditas* 136, 29–38. doi: 10.1034/j.1601-5223.2002.1360105.x
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Putman, A. I., and Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecol. Evol.* 4, 4399–4428. doi: 10.1002/ece3.1305
- Rambaut, A. (2014). *FigTree v1.4.2*. Available online at: <http://tree.bio.ed.ac.uk/software/figtree/>
- Ribeiro, M., Seabra, L., Ramos, A., Santos, S., Pinto-Carnide, O., Carvalho, C., et al. (2012). Polymorphism of the storage proteins in Portuguese rye (*Secale cereale* L.) populations. *Hereditas* 149, 72–84. doi: 10.1111/j.1601-5223.2012.02239.x
- Rice, W. (1989). Analyzing table of statistical tests. *Evolution* 43, 223–225. doi: 10.2307/2409177
- Richards, A. (1986). *Plant Breeding Systems*. Winchester, MA: Allen & Unwin Inc.
- Rizvi, S., and Scoles, G. (2014). “Small grains: barley, oat and rye,” in *Handbook on Agriculture, Biotechnology and Development*, eds S. Smyth, P. Philips, and D. Castle (Cheltenham: Edward Edgar Publishing), 770–792.
- Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol.* 4, 137–138. doi: 10.1046/j.1471-8286.2003.00566.x
- Roshevitz, R. (1947). A monograph of the wildweedy and cultivated species of rye. *Acta Inst. Bot. Nomine Acad. Sci. USSR Ser.* 6, 105–163.
- Rousset, F. (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* 8, 103–106. doi: 10.1111/j.1471-8286.2007.01931.x
- Saal, B., and Wricke, G. (1999). Development of simple sequence repeat markers in rye (*Secale cereale* L.). *Genome* 42, 964–972. doi: 10.1139/gen-42-5-964
- Santos, E., Matos, M., Silva, P., Figueiras, A. M., Benito, C., and Pinto-Carnide, O. (2016). Molecular diversity and genetic relationships in *Secale*. *J. Genet.* 95, 273–281. doi: 10.1007/s12041-016-0632-3
- Schlegel, R. H. J. (ed.). (2014). “Physiology,” in *Rye: Genetics, Breeding, and Cultivation* (Boca Raton, FL: CRC Press; Taylor & Francis Group) 51.
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18, 233–234. doi: 10.1038/72708
- Sencer, H. A., and Hawkes, J. G. (1980). On the origin of cultivated rye. *Biol. J. Linn. Soc.* 13, 299–313. doi: 10.1111/j.1095-8312.1980.tb00089.x
- Shang, H.-Y., Wei, Y.-M., Wang, X.-R., and 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, 685–691. doi: 10.1590/S1415-47572006000400018
- Sint, D., Raso, L., and Traugott, M. (2012). Advances in multiplex PCR: balancing primer efficiencies and improving detection success. *Methods Ecol. Evol.* 3, 898–905. doi: 10.1111/j.2041-210X.2012.00215.x
- Targońska, M., Bolibok-Bragoszewska, H., and Rakoczy-Trojanowska, M. (2016). Assessment of genetic diversity in *Secale cereale* based on SSR markers. *Plant Mol. Biol. Report.* 34, 37–51. doi: 10.1007/s11105-015-0896-4
- R Developmental Core Team (2015). *R Version 3.2.3*. Vienna R Found. Statistical Computing. Available online at: www.r-project.org
- Thomas, M. R., Matsumoto, S., Cain, P., and Scott, N. S. (1993). Repetitive DNA of grapevine: classes present and sequences suitable for cultivar identification. *Theor. Appl. Genet.* 86, 173–180. doi: 10.1007/bf00222076
- Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., and Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol.* 4, 535–538. doi: 10.1111/j.1471-8286.2004.00684.x
- Vaughan, J., and Geissler, C. (2009). “Grain Crops: Rye,” in *The New Oxford Book of Food Plants*, eds J. Vaughan and C. Geissler (New York, NY: Oxford University Press Inc.), 6–7.
- Veteläinen, M., Negri, V., and Maxted, N. (2009). *European Landraces on Farm Conservation, Management and Use. Bioversity Technical Bulletin No. 15*. Rome: Bioversity International.
- Weir, B. S., and Cockerham, C. C. (1984). Estimating *F*-Statistics for the analysis of population structure. *Evolution* 38, 1358. doi: 10.2307/2408641

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Monteiro, Vidigal, Barros, Monteiro, Oliveira and Viegas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.