



# MARTIN-LUTHER-UNIVERSITÄT HALLE-WITTENBERG

Naturwissenschaftliche Fakultät III  
Institut für Agrar- und Ernährungswissenschaften

## Masterarbeit

Über das Thema

---

# Population genetic analysis of self-incompatibility in *Secale*

---

zur Erlangung des akademischen Grades  
*Master of Science (MSc)* in Nutzplanzenwissenschaften

vorgelegt am 04.04.2024

vorgelegt von

cand. agr. Max Pfeifer,

geboren am 29.03.1997 in Attendorn

1. Gutachter: Prof. Dr. Klaus Pillen

2. Gutachter: Dr. Steven Dreißig

## Table of Contents

1. Introduction.....	1
1.1. Genus <i>Secale</i> L.....	1
1.1.1. Taxonomy.....	1
1.1.2. Origin and Domestication of <i>Secale</i> .....	3
1.1.3 Self-incompatibility under the influence of S- & Z-Locus .....	6
1.2 Outline through the history of population genetic analyses.....	8
1.2.1. Population structure analysis .....	10
1.2.2. Quantitative trait loci (QTL) analysis.....	10
1.3. Aims and objectives.....	12
2. Material and Methods.....	13
2.1. Plant Material, GBS Read Alignment and variant calling.....	13
2.2. Population Structure Analysis.....	14
2.2.1. Principal component analysis (PCA) .....	15
2.2.2. Cross-entropy criterion.....	15
2.2.3. Ancestry Coefficient.....	16
2.2.4. Nucleotide diversity ( $\pi$ ) .....	16
2.2.5. Fixation Index ( $F_{ST}$ ).....	17
2.3. Inbreeding coefficient ( $F$ ).....	19
2.4. Quantitative trait loci (QTL) analysis.....	20
2.4.1. Genome-wide association studies (GWAS) .....	20
2.4.2. Linkage Disequilibrium (LD) decay estimates and Gene Annotations .....	21
3. Results .....	23
3.1. Population Structure Analysis.....	23
3.1.1. Principal component analysis (PCA) .....	23
3.1.2. Cross-entropy criterion.....	25
3.1.3. Ancestry Coefficient analysis .....	26
3.1.4. Fixation Index ( $F_{ST}$ ).....	28
3.1.5. Nucleotide diversity.....	28
3.2. Inbreeding coefficient ( $F$ ).....	29

## Table of Contents

---

3.3. Quantitative trait loci (QTL) analysis.....	31
3.3.1. Genome-wide association studies (GWAS) .....	31
3.3.2. Linkage Disequilibrium (LD) decay estimates and Gene Annotations .....	34
4. Discussion .....	47
4.1. Understanding the Population Structure of <i>Secale</i> .....	47
4.2. Inbreeding coefficient as a measurement of self-incompatibility .....	50
4.3. Discussion of candidate genes for self-incompatibility .....	52
4.3. Suggestions for future work .....	55
5. Summary.....	56
5.1 Zusammenfassung.....	57
6. References .....	58
7. Appendix.....	69
A1. The S- & Z-locus.....	69
A2. Passport data.....	70
A3. Unfiltered single nucleotide variant data.....	91
A4. Principal component analysis (PCA) – Variance Proportion.....	92
A5. Calculation of Cross-entropy criterion .....	93
A6. LD decay – average Pair-wise correlation ( $r^2$ ) values.....	94
A7. LD decay at SNPs – average Pair-wise correlation ( $r^2$ ) values.....	96
8. Indices.....	100
8.1 List of abbreviations .....	100
8.2 List of figures .....	102
8.3 List of tables.....	102
9. Acknowledgements .....	103
10. Eidesstattliche Erklärung .....	104

## 1. Introduction

---

# 1. Introduction

## 1.1. Genus *Secale* L.

Rye stands out among other cereals like wheat and barley for its severe cold tolerance and higher yield in challenging conditions, such as poor and moderate soils and drought stress (Schittenhelm et al. 2014, Skuza et al. 2018). Its versatility is reflected in its primary uses, like bread making, alcohol production, and animal feed. Recently, it is also earning attention as a biomass crop (Bolibok-Brągoszewska et al. 2014). Wild rye species and subspecies serve as valuable research materials for research aimed at expanding genetic diversity in cultivated rye. Their distinctive genetics, marked by noble traits such as high protein content, disease resistance, and abiotic stress tolerance, make rye a valuable crop in many countries and a globally important genetic resource of genes for enhancing tetraploid and hexaploid wheat, as well as triticale improvement (Daskalova and Spetsov 2020, Li et al. 2021).

While rye was historically significant in the Middle Ages for sustaining European populations due to its robust winter hardiness (Schlegel 2014), it continues to play a pivotal role in global agriculture (Hagenblad et al. 2016). Agricultural production of rye spans six continents, with over 15 million metric tons grown on 4.4 million hectares in 2020 (Brzozowski et al. 2023).

### 1.1.1. Taxonomy

The genus *Secale* L. includes perennial and annual species, cross-pollinators and self-pollinators, as well as cultivated and wild species and species that are important as weeds (Vences et al. 1987). Different taxonomies have been proposed for the genus *Secale* over time (for example Vavilov 1926, Roshevitz 1947, Sencer and Hawkes 1980, Frederiksen and Petersen 1998), but the taxonomy remains inconclusive due to the lack of clearly assignable features (Al-Beyroutiová et al. 2016, Hagenblad et al. 2016), caused by the ability of each species to intercross and yielding (partially) fertile hybrids (Stutz 1972). Over and beyond, all taxa in the genus are diploids with  $2n = 14$  chromosomes (Bustos and Jouve 2002) with a haploid genome size between 7.6 Gb - 8 Gb, depending on the species (Rabanus-Wallace et al. 2021). For a taxonomic classification with synonyms of different *Secale* species according to Frederiksen and Petersen (1998) see Table 2.

Vavilov (1926) first classified the genus *Secale* L. by categorizing the putative species on the basis of morphological characteristics, life cycle and distribution area. As a result, Vavilov obtained four species that can be assigned to the genus *Secale*: *S. africanum* Stapf, *S. cereale* L., *S. fragile* M.Bieb., and *S. montanum* Guss.. According to Al-Beyroutiová et al. (2016), Roshevitz showed 1947, that there are also other criteria to be considered in taxonomic classification, dividing the genus *Secale* L. into fourteen species, in addition to over 20 intra-specific taxa, based on their crossability (Roshevitz 1947). He categorized several of the annual, weedy intermediates as separate species, namely, *S. ancestrale* (Zhuk.) Zhuk., *S. dighoricum* (Vav.) Roshev., *S. segetale* (Zhuk.) Roshev., *S. afghanicum* (Vav.) Roshev. and *S. vavilovii* Grossh. Later, the similar cytology between these weedy taxa and its resemblance to that of domesticated *Secale*, with whom the weeds are inter-fertile, led Sencer (1975), Kobyljanskij

## 1. Introduction

---

(1983), Hammer (1990), Frederiksen and Petersen (1998), Hancock (2004) to consider them subspecies of *S. cereale*, rather than independent species (Feldman and Levy 2023).

Khush published a further taxonomic classification of the genus *Secale* in 1962. In contrast to Vavilov, he divided the genus into five species with respective subspecies based on cytogenetic studies. That said, *S. anatolicum* Boiss., *S. kuprianovii* Grossh. & Nevski have been classified subspecies of *S. montanum*. *S. ancestrale* Zhuk., *S. afghanicum* (Vavilov) Roshev., *S. dighoricum* (Vavilov) Roshev., and *S. segetale* (Zhuk.) Roshev. however, were classified as subspecies of *S. cereale*. Finally, he named the independent species *S. sylvestre* Host, *S. africanum* Stampf and *S. vavilovii* Grossh. In 1987, a further classification system was published by Hammer, in which he divided the genus into four species, the annuals and autogamous *S. sylvestre* and *S. vavilovii*, the annual outbreeder *S. cereale* and the perennial *S. strictum*. So, you could say that the number of *Secale* species has fluctuated over the years in which other authors also made classifications for *Secale* (see Table 1).

**Table 1 Taxonomic relationships within the genus Secale.** (source: Daskalova and Spetsov 2020)

Vavilov, 1917 [4]	Roshevitz, 1947 [40]	Sencer & Hawkes, 1980 [6]; Fredriksen & Petersen, 1997 [14]; Jaaska, 1998 [38]; Chikmawati et al. 2005 [3]; Ren et al. 2011 [31]; Al-Beyrouti et al. 2016 [35]
<b>4 species</b>	<b>14 species in 3 sections:</b>	<b>3 species</b>
<i>S. africanum</i> Stapf	- <i>Cerealia</i> Rosh.	<i>S. cereale</i> L.
<i>S. cereale</i> L.	- <i>Sylvestria</i> Rosh.	<i>S. strictum</i> Presl
<i>S. fragile</i> Marsch	- <i>Kuprianovia</i> Rosh.	<i>S. sylvestre</i> Host
<i>S. montanum</i> Gussone		
Khush, 1962 [15]	Clayton et al. 2016 [41]	Vences et al. 1987 [1], Tang et al. 2011 [2], GRIN 2019 [37]
<b>5 species</b>	<b>8 species</b>	<b>4 species</b>
<i>S. cereale</i> L.	<i>S. cereale</i>	<i>S. cereale</i> (subsp. <i>afghanicum</i> , <i>ancestrale</i> , <i>dighoricum</i> , <i>segetale</i> , <i>cereale</i> )
<i>S. sylvestre</i> Host	<i>S. segetale</i>	
<i>S. montanum</i> Guss.	<i>S. sylvestre</i>	<i>S. strictum</i> (subsp. <i>africanum</i> , <i>anatolicum</i> , <i>ciliatoglume</i> , <i>kuprianovii</i> , <i>strictum</i> (syn. <i>S. montanum</i> ))
<i>S. africanum</i> Stapf	<i>S. vavilovii</i>	
<i>S. vavilovii</i> Grossh.	<i>S. africanum</i>	<i>S. sylvestre</i> Host
	<i>S. anatolicum</i>	<i>S. vavilovii</i> Grossh.
	<i>S. ciliatoglume</i>	
	<i>S. montanum</i>	

Most recently, the genus *Secale* was divided into the three species *S. strictum*, *S. sylvestre* and *S. cereale* by various authors (Sencer and Hawkes 1980, Frederiksen and Petersen 1998, Schreiber et al. 2019), although the classification with four species (*S. cereale*, *S. strictum*, *S. vavilovii*, *S. sylvestre*) according to Germplasm Resource Information Network (GRIN) is frequently used, too (Shang et al. 2006, Hawliczek et al. 2023, Bolibok-Brągoszewska and Rakoczy-Trojanowska 2015). In the classification with three species, *S. strictum* includes all wild perennial taxa exhibiting significant morphological similarities and cytogenetic affinity to each other. In contrast, *S. sylvestre*, a wild annual species, is geographically, ecologically, and reproductively isolated from *S. strictum*, despite demonstrating cytogenetic affinity under experimental conditions. Lastly, *S. cereale*, contains the annual wild, weedy, and domesticated types (*S. cereale* subsp. *cereale*) as well as *S. cereale* subsp. *vavilovii*, which was previously classified as an independent species in previous studies (Feldman and Levy 2023).

## 1. Introduction

---

And although taxonomists still disagree on the question of species classification (Al-Beyroutiová et al. 2016), the classification suggested by Frederiksen and Petersen (1998) is currently accepted by many taxonomists (Feldman and Levy 2023) and is used in this study.

**Table 2 The species and subspecies of *Secale* L. According to Frederiksen and Petersen (1998) (source: Feldman and Levy 2023)**

Species	Synonyms	Subspecies	Synonyms	Spike rachis	Pollination mode	Growth habit
<i>Strictum</i> (C. Presl) C. Presl	<i>Triticum strictum</i> C. Presl; <i>Secale montanum</i> Guss.; <i>T. cereale</i> var. <i>montanum</i> (Guss.) Kuntze; <i>S. cereale</i> var. <i>montanum</i> (Guss.) Fiori; <i>Frumentum Secale</i> E. H. L. Krause	<i>Strictum</i>	<i>Secale anabolism</i> Boiss.; <i>S. dalmaticum</i> Vis.; <i>S. serbicum</i> PanEic ex Griseb.; <i>S. kuprianovii</i> Grossh.; <i>S. chaldaicum</i> Fed.; <i>S. daralagesi</i> Tumanian; <i>S. rhodopaeum</i> Delip.; <i>S. perenne</i> Hortor in Fisch. & C. A. Mey.	Fragile	Allogamous	Perennial
		<i>Africanum</i> (Stapf) K. Hammer	<i>S. africanum</i> Stapf; <i>S. montanum</i> ssp. <i>africanum</i> (Stapf) Kobyl.	Fragile	Autogamous	Perennial
<i>Sylvestre</i> Host	<i>Triticum silvestre</i> (Host) Asch. & Graebn.; <i>T. campestre</i> (Kit.) Kit. ex Roem. & Schult.; <i>Secale fragile</i> M. Bieb.; <i>S. cereale</i> M. Bieb.; <i>S. glaucum</i> d'Urv.; <i>S. spontaneum</i> Fisch.	–	–	Fragile	Autogamous	Annual
<i>Cereale</i> L.		<i>Cereale</i>	<i>S. cereale</i> ssp. <i>indo-europaeum</i> Antropov & Antropova in Roshev; <i>S. trijlorum</i> P. Beauv.	Tough	Mostly allogamous	Annual
		<i>Ancestrale</i> Zhuk.	<i>S. ancestrale</i> (Zhuk.) Zhuk.	Fragile or tough	Mostly allogamous	Annual

### 1.1.2. Origin and Domestication of *Secale*

The origin of rye and its spread across the world is determined by complicated evolutionary processes and the gradual spread of agriculture at the time. Diverging from its wild ancestors, the *Secale* species, including the cultivated rye, exhibit recent divergence times compared to other cereals like barley (*Hordeum*) and wheat (*Triticum*).

The geographical origin of rye may align with the broader context of the Fertile Crescent, a region which extends across the Levant and Anatolia where wild progenitors of rye, barley, and wheat also share their centers of diversity (Zohary et al. 2012). The Russian botanist, geneticist, and research traveler Nikolai Ivanovič Vavilov (\*1887, †1943) played a pivotal role in understanding crop diversity. His field studies led him to develop the concept of centers of origin of our cultivated plants. These centers were identified in regions of the world where the genetic diversity in related wild species is highest (Schreiber et al. 2021). For Vavilov it was clear that the primary center of diversity could be regarded as the center of origin for both the genus *Secale* and for rye as a cereal cultivated by humankind (Hillman 1978).

## 1. Introduction

---

Vavilov's observations led to the hypothesis that cultivated rye emerged as a hybrid of *S. cereale* subsp. *vavilovii* and *S. strictum*, a theory later challenged by global ancestry analysis, which did not support the notion that domesticated rye possess a major ancestry component tracing back to *S. strictum* (Schlegel 2014, Schreiber et al. 2021). Vavilov reported that domestication of rye evolved only as a secondary crop in wheat and barley as wild populations of rye invaded fields growing these crops in southwestern Asia (Middle East) resulting in weedy ryes with varying degrees of rachis brittleness (Rebong et al. 2024). This phenomenon persisted through regions with poorer soils and colder climates, in which rye outperformed wheat and barley, eventually leading to intentional cultivation (Zohary et al. 2012). Relying on the observation that the most significant genetic diversity of *S. cereale* existed in southwest Asia, (Vavilov 1917, 1926) identified this region as the primary center of origin for domesticated rye. At the same time, he considered Afghanistan and Tadzhikistan as a secondary center of variation for this crop (Feldman and Levy 2023). Ultimately one suspects the center of origin of cultivated rye is located around Tabriz (Iran) towards the Black Sea, but also a region east of Iran toward Afghanistan has been considered as a potential center of origin (Rebong et al. 2024). This consideration arises from the correlation of these regions with the contemporary distribution of weedy rye, as indicated by (Schlegel 2014). A recent investigation utilizing simple sequence repeat marker analysis by (Maraci et al. 2018) revealed that domesticated populations originating from southwestern Asia (Middle East) exhibit the highest genetic diversity, supporting the idea of this area being the true center of origin for cultivated rye (Rebong et al. 2024).

According to (Sencer and Hawkes 1980), archaeological evidence indicates the early transportation of rye from northeastern Anatolia to the Black Sea region, intertwined with wheat and barley, around the 8th and 5th centuries BCE, by which it subsequently underwent domestication and became a useful grain (Rebong et al. 2024, Schlegel 2014). Based on their synthetic analysis that considered evidence from the fields of morphology, taxonomy, ecology, phytogeography, reproductive biology, genetics, cytology, palaeoethnobotany, philology, phylogeny and evolution, they delimited the area more closely to Mt. Ararat and Lake Van in eastern Turkey as the geographic origin of domesticated rye (Feldman and Levy 2023).

From the center of origin, two possibly parallel migration routes of the species into Europe have been suggested: the first route via Russia to Poland and Germany, leading to the subsequent distribution of rye across most of Europe, and the second route via Turkey and across the Balkan Peninsula (Isik et al. 2007, Ma et al. 2004). Rye was later introduced to America by the first European settlers, Chinese and subsequently Japanese ryes originated from Turkey (Bolibok-Brągoszewska and Rakoczy-Trojanowska 2015, Ma et al. 2004).

In conclusion, the intricate journey of rye from a weedy synanthrope to intentional cultivation spans millennia and multiple continents. The genetic diversity and evolutionary history of rye underscore the complex interplay between human agricultural practices and the adaptability of this vital cereal crop.

The domestication of rye was influenced by the northward and eastward movement of wheat and barley into regions with harsh climatic conditions and poorer soil, leading to the subsequent

## 1. Introduction

---

domestication of weedy rye (Feldman and Levy 2023). In the broader context of rye evolution, the species shares a common ancestor with wheat and diverged seven million years ago. Both lineages and the barley lineage diverged from a common Triticeae ancestor around 11 million years ago (Huang et al. 2002, Martis et al. 2013).

Over the course of history, almost all known *Secale* species were considered direct ancestors of today's cultivated rye (*Secale cereale* subsp. *cereale*). Zukovskij (1933) and (Schiemann 1948) initially classified *S. ancestrale* as a direct ancestor of *S. cereale* (Stutz 1972). Later Zukovskij (1950) revised his statement and now argued that *S. segetale* must be considered as the ancestor (Stutz 1972). Stutz (1957) assumed that *S. sylvestre* must be the ancestor of cultivated rye, while other authors believed *S. cereale* must have descended directly from *S. montanum* (now *S. strictum*), a perennial wild rye species inhabiting elevated continental areas in Anatolia and neighboring regions of southwest Asia (Zohary et al. 2012). These included, among others, Vavilov (1926), Percival (1949), Riley (1955), Kush and Stebbins (1961) and Nurnberg (1967) (see Stutz 1972).

Even today it is assumed that the original ancestor of both weedy races and domesticated rye can be identified as *S. strictum*. Speculations suggest that the impressive evolvement of weedy rye and the variation in domesticated rye were notably enhanced by introgressive hybridization with *S. strictum* (Feldman and Levy 2023). Based on cytological, ecological and morphological studies, Stutz (1972) further proposed that the hybridization took part between perennial *S. strictum* and annual, self-pollinated *S. vavilovii*, with the latter derived from *S. sylvestre* through chromosomal translocations. However, alternative views suggest that *S. cereale* evolved from *S. strictum* through progressive cytological and morphological differentiation, possibly facilitated by adaptive advantages of translocation heterozygotes and rearrangement homozygotes (Khush and Stebbins 1961, Nürnberg-Krüger 1960; see Feldman and Levy 2023), while Zohary (1971) ascribed sympatric speciation of *S. cereale* from *S. strictum* to disruptive selection (see Feldman and Levy 2023).

Recent research is particularly focusing on the role of *S. cereale* subsp. *vavilovii* as the immediate wild ancestor of cultivated rye (Sun et al. 2022b). The study confirms that cultivated rye originated directly from a weedy population. Despite shared genetic traits with cultivated rye, *S. cereale* subsp. *vavilovii* exhibits distinct characteristics, displaying both weedy-like and wild-like types.

Weedy taxa of *S. cereale* can be considered as younger from an evolutionary and phylogenetic standpoint, with the domesticated rye, *S. cereale* subsp. *cereale*, being the youngest of all *Secale* species (Sencer and Hawkes 1980) and the only domesticated of all 8 subspecies of *S. cereale*. All other subspecies of *S. cereale* are commonly named feral or weedy rye (Daskalova and Spetsov 2020). In addition, the weedy taxa of *S. cereale*, with varying seed dispersal patterns, likely originated from wild forms of *S. cereale* subsp. *ancestrale* invading wheat and barley fields and undergoing mutations (Zohary et al. 2012). The selection for non-brittle rachis and larger caryopsis by farmers during cultivation played a crucial role in the domestication of rye (Sencer and Hawkes 1980).

## 1. Introduction

---

### 1.1.3 Self-incompatibility under the influence of S- & Z-Locus

Self-incompatibility (SI) occurs in many flowering plants, whereby selfpollen is recognized and rejected by the stigma (Herridge et al. 2022). Physiological SI can be classified as gametophytic or sporophytic, but the specific mechanisms, as well as the number of genetic determinants involved, vary between families (Cropano et al. 2021). It can be advantageous for a plant to maintain SI to limit inbreeding depression (Barrett 2013) and promoting outcrossing at the same time, to generate and maintain genetic diversity within a species (Takayama and Isogai 2005), because self-compatibility (SC) and inbreeding have been proposed to lead to a high species extinction rate (Fujii et al. 2016, Takebayashi and Morrell 2001). Rye's strong SI, coupled with wind pollination, results in panmictic populations of traditional rye varieties characterized by high levels of heterozygosity and heterogeneity (Geiger and Miedaner 2009, Melonek et al. 2021). Generally, the SI system in grass species (Poaceae) is controlled gametophytically, which means that the genotype of the haploid pollen contributes to SI response, and SI determinants are produced by the pollen grain itself (Herridge et al. 2022), and current knowledge suggests that self-incompatible species of the entire Poaceae family share the same SI system (Rohner et al. 2023). Specifically, this means for Poaceae, that the interaction of selfpollen on the stigmatic exudate quickly leads to pollen tube arrest (Shivanna et al. 1982). It was shown that the SI system in this family is generally controlled by at least two multiallelic and independent loci, S and Z, and it is assumed to be conserved (Do Canto et al. 2016). Nonetheless the genes controlling SI have not been fully elucidated yet (Herridge et al. 2022).

Already in 1956 it was found out by Lundqvist that the SI in rye is controlled by two multiallelic loci, denoted as S- & Z-locus, which thought to have arisen from a duplication occurring in the common ancestor of grasses (Lundqvist 1962). Mutations at these loci result in self-fertility (SF), allowing for the autonomous fertilization of the plant (Lundqvist 1956, 1960). Over years and through genetic mapping, it was found out by using isozymes that the S- & Z- loci are localized on chromosomes 1R and 2R, respectively (Gertz and Wricke 1989, Wricke and Wehling 1985). Both loci are gametophytically expressed and cause an SI response when both the S and Z alleles of a haploid pollen grain match the same alleles in the diploid stigma (Hackauf and Wehling 2005, Melonek et al. 2021). Genes for self-fertility have been referred to the same loci (Fuong et al. 1993, Voylokov et al. 1998) which is why these self-fertility phenotypes can be interpreted as consequence of special allele "Sf" of the self-incompatibility locus (Voylokov et al. 1993, Gruner and Miedaner 2021). Furthermore, additional self-fertility genes on chromosomes 3R, 4R, 5R, and 6R were identified (Melz and Winkel 1987, Melz et al. 1990, Voylokov et al. 1998), as well as the self-fertility locus S5 (also referred as T-locus) on chromosome 5R (Voylokov et al. 1993). Further mapping efforts allowed to localize the S5 gene in the centromeric region of chromosome 5R (Egorova et al. 2000, Voylokov et al. 1998). It was speculated that SC arising from genes unlinked to the SI-determinant genes, can affect their expression, or mediate the up- or downstream pathways involved in the SI response. More specific, they could be part of a signal transduction cascade within the pollen grain causing the pollen tube arrest, which is triggered by the S- & Z-locus (Cropano et al. 2021, Do Canto et al. 2016, Wehling et al. 1995).

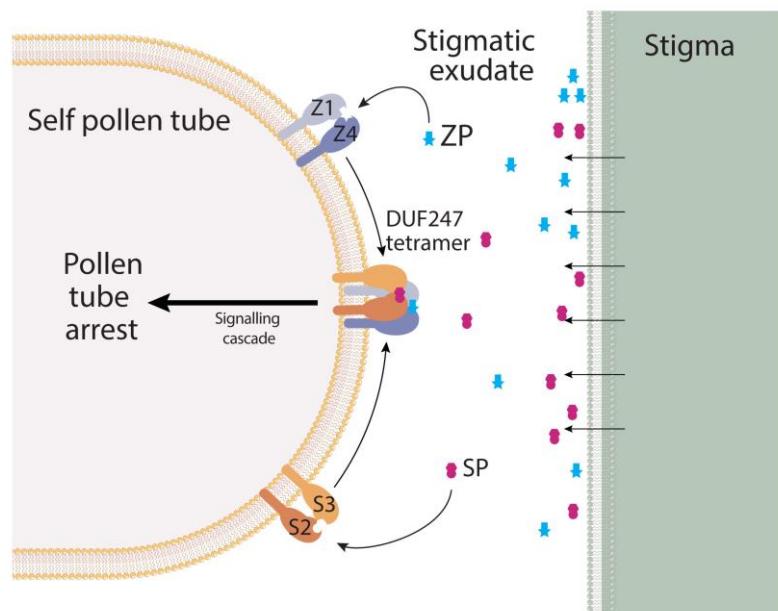
Both the S- & Z-genes have not yet been isolated in rye or any other grass family. However, in perennial ryegrass (*Lolium perenne L.*) a gene encoding a Domain of Unknown Function 247 (DUF247) protein as

## 1. Introduction

---

well as a ubiquitin-specific protease was found to co-segregate with the Z-locus (Manzanares et al. 2016, Shinozuka et al. 2010, Thorogood et al. 2017). This is also true for rye (Hackauf and Wehling 2005). Nonetheless, genetic markers enabled mapping of the rye S- & Z-locus, with a total base pair lengths of 3 Mbp with 19 annotated genes for the S-locus and 1.3 Mbp with 20 annotated genes in the rye ‘Lo7’ inbred line (Melonek et al. 2021, Rabanus-Wallace et al. 2021). A full list can be found in the appendix (see appendix A1).

In recent years it was found out, that the presence of two DUF247 genes and a short unstructured protein (SP/ZP) was identified in both the S- & Z-locus across some self-incompatible grass species, while in self-compatible species, these genes were frequently lost or had frame-shift mutation (Herridge et al. 2022). These genes exhibited high allelic diversity as well as tissue-specific gene expression, and therefore matching the expected characteristics of SI determinants known from other SI systems (Rohner et al. 2023). In addition, Herridge et al. (2022) proposed a model explaining how the proteins encoded at the S- & Z-loci might function to specify self-incompatibility (Figure 1). They assume that “a pollen extracellular receptor, comprising two S- and two Z-DUF247 proteins, specifically recognizes its own stigma excreted SP/ZP signal proteins, activating a pathway that prevents fertilization by its own pollen” (Herridge et al. 2022, p.2).



**Figure 1 Model for SI in grasses.** DUF247 proteins from S- & Z-loci form dimers at the pollen tube surface and are anchored to the membrane by a conserved transmembrane domain. SP/ZP peptides are free-floating in the stigmatic exudate, allowing pollen tube arrest to occur immediately upon pollen germination. Signals from S- & Z- must be incorporated, therefore Herridge et al. propose a tetramerization of DUF247 proteins, resulting in a downstream signaling cascade that results in pollen tube arrest. Absence of any component, or non-self components would result in a lack of signaling, thus allowing pollen tube growth (source: Herridge et al. 2022).

By analyzing two DUF247 DUF247/SP/ZP proteins in grass species, conspicuous features were found in rye: While the self-compatible rye genotype Lo7 had a mutation in one of the DUF genes of the Z-locus, the other self-compatible rye genotype Weining did not have such a mutation. Therefore, Herridge et

## 1. Introduction

---

al. (2022) speculated that in this case another locus, outside of the S & Z-locus, was responsible for the self-compatibility of Weining, but also recombination event between the SI determinants, silencing of the SI determinants, or a mutation interfering with the downstream cascade of SI unlinked to the S- & Z-locus could lead to self-compatibility (Cropano et al. 2021, Do Canto et al. 2016, Rohner et al. 2023).

Also, no DUF247 candidate gene at the S- & Z-locus are represented within the global transcriptional profiles of the anther/pollen, ovary and stigma concurrent developmental stages (Melonek et al. 2021, Tran et al. 2013). Throughout history a few candidate genes at S- & Z-locus were found for SC in grass species. These were also considered for rye. For example, Bian et al. (2004) identified through fine mapping of perennial grass *Phalaris coerulescens* a cosegregation of the markers Bm2 and BCD762 with the S-locus. Bm2 represents a thioredoxin-encoding gene and its rye ortholog maps 22.4 Mbp distal from a gene encoding a 40S ribosomal protein (Melonek et al. 2021). Studer and Asp (2014), reported two glycerol kinase-like linked genes (LpGK1 and LpGK2) as Z-candidates in perennial ryegrass (see Melonek et al. 2021). A glycerol kinase residing at the Z-locus with high allelic variability could be excluded due to observed recombination and the expression pattern (Melonek et al. 2021). In conclusion, it can be said that no homology is observed between any of the predicted genes located in the genomic regions of the S- & Z-loci in the 'Lo7' reference genome and the cloned S-genes from other species, except for a gene affiliated with the F-box family located at the Z-locus. This observation implies that the bifactorial self-incompatibility mechanism found in grasses constitutes an additional type within gametophytic self-incompatibility (GSI) systems (Melonek et al. 2021).

Over and beyond, Lundqvist (1962) theory of a duplication event of one of the two loci could be verified through evolutionary analysis of the DUF247/SP/ZP proteins, with the Z-locus being the original locus and S has resulted from a subsequent duplication (Herridge et al. 2022), with a prior duplication of one of the two DUF genes (Rohner et al. 2023). Interestingly, the sZ, ZDUF247-I, and ZDUF247-II at the Z-locus are the only three genes for which genes of similar sequence and structure were found at the S-locus (sS, SDUF247-I, and SDUF247-II) (Rohner et al. 2023).

### 1.2 Outline through the history of population genetic analyses

Generally speaking, population genetic analyses consist of more or less two independent things: describing the genetic structure of populations or theorizing on the evolutionary forces acting on populations (Gillespie 2004) through mathematical approaches (Müller et al. 2022). That means, population geneticists are interested in both inference and prediction – making inferences about demographic history of species on the one hand and detecting signatures of natural selection or assigning individuals to populations on the other. Nowadays is transitioning into a more and more data-driven discipline due to the availability of large-scale genomic data and the ability to study increasingly complex evolutionary scenarios (Korfmann et al. 2023). Broken down, population genetics is studying allele frequencies, and is therefore closely related to quantitative genetics (the study of allelic effects) and evolutionary genetics, which seeks to understand how they interact over time (Gibson 2018). The characterization of allele frequencies and distributions in populations enable inferences about

## 1. Introduction

---

processes. Namely genetic drift, mutation, gene flow, and natural selection, which in turn have shaped the patterns observed for a given population (Grünwald et al. 2017).

The beginnings of genetics and population genetics are inevitably connected (Crow 1987). Both started with early achievements by Mendel (1866), Hardy (1908), and Weinberg (1908) (see Crow 1987), even though the first generation of population geneticists arose in the early 20th century in the narrow sense (Charlesworth 2015). Thereby the population genetic analysis aimed at connecting two opposing views of evolution by focusing on genetic variants in the population – the Mendelian Genetics with the Darwinian theory of evolution (Korfmann et al. 2023, Okazaki et al. 2021). Especially R. A. Fisher (\*1890, †1962), S. Wright (\*1889, †1988), and J. B. S. Haldane (\*1892, †1964) have been pioneers of that time, with Wright spending a major part of his time on physiological genetics, Fisher laid the foundations for modern statistics, and Haldane multifarious activities included astronomy and many areas of biology (Crow 1987).

In recent years, a growing understanding has become that many molecular changes have no effects on phenotypes. Based on Wright's drift hypothesis (1938) and Haldane's model of advantageous mutations in 1927, Motoo Kimura introduced the neutral theory in 1964 (see Okazaki et al. 2021). He stated that in “sharp contrast to the Darwinian theory of evolution by natural selection, the neutral theory claims that the overwhelming majority of evolutionary changes at the molecular level are caused by random fixation (due to random sampling drift in finite populations) of selectively neutral (i.e., selectively equivalent) mutants under continued inputs of mutations.” (Kimura 1991). Kimura also created the infinite allele model (Kimura and Crow 1964) as well as the “infinite sites model” (Kimura 1969). The first is related to genetic variation in populations, which arises from the balance between mutation and genetic drift, whereas the “infinite sites model” states, that if the mutation rate is low and the effective population size is small, a mutant variant will always appear at a different site in the genome (Okazaki et al. 2021). This model also implies that each mutation creates a new allele in the population and that “backward” or “reverse” mutation do not occur (Barroso et al. 2020). Therefore, it represents one of the bases for genome-wide association studies using single nucleotide polymorphisms (SNPs) as genetic markers in unrelated individuals (Sella and Barton 2019). Based on Kimuras assumptions, Tomoko Ohta (Ohta 2017) proposed the “nearly neutral” theory, which describes natural selection at the molecular level. This theory emphasizes the importance of slightly deleterious mutations by recognizing their ability to segregate and eventually get fixed due to genetic drift despite the presence of purifying selection. For such mutations, interaction of drift and selection becomes important. Since genetic drift is stronger in smaller than in larger populations, a correlation between population size and molecular measures of natural selection is expected within the nearly neutral theory (Brenner 2013, Müller et al. 2022). The underlying mathematics of these models and many others can be read in the book “Theoretical Evolutionary genetic” by Joseph Felsenstein (2015).

Thus, the evolution of population genetic analysis over time can be said to have started as a conceptual framework and developed a rich body of mathematical theory that became a vast treasure trove of probabilistic models for developing sophisticated statistical methods as molecular data became available. This body of theory has continued to grow in complexity to accommodate more realistic

## 1. Introduction

---

evolutionary and genetic scenarios and more efficient computational algorithms (Korfmann et al. 2023).

### 1.2.1. Population structure analysis

Population structure analysis is an integral component of population genetic analysis and leads to systematic patterns in measures of mean relatedness, as an average of the relationships across different loci, between individuals in large genomic data sets. Mean relatedness can be strongly affected by linked selection and other factors (Li and Ralph 2019). Population genetic structure is the result of interaction between ecological and genetic processes, plays an important part in understanding the genetic characteristics and the dynamics of population, and gene flow is considered as one of the most important determinants of the genetic structure of plant populations (Cheng et al. 2020, Robledo-Arnuncio et al. 2014). In addition, historical events, ecological factors, natural selection factors, interference of human activities, historical events, overlapping generations and other factors will also lead to the formation of specific patterns of genetic structure in the population (Barroso et al. 2020, Cheng et al. 2020).

Analyzing population structure involves navigating challenges such as detecting subgroups within populations and assessing relatedness among individuals (Sul et al. 2018). Two different types of methods for analyzing population structure have prevailed: methods based on admixture models (Falush et al. 2003, Pritchard et al. 2000), and principal components analysis (PCA, Patterson et al. 2006). In admixture-based models it is assumed that each individual has inherited some proportion of its ancestry from a set of K distinct populations. These proportions are known as the admixture proportions of each individual. A key goal of these methods is to estimate these proportions and the allele frequencies of each population (Engelhardt and Stephens 2010). PCA can be thought of as projecting the individuals into a low-dimensional subspace, wherein their positions reflect genetic similarities among them (Engelhardt and Stephens 2010, Li and Ralph 2019). The results of PCA can be related to the genealogical history of the samples, such as time to most recent common ancestor and migration rate between populations (McVean 2009, Novembre and Stephens 2008). Not only can the analysis of population structure provide us with valuable information about the events described above, but it is also essential for controlling the stratification in Quantitative trait loci (QTL) analysis (Qu et al. 2020, Wang 2022).

### 1.2.2. Quantitative trait loci (QTL) analysis

Genetics concerns the study of heritably quantitative or complex traits because many agricultural traits of interest, such as the response to fertilizer or disease susceptibility in different plant species, are complex traits (Miles and Wayne 2008, Prins et al. 2019). Quantitative trait locus (QTL) analysis is a statistical method to investigate such traits by linking two types of information - phenotypic and genotypic data to explain the genetic basis of variation in these complex traits (Miles and Wayne 2008). But it can also be used for evolutionary studies to study the genetic basis of adaptation and therefore

## 1. Introduction

---

it can provide clues about the evolutionary history of populations, as well as causes of possible population differentiation (Calabrese 2019, Morton 2005, Zeng 2005).

Two predominant approaches have been extensively employed in QTL analysis: linkage mapping and linkage disequilibrium (LD) mapping, also known as association mapping. Both mapping strategies aim to leverage the concept that recombination break up the genome into small fragments which can be correlated with phenotypic variation (Myles et al. 2009). The linkage-based analysis assumes the generation of a mapping population through crossing of specific individuals, in which relatedness is known (Myles et al. 2009). While this has been proven to be effective for QTL detection, this approach tends to deliver low-resolution, population-specific QTLs (Zhu et al. 2008a) and samples only a limited portion of the allelic diversity within the available germplasm (Newell et al. 2011). It is also possible to use linkage disequilibrium (LD) to map QTL in outbred and natural populations (Prins et al. 2019), for example in association mapping (Morton 2005). LD describes the non-random association of alleles at two or more loci (Qu et al. 2020).

The degree of linkage disequilibrium (LD) and its decay with genetic distance are crucial parameters in assessing the requisite number of markers for effective QTL mapping and the resolution at which the trait can be successfully mapped (Vos et al. 2017). LD is inherently population specific, its decay over time within a population is influenced by factors such as the recombination rate between loci and the number of generations of recombination (Otyama et al. 2019). Estimates of LD and its decay with distance in any given population are influenced by various factors, including non-random mating, selection, mutation, migration or admixture, genetic drift, recombination, population structure associated with breeding history, and the effective population size (Flint-Garcia et al. 2003, Newell et al. 2011, Otyama et al. 2019). Even specific population structures and kinship between individuals can produce LD between unlinked loci, which could lead to increasing rates of false-positives in GWAS (Johnson et al. 2023, Kulkarni et al. 2020, Mangin et al. 2012). That said, self-fertilizing plants usually show less decay of LD, because the recombination events are ineffective to cause LD decay due to their homozygous genetic background (Auinger et al. 2016, Vos et al. 2017).

Genome-wide association studies (GWAS) quantify statistical association between genetic variation and phenotypes (=traits). In the field of quantitative genetics, the GWAS method rely on the existence of linkage disequilibrium (LD) between the molecular markers and the unobserved quantitative trait loci (QTL) close to the marker (Otyama et al. 2019, Qu et al. 2020). The ability of GWAS to deliver high-power, high-resolution results is largely dependent on the extent of LD within the working population (Newell et al. 2011). One important challenge is the inflated false-positive associations that arise in GWAS results when unrecognized population structure, kinship and cryptic relatedness exist (Cardon and Palmer 2003, Choi et al. 2009, Johnson et al. 2023). Cryptic relatedness exists when some individuals are closely related, but this shared ancestry is unknown to the investigators (Sul et al. 2018). To prevent this problem different computational approaches have been made. One of the first was the general linear model (GLM), capable of integrating population structure or principal components (PCs)

## 1. Introduction

---

as fixed effects to reduce the false positives caused by population stratification (Pritchard et al. 2000, Wang and Zhang 2021). To account for the relationships among individuals within subpopulations, kinship among individuals was introduced through the mixed linear model (MLM) by using genetic markers covering the entire genome (Wang and Zhang 2021, Zhu et al. 2008b), as well as enhancements of MLM by improving computing efficiency and remaining the statistical power (Wang and Zhang 2021).

### 1.3. Aims and objectives

The aims of the present study were:

1. Analyzing the population structure of *Secale* and
2. Determination of Quantitative trait loci (QTL) associated with self-incompatibility (SI)

Objectives:

1. Carrying out a Principal component analysis (PCA)
2. Calculation of the Cross-entropy criterion
3. Calculation of the ancestry coefficient
4. Calculation of the Fixation Index ( $F_{ST}$ )
5. Determination of nucleotide diversity ( $\pi$ )
6. Calculation of the inbreeding coefficient ( $F$ )
7. Implementation of genome-wide association studies (GWAS) for SI
8. Calculation of the LD decay
9. Searching for candidate genes

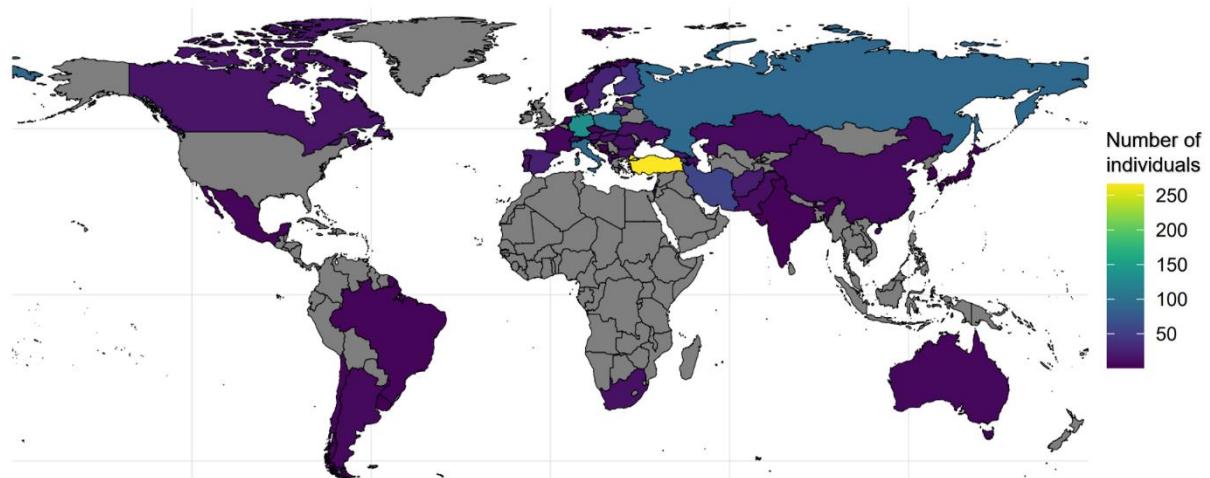
## 2. Material and Methods

---

### 2. Material and Methods

#### 2.1. Plant Material, GBS Read Alignment and variant calling

In this study, 356 genebank accessions from the *Secale* taxa were used. The total of 1,396 individuals can either be assigned to wild and domesticated rye (*S. cereale* subsp. *cereale*), belong to *S. cereale* subsp. *vavilovii*, *S. strictum*, or *S. sylvestre* and go back to the previous work of (Schreiber et al. 2022), (Rabanus-Wallace and Stein 2021) and (Schreiber et al. 2019). A few individuals are classified as hybrids between the different species. These will be marked as “NA”, if the samples were subdivided into subspecies for some calculations. On average, four plants can be counted for each accession with a number of plants per accession ranging from 1 to 52. Their origin lies in over 50 countries (see Figure 2).



**Figure 2 Geographical origin** of all 1,396 selected genotypes. Worldmap was created with R package “rnatural earth” (Massicotte and South 2024).

Further information on the plant panel, such as taxonomic status, country of origin, and collection site can be accessed via the genebank information system of the German Federal ex situ Genebank at IPK Gatersleben (GBIS; <https://gbis.ipk-gatersleben.de>, Oppermann et al. 2015).

The primary data analysis followed the procedures outlined by Mascher, Wu, Amand, Stein, and Poland (2013). Adapters were trimmed from raw reads with Cutadapt (Martin 2011), and the trimmed reads were mapped to the whole- genome shotgun sequence assembly of rye cultivar Lo7 reference genome sequence assembly (Rabanus-Wallace and Stein 2021) using BWA-MEM version 0.7.13 (Li 2013). Alignment records were converted to Binary/AlignmentMap (BAM) format with SAMtools (Li et al. 2009) and afterwards sorted with Novosort (<http://www.novocraft.com/products/novosort/>). Variant calling was performed with SAMtools and BCFtools version 1.3 (Li 2011) using a mapping quality threshold of 30 and a base quality threshold of 20 (see Schreiber et al. 2019, Schreiber et al. 2022).

## 2. Material and Methods

---

The resulting VCF file was filtered using VCFtools (v0.1.16) (Danecek et al. 2011) according to the following criteria (VCFtools command in brackets):

- Minor allele frequency (--maf): 0.01
- Minimum alleles per site (--min-alleles): 2
- Maximum alleles per site (--max-alleles): 2
- Minimum depth allowed for a genotype (--minDP): 4
- Maximum depth allowed for a genotype (--maxDP): 40
- Minimum mean depth for a site (--min-meanDP): 4
- Set minimum non-missing data (--max-missing): 0.9

Ultimately, of the 2,867,917 possible sites for the 1,396 individuals, 29,616 sites and all individuals were retained. Information on the unfiltered data can be found in the appendix A2.

The minor allele frequency (maf) is widely used in population genetics studies because it provides information to differentiate between common and rare variants in the population. The power to detect genetic effects through genome-wide association studies (GWAS) is dependent on maf and its distribution - from nearly monomorphic ( $maf < 0.5\%$ ) to very common ( $maf \approx 50\%$ ) (Tabangin et al. 2009). Since as many rare alleles as possible were to be retained in this study, a maf of 0.01 was used.

The use of 2 alleles per site (--min-alleles 2, --max-alleles 2) ensures the presence of only bi-allelic sites. Insertions and deletions have already been filtered in advance.

For the maximum and minimum depth allowed for a genotype it applies that any individual that has a depth outside the range of 4-40 is marked as a missing genotype. Similarly, the minimum mean depth for a site should not be less than 4. This number was chosen as the lower limit here, as this lies within the range of the average number of reads per position (see appendix A3). In this way, as many positions as possible can be included in subsequent calculations without any loss of data quality. The maximum depth allowed for a genotype was selected so that regions with very high coverage are excluded. These are very likely to be reflective mapping / assembly errors, paralogous or repetitive regions (Pereira et al. 2018).

In addition, all sites were removed where over 10% of individuals are missing a genotype to get robust results.

### 2.2. Population Structure Analysis

At first, the population structure was analyzed. To determine the number of subgroups within the association panel, two different approaches were implemented:

1. Principal component analysis (PCA) was computed based on SNP marker information
2. Population structure was calculated through cross-entropy criterion

Afterwards, the ancestry coefficient was visualized and the nucleotide diversity ( $\pi$ ) as well as the fixation index ( $F_{ST}$ ) have been calculated for the determined subgroups.

## 2. Material and Methods

---

### 2.2.1. Principal component analysis (PCA)

Population structure leads to systematic patterns in measures of mean relatedness among individuals in extensive genomic datasets. These patterns are frequently identified and visualized using dimension reduction techniques, with principal component analysis (PCA) being a common approach for exploring multilocus population genetic data (François et al. 2010, Li and Ralph 2019).

In general, the PCA uses an orthogonal (linear) transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables. These uncorrelated variables are called principal components (PCs; Mitra 2019). PCs are calculated for each genotype and can then be projected onto the space spanned by the PC axes, allowing the samples and their distances from each other to be visualized. One advantage is, that the distances between clusters could give hints about the genetic distances between them (Elhaik 2022).

The linear transformation of all 29,616 SNPs into a new coordinate system was done with the help of the snpgdsPCA() function of SNPrelate package (Zheng et al. 2012) in R (version 4.2.3; R Core Team 2023) for all 1,396 individuals. The greatest variance lies on the first coordinate (first PC), the second greatest variance on the second coordinate and so on. SNPrelate package uses the FastPCA algorithm from Galinsky et al. (2016). Here, a test for selection will be done, which uses the SNP weights from PCA to calculate the differentiation of each locus along top principal components. In addition, a separate PCA was performed for the *S. cereale* taxa to detect substructures. A total of 1,147 individuals were used, with 1028 individuals belonging to *S. cereale* subsp. *cereale* and 119 individuals belonging to *S. cereale* subsp. *vavilovii*.

One characteristic of a PCA interpretation in terms of ancestry is, that it can be confounded by demographic factors (McVean 2009) or distribution of sampling locations and amounts of data (Novembre and Stephens 2008). Therefore, a second approach to determine the number of subgroups within the association panel was made – calculating the cross-entropy criterion.

### 2.2.2. Cross-entropy criterion

The cross-entropy criterion will help to choose the most fitting number of ancestral populations (K) within the association panel, based on the prediction of masked genotypes to evaluate the prediction error of ancestry estimation (Carvalho-Madrigal and Sanín 2024, Fritchot et al. 2014). Basically, the cross-validation process divides the entries in the genotypic matrix into a learning set and a test set, designating 5% of all genotypes as masked entries. The genotype probabilities for these masked entries are predicted based on factor estimates derived from unmasked entries. Subsequently, the error between the predicted and observed genotype frequencies is calculated, with smaller values of this criterion indicating more optimal selections (Caye et al. 2018).

The cross-entropy criterion for K-values between 1 and 20 was calculated with 20 iterations using the sNMF algorithm. sNMF's coding enabled the estimation of homozygote and heterozygote frequencies by avoiding assumptions about Hardy-Weinberg equilibrium (Fritchot et al. 2014).

## 2. Material and Methods

---

The repetitions were then averaged using CLUMPP (v1.1.2; Jakobsson and Rosenberg 2007), a cluster matching and permutation program. CLUMPP provides three algorithms for optimal alignments of the individual runs of the same data. Here, the calculation was limited to the LargeKreedy algorithm (Jakobsson and Rosenberg 2007).

### 2.2.3. Ancestry Coefficient

In the next step, the individual ancestry coefficients were calculated for the test panel to infer the ancestry of each individual. The calculation was carried out using the sNMF function of the R package LEA (v3.10.2; (Frichot and François 2015), which is based on sparse non-negative matrix factorization (snmf) and least-squares optimization of ancestry proportions for diallelic marker (Frichot et al. 2014). The calculation for the number of ancestral populations K = 4-7 was carried out based on the PCA and cross-entropy criterion results.

On modelling ancestry coefficients using an admixture model based on snmf, Frichot et al. (2014) write: The probability that individual  $i$  carries  $j$  derived alleles at locus  $l$  can be displayed as:

$$P_{il}(j) = \sum_{k=1}^K q_{ik} g_{kl}(j), \quad j = 0,1,2$$

where  $q_{ik}$  is the fraction of individual  $i$ 's genome that originates from the ancestral population  $k$ , and  $g_{kl}(j)$  represents the homozygote ( $j = 0, 2$ ) or the heterozygote ( $j = 1$ ) frequency at locus  $l$  in population  $k$ .

### 2.2.4. Nucleotide diversity ( $\pi$ )

Nucleotide diversity  $\pi$  represents an important property of populations in terms of nucleic acid sequences (Nelson and Hughes 2015) and remains one of the central measures for describing sequence variation in genetic and genomic diversity studies (Konopiński 2023). For example, it can provide information about the demographic history of the different species (Charlesworth 2009). Therefore, the nucleotide diversity  $\pi$  was calculated for each genotype.

At first Nei and Li (1979) defined the nucleotide diversity ( $\pi$ ) as a measure for genetic variation by the average number of nucleotide differences per site between two randomly chosen DNA sequences. It can be calculated with:

$$\pi = \sum_{ij} x_i x_j \pi_{ij}$$

With  $x_i$  as the frequency of the  $i^{th}$  sequence in the population and  $x_j$  as the frequency of the  $j^{th}$  sequence in the population.  $\pi_{ij}$  is the number of nucleotide differences per nucleotide site between the  $i^{th}$  and  $j^{th}$  sequences.

## 2. Material and Methods

---

One year later it was modified to guarantees an unbiased estimator, which does not depend on sequence size of the samples (Nei and Tajima 1980), especially for very small number of nucleons sampled ( $n$ ). In consequence  $\pi$  should be estimated by:

$$\hat{\pi} = \frac{n}{n-1} \sum_{i \neq j} x_i x_j \hat{\pi}_{ij}$$

Where  $\hat{\pi}_{ij}$  is the estimate of  $\pi_{ij}$ .

These formulas do not consider the possibility of missing nucleotides occurrence in one or more sequences in the data set which can lead to problems especially in massive-sequencing data due to their currently used technologies for generation, which often introduce substantial amounts of missing nucleotides in their output (Konopiński 2023).

A more modern approach for estimating the nucleotide diversity in a population is to first take a random sample of  $n$  sequences from the population. In the next step the number of nucleotide substitutions per site ( $d_{ij}$ ) will be estimated between sequences  $i$  and sequences  $j$  (Hubisz and Siepel 2020, Nelson and Hughes 2015). Several models exist for calculating  $d_{ij}$ , which address issues such as multiple hits, base composition bias, and transitional bias (Felsenstein 1981, Jukes and Cantor 1969, Tajima and Nei 1982) and much mor (see Nei and Kumar 2000). Nucleotide diversity ( $\pi$ ) is estimated by the mean  $d_{ij}$  for all  $(n^2 - n)/2$  possible pairwise comparisons among sequences (Nelson and Hughes 2015):

$$\pi = \sum_{i < j} \frac{d_{ij}}{(n^2 - n)/2}$$

In this study the nucleotide diversity ( $\pi$ ) was calculated in 1 Mb windows with steps of 0.25 Mb for overlapping windows to account for genome-wide differences between subpopulations. These was done using VCFtools (Danecek et al. 2011).

### 2.2.5. Fixation Index ( $F_{ST}$ )

The fixation index  $F_{ST}$  can provide insights into the evolutionary processes shaping the structure of genetic variation in a subdivided population (Holsinger and Weir 2009, Ochoa and Storey 2021).  $F_{ST}$  estimates are among the most widely employed descriptive statistics in the fields of population and evolutionary genetics (Holsinger and Weir 2009) and it describes the probability that alleles drawn randomly from a subpopulation are “identical by descent” (IBD) relative to an ancestral population. Its value varies between 0 and 1, with  $F_{ST} = 0$  for an unstructured population and  $F_{ST} = 1$  if every locus has become fixed for some allele in each subpopulation (Ochoa and Storey 2021). Estimates of  $F_{ST}$  can

## 2. Material and Methods

---

identify regions of the genome that have been the target of selection, and comparisons of  $F_{ST}$  from different parts of the genome can provide insights into the demographic history of populations (Holsinger and Weir 2009).

The calculation of the  $F_{ST}$  statistics was done using the --weir-fst-pop command of VCFtools (Danecek et al. 2011) for predefined groups of *Secale* (see end of chapter). It is a calculated estimate following the methodology outlined in Weir and Cockerham's paper from 1984, which build on the calculation to  $F_{ST}$  from Wright (1951). According to (Willing et al. 2012), Wright defined  $F_{ST}$  as a measure of correlation of gene frequencies and further suggested the first and simplest estimator,  $F_{ST}^W$  for one allele at locus  $k$ :

$$\hat{F}_{ST}^{[k]} = \frac{s^2}{\bar{p}(1-\bar{p})}$$

with

$$s^2 = \sum_i (p_i - \bar{p})^2 / (r - 1)$$

as the observed variance of allele frequencies  $p_i$  among the sampled populations  $i$  ( $i = 1 \dots, r$ ).  $\bar{p}$  is the mean allele frequency over all populations.

The estimation of  $F_{ST}^W$  for multiple loci is calculated by taking the mean across  $k$  loci.

$$\hat{F}_{ST} = \frac{1}{k} \sum_k \hat{F}_{ST}^{[k]}$$

But because  $F_{ST}^W$  tends to overestimate the level of genetic differentiation at low values, Weir and Cockerham (1984) came up with the  $F_{ST}^{W&C}$  estimator, which preserves Wright's definition of  $F_{ST}$  in terms of correlation of gene frequencies (Willing et al. 2012).

As a consequence, at a single locus  $k$ ,  $F_{ST}^{W&C}$  is defined as:

$$\hat{F}_{ST}^{[k]} = \frac{\hat{N}^{[k]}}{\hat{D}^{[k]}}$$

with

$$\hat{N}^{[k]} = s^2 - \frac{1}{2n-1} \left[ \bar{p}(1-\bar{p}) - \frac{r-1}{r} s^2 - \frac{\bar{h}}{4} \right]$$

and

$$\hat{D}^{[k]} = \bar{p}(1-\bar{p}) + \frac{s^2}{r}$$

Here,  $s^2$  is the observed variance of allele frequencies,  $n$  is the number of individuals per population,  $p$  is the mean allele frequency over all populations,  $r$  is the number of sampled populations and  $h$  is

## 2. Material and Methods

---

the mean observed heterozygosity (Willing et al. 2012). The overall estimate from all  $k$  loci is derived by:

$$\hat{F}_{ST} = \frac{\sum_k \hat{N}^{[k]}}{\sum_k \hat{D}^{[k]}}$$

As mentioned above, when calculating  $F_{ST}$  with VCFtools (Danecek et al. 2011), the populations of *S. cereale* subsp. *cereale* were not subdivided into different domestication stages and brittleness, as further subdivision of these individual groups does not lead to significant differences in the  $F_{ST}$  statistics. Thus, the final populations included in the analysis were as follows: *S.cereale*:dom (for *S.cereale* domesticated:top, domesticated:middle and domesticated:bottom), *S.cereale*:feral (for *S.cereale* feral:brittle and feral:nonbrittle), *S.cereale*:weed, *S.vavilovii* (for *S. cereale* subsp. *vavilovii*), *S. strictum* and *S. sylvestre*.

### 2.3. Inbreeding coefficient ( $F$ )

Now that the population structure for this experimental panel has been calculated and analyzed, the inbreeding coefficient ( $F$ ) was calculated according to the predominant subpopulations. The inbreeding coefficient ( $F$ ) was first described as the correlation between homologous alleles of two gametes that unite to form the individual (Wright 1921). In 1948, Malécot G. defined it as the probability that two homologous alleles at a given locus being identical by decent (see Villanueva et al. 2021). In addition, the inbreeding coefficient gives the proportional loss of genetic variation in the form of loss of heterozygosity by inbreeding (Villanueva et al. 2021).

In this project, the inbreeding coefficient was calculated using the --het function of VCFtools (Danecek et al. 2011). Therefore, the heterozygosity gets measured on a per-individual basis. Specifically,  $F$  is estimated for each genotype using a method of moments (see Wang 2011). The data was then grouped and visualized in R (version 4.2.3; R Core Team 2023).

Wang (2011) summarized one of the first estimator of the inbreeding coefficient for multi locus genotype data (by Ritland 1996) as follows:

$$\hat{F} = \left( \sum_{l=1}^L (k_l - 1) \right)^{-1} \sum_{l=1}^L \sum_{i=1}^{k_l} \frac{S_{il} - P_{il}^2}{p_{il}}$$

Where  $p_{il}$  is the frequency of allele  $i$  ( $= 1, 2, \dots, k_l$ ) at locus  $l$  ( $= 1, 2, \dots, L$ ) and  $S_{il}$  is an indicator variable taking a value of 1 if the individual is homozygous for allele  $i$  at locus  $l$  or 0 if otherwise. If it is assumed that there is a single locus  $L = 1$  this estimator is the same as that derived by Li and Horvitz (1953).

But because Ritlands' estimator (1996) is very sensitive to allele frequencies it can be affected by misspecification of allele frequencies, mutations, and genotyping errors (Wang 2011), which is why Lynch and Ritland have refined this estimator in 1999 for pairwise relatedness with molecular markers. To provide a comprehensive overview, it is important to mention that other authors have defined more

## 2. Material and Methods

---

estimators for the inbreeding coefficient by using a method of moments (for example Carothers et al. 2006, Ritland and Travis 2004, Wang 2011).

### 2.4. Quantitative trait loci (QTL) analysis

#### 2.4.1. Genome-wide association studies (GWAS)

The next step was to examine whether and which gene regions correlate with the self-incompatibility trait, expressed here as an inbreeding coefficient. Therefore, a GWAS analysis was performed with FarmCPU (Fixed and random model Circulating Probability Unification; Liu et al. 2016) implemented in the R package GAPIT v3.1.0 (Genomic Association and Prediction Integrated Tool; Lipka et al. 2012, Wang and Zhang 2021) using 29,128 out of the 29,616 SNPs. Marker trait associations (MTA) were declared if their  $-\log_{10}(p)$  value was higher than 5.765. The first five PCs were used for the calculation, based on the previous results, as well as on Pardiñas et al. (2018) and Tucker et al. (2014), who recommend using the first five PCs for most GWAS approaches. Two different GWASs were performed. One included all 1396 individuals from all *Secale* taxa. On the other hand, a GWAS was performed without the perennial species *S. sylvestre*, so that a total of 1372 individuals belonging to *S. cereale* subsp. *cereale*, *S. cereale* subsp. *vavilovii* and *S. strictum* were included in this analysis. This was done to investigate the dependence of the inbreeding coefficient on perenniability. The Significance threshold for marker trait associations (MTA) was elaborated based on Bonferroni correction of  $p \leq 0.05 / n$ , where  $n$  is the number of SNP markers used. To validate the GWAS output, 100 permutations were done in which genotype and phenotype values were randomized, without detecting any significant associations.

FarmCPU takes up the Multiple Loci Linear Mixed Model (MLMM; Segura et al. 2012), which simultaneously incorporate multiple markers as covariates in a stepwise Mixed Linear Model (MLM) approach, and divides it into two components: the Fixed Effect Model (FEM) and the Random Effect Model (REM), which are used iteratively. Thereby FEM contains testing markers one at a time, with multiple associated markers as covariates (pseudo quantitative trait nucleotides (QTNs)), to control false positives. To prevent overfitting in FEM, the associated markers are estimated in REM, utilizing them to define kinship. P values of testing markers and associated markers are unified at each iteration (Liu et al. 2016).

Liu et al. (2016) summarized the FEM model as:

$$y_i = M_{i1}b_1 + M_{i2}b_2 + \dots + M_{it}b_t + S_{ij}d_j + e_i$$

with  $y_i$  as the observation of the  $i^{th}$  individual and  $M_{i1}, M_{i2}, \dots, M_{it}$  are the genotypes of  $t$  pseudo QTNS, initiated as an empty set.  $b_1, b_2, \dots, b_j$  are the corresponding effects of the pseudo QTNS whereas  $S_{ij}$  is the genotype of the  $i^{th}$  individual and  $j^{th}$  genetic marker with  $d_j$  as the corresponding effect of the  $j^{th}$  genetic marker. Lastly,  $e_i$  is the residuals having a distribution with zero mean and variance of  $\sigma^2_e$ .

## 2. Material and Methods

---

P values are assigned to each testing marker, except those identified as pseudo QTNs, which are utilized as covariates. Initially, these pseudo QTN markers are marked as "NA" (Not Available) for their P value. Through a substitution process, each of these is individually examined for each testing marker, and the "NA" is replaced with the most significant P value associated with that pseudo QTN. Consequently, this newly assigned P value becomes the P value for the corresponding marker. This process is referred to as substitution (Liu et al. 2016).

Following the substitution, each marker is associated with its unique P value. The P values, along with the associated marker map, are then employed to update the selection of pseudo QTNs using the SUPER (Settlement of MLM Under Progressively Exclusive Relationship; Wang et al. 2014b) algorithm (Liu et al. 2016) in a REM as follows:

$$y_i = u_i + e_i$$

Where  $y_i$  and  $e_i$  remain unchanged, and  $u_i$  represents the total genetic effect of the  $i^{th}$  individual. The expectations of individuals' total genetic effects are assumed to be zeros. The variance and covariance matrix of the individuals' total genetic effects, denoted as G, is given by  $G = 2K\sigma_a^2$ , where  $\sigma_a^2$  is an unknown genetic variance, and K is the kinship derived from the pseudo QTNs (Liu et al. 2016).

The selection of pseudo QTNs that maximizes the likelihood of the REM is used to replace the pseudo QTNs in the FEM. The iterative process continues until no further changes occur in the estimated set of pseudo QTNs (Liu et al. 2016).

### 2.4.2. Linkage Disequilibrium (LD) decay estimates and Gene Annotations

To define the physical length of the region of the QTL within which gene annotations can be sought, the Linkage Disequilibrium (LD) was calculated. Two extensively used statistics for quantifying LD between allelic values at two loci are: their covariance ( $D$ ) or their squared correlation ( $r^2$ ) (Hill and Robertson 1968). According to Qu et al. (2020), Hill and Robertson (1968) defined these statistics in a biallelic system as:

$$D = p_{A_1 B_1} - p_{A_1} p_{B_1}$$

and

$$r^2 = \frac{D^2}{p_{A_1}(1-p_{A_1})p_{B_1}(1-p_{B_1})}$$

Where  $p_{A_i}$  represents the frequency of the  $i^{th}$  allele at locus A,  $p_{B_j}$  represents the frequency of the  $j^{th}$  allele at locus B, and  $p_{A_i B_j}$  represents the frequency of haplotype  $A_i B_j$  in the population (Qu et al. 2020). As the square of correlation coefficient,  $r^2$  can range from 0 to 1 (VanLiere and Rosenberg 2008).

The statistic  $r^2$  provides a summary of both recombinational and mutational history, while  $D$  measures only recombinational history. Therefore,  $D$  is considered the more accurate statistic for estimating

## 2. Material and Methods

---

recombination differences. However, for the purpose of evaluating the resolution of association studies, the  $r^2$  statistic is often favored. This is because  $r^2$  indicates how markers might correlate with the QTL of interest (Flint-Garcia et al. 2003).

For calculating LD decay, the pair-wise correlation ( $r^2$ ) between marker pairs of all 29,616 SNPs was determined in PLINK v1.90b6.21 (Purcell et al. 2007) and the results were plotted for an area of 200 megabases (Mb) for all chromosomes. Average  $r^2$  values were formed for intervals of 100 kilobases (kb) for the first 1 Mb, and 1 Mb intervals were formed for subsequent distances up to 200 Mb. These were merged and plotted.  $r^2 = 0.1$  was used as a threshold to define at which LD stops to exist and equilibrium was reached. According to Vos et al. (2017), it is the most used threshold for LD decay estimation. For a more detailed resolution around the threshold, LD decay was also calculated for a region of 10 kb with an interval of 100 bp. The LD was then calculated for each QTL in a range of 20 Mb upstream and downstream and summarized in 1 Mb intervals. Exceedances of the threshold of  $r^2 = 0.1$  were used to define the range of gene annotations. If  $r^2$  values within the 20 Mb were below the threshold, the global LD was used as the limit.

The structural gene prediction and annotation for *S. cereale* refers to Lux (2020) and was created with the help of the rye 'Lo7' inbred line (Rabanus-Wallace et al. 2021). In this work, the interval determined by LD decay was searched for genes with low confidence (LC) and high confidence (HC).

### 3. Results

---

## 3. Results

The aim of this study was to investigate the self-incompatibility (SI) of *Secale* in the overall context of population genetic analyses. The focus was on analyzing the population structure within a diversity panel, followed by a genome wide association study (GWAS) to identify quantitative trait loci (QTLs) associated with SI. Previously identified SI QTLs from the literature can serve as a reference point for the evaluation of the results.

Based on the research of Schreiber et al. (2019), Schreiber et al. (2022), Rabanus-Wallace and Stein (2021), 1,396 individuals of different *Secale* species from 356 genebank accessions were used to achieve this goal. They originated from more than 50 countries worldwide and included the following species: *S. cereale* subsp. *cereale*, *S. cereale* subsp. *vavilovii*, *S. strictum*, *S. sylvestre*, as well as some hybrids between different species.

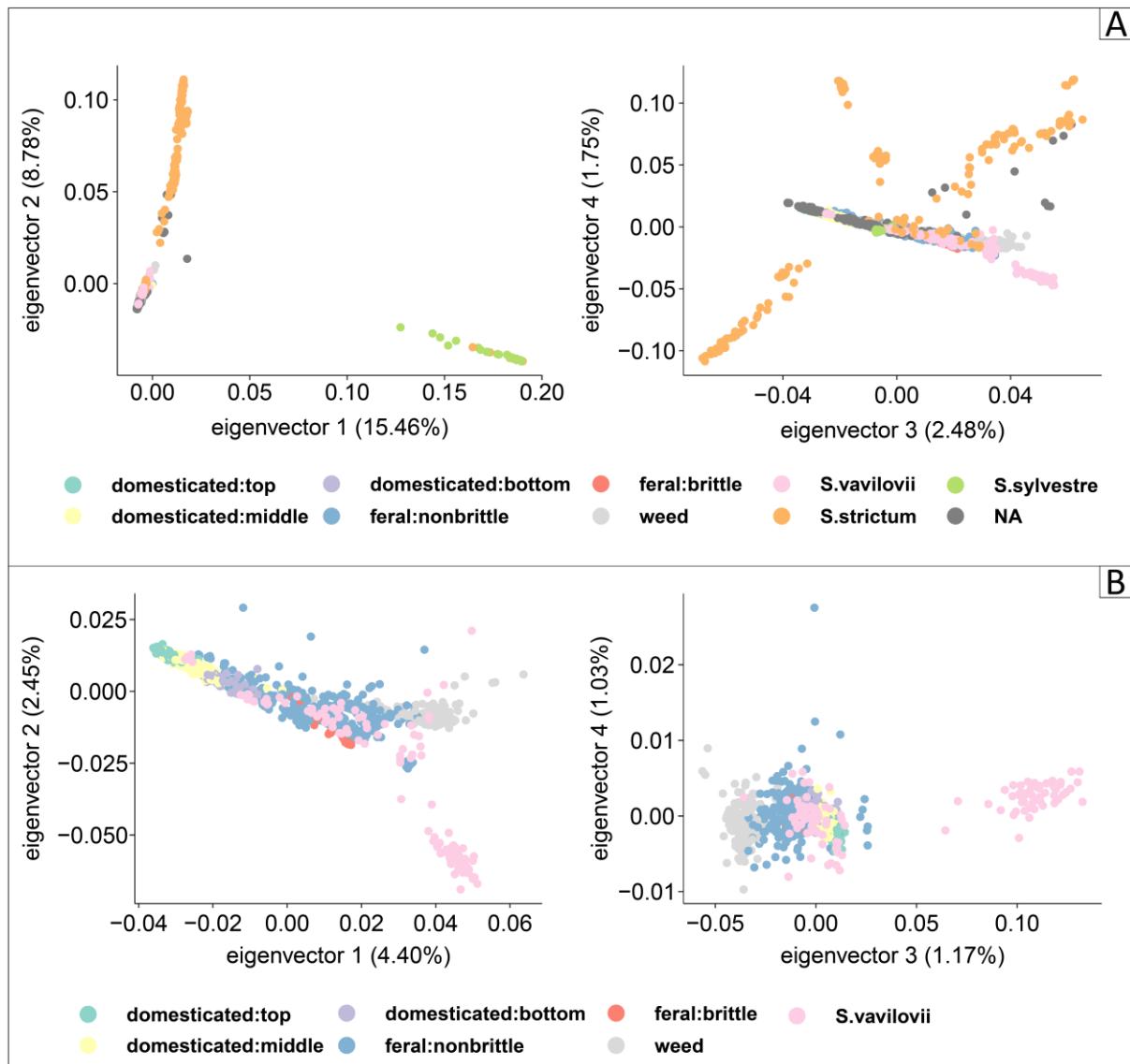
### 3.1. Population Structure Analysis

#### 3.1.1. Principal component analysis (PCA)

At first, a principal component analysis (PCA) was computed for all 1,396 genotypes based on 29,616 SNPs to study population structure.

Based on the first two principal components (PCs) 24.24 % of the total variation can be explained, with eigenvector (EV) 1 explaining 15.46 % and EV 2 explaining 8.78 % (Figure 3A). The variation explained by the first 32 EVs is shown in the appendix A4. In general, by visualizing the first two PCs, three clusters could be clearly assigned among all accessions, separating the three species, *S. cereale*, *S. strictum*, and *S. sylvestre* from each other. Of the three clusters formed, one is exclusively formed by *S. strictum* and in particular EV 2 is responsible for differentiation of the individuals in this cluster. The second cluster comprises primarily genotypes of *S. sylvestre*, with a few individuals of *S. strictum*. Compared to the other two clusters, EV 1 plays an important role in this cluster. Genotypes of wild and domesticated rye, but also *S. cereale* subsp. *vavilovii* clustered closely together and form the third and last cluster. This cluster also includes a few genotypes that are classified as *S. strictum* or *S. sylvestre*. Looking at the distribution of individuals along EV 3 & EV 4, no clear clustering of *S. strictum* and *S. sylvestre* can be observed, in contrast to the observation of the plot of EV 1 & EV 2. On the other hand, *S. strictum* is mostly plotted separately from the other species but shows a high dispersion along both EVs. All individuals of *S. sylvestre* cluster closely together but they lie in another cluster that includes most genotypes of *S. cereale*, but also from the other *Secale* species. This cluster shows a high variance along EV3. In addition, a segregation of some individuals of *S. cereale* subsp. *vavilovii* can be detected here. The cluster formed only by genotypes of *S. cereale* subsp. *vavilovii* is in close proximity to the highly diversified *S. cereale* cluster.

### 3. Results



**Figure 3 Principal component analysis (PCA) of different Secale taxa.** The proportion of variance explained by each principal components (PC) is indicated in parentheses in the axis labels. Dots correspond to individual samples. Coloration according to taxonomy. Number of genotypes per group is given in square brackets. **A: PCA for all Secale taxa.** The first two PCs and third and fourth PC are plotted against each other for *S. cereale* subsp. *cereale* (domesticated:top [93], domesticated:middle [254], domesticated:bottom [111], feral:nonbrittle [272], feral:brittle [19], weed [158]), *S. cereale* subsp. *vavilovii* (*S. vavilovii*) [119], *S. strictum* [154], *S. sylvestre* [24], NA [192]. **B: PCA for Secale cereale.** The first two PCs and third and fourth PC are plotted against each other for *S. cereale* subsp. *cereale* (domesticated:top [93], domesticated:middle [254], domesticated:bottom [111], feral:nonbrittle [272], feral:brittle [19], weed [158]) and *S. cereale* subsp. *vavilovii* (*S. vavilovii*) [119].

To identify if subgroups are within the clusters formed by mostly genotypes of *S. cereale*, a second PCA was performed exclusively with representatives of this species (see Figure 3B). Here, EV 1 (4.40 %) and EV 2 (2.45 %) explain only 6.85 % of the total variance. For this PCA variation explained by the first 32 EVs can be seen in appendix A4, too. Upon closer inspection, one can see that both domesticated and wild rye appear to be separate from each other, with smooth transitions from “domesticated:top” to “domesticated:bottom” and “feral:nonbrittle” to “weed”. Some individuals of *S. cereale* subsp. *vavilovii*

### 3. Results

---

are distributed along the entire length of this cluster, while others tend to form a separated cluster. Looking at the plot of EV 3 (1.17 %) & EV 4 (1.03 %), it reflects the previous findings. Both cultivated and wild rye plot closely together with visible transitions from wild to cultivated rye. *S. cereale* subsp. *vavilovii* can also be found in this cluster, with the main genotypes of this subspecies lying between wild and cultivated rye. However, a second cluster can also be detected in this plot, which consists exclusively of *S. cereale* subsp. *vavilovii*.

In summary, especially when looking at the first plot, the taxonomic classification according to Frederiksen and Petersen (1998) can be understood, in which the genus *Secale* is divided into the three species marked here (*S. sylvestre*, *S. strictum*, and *S. cereale*). The evolutionary closeness of *S. strictum* to *S. cereale* can also be clearly seen here. This appears to contradict the plot of EV 3 & 4, where the species *S. sylvestre* is found immediately within the broad cluster of *S. cereale*. But individuals of *S. strictum* can also be found there, as well as heterogeneously distributed throughout the plot. However, one cluster that stands out includes some individuals from *S. cereale* subsp. *vavilovii*. This observation, in connection with the consistent distribution of domesticated and wild rye observed across both plots, underlines the thesis that domesticated rye, *S. cereale* subsp. *cereale*, being the youngest of all *Secale* species (Sencer and Hawkes 1980) with *S. cereale* subsp. *vavilovii* as the immediate wild ancestor of cultivated rye (Sun et al. 2022a).

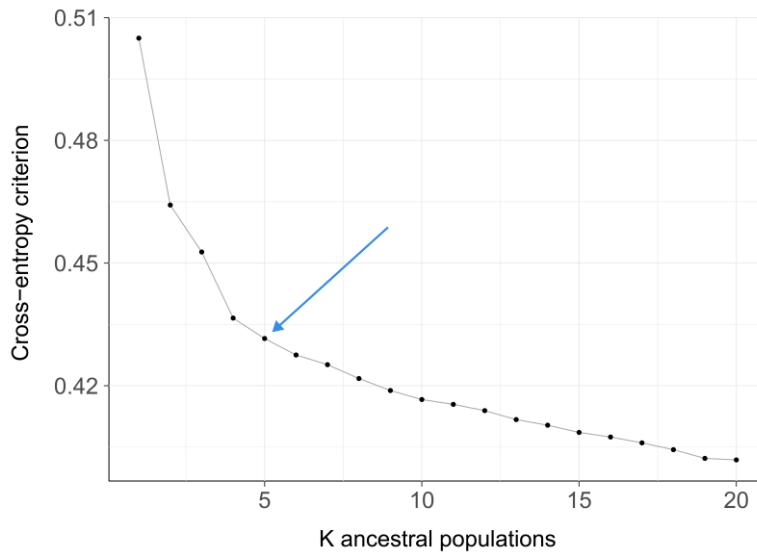
#### 3.1.2. Cross-entropy criterion

To obtain reference points for the calculation of the ancestry coefficient (see 3.1.3), a population structure analysis based on the prediction of masked genotypes was carried out using the cross-entropy criterion. This involves calculating a number (K) of ancestral populations that best matches the underlying genotypic data.

As can be seen in Figure 6, the cross-entropy criterion is 0.505 at a K value of one and continues to fall with increasing ancestral populations until it finally assumes a value of 0.402 at K = 20. Nevertheless, K = 5 (0.432) is considered to be the optimum number of ancestral populations, as for data sets including 1000 individuals genotyped at over 20,000 SNPs, the third digit of the cross-entropy criterion can be significant (Frichot et al. 2014) - a mathematical plateau is reached from this point onwards. The complete list of values can be found in appendix A5.

### 3. Results

---



**Figure 4 Cross-entropy criterion.** Estimation of cross-entropy criterion for varying numbers of ancestral populations ranging from  $K = 1$  to  $K = 20$ . The blue arrow ( $K = 5$ ) marks the optimal number of ancestral populations. The complete list of values can be found in appendix A5.

#### 3.1.3. Ancestry Coefficient analysis

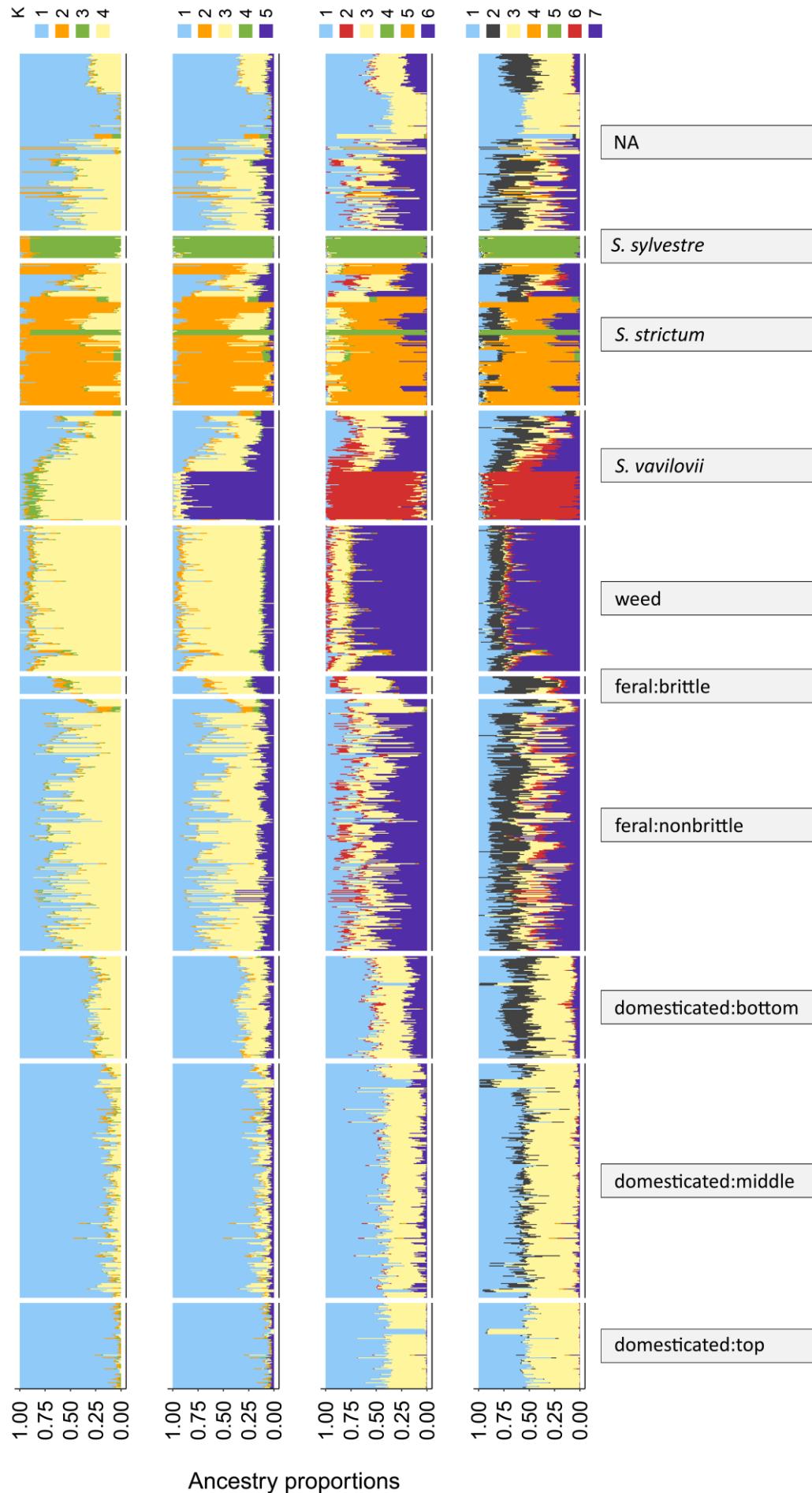
Based on the previous calculation of the optimal number of ancestral populations and taking into account the PCA results, an ancestry coefficient analysis was performed assuming 4-7 ancestral populations (Figure 7) to infer the ancestry of each individual genotype.

Across all four calculations, the genome of the *S. sylvestre* genotypes appears to be largely formed by an ancestral population. This clearly distinguishes itself from the other populations studied, as it accounts for only a small proportion of their genomes. Exceptions to this are a few genotypes of *S. strictum*, which also have a very high proportion of this ancestral population. In addition, *S. strictum* has a high proportion of a second ancestral population, which, like the one described above, is mainly found in this species.

In *S. cereale* subsp. *vavilovii*, the calculations of 5-7 factors show a subdivision into two groups, one of which is primarily characterized by another ancestral population, while the other group has almost balanced genome proportions of different ancestral populations. The calculation with 4 factors does not show this pattern.

For the domesticated subpopulations of *S. cereale*, the calculations show a strong dominance of an ancestral population with 4 or 5 factors, whose genome proportion also decreases with decreasing domestication. This trend continues in the increasingly wild rye. A similar pattern can be observed for the calculation with 6 factors. However, the dominance of an ancestral population within the domesticated rye is not quite as pronounced. Also for the calculation with 7 factors, the genome proportion of two ancestral populations of wild to domesticated rye increases increasingly until finally the genome of the subpopulation domesticated:top is formed almost exclusively in the ratio 1:1 by these two groups.

### 3. Results



**Figure 5 Ancestry coefficient analysis.** Number of ancestral populations ( $K$ ) ranging from 4 to 7 for *S. cereale* subsp. *cereale* (*domesticated:top* [93], *domesticated:middle* [254], *domesticated:bottom* [111], *feral:brittle* [272], *feral:nonbrittle* [272], *weed* [158]), *S. cereale* subsp. *vavilovii* (*S. vavilovii*) [119], *S. strictum* [154], *S. sylvestre* [24], *NA* [192]. Number of genotypes per group is given in square brackets.

### 3. Results

---

#### 3.1.4. Fixation Index ( $F_{ST}$ )

The calculation of the Fixation Index ( $F_{ST}$ ) corroborated the clear separation between *Secale* species and the weak differentiation between *S. cereale* subsp. *cereale* and *S. cereale* subsp. *vavilovii* (Table 3). It can also be seen that the differentiation from domesticated rye to increasingly wild rye increases until it reaches similar values for differentiation to *S. cereale* subsp. *vavilovii*. In addition,  $F_{ST}$  values of *S. cereale* show a closer affinity to *S. strictum* than to *S. sylvestre* and thus confirm the results of the previous calculations of population structure.

**Table 3 Fixation Indices ( $F_{ST}$ ).** between different *Secale* species (*S. cereale* subsp. *cereale* (*S. cereale*), *S. cereale* subsp. *vavilovii* (*S. vavilovii*), *S. strictum* and *S. sylvestre*) and previously defined subgroups. *S. cereale* (dom) for subgroups “domesticated:top”, “domesticated:middle”, “domesticated:bottom”; *S. cereale* (feral) for subgroups “feral:nonbrittle”, “feral:brittle”; *S. cereale* (weed) for subgroup “weed”

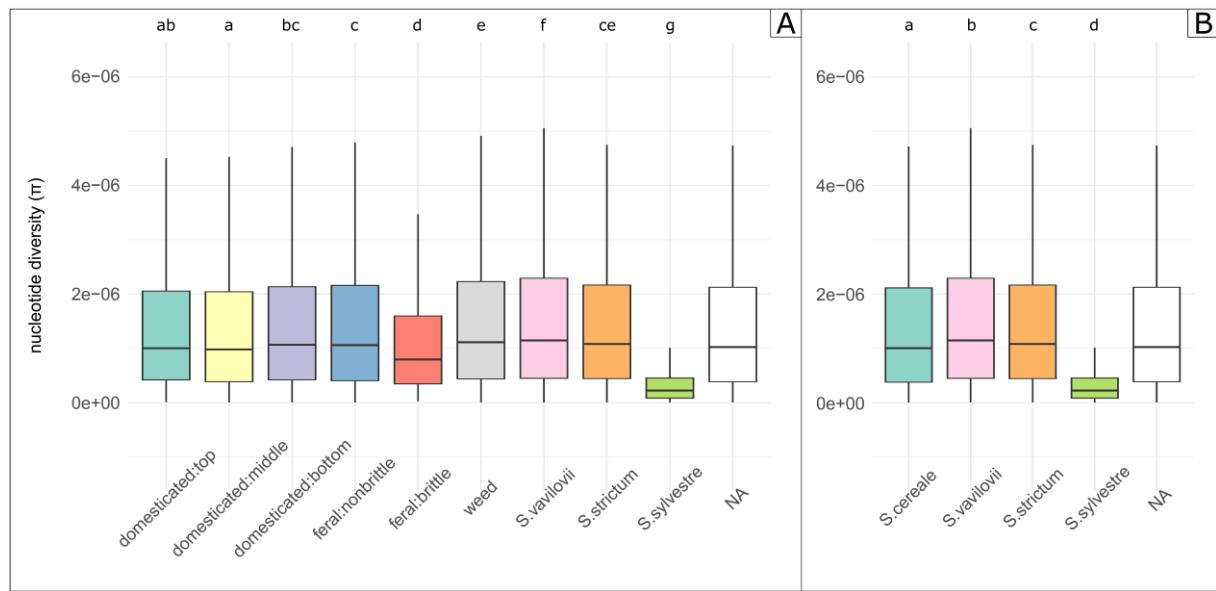
	<i>S. cereale</i> (feral)	<i>S. cereale</i> (weed)	<i>S. vavilovii</i>	<i>S. strictum</i>	<i>S. sylvestre</i>
<i>S. cereale</i> (dom)	0.03	0.07	0.06	0.16	0.32
<i>S. cereale</i> (feral)		0.02	0.02	0.12	0.28
<i>S. cereale</i> (weed)			0.03	0.10	0.28
<i>S. vavilovii</i>				0.12	0.31
<i>S. strictum</i>					0.25

#### 3.1.5. Nucleotide diversity

When analyzing nucleotide diversity, there are no significant differences for the domesticated rye in the subgroups of *S. cereale* subsp. *cereale* (domesticated:top, domesticated:middle, domesticated:bottom), whereas weedy rye is significantly higher in the median (Figure 8). Further on there are significant differences between the individual species of *S. cereale* subsp. *cereale*, *S. cereale* subsp. *vavilovii* and *S. sylvestre* without exception, *S. strictum* shows a similar nucleotide diversity to the wilder members of *S. cereale* subsp. *cereale* (subgroups feral:nonbrittle and weed).

The nucleotide diversity is significantly lower in the feral:brittle and *S. sylvestre* subgroups compared to the other subgroups, which is noticeable in the median as well as in the value range of the first and third quartile. *S. sylvestre* stands out with a particular small nucleotide diversity, as well as a small scattering in it.

### 3. Results



**Figure 6 Nucleotide diversity ( $\pi$ ) of different *Secale* taxa and subgroups.** Number of genotypes per group is given in square brackets. Letters (a-g) indicate significance at the  $P < 0.05$  level using a Bonferroni corrected p-value for pairwise comparisons using t-tests with pooled SD. **A: Nucleotide diversity ( $\pi$ ) for *Secale* taxa and subgroups of *S. cereale* subsp. *cereale*.**  $\pi$  values shown for *S. cereale* subsp. *cereale* (domesticated:top [93], domesticated:middle [254], domesticated:bottom [111], feral:nonbrittle [272], feral:brittle [19], weed [158]), *S. cereale* subsp. *Vavilovii* (*S. vavilovii*) [119], *S. strictum* [154], *S. sylvestre* [24], NA [192]. **B: Nucleotide diversity ( $\pi$ ) for *Secale* taxa.**  $\pi$  values shown for *S. cereale* subsp. *cereale* (*S. cereale*) [1027], *S. cereale* subsp. *vavilovii* (*S. vavilovii*) [119], *S. strictum* [154], *S. sylvestre* [24], NA [192].

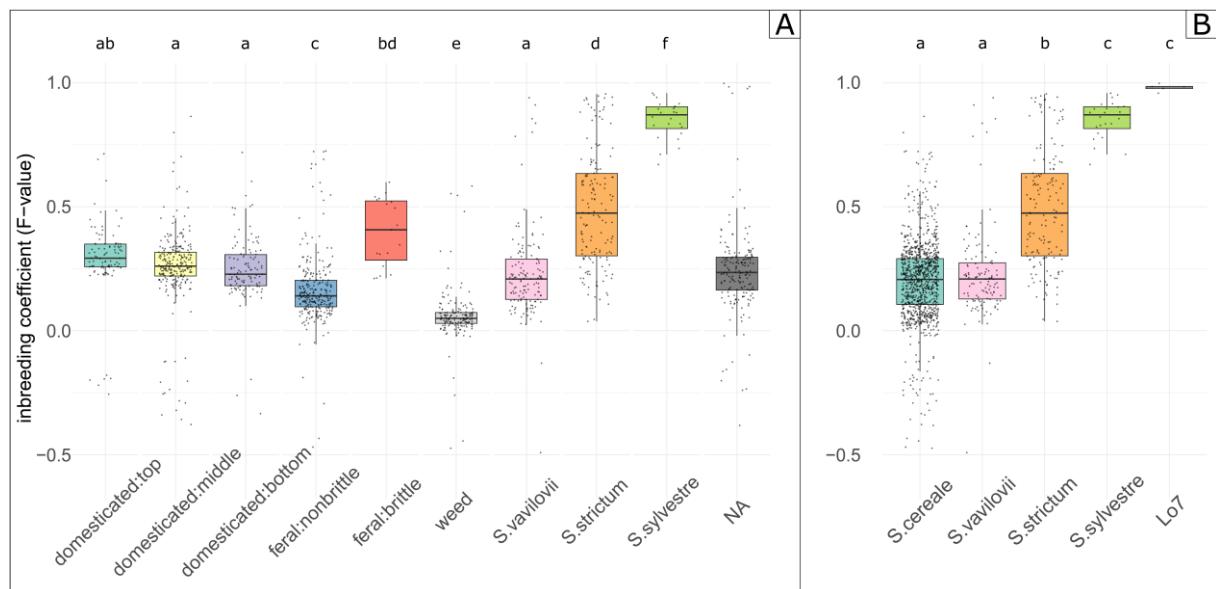
### 3.2. Inbreeding coefficient ( $F$ )

Now that the population structure of the present experimental panel has been analyzed, the next step was to examine the inbreeding coefficient ( $F$ ) of different subpopulations of *Secale*. Based on the previous findings, the rye was further subdivided according to its species (*S. cereale*, *S. strictum* and *S. sylvestre*). In addition, *S. cereale* was subdivided into its subspecies *S. cereale* subsp. *vavilovii* and *S. cereale* subsp. *cereale*. For the latter, further subgroups were made according to the progress of domestication.

Looking at Figure 7A, it can be observed that the median of inbreeding coefficient for domesticated and wild rye is essentially between 0 and 0.4, indicating a lower frequency of heterozygote alleles than expected by the Hardy-Weinberg-Equilibrium (HWE). The inbreeding coefficient tends to increase with increasing domestication, with the exception of the feral:brittel group, whose median is still above the domesticated:top group (0.4 vs 0.3) and whose boxplot also covers a higher value range. The species *S. cereale* subsp. *vavilovii*, on the other hand, has a similar median with 0.2 to the domesticated:bottom and feral:nonbrittle groups, while this is almost 0.5 for *S. strictum*, with a very high dispersion of the individual values. The species *S. sylvestre* has the highest inbreeding coefficient (0.8).

### 3. Results

However, if we look at more than just the values between the first and third quartile, it is noticeable that the inbreeding coefficient fluctuates around the value range of 1 in each case, with the exception of the subgroups feral:brittle and *S. sylvestre* (inbreeding coefficient fluctuates around the value of 0.3). Both subgroups also have the lowest number of individuals in their subgroups, with feral:brittle having 19 individuals and *S. sylvestre* having 24. The inbreeding coefficient reaches negative values for some individuals of different subgroups especially of *S. cereale* subsp. *cereale* ( $F \approx -0.4$ ), while it is characterized by particularly high values ( $F \approx 0.9$ ) in *S. cereale* subsp. *vavilovii*, *S. strictum* and *S. sylvestre*. A high degree of heterogeneity in the individual subgroups can therefore be assumed.



**Figure 7 Inbreeding coefficient (F) of different *Secale* taxa and subgroups. Number of genotypes per group is given in square brackets. Letters (a-f) indicate significance at the  $P < 0.05$  level using a Bonferroni corrected p-value for pairwise comparisons using t-tests with pooled SD. A: Inbreeding coefficient (F) for *Secale* taxa and subgroups of *S. cereale* subsp. *cereale*. F values shown for *S. cereale* subsp. *cereale* (domesticated:top [93], domesticated:middle [254], domesticated:bottom [111], feral:nonbrittle [272], feral:brittle [19], weed [158]), *S. cereale* subsp. *vavilovii* (*S. vavilovii*) [119], *S. strictum* [154], *S. sylvestre* [24], NA [192]. B: Inbreeding coefficient (F) for *Secale* taxa and the inbreeding line “Lo7”. F values shown for *S. cereale* subsp. *cereale* (*S. cereale*) [1027], *S. cereale* subsp. *vavilovii* (*S. vavilovii*) [119], *S. strictum* [154], *S. sylvestre* [24], Lo7 [5].**

It should be noted that a few genotypes that are counted here as NA stand out from this group, as their inbreeding coefficient is almost 1. These belong to the Lo7 inbred line and can be seen as a verification of the correct calculation of the inbreeding coefficient in the experimental panel. An alternative representation of the inbreeding coefficient, divided into only the four different species mentioned here, including Lo7, can be found in Figure 7B. Using this method of presentation no significant difference between *S. cereale* and *S. vavilovii* is recognizable, while significant differences occur between the three species (*S. cereale*, *S. strictum* and *S. sylvestre*).

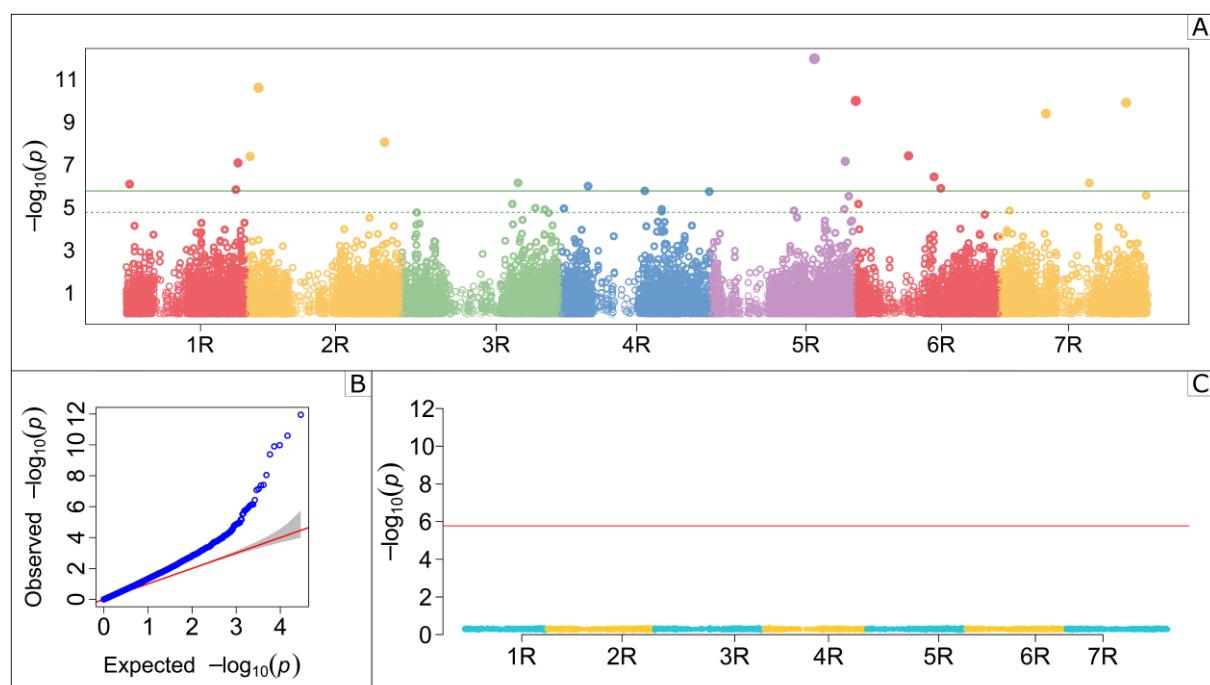
### 3. Results

#### 3.3. Quantitative trait loci (QTL) analysis

##### 3.3.1. Genome-wide association studies (GWAS)

By carrying out a genome-wide association studies (GWAS) with 29,128 SNPs of the 1,396 different rye genotypes, including the inbreeding coefficient ( $F$ ) as a trait, it should be tested whether and to what extent the inbreeding coefficient can be used as an indicator for self-incompatibility (SI). In doing so, attention will be paid to the detection of QTLs that are in close proximity to already known SI loci of rye (see 4.2).

A threshold of the genome-wide significance ( $p$ -value =  $p < 0.05/29,128$  ( $-\log P = 5.765$ )) and a false discovery rate (FDR) cut-off with an alpha of 0.05 were determined. A total of 18 significant makers were detected across the 7 chromosomes of rye (Figure 8A), of which no quantitative trait loci (QTL) is close to the known self-incompatibility S- & Z-loci of rye. Figure 8C also shows the validation of the GWAS in a Manhattan plot. In this plot, the results of 100 GWAS calculations with randomly mixed genotype and phenotype data were summarized. No significant marker-trait associations were found.



**Figure 8 Genome-Wide Association Study (GWAS) for all Secale taxa.** Depicting GWAS single nucleotide polymorphism (SNP) association results of fixed and random model circulating probability unification (FarmCPU). **A:** **Manhattan Plot.** The trait used is inbreeding coefficient ( $F$ ). Chromosome numbers are across the x-axis with the  $-\log(p\text{-value})$  on the y-axis. Points represent SNPs associated with each trait and are colored according to the chromosome on which they reside. The green horizontal line represents the genome-wide significance threshold of Bonferroni adjusted  $p$ -value =  $p < 0.05/29,128$  ( $-\log P = 5.765$ ). The dotted line represents the threshold of false discovery rate adjusted  $p$ -value. **B:** **Quantile-quantile (QQ) plots** of FarmCPU for the inbreeding coefficient ( $F$ ) as the trait. The solid black line showing the expected null distribution of the  $P$  value assuming no associations. The grey area represents the 95% confidence interval for the QQ plot under the null hypothesis of no association between the SNP and the trait. Blue dots showing the observed  $P$  values.

### 3. Results

---

**C: GWAS permutation test.** GWAS performed on randomly mixed genotype and phenotype data. Manhattan plot is showing the mean of a total of 100 iterations without significant marker-trait associations. The red horizontal line represents the genome-wide significance threshold of Bonferroni adjusted p-value ( $-\log P = 5.765$ ).

The percentage of phenotypic variation explained by significant markers ranged from near to 0% to 8%. The SNP “19221” was the most significant association (LOD = 11.952) but only responsible for around 1 % of phenotypic variation (Table 4). The second highest SNP (“4667”, LOD = 10.594) on the other hand is responsible for 5.8 % of the proportion of phenotypic variation. Minor allele frequencies for the single SNP ranged from over 2 % to over 30 %, with mostly being between 10-20 %.

**Table 4 List of all QTLs for all Secale taxa.** Each significant SNP associated with inbreeding coefficient ( $F$ ) implemented by FarmCPU (fixed and random model circulating probability unification) models. Position (Pos) on Chromosome (Chr) enumerated in mega base pairs (Mb). Minor allele frequency = MAF. R2 = adjusted  $r^2$  for linear regression analysis calculated in R (version 4.2.3; R Core Team 2023).

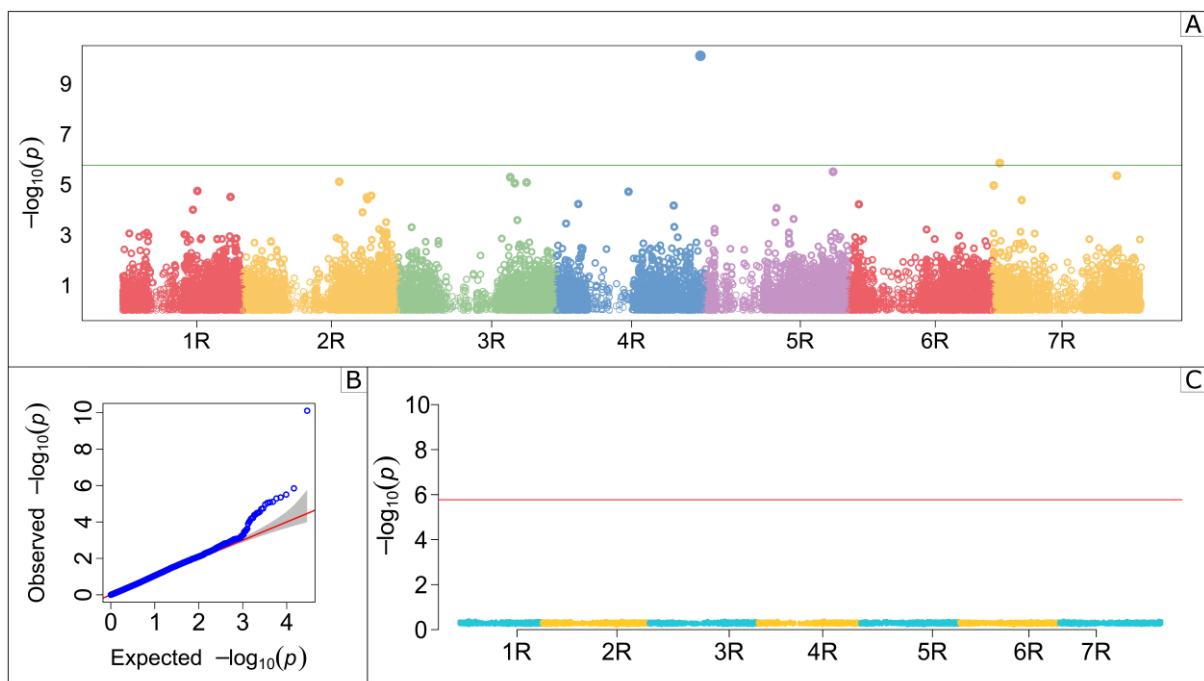
SNP	Chr	Pos (Mb)	MAF	value -log10(p)	R2
116	1R	20,488	0.217	6.080	0.00006
3467	1R	664,908	0.274	5.826	0.04651
3568	1R	678,161	0.204	7.079	0.01180
4224	2R	23,749	0.114	7.378	0.03467
4667	2R	74,986	0.121	10.594	0.05807
8211	2R	840,372	0.212	8.047	0.02315
11215	3R	704,098	0.083	6.146	0.01007
14380	4R	163,700	0.094	5.990	0.00060
14715	4R	507,660	0.193	5.766	0.00698
19221	5R	633,410	0.080	11.952	0.00715
21013	5R	819,890	0.044	7.152	0.00803
21698	6R	8,511	0.335	9.981	0.00138
22512	6R	328,381	0.256	7.408	0.02902
22876	6R	482,972	0.257	6.422	0.08080
22999	6R	523,767	0.072	5.887	0.01122
27124	7R	280,672	0.244	9.383	0.04079
27431	7R	543,612	0.020	6.136	0.04555
28497	7R	767,246	0.148	9.898	0.01868

In the next step, it should be tested whether the significant QTLs from the comprehensive GWAS can also be detected excluding the perennial species *S. strictum*, or whether other associations of certain SNPs with the inbreeding coefficient occur. Perenniality can lead to increased observed homozygosity, and thus to increased inbreeding coefficients causing false-positive significant QTLs.

Figure 9A shows the Manhattan plot for the GWAS with different *S. cereale* taxa excluding *S. strictum*. 29,128 SNPs and 1,243 different rye genotypes, including the inbreeding coefficient as a trait were used. A threshold of the genome-wide significance ( $p\text{-value} = p < 0.05/29,128$  ( $-\log P = 5.765$ )) and a

### 3. Results

false discovery rate (FDR) cut-off with an alpha of 0.05 were determined. A total of 2 significant makers on chromosome 4 and chromosome 7 were detected across the 7 chromosomes of rye, of which no QTL is close to the known self-incompatibility loci (S- or Z-locus). None of the two QTLs occurred in the GWAS with all *Secale* taxa, indicating that the inbreeding coefficient cannot be used exclusively to test the panel of different *Secale* genotypes for self-incompatibility. Figure 9C also shows the validation of the GWAS in a Manhattan plot. In this plot, the results of 100 GWAS calculations with randomly mixed genotype and phenotype data were summarized. No significant marker-trait associations were found.



**Figure 9 Genome-Wide Association Study (GWAS) for *Secale* taxa without *S. strictum*.** Depicting GWAS single nucleotide (SNP) association results of fixed and random model circulating probability unification (FarmCPU). **A: Manhattan Plot.** The trait used is the inbreeding coefficient ( $F$ ). Used *Secale* taxa: *S. cereale* subsp. *cereale*, *S. cereale* subsp. *vavilovii*, *S. sylvestre*. Chromosome numbers are across the x-axis with the  $-\log(p\text{-value})$  on the y-axis. Points represent SNPs associated with each trait and are colored according to the chromosome on which they reside. The green horizontal line represents the genome-wide significance threshold of Bonferroni adjusted  $p\text{-value} = p < 0.05/29,128$  ( $-\log P = 5.765$ ). The dotted line represents the threshold of false discovery rate adjusted  $p\text{-value}$ . **B: Quantile-quantile (QQ) plots** of FarmCPU for the the inbreeding coefficient ( $F$ ) as the trait. The solid black line showing the expected null distribution of the  $P$  value assuming no associations. The grey area represents the 95% confidence interval for the QQ plot under the null hypothesis of no association between the SNP and the trait. Blue dots showing the observed  $P$  values. **C: GWAS permutation test.** GWAS performed on randomly mixed genotype and phenotype data. Manhattan plot is showing the mean of a total of 100 iterations without significant marker-trait associations. The red horizontal line represents the genome-wide significance threshold of Bonferroni adjusted  $p\text{-value}$  ( $-\log P = 5.765$ ).

The percentage of phenotypic variation explained by these two significant markers for *Secale* taxa without *S. strictum* are 2 % and 3%. Both Minor allele frequencies are under 3 %, and thus suggest rare alleles (Table 5).

### 3. Results

---

**Table 5 List of all QTLs for Secale taxa without *S. strictum*.** Each significant SNP associated with inbreeding coefficient ( $F$ ) implemented by FarmCPU (fixed and random model circulating probability unification) models. Position (Pos) on Chromosome (Chr) enumerated in mega bases (Mb). Minor allele frequency = MAF. R2 = adjusted  $r^2$  for linear regression analysis calculated in R (version 4.2.3; R Core Team 2023).

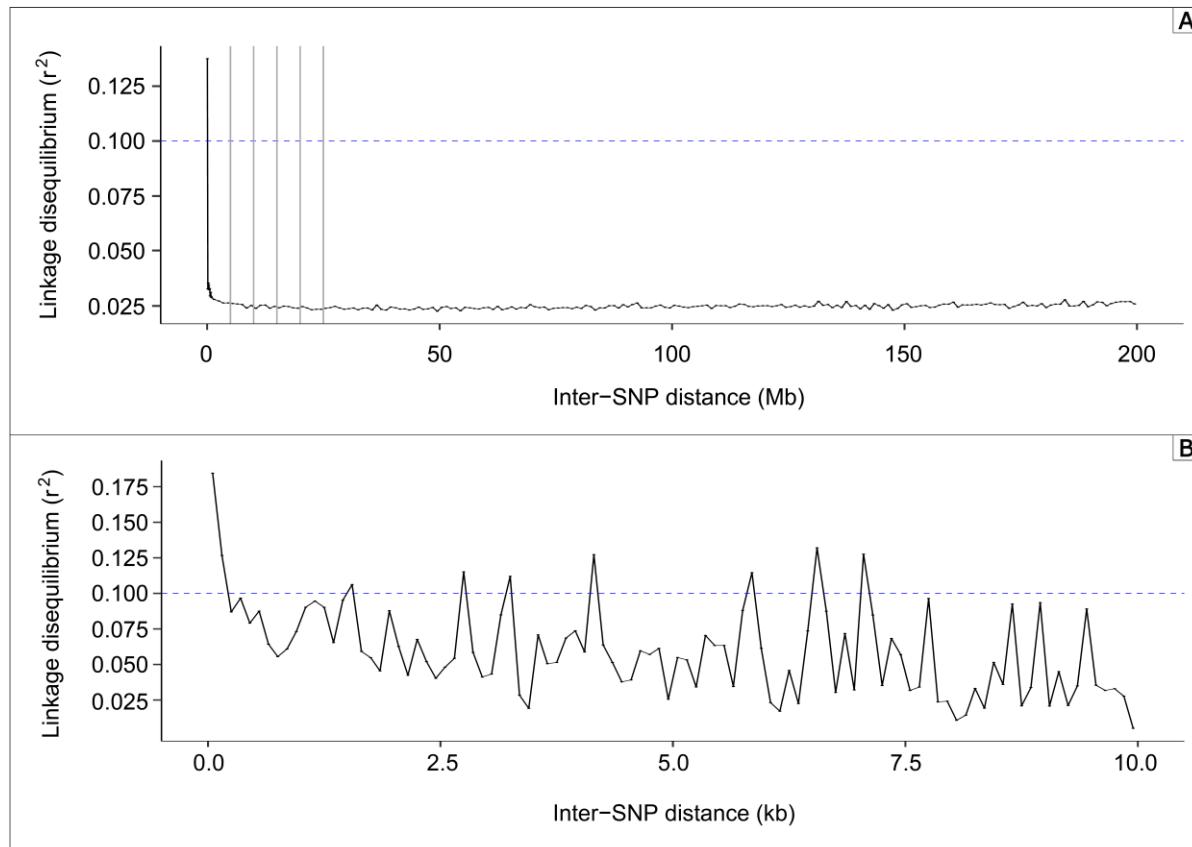
SNP	Chr	Pos (Mb)	MAF	value -log10(p)	R2
16769	4R	878,038	0.005	10.107	0.03388
25651	7R	38,865	0.029	5.847	0.02090

#### 3.3.2. Linkage Disequilibrium (LD) decay estimates and Gene Annotations

To determine the linkage disequilibrium (LD) decay, the pair-wise correlation ( $r^2$ ) between marker pairs of all 29,616 SNPs was determined and the results were plotted for an area of 200 mega bases (Mb) for all chromosomes. Average values for 100 kilo bases (kb) intervals for the first 1 Mb were calculated, followed by 1 Mb intervals for the length of 1 Mb to 200 Mb. This can be seen in Figure 10A, with the blue dashed line representing the LD threshold of  $r^2 = 0.1$ . For a more detailed examination, the LD decay was also calculated for a region of 10 kb with an interval of 100 base pairs (bp) (see Figure 10B). The complete list of average  $r^2$  values for the global LD can be found in the appendix A6.

The LD decreases rapidly within 10 kb, so that it drops from around 0.175 on average to around 0.025 and remains stable for the further course of the calculated area of 200 Mb.

### 3. Results



**Figure 10 Linkagen disequilibrium (LD) decay for all rye chromosomes.** Pair-wise correlation ( $r^2$ ) between marker pairs with averaged values at defined intervals. Blue dashed line: representing the LD threshold of  $r^2 = 0.1$ . **A:** LD decay for 200 mega bases (Mb). Average  $r^2$  values were calculated in 0.1 Mb intervals for the first 1 Mb and 1 Mb intervals for the length of 1 Mb to 200 Mb. Gray, vertical line define range of 5, 10, 15, 20 and 25 Mb. **B:** LD decay for 10 kilo bases (kb). Average  $r^2$  values were calculated in 0.1 kb intervals.

To limit the area for the QTLs in which the desired genes can be searched, the LD decay at the QTLs was calculated in an interval of 20 Mb upwards and downwards of the specific SNP (Figure 11). The yellow circles represent the average values of pair-wise correlation ( $r^2$ ) to the surrounding markers in 1 Mb intervals. It can be observed that for almost all SNPs there is little or no correlation to the surrounding markers. The exceptions are the SNPs “22512” and “22876” near the centromere on chromosome 6R. The former also shows increased  $r^2$  values over a length of 15 Mb downstream, while these fluctuate both upstream and downstream for SNP “22876” over the length of 20 Mb. The complete list of average  $r^2$  values for the LD at QTLs can be found in the appendix A7.

### 3. Results

---

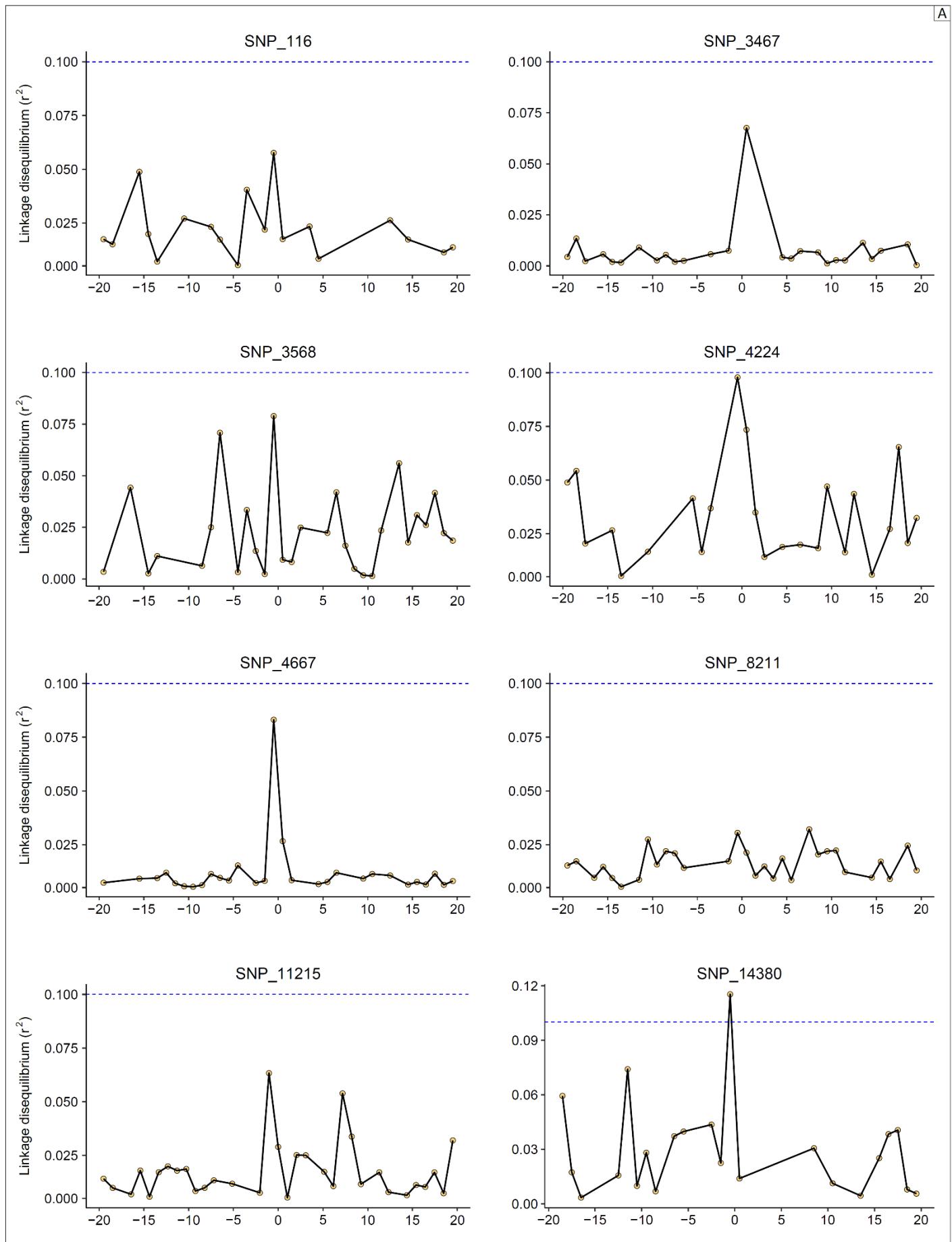


Figure continued on next page

### 3. Results

---

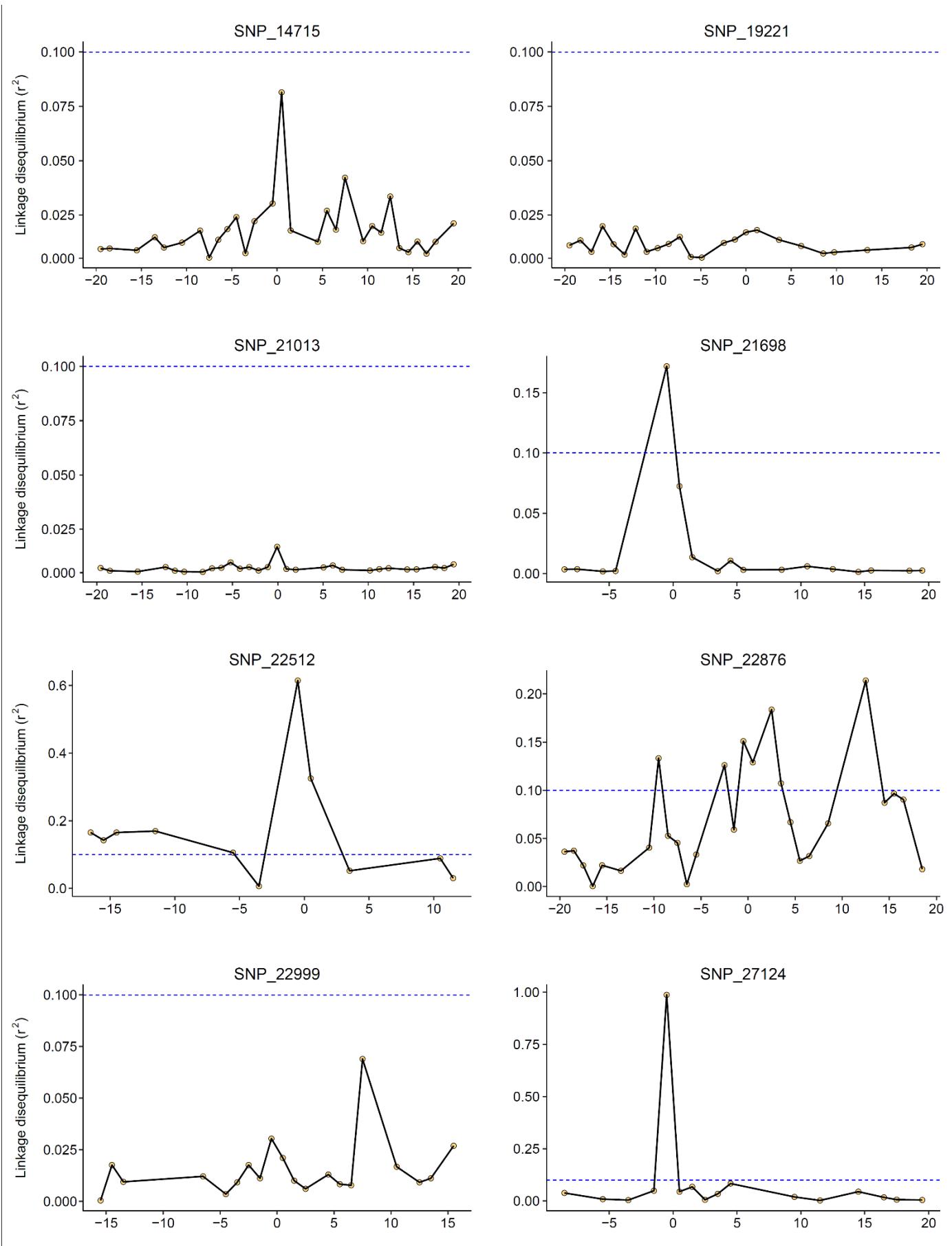
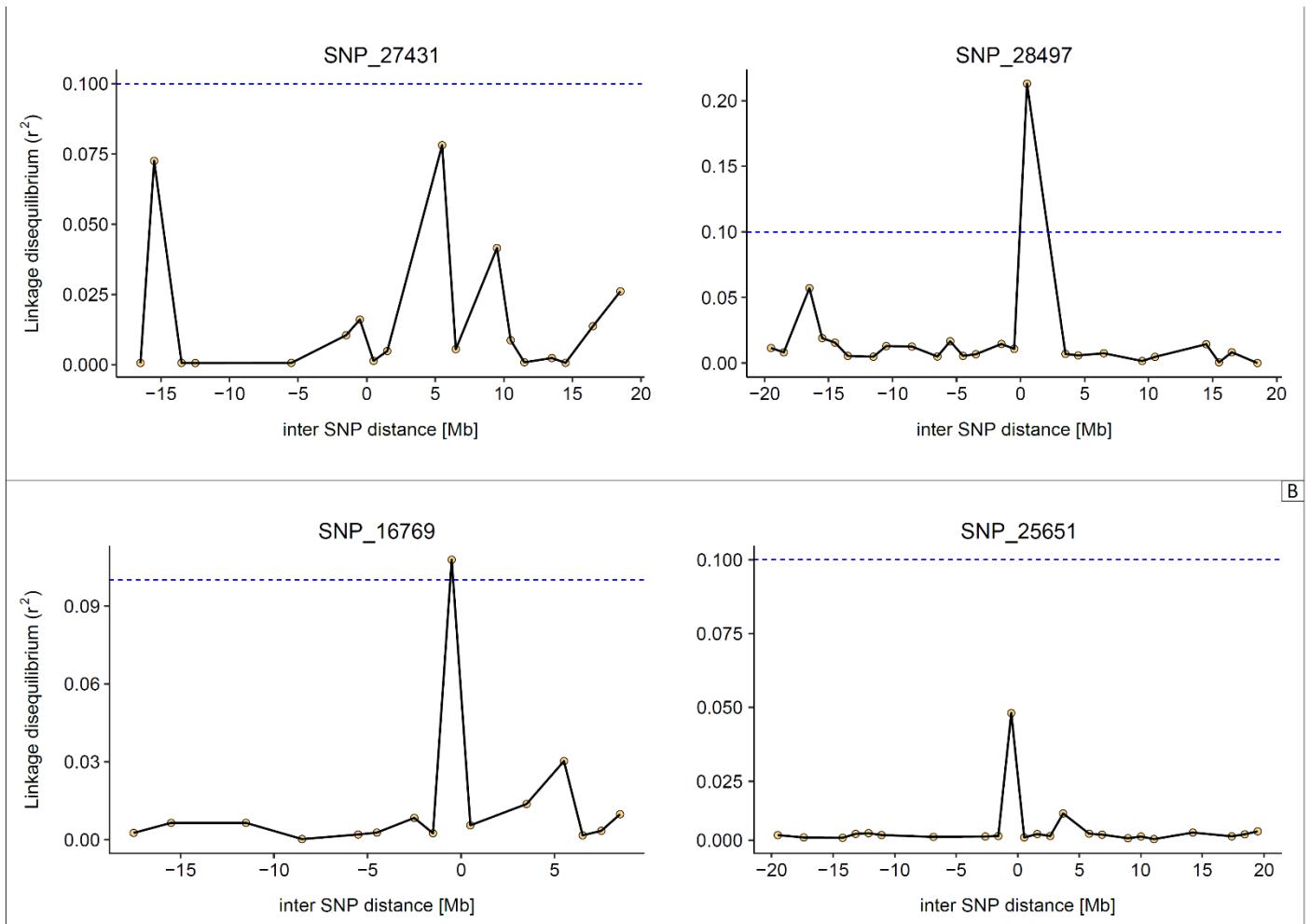


Figure continued on next page

### 3. Results



**Figure 11 Linkage disequilibrium (LD) decay for QTLs.** Pair-wise correlation ( $r^2$ ) between marker pairs for the SNP found out in the GWAS and surrounding SNPs over a total length of 20 mega base pairs (Mb) downstream and upstream. Average  $r^2$  value for 1 Mb intervals was calculated. **A: LD decay for QTLs of GWAS for all Secale taxa.** 18 SNPs in total distributed over alle 7 rye chromosomes. Position is mentioned in Table 4. **B: LD decay for QTLs of GWAS for Secale taxa without S. strictum.** The 2 SNPs map on chromosome 2R and 4R. Position is mentioned in Table 5.

Based on the LD decay, gene annotations were made to the Lo7 rye reference genome according to Lux (2020) for each SNP. Gene annotations for both GWAS “for all *Secale* taxa” and GWAS “for *Secale* taxa without *S. strictum*” are shown in Table 6. The interval range was selected based on the LD decay of the considered SNP (see Figure 11). If a threshold of von  $r^2 = 0.1$  was exceeded in a surrounding interval of 1 Mb of the SNP, this was defined as the region to be analyzed. If no interval range was available, the previously calculated general LD decay was used to determine the interval for gene annotation (10 kb). The distance to the SNP was calculated based on the average distance between the start and end sequence of each gene.

For most SNPs of both GWAS, the interval to be analyzed was determined by the global LD decay. For a total of 5 QTLs, at least one of the two boundaries was defined by the specific 1 Mb upstream or downstream interval. And for 2 of the 5 QTLs on chromosome 4R, the boundaries were several Mb away. For both GWAS together, a total of 276 gene annotations were made for all 20 QTLs, with 262

### 3. Results

---

gene annotations in the first GWAS "for all *Secale* taxa" for the 18 QTLs across all 7R rye chromosomes and 14 gene annotations for the 2 QTLs on chromosomes 4R and 7R in the GWAS "for *Secale* taxa excluding *S. strictum*".

**Table 6 Gene Annotation** to rye reference genome of "Lo7" for GWAS results for all *Secale* taxa (upper table) and GWAS results for *Secale* taxa without *S. strictum* (lower table). The interval range was selected based on the LD decay of the considered SNP (see Figure 11). If no interval range was available, the previously calculated general LD decay was used to determine the interval for gene annotation. Distance to SNP calculated based on the average distance between starting sequence and ending sequence of each gene. Chr = Chromosome, (bp) = base pair. Gene annotations according to Lux (2020). CC= Confidence class. HC = high confidence. LC = low confidence.

Chr	start [bp]	end [bp]	Gene Description	Gene ID	CC	Distance to SNP [bp]	SNP
1R	664,907,667	664,911,334	Kinase family protein	SECCE1Rv1G0052850.1	HC	1,462	3467
	678,153,889	678,154,409	60S ribosomal protein L35a-like protein	SECCE1Rv1G0055080.1	HC	-6,865	3568
	678,160,872	678,162,347	transmembrane protein, putative (DUF247)	SECCE1Rv1G0055090.1	HC	596	
2R	22,751,926	22,752,519	Transposon protein, putative, CACTA, En/Spm sub-class, expressed	SECCE2Rv1G0068350.1	LC	-996,688	4224
	22,759,986	22,760,567	Glycine-rich cell wall structural protein 2	SECCE2Rv1G0068360.1	HC	-988,634	
	22,939,480	22,940,602	Strictosidine synthase	SECCE2Rv1G0068370.1	HC	-808,869	
	22,983,917	22,985,155	Maturase K	SECCE2Rv1G0068380.1	HC	-764,374	
	22,999,599	22,999,919	Glutamate 5-kinase	SECCE2Rv1G0068390.1	LC	-749,151	
	23,052,145	23,053,259	Strictosidine synthase	SECCE2Rv1G0068400.1	HC	-696,208	
	23,143,959	23,145,082	Strictosidine synthase	SECCE2Rv1G0068410.1	HC	-604,390	
	23,149,461	23,151,105	Cytochrome P450	SECCE2Rv1G0068420.1	HC	-598,627	
	23,159,812	23,161,467	Chloride channel protein	SECCE2Rv1G0068430.1	LC	-588,270	
	23,176,226	23,176,771	Cytochrome P450, putative	SECCE2Rv1G0068440.1	HC	-572,412	
	23,184,085	23,185,356	F-box protein skip23	SECCE2Rv1G0068450.1	HC	-564,190	
	23,518,356	23,518,631	RNA polymerase II transcription mediator	SECCE2Rv1G0068460.1	LC	-230,416	
	23,549,981	23,550,415	hAT transposon superfamily	SECCE2Rv1G0068470.1	LC	-198,712	
	23,564,754	23,566,459	Protein trichome birefringence	SECCE2Rv1G0068480.1	HC	-183,304	
	23,575,857	23,576,477	Mitochondrial transcription termination factor-like	SECCE2Rv1G0068490.1	LC	-172,743	
	23,579,983	23,580,306	Retrotransposon protein, putative, Ty1-copia subclass	SECCE2Rv1G0068500.1	LC	-168,766	
	23,619,670	23,620,192	DNA topoisomerase	SECCE2Rv1G0068510.1	LC	-128,979	
	23,625,894	23,626,364	Pentatricopeptide repeat-containing protein	SECCE2Rv1G0068520.1	LC	-122,781	
	23,630,047	23,631,522	Retrotransposon protein, putative, Ty1-copia subclass	SECCE2Rv1G0068530.1	LC	-118,126	
	23,713,589	23,714,230	Omega-hexatoxin-Hi1c	SECCE2Rv1G0068540.1	LC	-35,000	
	23,728,335	23,730,158	Repetitive proline-rich cell wall protein	SECCE2Rv1G0068550.1	LC	-19,664	
	23,732,664	23,733,152	Cell surface antigen I/II	SECCE2Rv1G0068560.1	LC	-16,002	
	23,740,035	23,742,592	FBD-associated F-box protein	SECCE2Rv1G0068570.1	HC	-7,596	
	23,753,434	23,754,420	Dof zinc finger protein	SECCE2Rv1G0068580.1	HC	5,017	
	74,986,120	74,987,682	WD repeat domain phosphoinositide-interacting protein	SECCE2Rv1G0074200.1	HC	673	4667
	840,369,758	840,371,356	Pentatricopeptide repeat-containing protein	SECCE2Rv1G0125370.1	HC	-1,268	8211

Table continued on next page

### 3. Results

3R	704,093,536	704,095,699	Auxin-responsive protein	SECCE3Rv1G0187710.1	HC	-3,082	11215
4R	162,930,288	162,933,095	Type I inositol-1, 4, 5-trisphosphate 5-phosphatase	SECCE4Rv1G0230340.1	HC	-768,116	14380
	163,117,545	163,118,024	Serine/threonine-protein phosphatase 4 regulatory subunit 3	SECCE4Rv1G0230350.1	HC	-582,024	
	163,118,438	163,123,291	Protein DETOXIFICATION	SECCE4Rv1G0230360.1	HC	-578,944	
	163,224,379	163,224,717	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE4Rv1G0230370.1	LC	-475,260	
	163,371,286	163,371,759	DEK domain-containing chromatin associated protein	SECCE4Rv1G0230380.1	LC	-328,286	
	163,425,391	163,430,470	Methyl esterase	SECCE4Rv1G0230390.1	HC	-271,878	
	163,665,175	163,667,726	ELMO domain containing protein	SECCE4Rv1G0230400.1	LC	-33,358	
	163,685,325	163,685,693	FKBP-like peptidyl-prolyl cis-trans isomerase family protein	SECCE4Rv1G0230410.1	HC	-14,299	
	163,686,676	163,701,972	Regulation of nuclear pre-mRNA domain-containing protein 1B	SECCE4Rv1G0230420.1	HC	-5,484	
	163,706,790	163,707,669	DUF309 domain protein	SECCE4Rv1G0230430.1	HC	7,422	
5R	507,658,415	507,659,927	extracellular ligand-gated ion channel protein (DUF3537)	SECCE4Rv1G0243890.1	HC	-523	14715
	633,410,105	633,411,364	F-box protein	SECCE5Rv1G0342680.1	HC	898	19221
	8,503,936	8,519,237	Endoglucanase 11	SECCE6Rv1G0378440.1	LC	362	21698
	311,425,354	311,425,593	PIF1-like helicase	SECCE6Rv1G0394920.1	LC	-16,955,436	22512
	311,615,676	311,643,320	CMP-sialic acid transporter 5	SECCE6Rv1G0394930.1	HC	-16,751,412	
	312,377,617	312,378,105	Transposon Ty3-I Gag-Pol polyprotein	SECCE6Rv1G0394940.1	LC	-16,003,049	
	312,378,289	312,379,518	Retrotransposon protein, putative, Ty3-gypsy subclass	SECCE6Rv1G0394950.1	LC	-16,002,006	
	312,836,544	312,855,196	Apoptosis inhibitor 5	SECCE6Rv1G0394960.1	HC	-15,535,040	
	312,956,097	312,956,342	3-hydroxyacyl-[acyl-carrier-protein] dehydratase FabZ	SECCE6Rv1G0394970.1	HC	-15,424,690	
	313,995,748	313,998,465	mesoderm induction early response protein	SECCE6Rv1G0394980.1	HC	-14,383,804	
	315,506,707	315,507,014	Chaperone protein DnaK	SECCE6Rv1G0394990.1	LC	-12,874,050	
	315,512,180	315,512,578	BTB/POZ/MATH-domains containing protein	SECCE6Rv1G0395000.1	LC	-12,868,531	
	315,673,973	315,706,195	GPI ethanolamine phosphate transferase 1	SECCE6Rv1G0395010.1	HC	-12,690,826	
	315,712,741	315,719,348	2, 3-bisphosphoglycerate-independent phosphoglycerate mutase	SECCE6Rv1G0395020.1	HC	-12,664,866	
	316,446,756	316,447,175	alpha/beta-Hydrolases superfamily protein	SECCE6Rv1G0395030.1	LC	-11,933,944	
	316,646,276	316,647,148	Retrotransposon protein, putative, Ty1-copia subclass	SECCE6Rv1G0395040.1	LC	-11,734,198	
	316,647,167	316,647,625	Retrotransposon protein, putative, Ty1-copia subclass	SECCE6Rv1G0395050.1	LC	-11,733,514	
	316,647,683	316,647,949	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0395060.1	LC	-11,733,094	
	316,842,156	316,843,952	NFX1-type zinc finger-containing protein 1	SECCE6Rv1G0395070.1	LC	-11,537,856	
	316,901,880	316,912,704	intracellular protein transporter	SECCE6Rv1G0395080.1	HC	-11,473,618	
	316,916,699	316,917,052	Actin-binding FH2/DRF autoregulatory protein	SECCE6Rv1G0395090.1	LC	-11,464,034	
	317,078,425	317,080,055	U6 snRNA-associated Sm-like protein LSm7	SECCE6Rv1G0395100.1	LC	-11,301,670	
	317,410,666	317,411,133	BED zinc finger, hAT family dimerization domain	SECCE6Rv1G0395110.1	LC	-10,970,010	

Table continued on next page

### 3. Results

---

317,411,261	317,411,713	BED zinc finger, hAT family dimerization domain	SECCE6Rv1G0395120.1	LC	-10,969,423
317,428,281	317,428,790	Oxygen-evolving enhancer protein 3-2	SECCE6Rv1G0395130.1	LC	-10,952,374
317,469,219	317,469,425	DNA-directed RNA polymerase subunit beta''	SECCE6Rv1G0395140.1	LC	-10,911,588
318,150,329	318,150,841	L-threonine dehydratase biosynthetic IlvA	SECCE6Rv1G0395150.1	LC	-10,230,325
318,154,905	318,157,883	18S pre-ribosomal assembly protein gar2-related protein	SECCE6Rv1G0395160.1	HC	-10,224,516
318,167,616	318,167,852	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0395170.1	LC	-10,213,176
319,255,013	319,257,023	Nucleosome assembly protein 1-1	SECCE6Rv1G0395180.1	LC	-9,124,892
319,888,423	319,893,572	Ras-related protein	SECCE6Rv1G0395190.1	HC	-8,489,912
319,894,340	319,897,003	Small nuclear ribonucleoprotein Sm D1	SECCE6Rv1G0395200.1	HC	-8,485,238
320,307,738	320,308,145	Calcium-transporting ATPase	SECCE6Rv1G0395210.1	LC	-8,072,968
320,919,983	320,941,234	ATP-dependent Clp protease ATP-binding subunit ClpX	SECCE6Rv1G0395220.1	HC	-7,450,302
321,746,304	321,749,022	Protein phosphatase 2A regulatory B subunit family protein	SECCE6Rv1G0395230.1	HC	-6,633,247
322,153,951	322,154,820	ATP synthase subunit alpha	SECCE6Rv1G0395240.1	LC	-6,226,524
322,155,925	322,156,167	ATP synthase subunit 9, mitochondrial	SECCE6Rv1G0395250.1	LC	-6,224,864
322,660,277	322,660,669	Transposon Ty3-I Gag-Pol polyprotein	SECCE6Rv1G0395260.1	LC	-5,720,437
323,274,586	323,275,047	serine-rich adhesin for platelets-like protein	SECCE6Rv1G0395270.1	LC	-5,106,094
323,279,509	323,281,527	Cytochrome b-c1 complex subunit 6	SECCE6Rv1G0395280.1	LC	-5,100,392
323,583,373	323,589,186	Vacuolar-processing enzyme	SECCE6Rv1G0395290.1	LC	-4,794,630
323,590,703	323,591,272	Vacuolar-processing enzyme	SECCE6Rv1G0395300.1	LC	-4,789,922
323,637,867	323,638,376	5-dehydro-2-deoxygluconokinase	SECCE6Rv1G0395310.1	LC	-4,742,788
323,805,670	323,806,809	Transposon Ty3-I Gag-Pol polyprotein	SECCE6Rv1G0395320.1	LC	-4,574,670
323,855,771	323,856,082	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0395330.1	LC	-4,524,984
324,215,350	324,215,541	L-fucokinase/GDP-L-fucose pyrophosphorylase	SECCE6Rv1G0395340.1	LC	-4,165,464
324,638,661	324,639,017	autoinhibited Ca2+-ATPase 1	SECCE6Rv1G0395350.1	LC	-3,742,071
325,024,479	325,024,787	E3 ubiquitin-protein ligase Arkadia	SECCE6Rv1G0395360.1	LC	-3,356,277
325,243,795	325,244,211	Calcium-transporting ATPase	SECCE6Rv1G0395370.1	LC	-3,136,907
325,244,335	325,244,571	Calcium-dependent phosphotriesterase superfamily protein	SECCE6Rv1G0395380.1	LC	-3,136,457
325,322,979	325,329,288	MutL DNA mismatch repair protein	SECCE6Rv1G0395390.1	HC	-3,054,776
326,516,078	326,516,255	Proteasome-associated protein ECM29-like protein	SECCE6Rv1G0395400.1	LC	-1,864,744
327,070,796	327,071,116	3-phosphoshikimate 1-carboxyvinyltransferase	SECCE6Rv1G0395410.1	LC	-1,309,954
327,162,350	327,163,009	Transposon Ty3-I Gag-Pol polyprotein	SECCE6Rv1G0395420.1	LC	-1,218,230
328,397,962	328,402,367	Choline transporter-like protein	SECCE6Rv1G0395430.1	HC	19,254
328,847,900	328,848,123	Disease resistance protein (TIR-NBS-LRR class) family	SECCE6Rv1G0395440.1	LC	467,102
328,849,241	328,849,441	Transposon protein, putative, CACTA, En/Spm sub-class	SECCE6Rv1G0395450.1	LC	468,431
329,329,945	329,330,548	2-oxoglutarate-dependent dioxygenase	SECCE6Rv1G0395460.1	LC	949,336
329,341,344	329,342,794	Cortactin-binding protein 2	SECCE6Rv1G0395470.1	HC	961,159
329,351,257	329,356,287	Alkaline phosphatase D	SECCE6Rv1G0395480.1	HC	972,862

Table continued on next page

### 3. Results

473,042,621	473,044,372	Zinc finger family protein	SECCE6Rv1G0401680.1	HC	-9,928,402
473,047,938	473,052,796	Protein transport sec20	SECCE6Rv1G0401690.1	HC	-9,921,532
473,058,737	473,059,924	Expansin protein	SECCE6Rv1G0401700.1	HC	-9,912,568
473,131,364	473,131,980	Beta-galactosidase BgaP	SECCE6Rv1G0401710.1	LC	-9,840,227
473,363,724	473,364,902	Expansin protein	SECCE6Rv1G0401720.1	HC	-9,607,586
473,467,024	473,467,353	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0401730.1	LC	-9,504,710
473,666,817	473,668,025	Expansin protein	SECCE6Rv1G0401740.1	HC	-9,304,478
473,878,390	473,878,683	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0401750.1	LC	-9,093,362
474,012,584	474,018,339	F-box protein-like	SECCE6Rv1G0401760.1	HC	-8,956,438
474,019,157	474,021,722	Remorin	SECCE6Rv1G0401770.1	HC	-8,951,460
474,078,581	474,080,347	nuclear fusion defective 6	SECCE6Rv1G0401780.1	LC	-8,892,435
474,287,243	474,287,662	Glutamate receptor 2.7	SECCE6Rv1G0401790.1	LC	-8,684,446
474,451,225	474,451,669	Zinc finger protein	SECCE6Rv1G0401800.1	HC	-8,520,452
474,561,736	474,563,637	Retrotransposon protein, putative, Ty3-gypsy subclass	SECCE6Rv1G0401810.1	LC	-8,409,212
474,604,873	474,608,548	WD-repeat protein	SECCE6Rv1G0401820.1	HC	-8,365,188
474,612,906	474,613,837	Aquaporin	SECCE6Rv1G0401830.1	HC	-8,358,528
474,760,403	474,761,739	auxin canalization protein (DUF828)	SECCE6Rv1G0401840.1	HC	-8,210,828
474,762,439	474,762,771	PPPDE peptidase domain-containing 2	SECCE6Rv1G0401850.1	LC	-8,209,294
474,776,773	474,777,669	Retrotransposon protein, putative, Ty1-copia subclass	SECCE6Rv1G0401860.1	LC	-8,194,678
474,777,904	474,778,248	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0401870.1	LC	-8,193,823
474,793,456	474,793,803	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0401880.1	LC	-8,178,270
474,890,825	474,901,665	transmembrane protein	SECCE6Rv1G0401890.1	HC	-8,075,654
474,902,142	474,903,149	Acetylglutamate kinase-like protein	SECCE6Rv1G0401900.1	HC	-8,069,254
475,097,289	475,100,504	Chaperone protein dnaJ, putative	SECCE6Rv1G0401910.1	HC	-7,873,002
475,101,718	475,102,470	ethylene-responsive transcription factor	SECCE6Rv1G0401920.1	HC	-7,869,805
475,115,971	475,116,660	ethylene-responsive transcription factor	SECCE6Rv1G0401930.1	HC	-7,855,584
475,472,248	475,472,919	ethylene-responsive transcription factor	SECCE6Rv1G0401940.1	HC	-7,499,316
475,620,733	475,632,083	E3 ubiquitin-protein ligase BRE1-like 1	SECCE6Rv1G0401950.1	HC	-7,345,491
476,207,949	476,208,137	Leucine-rich repeat protein kinase family protein	SECCE6Rv1G0401960.1	HC	-6,763,856
476,208,414	476,211,632	Anaphase-promoting complex subunit 8-like protein	SECCE6Rv1G0401970.1	HC	-6,761,876
476,470,676	476,472,952	UDP-N-acetylglucosamine transferase subunit ALG14	SECCE6Rv1G0401980.1	HC	-6,500,085
476,474,722	476,476,625	Chaperone protein dnaJ	SECCE6Rv1G0401990.1	HC	-6,496,226
476,509,237	476,509,833	DUF740 family protein	SECCE6Rv1G0402000.1	HC	-6,462,364
477,188,485	477,189,537	PLATZ transcription factor	SECCE6Rv1G0402010.1	HC	-5,782,888
477,456,621	477,457,766	Ring finger protein, putative	SECCE6Rv1G0402020.1	HC	-5,514,706
477,710,097	477,712,787	Basic helix-loop-helix transcription factor	SECCE6Rv1G0402030.1	HC	-5,260,457
478,273,065	478,273,391	ATP-dependent DNA helicase PIF1	SECCE6Rv1G0402040.1	LC	-4,698,671
478,999,749	479,000,111	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0402050.1	LC	-3,971,969

Table continued on next page

### 3. Results

479,060,215	479,063,762	2-aminoethanethiol dioxygenase	SECCE6Rv1G0402060.1	HC	-3,909,910
479,588,154	479,588,861	2-methylene-furan-3-one reductase	SECCE6Rv1G0402070.1	HC	-3,383,392
479,820,187	479,823,759	Acetylornithine deacetylase or succinyl-diaminopimelate desuccinylase	SECCE6Rv1G0402080.1	HC	-3,149,926
479,829,847	479,835,519	AP-2 complex subunit mu	SECCE6Rv1G0402090.1	HC	-3,139,216
479,841,605	479,841,799	Retrotransposon protein, putative, unclassified	SECCE6Rv1G0402100.1	LC	-3,130,197
480,093,208	480,094,488	U-box domain-containing protein	SECCE6Rv1G0402110.1	HC	-2,878,051
480,246,813	480,250,208	Protein phosphatase 2C	SECCE6Rv1G0402120.1	HC	-2,723,388
480,624,570	480,627,362	Protein NRT1/ PTR FAMILY 5.1	SECCE6Rv1G0402130.1	HC	-2,345,933
480,790,619	480,790,951	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0402140.1	LC	-2,181,114
480,790,955	480,791,491	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0402150.1	LC	-2,180,676
480,797,146	480,798,337	Glutathione S-transferase T3	SECCE6Rv1G0402160.1	LC	-2,174,158
480,866,529	480,866,795	extra-large GTP-binding protein 3	SECCE6Rv1G0402170.1	HC	-2,105,237
480,941,273	480,941,830	50S ribosomal protein L18	SECCE6Rv1G0402180.1	HC	-2,030,348
481,241,513	481,242,358	Transposon Ty1-LR4 Gag-Pol polyprotein	SECCE6Rv1G0402190.1	HC	-1,729,964
481,495,001	481,495,489	glucan synthase-like 9	SECCE6Rv1G0402200.1	HC	-1,476,654
481,499,932	481,501,810	Purine permease family protein	SECCE6Rv1G0402210.1	HC	-1,471,028
481,560,953	481,561,567	Retrotransposon protein, putative, unclassified	SECCE6Rv1G0402220.1	LC	-1,410,639
481,584,257	481,584,814	Invertase/pectin methylesterase inhibitor family protein	SECCE6Rv1G0402230.1	HC	-1,387,364
482,000,532	482,001,629	Ring finger protein, putative	SECCE6Rv1G0402240.1	HC	-970,818
482,011,768	482,014,591	Alpha-1, 3-glucosyltransferase	SECCE6Rv1G0402250.1	HC	-958,720
482,204,868	482,205,194	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0402260.1	LC	-766,868
482,458,859	482,460,173	SBP (S-ribonuclease-binding protein) family protein	SECCE6Rv1G0402270.1	HC	-512,383
482,540,823	482,541,029	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0402280.1	LC	-430,973
482,563,709	482,564,171	senescence-associated family protein (DUF581)	SECCE6Rv1G0402290.1	HC	-407,959
482,871,713	482,872,204	senescence-associated family protein (DUF581)	SECCE6Rv1G0402300.1	HC	-99,940
482,971,785	482,974,154	Pentatricopeptide repeat-containing protein	SECCE6Rv1G0402310.1	HC	1,070
483,144,937	483,145,474	senescence-associated family protein (DUF581)	SECCE6Rv1G0402320.1	HC	173,306
483,254,852	483,255,211	FORMS APLOID AND BINUCLEATE CELLS 1A	SECCE6Rv1G0402330.1	LC	283,132
483,256,339	483,260,258	UPF0160 protein MYG1, mitochondrial	SECCE6Rv1G0402340.1	HC	286,400
483,260,715	483,261,532	F-box-like protein	SECCE6Rv1G0402350.1	HC	289,224
483,262,227	483,264,688	Glycosyltransferase	SECCE6Rv1G0402360.1	HC	291,558
483,270,432	483,270,941	Ring finger protein	SECCE6Rv1G0402370.1	HC	298,788
483,575,459	483,579,832	Calcium-dependent protein kinase	SECCE6Rv1G0402380.1	HC	605,746
483,669,300	483,669,539	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0402390.1	LC	697,520
483,714,109	483,714,390	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0402400.1	LC	742,350

Table continued on next page

### 3. Results

483,738,162	483,740,889	Homeodomain-like superfamily protein	SECCE6Rv1G0402410.1	HC	767,626
483,742,537	483,746,702	Regulator of chromosome condensation (RCC1) family with FYVE zinc finger domain-containing protein	SECCE6Rv1G0402420.1	HC	772,720
483,990,538	483,990,798	B3 domain-containing protein family	SECCE6Rv1G0402430.1	LC	1,018,769
484,047,277	484,048,602	Hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase	SECCE6Rv1G0402440.1	HC	1,076,040
484,257,939	484,258,283	RING/U-box superfamily protein	SECCE6Rv1G0402450.1	LC	1,286,212
484,259,885	484,264,046	DHHA1 domain protein	SECCE6Rv1G0402460.1	LC	1,290,066
484,298,708	484,299,019	Histone H4	SECCE6Rv1G0402470.1	HC	1,326,964
484,469,277	484,471,847	Lipid transfer protein-like	SECCE6Rv1G0402480.1	HC	1,498,663
484,473,905	484,475,119	Retrotransposon protein, putative, Ty1-copia subclass	SECCE6Rv1G0402490.1	LC	1,502,613
484,475,402	484,475,620	Retrotransposon protein, putative, Ty1-copia subclass	SECCE6Rv1G0402500.1	LC	1,503,612
484,515,658	484,516,092	Dimeric alpha-amylase inhibitor	SECCE6Rv1G0402510.1	HC	1,543,976
484,516,453	484,519,099	Eukaryotic translation initiation factor 3 subunit A	SECCE6Rv1G0402520.1	LC	1,545,877
484,522,362	484,523,252	RNA-directed DNA polymerase (reverse transcriptase)-related family protein	SECCE6Rv1G0402530.1	HC	1,550,908
484,700,378	484,700,999	Oxidative stress 3, putative isoform 1	SECCE6Rv1G0402540.1	HC	1,728,790
484,702,221	484,707,064	Nucleoporin protein Ndc1-Nup	SECCE6Rv1G0402550.1	HC	1,732,744
484,985,068	484,987,758	p-loop containing nucleoside triphosphate hydrolases superfamily protein, putative	SECCE6Rv1G0402560.1	LC	2,014,514
484,987,940	484,988,944	p-loop containing nucleoside triphosphate hydrolases superfamily protein, putative	SECCE6Rv1G0402570.1	LC	2,016,543
484,989,463	484,990,821	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	SECCE6Rv1G0402580.1	HC	2,018,243
484,991,201	484,991,890	Protein SLOW GREEN 1, chloroplastic	SECCE6Rv1G0402590.1	HC	2,019,646
484,993,828	484,996,547	Tobamovirus multiplication-like protein	SECCE6Rv1G0402600.1	HC	2,023,288
485,003,027	485,003,663	Heavy metal transport/detoxification superfamily protein	SECCE6Rv1G0402610.1	HC	2,031,446
485,748,292	485,749,533	B3 domain-containing protein	SECCE6Rv1G0402620.1	HC	2,777,014
486,220,712	486,223,818	Mitochondrial import inner membrane translocase subunit	SECCE6Rv1G0402630.1	HC	3,250,366
486,226,294	486,227,364	WD-repeat 1	SECCE6Rv1G0402640.1	HC	3,254,930
486,234,697	486,235,134	RING finger protein	SECCE6Rv1G0402650.1	HC	3,263,016
486,446,719	486,453,705	MADS box transcription factor	SECCE6Rv1G0402660.1	HC	3,478,313
486,744,750	486,748,821	Receptor-like protein kinase	SECCE6Rv1G0402670.1	HC	3,774,886
486,971,781	486,972,404	RING finger protein	SECCE6Rv1G0402680.1	HC	4,000,194
487,212,322	487,214,288	Beta-1, 3-N-acetylglucosaminyltransferase lunatic fringe	SECCE6Rv1G0402690.1	HC	4,241,406
487,370,448	487,374,602	Myb family transcription factor-like	SECCE6Rv1G0402700.1	LC	4,400,626
487,581,767	487,583,747	2-C-methyl-D-erythritol 2, 4-cyclodiphosphate synthase	SECCE6Rv1G0402710.1	HC	4,610,858
487,585,817	487,587,919	Pentatricopeptide repeat-containing protein	SECCE6Rv1G0402720.1	HC	4,614,969
487,589,492	487,589,845	Membrane protein of er body-like protein	SECCE6Rv1G0402730.1	HC	4,617,770
487,592,688	487,593,480	Calcium-dependent protein kinase	SECCE6Rv1G0402740.1	LC	4,621,185
487,659,911	487,666,734	Protease HtpX	SECCE6Rv1G0402750.1	HC	4,691,424

Table continued on next page

### 3. Results

488,085,743	488,086,645	transcription repressor	SECCE6Rv1G0402760.1	HC	5,114,295
488,508,584	488,508,888	Serine/threonine protein phosphatase 7 long form isogeny	SECCE6Rv1G0402770.1	LC	5,536,837
488,512,242	488,512,700	Tetratricopeptide repeat-containing family protein	SECCE6Rv1G0402780.1	LC	5,540,572
488,684,951	488,685,661	Transposon Ty3-I Gag-Pol polyprotein	SECCE6Rv1G0402790.1	LC	5,713,407
488,736,885	488,743,186	Rho GTPase-activating protein	SECCE6Rv1G0402800.1	HC	5,768,136
489,246,964	489,248,460	Pentatricopeptide repeat-containing protein	SECCE6Rv1G0402810.1	HC	6,275,813
489,595,437	489,596,287	Growth-regulating factor	SECCE6Rv1G0402820.1	HC	6,623,963
490,027,042	490,030,170	3, 5-epimerase/4-reductase	SECCE6Rv1G0402830.1	HC	7,056,707
490,034,059	490,036,125	Oxygen-dependent choline dehydrogenase	SECCE6Rv1G0402840.1	HC	7,063,193
490,086,599	490,089,682	SPX domain-containing family protein	SECCE6Rv1G0402850.1	HC	7,116,242
490,255,116	490,262,399	zinc finger WD40 repeat protein 1	SECCE6Rv1G0402860.1	HC	7,286,858
490,573,116	490,573,313	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0402870.1	LC	7,601,316
490,689,163	490,690,165	ATP-dependent Clp protease ATP-binding subunit	SECCE6Rv1G0402880.1	HC	7,717,765
490,915,947	490,916,603	Dehydration-responsive element-binding factor 1	SECCE6Rv1G0402890.1	HC	7,944,376
491,157,439	491,166,251	DWNN domain, A CCHC-type zinc finger protein	SECCE6Rv1G0402900.1	HC	8,189,946
491,421,548	491,422,771	U-box domain-containing family protein	SECCE6Rv1G0402910.1	HC	8,450,260
491,432,482	491,433,491	Glycerol kinase	SECCE6Rv1G0402920.1	LC	8,461,088
491,871,373	491,872,668	Myb/SANT-like DNA-binding domain protein	SECCE6Rv1G0402930.1	LC	8,900,122
491,873,665	491,874,062	Transposon protein, putative, CACTA, En/Spm sub-class	SECCE6Rv1G0402940.1	LC	8,901,964
491,903,174	491,903,959	Dehydration-responsive element binding factor	SECCE6Rv1G0402950.1	HC	8,931,668
492,842,943	492,846,054	RING/U-box superfamily protein	SECCE6Rv1G0402960.1	HC	9,872,600
493,593,460	493,595,181	Protein DETOXIFICATION	SECCE6Rv1G0402970.1	HC	10,622,422
493,860,534	493,869,242	Lysophospholipid acyltransferase	SECCE6Rv1G0402980.1	HC	10,892,989
494,427,300	494,430,751	Transcription initiation factor IIE subunit beta	SECCE6Rv1G0402990.1	HC	11,457,126
494,432,527	494,435,069	F-box protein PP2-A13	SECCE6Rv1G0403000.1	HC	11,461,899
495,051,531	495,054,767	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	SECCE6Rv1G0403010.1	HC	12,081,250
495,065,777	495,066,289	Nucleotide/sugar transporter family protein	SECCE6Rv1G0403020.1	LC	12,094,134
495,084,532	495,087,954	Cytochrome P450 family protein	SECCE6Rv1G0403030.1	HC	12,114,344
495,328,227	495,334,271	Homeobox-leucine zipper protein	SECCE6Rv1G0403040.1	HC	12,359,350
495,677,292	495,681,070	E3 ubiquitin-protein ligase	SECCE6Rv1G0403050.1	HC	12,707,282
495,792,674	495,792,961	Cold regulated protein 27	SECCE6Rv1G0403060.1	HC	12,820,918
495,934,848	495,936,945	Dof zinc finger protein	SECCE6Rv1G0403070.1	HC	12,963,998
495,944,743	495,944,955	Transducin/WD40 repeat-like superfamily protein	SECCE6Rv1G0403080.1	LC	12,972,950
496,304,991	496,309,685	BHLH transcription factor-like protein	SECCE6Rv1G0403090.1	HC	13,335,439
496,388,857	496,389,120	Transposase	SECCE6Rv1G0403100.1	LC	13,417,090
496,621,226	496,622,925	Aluminum-activated malate transporter-like	SECCE6Rv1G0403110.1	HC	13,650,176
496,715,261	496,715,605	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0403120.1	LC	13,743,534

Table continued on next page

### 3. Results

6R	497,268,016	497,271,391	Kinase, putative	SECCE6Rv1G0403130.1	HC	14,297,804	
	497,273,430	497,274,908	DNA-directed RNA polymerase subunit	SECCE6Rv1G0403140.1	HC	14,302,270	
	497,277,461	497,282,513	N6-adenosine-methyltransferase MT-A70-like protein	SECCE6Rv1G0403150.1	HC	14,308,088	
	497,492,238	497,496,888	THUMP domain-containing protein 1	SECCE6Rv1G0403160.1	HC	14,522,664	
	497,498,362	497,502,884	MYB transcription factor-like	SECCE6Rv1G0403170.1	HC	14,528,724	
	497,682,287	497,682,703	Beta-galactosidase BgaP	SECCE6Rv1G0403180.1	LC	14,710,596	
	498,033,574	498,040,653	Argonaute	SECCE6Rv1G0403190.1	HC	15,065,214	
	498,207,607	498,214,580	Zinc finger protein, putative	SECCE6Rv1G0403200.1	HC	15,239,194	
	498,922,091	498,922,510	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE6Rv1G0403210.1	LC	15,950,402	
	523,758,873	523,760,348	Pentatricopeptide repeat-containing protein	SECCE6Rv1G0404780.1	HC	-7,618	
	523,764,183	523,767,314	kinetochore protein	SECCE6Rv1G0404790.1	HC	-1,480	
	523,769,469	523,769,885	VQ motif family protein	SECCE6Rv1G0404800.1	HC	2,449	
7R	280,453,849	280,467,090	calmodulin 1	SECCE7Rv1G0481110.1	HC	-211,904	
	280,467,698	280,468,406	Glycine-rich protein	SECCE7Rv1G0481120.1	HC	-204,322	
	280,496,379	280,496,627	hAT transposon superfamily protein	SECCE7Rv1G0481130.1	LC	-175,871	
	280,672,114	280,674,397	Chaperone protein DnaJ	SECCE7Rv1G0481140.1	HC	882	
	543,611,496	543,615,169	Sugar phosphate/phosphate translocator	SECCE7Rv1G0492090.1	HC	1,426	
	543,616,306	543,619,383	Ribosomal RNA small subunit methyltransferase F	SECCE7Rv1G0492100.1	HC	5,938	
	767,243,012	767,246,117	Inter-alpha-trypsin inhibitor heavy chain H3, putative	SECCE7Rv1G0508010.1	HC	-1,122	
	767,519,425	767,519,976	Retrovirus-related Pol polyprotein from transposon TNT 1-94	SECCE7Rv1G0508020.1	LC	274,014	
	767,909,181	767,910,231	Xyloglucan endotransglucosylase/hydrolase	SECCE7Rv1G0508030.1	HC	664,020	
	767,916,469	767,917,316	Xyloglucan endotransglucosylase/hydrolase	SECCE7Rv1G0508040.1	HC	671,206	
Chr	start [bp]	end [bp]	Gene Description	Gene ID	CC	Distance to SNP [bp]	SNP
4R	877,048,012	877,049,500	Glycosyltransferase	SECCE4Rv1G0291590.1	HC	-988,897	
	877,081,348	877,082,037	Germin-like protein	SECCE4Rv1G0291600.1	HC	-955,960	
	877,087,417	877,090,948	Subtilisin-like protease	SECCE4Rv1G0291610.1	HC	-948,470	
	877,111,318	877,114,280	Serine carboxypeptidase family protein, expressed	SECCE4Rv1G0291620.1	HC	-924,854	
	877,239,897	877,242,116	Subtilisin-like protease	SECCE4Rv1G0291630.1	HC	-796,646	
	877,246,682	877,248,692	F-box family protein	SECCE4Rv1G0291640.1	HC	-789,966	
	877,250,912	877,254,635	Protein kinase	SECCE4Rv1G0291650.1	HC	-784,880	
	877,257,202	877,260,235	Protein kinase family protein	SECCE4Rv1G0291660.1	HC	-778,934	
	877,261,497	877,265,207	Mannosyltransferase	SECCE4Rv1G0291670.1	HC	-774,301	
	877,544,523	877,545,974	F-box family protein	SECCE4Rv1G0291680.1	HC	-492,404	
	877,633,800	877,635,335	Pentatricopeptide repeat-containing protein	SECCE4Rv1G0291690.1	HC	-403,086	
	877,643,767	877,656,238	Pre-mRNA-processing protein 40A	SECCE4Rv1G0291700.1	HC	-387,650	
	878,023,390	878,038,264	chromatin remodeling factor	SECCE4Rv1G0291710.1	HC	-6,826	
7R	38,860,425	38,865,706	Pumilio	SECCE7Rv1G0461650.1	HC	-1,534	25651

## 4. Discussion

---

# 4. Discussion

## 4.1. Understanding the Population Structure of *Secale*

The population structure of rye, as revealed through various studies employing molecular markers and population genetic analyses, presents a complex picture of genetic relationships within the *Secale* genus. Using different methods, researchers have gained insights into the evolutionary history, genetic diversity, and relationships among different rye taxa.

Early investigations by Vences et al. (1987) laid the groundwork for comprehending the population structure of *Secale*. By analyzing isozyme patterns as well as allelic and genotypic frequencies of isozyme loci in 23 samples of *Secale*, Vences et al. were able to distinguish *S. sylvestre* from *S. vavilovii*, and both from *S. cereale* and *S. montanum* (*S. cereale* subsp. *vavilovii*). This allowed them to confirm one of the prevailing hypotheses about the existence of 4 species in the genus *Secale* by Khush and Stebbins (1961), although he mentioned that there was no clear differentiation between *S. cereale* and *S. montanum* (*S. cereale* subsp. *vavilovii*) recognizable.

Many other studies based on molecular markers have found similar patterns. A clear separation between *S. sylvestre* and other *Secale* species was noted, with a further separation between perennial (*S. strictum*) and annual (*S. cereale*) species (for example Bolibok-Bragoszewska et al. 2014, Chikmawati et al. 2005, Shang et al. 2006), with *S. sylvestre* as the most ancient *Secale* species and *S. cereale* evolved most recently (Chikmawati et al. 2005). In general, wild, and cultivated rye showed a high similarity with *S. cereale* subsp. *vavilovii* (Bolibok-Bragoszewska and Rakoczy-Trojanowska 2015, Shang et al. 2006).

In a study by Chikmawati et al. in 2012 over 100 *Secale cereale* ssp. Accessions were analyzed regarding genetic relationships through Principal coordinate (PCO) analysis and neighbor joining. They confirmed, that wild *S. cereale* accessions were separated from cultivated accessions, which in turn could be confirmed by a genome-wide study using over 1000 Diversity Array Technology (DArT) markers to characterize genetic diversity and population structure in a collection of 379 rye accessions including wild species, landraces, cultivated materials, historical and contemporary rye varieties (Bolibok-Bragoszewska et al. 2014).

Several workers have sought to understand the geographic structure of rye genetic variation, and although researchers have noted grouping by geographic region (Bolibok-Bragoszewska et al. 2014, Daskalova and Spetsov 2020, Hagenblad et al. 2016, Maraci et al. 2018, Parat et al. 2016, Schreiber et al. 2019), previous as well as latest findings in separate studies with over 100 accessions from all around the world speak against it (Chikmawati et al. 2012, Isik et al. 2007, Monteiro et al. 2016, Schreiber et al. 2022, Sidhu et al. 2019, Targonska-Karasek et al. 2020). For this reason, the geographical origin is not considered in this analysis of population structure.

Overall, the results from this experimental panel show a similar population structure to the findings described above. In the PCA, a separation of the three *Secale* species could be identified, with no clear separation of the individual subgroups of *S. cereale*. Schreiber et al. (2019) and Sun et al. (2022b) came

#### 4. Discussion

---

to the same result both using PCA with over 100 rye accessions and 55,744 SNPs or 908,599 synonymous SNPs. Furthermore, Sun et al. (2022b) upon further analysis, identify the presence of three subclusters in the *S. cereale* group. One consisting primarily of cultivated rye, while the other two consist of cultivated and weedy rye, and *S. cereale* subsp. *vavilovii*. It was also found that most *S. cereale* subsp. *vavilovii* accessions and weedy rye accessions were grouped into two of the three subclusters, suggesting that *S. cereale* subsp. *vavilovii* accessions share a much closer relationship with weedy than with cultivated rye (Sun et al. 2022b). When looking at the plot of EV 1 & 2, as well as EV 3 & 4, you can recognize indications of this structure for the experimental panel. Further on, as in this study, Schreiber et al. (2019) recognized a subcluster formed exclusively by *S. cereale* subsp. *vavilovii*.

However, when looking at the genus as a whole, this separation cannot be observed between weedy and cultivated forms of *Secale cereale* based on different molecular analyses (Al-Beyroutiová et al. 2016, Hawliczek et al. 2023, Schreiber et al. 2019, Skuza et al. 2019). Strong gene flow and a late origin of rye as a secondary crop have been suggested as causes for the weak genetic differentiation between these two groups (Hawliczek et al. 2023, Schreiber et al. 2019).

The pattern of the PCA can be transferred to the ancestry coefficient analysis. Thus, the genomic composition of *S. sylvestre* is primarily characterized by the same ancestral population. Indications that a few genotypes of *S. strictum* are characterized by the same ancestral population can already be found in the plot of the PCA of EV 1 & EV 2, where a few *S. strictum* members were also plotted in the *S. sylvestre* cluster. It is also noticeable that as soon as more than 5 ancestral populations are considered, *S. cereale* subsp. *vavilovii* is split into two groups and the genomic composition of one group follows the trend of cluster formation and relationships within the genus *Secale* as described above, while the other group is significantly characterized by one and the same ancestral population.

A possible explanation is delivered by Zohary et al. (2012). Here the genus *Secale* is divided into four species - *S. cereale*, *S. strictum*, *S. iranicum*, and *S. sylvestre*. The authors note that "True" *S. vavilovii* forms belong to *S. cereale* subsp. *vavilovii*, which is supported by extensive molecular data mentioned earlier. *S. iranicum* (Kolbylansky) is poorly known and at one time was erroneously described as *S. vavilovii* and sent to several germplasm collections under this description, causing confusion, with some researchers working on *S. cereale* subsp. *vavilovii* and others on *S. iranicum*, erroneously described only as *S. cereale* subsp. *vavilovii* (Hawliczek et al. 2023). The genetic proximity of these two groups could be confirmed by the analysis with K = 4 ancestral populations.

However, the hypothesis is not considered in the current studies and therefore no further evidence can be found. Recalculations separating the two structures of *S. cereale* subsp. *vavilovii* could provide further evidence for or against the hypothesis of Zohary et al. (2012).

The same pattern as in the PCA is found in the results of the fixation indices ( $F_{ST}$ ). This does also not contradict previous findings and the current literature. Schreiber et al. (2019), on whose test panel the

#### 4. Discussion

---

majority of the plants used here are based, confirm the clear separation between *Secale* species and the weak differentiation between *S. cereale* subsp. *vavilovii* and the other *S. cereale* subtaxa. Nevertheless, it is noticeable that the differentiation between the individual taxa is generally higher in the study of Schreiber et al. (2019). For example, an  $F_{ST}$  value of 0.50 between *S. cereale* subsp. *cereale* and *S. sylvestre* in their study contrasts with an  $F_{ST}$  value of 0.30 calculated here. This can also be observed when looking at *S. cereale* subsp. *vavilovii* and *S. sylvestre* (0.43 in Schreiber et al. (2019), 0.3 in this study). One possible explanation for this is the use of a larger test panel with additional representatives of the respective species, so that greater diversity can be observed for the individual species. According to Wright (1978),  $F_{ST}$  values of the populations between 0 - 0.05 will indicate no genetic differentiation among populations, if the  $F_{ST}$  value is between 0.05 - 0.15, it is moderately differentiated, if the  $F_{ST}$  value is between 0.15 - 0.25, then it is highly differentiated (see Cheng et al. 2020), so that Schreiber et al. (2019) statement is retained despite generally lower values in this study.

A closer look at the subgroups of *S. cereale* also confirms the pattern of population structure already observed in the PCA. Thus, a minimal genetic differentiation of domesticated rye to weedy rye and *S. cereale* subsp. *vavilovii* ( $F_{ST} > 0.05$ ) can be seen, while no differentiation between the latter two groups among each other, as well as to feral rye, can be recognised ( $F_{ST} < 0.05$ ). This is consistent with the results of Sun et al. (2022b), which noticed a weak genomic differentiation between weedy and the cultivated rye populations ( $F_{ST} = 0.013$ ). Also, Adamo et al. (2021) did not observe a clear structure on different *S. cereale* landraces from the Western Alps, with  $F_{ST}$  values ranging mostly between 0.00 - 0.15, indicating no clear differentiation between these on a smaller geographical scale.

Already Hawliczek et al. (2023) has mentioned the relatively narrow genomic diversity of *S. sylvestre*, but the very high diversity of *S. strictum* and wild *S. cereale* subsp. If you now compare this with the nucleotide diversity ( $\pi$ ) of the respective species in the experimental panel, these observations also apply here.

At a closer look, no significant difference can be found between some subgroups of *S. cereale*. This can be explained by the outbreeding nature of rye, resulting in high intraspecific diversity (Monteiro et al. 2016), and combined with continued gene flow between wild and cultivated plants (Hagenblad et al. 2016, Maraci et al. 2018, Skuza et al. 2019, Schreiber et al. 2019), this can lead to absence of reduction in genetic diversity (Monteiro et al. 2016, Parat et al. 2016).

Nevertheless, there is a trend of decline in nucleotide diversity towards more domesticated subspecies, which is in accordance with the assumption that domestication and improvement result in a decrease in diversity (Hawliczek et al. 2023). One reason could be, that positive selection can lead to reductions in nucleotide diversity and increased linkage disequilibrium within populations (Delmore and Liedvogel 2020, Yuan et al. 2017).

Overall, the study of the population structure confirmed the current scientific consensus. PCA plots, fixation indices of the individual species or subspecies, as well as the nucleotide diversity, follow the view of a three-species genus, according to the taxonomic classification of the genus *Secale* by Frederiksen and Petersen (1998). The evolutionary history could also be traced in parts, with

## 4. Discussion

---

*S. sylvestre* as the most ancient *Secale* species and *S. cereale* evolved most recently with a high degree of gene flow between wild and cultivated plants.

### 4.2. Inbreeding coefficient as a measurement of self-incompatibility

With the calculation of the inbreeding coefficient ( $F$ ), an easy-to-calculate variable should be created with which a diversity panel can be scanned for possible QTL for self-incompatibility (SI). Since such an approach has not yet been discussed in the literature, the identification of already known SI loci of rye should serve as verification, in addition to analyzing the gene annotations.

In contrast to the findings of Hagenblad et al. (2016), who primarily calculated averaged negative  $F$  values within and across different rye accessions, using a panel with 434 individuals from 76 accessions of wild, feral, and cultivated rye, all species and subgroups in this study showed positive  $F$  values on average. Their study also showed that the inbreeding coefficient in cultivated rye was lower than in feral and wild rye. This is also contrary to the findings of this test panel. On the other hand, SSR studies by Parat et al. (2016) in the same year, carried out with 620 individuals from 14 global rye populations (*S. cereale* subsp.), also showed positive inbreeding coefficients.

And although small to negative  $F$  values are to be expected due to the outcrossing nature of rye (Collevatti et al. 2010, Hagenblad et al. 2016), a possible cause for the positive  $F$  values observed in Parat et al. (2016) and this study include strong population substructure (Monteiro et al. 2016). Population structure will therefore lead to deviations from the expected Hardy-Weinberg (HW) proportions. On the one hand, it can increase the observed homozygosity due to the Wahlund effect (Mixed populations are usually out of Hardy-Weinberg proportions immediately after the mixture; Felsenstein 2015), but on the other hand, it can also increase the observed heterozygosity due to recent admixture (Meisner and Albrechtsen 2019).

None of the studies included the species *S. sylvestre* in their analyses. Here it was shown that it has a particularly high inbreeding coefficient compared to other *Secale* species. Due to its property of self-compatibility, a high homozygosity (Stratton 2008) and thus a high inbreeding coefficient is to be expected. The inbreeding coefficient of the subgroup "feral:brittle" is strikingly high, which based on previous findings should be in the range of the subgroups of *S. cereale*. The small sample size of 19 individuals, coupled with the fact that they all originate from the same country from the same collecting trip, leads to the assumption that the small population size or a sampling bias led to its high inbreeding coefficient.

Negative inbreeding coefficients for single individuals can be observed especially in the species *S. cereale*. As Moussouni et al. (2017) already suspected, this could be the case due to active farmer selection for heterozygous individuals. This is reinforced by the fact that these individuals can primarily be assigned to the domesticated subgroups. As different ploidy levels were detected in the GBS data, some of the genotypes with strong negative inbreeding coefficients are most likely tetraploid and therefore exhibit increased heterozygosity (Dreißig 2024, pers. commun.).

#### 4. Discussion

---

The high intraspecific diversity (Monteiro et al. 2016) forms most likely the high spread of  $F$  values in the individual species. Also, strong gene flow and a late origin of rye as a secondary crop can be cause for the weak genetic differentiation (Hawliczek et al. 2023, Schreiber et al. 2019), especially in *S. cereale*.

After the GWAS had been carried out, the detected QTLs were compared with already known SI loci. Three different SI loci (S- & Z-loci) on chromosomes 1R and 2R, as well as the SC locus (T-locus) S5 on chromosome 5R are relate to the SI of rye (Egorova et al. 2000, Melonek et al. 2021, Voylokov et al. 1998). In addition, further dominant self-fertility genes were found on chromosomes 1R, 4R, 5R and 6R (Melz and Winkel 1987, Melz et al. 1990, Voylokov et al. 1998). Through the pattern of segregation distortion observed for 28 out of 30 genetic markers, the S-locus has already been mapped for the rye reference genome "Lo7" to an interval of 3.2 Mbp delimited by the genes SECCE1Rv1G0014520 and SECCE1Rv1G0014770 on chromosome 1R (Melonek et al. 2021, see appendix A1). The interval covers the range from 115.1 to 118.3 Mb. None of the three SNPs on chromosome 1R identified in the GWAS is in close proximity to the S-locus (SNP\_116: 20.5 Mb, SNP\_3467: 665.0 Mb, SNP\_3568: 678.2 Mb). The same applies to the Z-locus on chromosome 2R. Three SNPs were also identified here, which were mapped to the positions 23.7 Mb (SNP\_4224), 75.0 Mb (SNP\_4667) and 840.4 Mb (SNP\_8211). Conversely, the Z-locus lies in a range of 878.2 - 879.4 Mb, covers a range of 1.2 Mb and could be anchored to the genomic region by the two flanking markers TC89057 and TC101821 (Melonek et al. 2021).

It is unlikely that the QTLs on chromosome 5R could be the T-locus, as the T-locus is located in the centromeric region (Egorova et al. 2000, Voylokov et al. 1998), which is located in the region of <round 180 Mb (Martis et al. 2013). The QTLs, on the other hand, are found at the positions 663.4 Mb (SNP\_19221) and 819,9 Mb (SNP\_21013).

As S- & Z-locus in particular are considered to be highly relevant (e.g. Gertz and Wricke 1989, Hackauf and Wehling 2005, Melonek et al. 2021, Voylokov et al. 1998), the results here suggest that the inbreeding coefficient cannot be used exclusively to identify SI loci. Since perenniability can also lead to higher levels of inbreeding (Duminil et al. 2009), and thus to a higher inbreeding coefficient, a further GWAS was performed excluding the perennial species *S. strictum*. Only two QTLs were identified on chromosomes 4R at position 878.0 Mb and 7R at position 38.9 Mb, which are unrelated to the previously identified QTLs. They are also not assigned to any relevant SI loci found in the literature.

Nevertheless, the results of the first GWAS in combination with the second one raise the question of whether the inbreeding coefficient as a trait could identify loci for perenniability. According to Gruner and Miedaner (2021), who proposed the first mapping study for perenniability in rye, perenniability is a complex trait with several QTLs involved. In their study, three of five QTLs were consistently found by several methods and were located on chromosome 4R (QTL-P4, position: 73,3 Mb), chromosome 5R (QTL-P5, position: 77 Mb) and chromosome 7R (QTL-P7 position: 0.6 Mb ), while the other QTLs can be

## 4. Discussion

---

found on chromosome 2R (QTL-P2, position: 121 Mb) and 3R (QTL-P3, position: 71.1 Mb). Again, none of the 18 QTLs in the first GWAS calculation is in close proximity to one of the perenniability QTLs.

Gruner and Miedaner (2021) also list several other QTLs that have been identified in other cereals and some of whose QTLs have already been connected across species. They further state, that through the highly syntenic character of rye to some of those species they are promising that these could also be mapped in rye and beyond could be relevant for perenniability.

So far, there is little evidence as to whether many of the QTLs identified in this study are responsible for SI or perenniability. However, the literature and the performance of GWAS without *S. strictum* suggest that the inbreeding coefficient cannot be readily used to detect QTLs associated with SI. A final clarification of this issue will be provided by gene annotations.

### 4.3. Discussion of candidate genes for self-incompatibility

To identify potential candidate genes, the LD decay was first calculated for the entirety of the SNPs. A drop of the pair-wise correlation ( $r^2$ ) of SNPs below the threshold of  $r^2 = 0.1$  was observed within the first 10 kb. Since rye is an outcrossing species with low ancestral LD, a low level of LD with a rapid decay is expected (Auinger et al. 2016, Li et al. 2011) compared to self-pollinating crops such as barley and wheat. For example, in elite barley cultivars an LD decay between and within four gene loci extends to 212 kb (Caldwell et al. 2006), whereas for 150 wheat varieties LD fell below  $r^2$  threshold (95%  $r^2$  value) at approximately 9 Mb (He et al. 2023).

Studying rye with 96 *S. cereale* genotypes and 10,244 SNP markers, estimations of global LD yielded an average LD value of 4.3 kb for the whole genome, by using a threshold of  $r^2 = 0.2$  Båga et al. (2022). Other authors have also observed a rapid drop in LD. By analyzing over 1,400 genotypes of a commercial hybrid rye breeding program, Auinger et al. (2016) noticed a rapid LD decay with 68 % of the marker pairs showing  $r^2 \leq 0.2$  within 1 cM. Siekmann et al. (2021) calculated the mean LD decay over the whole genome as around 2.5 cM for inbred lines of a commercial hybrid rye breeding program. In a publication by Li et al. (2011) the extend of LD for eleven candidate genes was approximately 520 bp using an  $r^2 = 0.16$ , by testing five open-pollinated winter rye breeding populations (*S. cereale*) from different Europe countries. In addition, they were able to detect LD blocks, which can be a few hundred base pairs long.

The LD in the QTL regions analyzed in this study decayed mostly in the first 1 Mb interval, which is why the globally calculated LD decay was assumed for the majority of SNPs, and thus meets the expectation from the literature. However, there are two exceptions on chromosome 6R where the LD is very pronounced and lies over an area of several Mb above the threshold (SNP\_22512, position: 328.4 Mb; SNP\_22876, position: 483.0 Mb). These two SNPs are close to the centromere, which can lead to a reduction in LD decay (Vos et al. 2017).

#### 4. Discussion

---

In recent years, many attempts have been made to identify genes associated with SI in rye, but also in other species. Due to the distinct SI mechanism within Poaceae, the focus has naturally been on identifying the candidate genes of the S- & Z-locus. A linkage between the isozyme glutamate oxaloacetate transaminase GOT/3 (Thorogood and Hayward 1991) and either the S-locus or the Z-locus has already been demonstrated. However, in 1993, Fuong et al. were also able to demonstrate a linkage between the Z-locus and beta-glucosidase and esterase 4/11 genes. In addition, there was evidence that the self-fertility locus S5 (T-locus) on chromosome 5R cosegregates with an esterase 5-7 complex (Est5-7 isozymes; Fuong et al. 1993), and it was speculated that their gene products, if functional, may be part of a signal transduction cascade within the pollen grain triggered by S- & Z-locus that causes pollen tube arrest (Do Canto et al. 2016, Wehling et al. 1995). A leaf peroxidase isozyme (PER) was also considered as SC-factor in *P. coerulescens*. The most recent findings refer to two genes encoding DUF247 domain proteins (sDUF247-I and sDUF247-II / zDUF247-I and zDUF247-II) and one gene encoding a short unstructured protein (sP/zP) as pollen and stigma components for SI in different Poaceae representants (Cropano et al. 2021, Herridge et al. 2022, Rohner et al. 2023). In 2016, Manzanares et al. 2016 showed that a DUF247 protein cosegregated with the S-locus in ryegrass (*Lolium perenne*), while this gene showed a frameshift mutation in self-compatible darnel *Lolium temulentum* and was therefore predicted to be non-functional. Corresponding gene counterparts were also found in a wild SI rice species (*Oryza longistaminata*). Thus, chromosome alignment with LpSDUF247 identified *O/SS1* and *O/SS2* as the SI stamen candidate genes and *O/SP* as the SI pistil candidate gene, and their respective stamen or pistil affiliation was verified by expression analyses (Lian et al. 2021).

As mentioned above, the GWAS did not detect any QTLs in close proximity neither to the T-locus nor to the S- or Z-locus. Of the previously documented genes and proteins associated with SI, no corresponding candidate gene could be annotated at the individual QTLs. The exception is a promising candidate gene of SNP\_3568 on chromosome 1R, which encodes the transmembrane protein DUF247 with high confidence. A second DUF247 gene is located only 60 kb upstream. According to the S- & Z-locus anchoring (Melonek et al. 2021), the distance between these two genes within the S- & Z-locus is also less than 100 kb. As the second DUF247 gene is not considered here due to the interval definition, this region offers the possibility for further investigation such as a sequence comparison between the different locations. The fact that only one DUF gene was found in this region and that no gene annotations could be made for certain SNPs (SNP\_116, SNP\_21013) may be due to the choice of interval considered for gene annotations, although the percentage of phenotypic variation explained by these two SNPs is below < 1 %. For future investigations, average  $r^2$  values for LD should be calculated at smaller intervals to allow for more precise determination to examine an area for gene annotations, if the marker coverage allows. Nevertheless, candidate genes for SI can also be found among the gene annotations for different SNPs across the chromosomes.

For example, at SNP\_4224 on chromosome 2R, a strictosidine synthase is present three times within a 1 Mb window upstream of the SNP. Strictosidine synthase enzymes are produced by higher plants and are a key enzyme in alkaloid biosynthesis pathway. They catalyze the metal independent condensation of tryptamine and secologanin to generate strictosidine (Hicks et al. 2011). In the early 2000s, Cigan et al. (2001) speculated that the anther-specific maize (*Zea mays*) fertility gene *Ms45*, which codes for a

#### 4. Discussion

---

strictosidine synthase, could be associated with formation of the pollen cell wall. An ortholog of this gene is involved in male fertility by encoding the strictosidine synthase-like enzyme in wheat (Awan 2023) and by producing mutations in the A, B and D homeologs of wheat with CRISPR-Cas9 led to pollen development abortion in triple homozygous mutants (Singh et al. 2018). Also, in Chinese cabbage (*Brassica campestris*) Wang et al. (2014a) showed, that Strictosidine synthase was one of three altered proteins in abundance in SI pistils after self-pollination.

The only candidate gene on chromosome 3R, a gene for an auxin-responsive protein, could also be associated with SI, as a study in Tomato (*Solanum lycopersicum*) showed in 2015 (Zhao et al.). The authors stated that a result of a SI reaction is a reduced number of auxin influx carrier (AUX1) resulting in many auxin-responsive genes showing altered expression after incompatible pollination.

SNP\_14380 reveals an engulfment and cell motility (ELMO) domain containing protein about 30 kb upstream on chromosome 4R. Its importance is particularly well known in mammals, where it stimulates a Rac-GEF, leading to Rac1 activation and cytoskeletal rearrangements (Yang et al. 2009) and functions as an upstream regulator of Rac1 GTPase (Gumienny et al. 2001). However, in plants, Rac-like GTPases have been shown to have profound effects on the actin cytoskeleton and regulation of pollen tube growth. Over-production led to abnormal pollen tube tip growth in *Arabidopsis* and tobacco (Yang et al. 2009).

Candidate genes associated with SI were also found on chromosome 7R. For example, a gene for calmodulin 1 is found approximately 200 kb upstream of SNP\_27124. Calmodulin (CaM) has been associated with various physiological and developmental processes in plants. It has also been detected in the extracellular matrix of transmitting tissue after pollination. There, it was found especially in the extracellular matrix surrounding the tips of growing pollen tubes, suggesting that sufficient apoplastic CaM is needed for pollen germination and tube growth (Jiang et al. 2014). Two candidate genes for Xyloglucan endotransglucosylase/hydrolase (XTH) can be found around 670 kb downstream of SNP\_28497 on the same chromosome. XTHs are a family of enzymes that are involved in cell wall biogenesis by mediating the construction and restructuring of xyloglucan cross-links, thereby controlling the extensibility or mechanical properties of the cell wall (Kurasawa et al. 2009, Wollenweber et al. 2021). By biochemical and immunocytological characterizations of *Arabidopsis thaliana* pollen tube cell wall, Dardelle et al. (2010), stated xyloglucan controls cell expansion in the pollen tube cell wall. A year earlier, Kurasawa et al. (2009) also concluded that XTH is involved in automatic self-pollination in *Arabidopsis thaliana*. Through reverse genetic analysis they were able to prove, that a loss-of-function mutation of an *Arabidopsis* XTH family gene (*AtXTH28*) led to a decrease in capability for self-pollination. They hypothesized that this decrease occurred probably due to inhibition of stamen filament growth.

And although the relevant SI loci for the Poaceae family cannot be found for this test panel by GWAS using the inbreeding coefficient ( $F$ ) as a trait, many other promising QTLs are mentioned here. Some of the orthologs of the candidate genes of this study have been previously associated with SI in the literature for different species. However, a connection with rye must first be found. The hypothesis put forward during the analyses that the inbreeding coefficient is not only a trait for SI, but is largely influenced by perenniability, could not be confirmed by the discovery of many QTLs associated with SI.

## 4. Discussion

---

Nevertheless, it is surprising that when a second GWAS was performed without perennial genotypes, hardly any QTLs were found whose gene annotation was not associated with SI. The minor allele frequency (maf) determined in the GWAS for these two QTLs shows very low values, with one of the two mafs being under 0.6 %. This raises the question of whether, due to the superior number of *S. cereale* subsp. *cereale* genotypes with similarly low inbreeding coefficients, certain alleles responsible for self-fertility, which should occur mainly in *S. sylvestre*, are not considered because their frequencies are too low.

### 4.3. Suggestions for future work

When analyzing the population structure, the occurrence of a subgroup of *S. cereale* subsp. *vavilovii* within *S. cereale* was observed both in the PCA and when looking at the ancestry coefficient. More detailed analyses of these genotypes could provide information on their separate grouping. As geographical influences have been repeatedly discussed in the literature, the geographical origin of these genotypes should be further investigated. However, this group could also be considered separately in the calculation of the fixation index ( $F_{ST}$ ).

Regarding the inbreeding coefficient, it is not entirely clear why the exclusion of the perennial species *S. strictum* from the GWAS led to the detection of fewer QTLs in which no SI components were found, whereas the inclusion of this species led to the opposite result. One hypothesis is that due to the small number of individuals of the self-compatible species *S. sylvestre* (24) compared to the > 1,000 *S. cereale* individuals in the GWAS, no significant associations with SI could be made. Increasing the number of individuals of both *S. strictum* and *S. sylvestre* might lead to clearer results. Nevertheless, the SI loci already known from the literature did not appear in the analyses. It is therefore necessary to critically analyze whether the inbreeding coefficient is strongly influenced by other factors and whether these can be eliminated.

Further analyses and practical approaches to investigate the candidate genes should be considered. In addition, a sequence alignment of the candidate gene for DUF247 with the predominant sequences of the DUFs known from the S- & Z-loci should be performed.

## 5. Summary

---

### 5. Summary

This study investigated self-incompatibility (SI) in the genus *Secale* using population genetic analyses. First, the population structure of a diversity panel consisting of more than 1,000 representatives of the taxa *S. cereale*, *S. strictum* and *S. sylvestre* from more than 50 countries was analyzed. Subsequently, quantitative trait loci (QTL) related to SI in terms of inbreeding coefficient ( $F$ ) were defined by performing genome-wide association studies (GWAS).

The current population structure was determined by principal component analysis (PCA) and the calculation of ancestry coefficients for a given number of ancestral populations (K). The latter was defined by calculating the cross-entropy criterion and the results of the PCA. Here, the three taxa *S. sylvestre*, *S. strictum* and *S. cereale* could be clearly distinguished from each other, confirming a classification according to one of the common classifications of Frederiksen and Petersen (1998) with the differentiation of these three species. The presence of a subgroup within *S. cereale*, which has also been reported in the literature, was also demonstrated here. This subgroup consists of some representatives of *S. cereale* subsp. *vavilovii*. The subsequent calculation of the fixation index ( $F_{ST}$ ) and nucleotide diversity ( $\pi$ ) also confirmed the classification described above, with the self-compatible species *S. sylvestre* as the most isolated and oldest species, the perennial species *S. strictum*, as well as representatives of wild and domesticated rye within *S. cereale*.

The inbreeding coefficient ( $F$ ) was then calculated to determine the homozygosity of each individual. As expected, this is particularly pronounced in the self-compatible species *S. sylvestre* and decreases from *S. strictum* to *S. cereale*. Positive median  $F$  values were calculated for all three species. The inbreeding coefficient was then used as a trait in a GWAS. For all species, 18 QTLs were found across all 7 rye chromosomes. Determination of the global linkage disequilibrium (LD) decay revealed a decay in LD within the first 10 kilobases (kb). This was used for gene annotation within the QTL regions, unless separate calculation of the LD decay for each single nucleotide polymorphism (SNP) allowed other interval limits. None of the identified QTLs could be linked to SI loci from the literature. However, the candidate genes included a gene for a transmembrane protein (DUF247), which is mainly mentioned in the literature as a candidate gene for the two highly relevant SI loci (S- & Z-loci). Other genes, such as a strictosidine synthase gene or an engulfment and cell motility (ELMO) domain containing protein, are also among the candidate genes. The suitability of the inbreeding coefficient for the determination of SI loci needs to be critically examined. Further investigation of the QTLs may provide information.

## 5. Summary

---

### 5.1 Zusammenfassung

In dieser Arbeit wurde die Selbstinkompatibilität (SI) im Genus *Secale* mittels populationsgenetischer Analysen untersucht. Dazu wurde zunächst die Populationsstruktur eines Diversitätspanels, bestehend aus über 1.000 Vertretern der Taxa *S. cereale*, *S. strictum* und *S. sylvestre* mit dem Ursprung in über 50 Ländern, analysiert. Anschließend wurden durch die Durchführung von genomweiten Assoziationsstudien (GWAS) quantitative trait loci (QTL) definiert, die im Zusammenhang mit der SI in Form des Inzuchtkoeffizienten ( $F$ ) stehen.

Die vorliegende Populationsstruktur wurde anhand einer Hauptkomponenten-Analyse (PCA), sowie der Berechnung eines Verwandtschaftskoeffizienten für eine bestimmte Anzahl an Ahnen-Populationen ( $K$ ) vorgenommen. Letzteres wurde durch die Berechnung des Kreuzentropie-Kriteriums und den Ergebnissen der PCA definiert. Hierbei konnten die drei Taxa *S. sylvestre*, *S. strictum* und *S. cereale* klar voneinander unterschieden werden, sodass eine Einordnung nach einer der gängigen Klassifizierungen nach Frederiksen and Petersen (1998), mit der Unterscheidung dieser drei Arten, bestätigt werden konnte. Das ebenfalls in der Literatur beobachtete Auftreten von einer Subgruppe innerhalb *S. cereale* konnte auch hier gezeigt werden. Bei dieser Subgruppe handelt es sich um einige Vertreter von *S. cereale* subsp. *vavilovii*. Die anschließende Berechnung des Fixierungsindex ( $F_{ST}$ ), der Nukleotiddiversität ( $\pi$ ) verifizierten ebenfalls die oben beschriebene Klassifizierung mit der selbstkompatiblen Art *S. sylvestre* als isolierteste und älteste Art, der mehrjährigen Art *S. strictum*, sowie Vertretern von wildem und domestiziertem Roggen innerhalb von *S. cereale*.

Darauf folgend wurde der Inzuchtkoeffizient ( $F$ ) berechnet, um die Homozygotie einzelner Individuen zu ermitteln. Wie zu erwarten war, ist diese in der selbstkompatiblen Art *S. sylvestre* besonders stark ausgeprägt, und nimmt von *S. strictum* zu *S. cereale* weiterhin ab. Für alle drei Arten sind im Median positive  $F$ -values berechnet worden. Der Inzuchtkoeffizient wurde im nächsten Schritt bei der Durchführung einer GWAS als Merkmal verwendet. Es konnten über alle 7 Roggenchromosomen für die Berücksichtigung aller Arten 18 QTLs gefunden werden. Die Bestimmung des Abfalls des globalen Kopplungsungleichgewichts (LD) ergab ein Abfall des LDs innerhalb der ersten 10 Kilobasen (kb). Dies wurde zur Genannotation innerhalb der QTL-Regionen herangezogen, sofern die gesonderte Berechnung des LD-Abfalls für jedes Einzelnukleotid-Polymorphismus (SNP) keine anderen Intervallgrenzen zuließ. Keines der ermittelten QTLs konnte mit SI-Loci aus der Literatur in Verbindung gebracht werden. Dennoch fanden sich unter den Kandidatengenen unter anderem ein Gen für ein Transmembranprotein (DUF247), dass vor allem als Kandidatengen der beiden hochrelevanten SI-Loci (S- & Z-Locus) in der Literatur Erwähnung findet. Auch andere Gene, wie etwa ein Gen für Strictosidinsynthase oder ein „engulfment and cell motility (ELMO) domain containing“ Protein finden sich unter den Kandidatengenen. Ob der Inzuchtkoeffizient zur Bestimmung von SI-Loci geeignet ist, ist kritisch zu hinterfragen. Weitere Untersuchungen zu den QTLs können Aufschluss darüber geben.

## 6. References

---

### 6. References

- Adamo M, Blandino M, Capo L, Ravetto Enri S, Fusconi A, Lonati M, Mucciarelli M (2021) A ddRADseq Survey of the Genetic Diversity of Rye (*Secale cereale* L.) Landraces from the Western Alps Reveals the Progressive Reduction of the Local Gene Pool. *Plants* (Basel, Switzerland) 10(11):2415. doi: 10.3390/plants10112415
- Al-Beyroutiová M, Sabo M, Slezák P, Dušinský R, Birčák E, Hauptvogel P, Kilian A, Švec M (2016) Evolutionary relationships in the genus *Secale* revealed by DArTseq DNA polymorphism. *Pl Syst Evol* 302(8):1083–1091. doi: 10.1007/s00606-016-1318-2
- Auinger H-J, Schönleben M, Lehermeier C, Schmidt M, Korzun V, Geiger HH, Piepho H-P, Gordillo A, Wilde P, Bauer E, Schön C-C (2016) Model training across multiple breeding cycles significantly improves genomic prediction accuracy in rye (*Secale cereale* L.). *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* 129(11):2043–2053. doi: 10.1007/s00122-016-2756-5
- Båga M, Bahrani H, Larsen J, Hackauf B, Graf RJ, Laroche A, Chibbar RN (2022) Association mapping of autumn-seeded rye (*Secale cereale* L.) reveals genetic linkages between genes controlling winter hardiness and plant development. *Sci Rep* 12(1):5793. doi: 10.1038/s41598-022-09582-2
- Barrett SCH (2013) The evolution of plant reproductive systems: how often are transitions irreversible? *Proceedings. Biological sciences* 280(1765):20130913. doi: 10.1098/rspb.2013.0913
- Barroso GV, Moutinho AF, Dutheil JY (2020) A Population Genomics Lexicon. In: Dutheil JY (ed) Statistical population genomics. Open Humana Press, New York, NY, pp 3–17
- Bauer E, Schmutzler T, Barilar I, Mascher M, Gundlach H, Martis MM, Twardziok SO, Hackauf B, Gordillo A, Wilde P, Schmidt M, Korzun V, Mayer KFX, Schmid K, Schön C-C, Scholz U (2017) Towards a whole-genome sequence for rye (*Secale cereale* L.). *The Plant journal : for cell and molecular biology* 89(5):853–869. doi: 10.1111/tpj.13436
- Bian X-Y, Friedrich A, Bai J-R, Baumann U, Hayman DL, Barker SJ, Langridge P (2004) High-resolution mapping of the S and Z loci of *Phalaris coerulescens*. *Genome* 47(5):918–930. doi: 10.1139/g04-017
- Bolibok-Brągoszewska H, Rakoczy-Trojanowska M (2015) Molecular Marker Based Assessment of Genetic Diversity in Rye. In: Ahuja MR, Jain SM (eds) Genetic Diversity and Erosion in Plants, vol 7. Springer International Publishing, Cham, pp 105–123
- Bolibok-Brągoszewska H, Targońska M, Bolibok L, Kilian A, Rakoczy-Trojanowska M (2014) Genome-wide characterization of genetic diversity and population structure in *Secale*. *BMC Plant Biol* 14:184. doi: 10.1186/1471-2229-14-184
- Brenner S (2013) Adaptor Hypothesis. In: Brenner's Encyclopedia of Genetics. Elsevier, p 18
- Brzozowski LJ, Szuleta E, Phillips TD, van Sanford DA, Clark AJ (2023) Breeding cereal rye (*Secale cereale*) for quality traits. *Crop Science* 63(4):1964–1987. doi: 10.1002/csc2.21022
- Bustos A de, Jouve N (2002) Phylogenetic relationships of the genus *Secale* based on the characterisation of rDNA ITS sequences. *Plant Syst Evol* 235(1):147–154. doi: 10.1007/s00606-002-0215-z
- Calabrese B (2019) Linkage Disequilibrium. In: Encyclopedia of Bioinformatics and Computational Biology. Elsevier, pp 763–765
- Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics* 172(1):557–567. doi: 10.1534/genetics.104.038489
- Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. *Lancet* (London, England) 361(9357):598–604. doi: 10.1016/S0140-6736(03)12520-2
- Carothers AD, Rudan I, Kolcic I, Polasek O, Hayward C, Wright AF, Campbell H, Teague P, Hastie ND, Weber JL (2006) Estimating human inbreeding coefficients: comparison of genealogical and marker heterozygosity approaches. *Annals of human genetics* 70(Pt 5):666–676. doi: 10.1111/j.1469-1809.2006.00263.x

## 6. References

---

- Carvalho-Madrigal S, Sanín MJ (2024) The role of introgressive hybridization in shaping the geographically isolated gene pools of wax palm populations (genus *Ceroxylon*). *Molecular phylogenetics and evolution* 193:108013. doi: 10.1016/j.ympev.2024.108013
- Caye K, Jay F, Michel O, François O (2018) Fast inference of individual admixture coefficients using geographic data. *Ann. Appl. Stat.* 12(1). doi: 10.1214/17-AOAS1106
- Charlesworth B (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature reviews. Genetics* 10(3):195–205. doi: 10.1038/nrg2526
- Charlesworth B (2015) What Use Is Population Genetics? *Genetics* 200(3):667–669. doi: 10.1534/genetics.115.178426
- Cheng J, Kao H, Dong S (2020) Population genetic structure and gene flow of rare and endangered *Tetraena mongolica* Maxim. revealed by reduced representation sequencing. *BMC plant biology* 20(1):391. doi: 10.1186/s12870-020-02594-y
- Chikmawati T, Miftahudin M, Skovmand B, Gustafson JP (2012) Amplified fragment length polymorphism-based genetic diversity among cultivated and weedy rye (*Secale cereale* L.) accessions. *Genet Resour Crop Evol* 59(8):1743–1752. doi: 10.1007/s10722-012-9796-8
- Chikmawati T, Skovmand B, Gustafson JP (2005) Phylogenetic relationships among *Secale* species revealed by amplified fragment length polymorphisms. *Genome* 48(5):792–801. doi: 10.1139/g05-043
- Choi Y, Wijsman EM, Weir BS (2009) Case-control association testing in the presence of unknown relationships. *Genetic Epidemiology* 33(8):668–678. doi: 10.1002/gepi.20418
- Cigan AM, Unger E, Xu R, Kendall T, Fox TW (2001) Phenotypic complementation of ms45 maize requires tapetal expression of MS45. *Sexual Plant Reproduction* 14(3):135–142. doi: 10.1007/s004970100099
- Collevatti RG, Lima JS, Soares TN, Telles MPdC (2010) Spatial Genetic Structure and Life History Traits in Cerrado Tree Species: Inferences for Conservation. *Nat. Conserv.* 08(01):54–59. doi: 10.4322/natcon.00801008
- Cropano C, Place I, Manzanares C, Do Canto J, Lübbertedt T, Studer B, Thorogood D (2021) Characterization and practical use of self-compatibility in outcrossing grass species. *Annals of botany* 127(7):841–852. doi: 10.1093/aob/mcab043
- Crow J (1987) Population Genetics History: A Personal View. *Annual Review of Genetics* 21(1):1–22. doi: 10.1146/annurev.genet.21.1.1
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R (2011) The variant call format and VCFtools. *Bioinformatics* (Oxford, England) 27(15):2156–2158. doi: 10.1093/bioinformatics/btr330
- Dardelle F, Lehner A, Ramdani Y, Bardor M, Lerouge P, Driouich A, Mollet J-C (2010) Biochemical and immunocytological characterizations of *Arabidopsis* pollen tube cell wall. *Plant physiology* 153(4):1563–1576. doi: 10.1104/pp.110.158881
- Daskalova N, Spetsov P (2020) Taxonomic Relationships and Genetic Variability of Wild *Secale* L. Species as a Source for Valued Traits in Rye, Wheat and Triticale Breeding. *Cytol. Genet.* 54(1):71–81. doi: 10.3103/S0095452720010041
- Delmore KE, Liedvogel M (2020) Avian Population Genomics Taking Off: Latest Findings and Future Prospects. In: Dutheil JY (ed) Statistical population genomics. Open Humana Press, New York, NY, pp 413–433
- Do Canto J, Studer B, Lubberstedt T (2016) Overcoming self-incompatibility in grasses: a pathway to hybrid breeding. TAG. Theoretical and applied genetics. *Theoretische und angewandte Genetik* 129(10):1815–1829. doi: 10.1007/s00122-016-2775-2
- Duminil J, Hardy OJ, Petit RJ (2009) Plant traits correlated with generation time directly affect inbreeding depression and mating system and indirectly genetic structure. *BMC evolutionary biology* 9:177. doi: 10.1186/1471-2148-9-177
- Egorova IA, Peneva TI, Baranova OA, Voylokov AV (2000) Analysis of Linkage Between Biochemical and Morphological Markers of Rye Chromosomes 1R, 2R, and 5R and Mutations of Self-Fertility

## 6. References

---

- at the Main Incompatibility Loci. *Russian Journal of Genetics* 36(12):1423–1430. doi: 10.1023/A:1009075408921
- Elhaik E (2022) Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific reports* 12(1):14683. doi: 10.1038/s41598-022-14395-4
- Engelhardt BE, Stephens M (2010) Analysis of population structure: a unifying framework and novel methods based on sparse factor analysis. *PLoS genetics* 6(9):e1001117. doi: 10.1371/journal.pgen.1001117
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164(4):1567–1587. doi: 10.1093/genetics/164.4.1567
- Feldman M, Levy AA (2023) Secale L. In: Feldman M, Levy AA (eds) *Wheat Evolution and Domestication*. Springer International Publishing, Cham, pp 159–195
- Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach
- Felsenstein J (2015) *Theoretical evolutionary genetics* Joseph Felsenstein
- Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. *Annual review of plant biology* 54:357–374. doi: 10.1146/annurev.arplant.54.031902.134907
- François O, Currat M, Ray N, Han E, Excoffier L, Novembre J (2010) Principal component analysis under population genetic models of range expansion and admixture. *Molecular biology and evolution* 27(6):1257–1268. doi: 10.1093/molbev/msq010
- Frederiksen S, Petersen G (1998) A taxonomic revision of Secale (Triticeae, Poaceae). *Nordic Journal of Botany* 18(4):399–420. doi: 10.1111/j.1756-1051.1998.tb01517.x
- Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. *Methods Ecol Evol* 6(8):925–929. doi: 10.1111/2041-210X.12382
- Frichot E, Mathieu F, Trouillon T, Bouchard G, François O (2014) Fast and efficient estimation of individual ancestry coefficients. *Genetics* 196(4):973–983. doi: 10.1534/genetics.113.160572
- Fujii S, Kubo K-I, Takayama S (2016) Non-self- and self-recognition models in plant self-incompatibility. *Nature plants* 2(9):16130. doi: 10.1038/nplants.2016.130
- Fuong FT, Voylokov AV, Smirnov VG (1993) Genetic studies of self-fertility in rye (*Secale cereale* L.). 2. The search for isozyme marker genes linked to self-incompatibility loci. *Theor Appl Genet* 87(5):619–623. doi: 10.1007/BF00221888
- Galinsky KJ, Bhatia G, Loh P-R, Georgiev S, Mukherjee S, Patterson NJ, Price AL (2016) Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. *American journal of human genetics* 98(3):456–472. doi: 10.1016/j.ajhg.2015.12.022
- Geiger HH, Miedaner T (2009) Rye (*Secale cereale* L.). In: Carena MJ (ed) *Cereals*. Springer Science+Business Media, LLC, New York, pp 157–181
- Gertz A, Wricke G (1989) Linkage Between the Incompatibility Locus Z and a  $\beta$ -Glucosidase Locus in Rye. *Plant Breeding* 102(3):255–259. doi: 10.1111/j.1439-0523.1989.tb00344.x
- Gibson G (2018) Population genetics and GWAS: A primer. *PLoS biology* 16(3):e2005485. doi: 10.1371/journal.pbio.2005485
- Gillespie JH (2004) Population genetics. A concise guide, 2nd ed. Johns Hopkins University Press, Baltimore, Md., London
- Gruner P, Miedaner T (2021) Perennial Rye: Genetics of Perenniality and Limited Fertility. *Plants* (Basel, Switzerland) 10(6). doi: 10.3390/plants10061210
- Grünwald NJ, Everhart SE, Knaus BJ, Kamvar ZN (2017) Best Practices for Population Genetic Analyses. *Phytopathology* 107(9):1000–1010. doi: 10.1094/PHYTO-12-16-0425-RVW
- Gumienny TL, Brugnera E, Tosello-Trampont AC, Kinchen JM, Haney LB, Nishiwaki K, Walk SF, Nemergut ME, Macara IG, Francis R, Schedl T, Qin Y, van Aelst L, Hengartner MO, Ravichandran KS (2001) CED-12/ELMO, a novel member of the CrkII/Dock180/Rac pathway, is required for phagocytosis and cell migration. *Cell* 107(1):27–41. doi: 10.1016/S0092-8674(01)00520-7

## 6. References

---

- Hackauf B, Wehling P (2005) Approaching the self-incompatibility locus Z in rye (*Secale cereale* L.) via comparative genetics. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik 110(5):832–845. doi: 10.1007/s00122-004-1869-4
- Hagenblad J, Oliveira HR, Forsberg NEG, Leino MW (2016) Geographical distribution of genetic diversity in *Secale* landrace and wild accessions. BMC plant biology 16:23. doi: 10.1186/s12870-016-0710-y
- Hammer K (1987) Resistenzmerkmale und Reproduktionssystem als Indikatoren für evolutionäre Tendenzen in der Gattung *Aegilops* L
- Hammer K (1990) Breeding system and phylogenetic relationships in *Secale* L
- Hancock J (2004) Plant evolution and the origin of crop species. CABI Publishing, Wallingford
- Hawliczek A, Borzęcka E, Tofil K, Alachiotis N, Bolibok L, Gawroński P, Siekmann D, Hackauf B, Dušinský R, Švec M, Bolibok-Bragoszewska H (2023) Selective sweeps identification in distinct groups of cultivated rye (*Secale cereale* L.) germplasm provides potential candidate genes for crop improvement. BMC plant biology 23(1):323. doi: 10.1186/s12870-023-04337-1
- He X, Lu M, Cao J, Pan X, Lu J, Zhao L, Zhang H, Chang C, Wang J, Ma C (2023) Genome-Wide Association Analysis of Grain Hardness in Common Wheat. Genes 14(3). doi: 10.3390/genes14030672
- Herridge R, McCourt T, Jacobs JME, Mace P, Brownfield L, Macknight R (2022) Identification of the genes at S and Z reveals the molecular basis and evolution of grass self-incompatibility. Frontiers in plant science 13:1011299. doi: 10.3389/fpls.2022.1011299
- Hicks MA, Barber AE, Giddings L-A, Caldwell J, O'Connor SE, Babbitt PC (2011) The evolution of function in strictosidine synthase-like proteins. Proteins 79(11):3082–3098. doi: 10.1002/prot.23135
- Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theoretical and Applied Genetics (38):226–231
- Hillman G (1978) On the Origins of Domestic Rye— *Secale Cereale* : the Finds from Aceramic Can Hasan III in Turkey. Anatol. Stud. 28:157–174. doi: 10.2307/3642748
- Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST
- Huang S, Sirikhachornkit A, Faris JD, Su X, Gill BS, Haselkorn R, Gornicki P (2002) Phylogenetic analysis of the acetyl-CoA carboxylase and 3-phosphoglycerate kinase loci in wheat and other grasses. Plant molecular biology 48(5-6):805–820. doi: 10.1023/a:1014868320552
- Hubisz M, Siepel A (2020) Inference of Ancestral Recombination Graphs Using ARGweaver. In: Dutheil JY (ed) Statistical population genomics. Open Humana Press, New York, NY, pp 231–266
- Isik Z, Parmaksiz I, Coruh C, Geylan-Su YS, Cebeci O, Beecher B, Budak H (2007) Organellar genome analysis of rye (*Secale cereale*) representing diverse geographic regions. Genome 50(8):724–734. doi: 10.1139/g07-052
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics (Oxford, England) 23(14):1801–1806. doi: 10.1093/bioinformatics/btm233
- Jiang X, Gao Y, Zhou H, Chen J, Wu J, Zhang S (2014) Apoplastic calmodulin promotes self-incompatibility pollen tube growth by enhancing calcium influx and reactive oxygen species concentration in *Pyrus pyrifolia*. Plant cell reports 33(2):255–263. doi: 10.1007/s00299-013-1526-y
- Johnson N, Boatwright JL, Bridges W, Thavarajah P, Kumar S, Thavarajah D (2023) Targeted improvement of plant-based protein: Genome-wide association mapping of a lentil (*Lens culinaris* Medik.) diversity panel. Plants People Planet. doi: 10.1002/ppp3.10470
- Jukes TH, Cantor CR (1969) Evolution of protein molecules. In Munro HN, editor, Mammalian Protein Metabolism
- Khush GS (1962) Cytogenetic and Evolutionary Studies in *Secale*. II. Interrelationships of the Wild Species. Evolution 16(4):484. doi: 10.2307/2406180

## 6. References

---

- Khush GS, Stebbins GL (1961) Cytogenetic and Evolutionary Studies in Secale. I. Some new data on the ancestry of *S. cereale*. *American Journal of Botany* 48(8):723–730. doi: 10.1002/j.1537-2197.1961.tb11703.x
- Kimura M (1969) The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* 61(4):893–903. doi: 10.1093/genetics/61.4.893
- Kimura M (1991) The neutral theory of molecular evolution: a review of recent evidence. *Idengaku zasshi* 66(4):367–386. doi: 10.1266/jgg.66.367
- Kimura M, Crow J (1964) The Number of Alleles That Can Be Maintained in a Finite Population. *Genetics* 49(4):725–738. doi: 10.1093/genetics/49.4.725
- Kobyljanskij VD (1983) The system of the genus *Secale* L
- Konopiński MK (2023) Average weighted nucleotide diversity is more precise than pi<sub>w</sub> in estimating the true value of π from sequence sets containing missing data. *Molecular ecology resources* 23(2):348–354. doi: 10.1111/1755-0998.13707
- Korfmann K, Gaggiotti OE, Fumagalli M (2023) Deep Learning in Population Genetics. *Genome biology and evolution* 15(2). doi: 10.1093/gbe/evad008
- Kulkarni KP, Vorsa N, Natarajan P, Elavarthi S, Iorizzo M, Reddy UK, Melmaiee K (2020) Admixture Analysis Using Genotyping-by-Sequencing Reveals Genetic Relatedness and Parental Lineage Distribution in Highbush Blueberry Genotypes and Cross Derivatives. *International journal of molecular sciences* 22(1). doi: 10.3390/ijms22010163
- Kurasawa K, Matsui A, Yokoyama R, Kuriyama T, Yoshizumi T, Matsui M, Suwabe K, Watanabe M, Nishitani K (2009) The AtXTH28 gene, a xyloglucan endotransglucosylase/hydrolase, is involved in automatic self-pollination in *Arabidopsis thaliana*. *Plant & cell physiology* 50(2):413–422. doi: 10.1093/pcp/pcp003
- Li CC, Horvitz DG (1953) Some methods of estimating the inbreeding coefficient. *Am J Hum Genet*(73):516–523
- Li G, Wang L, Yang J, He H, Jin H, Li X, Ren T, Ren Z, Li F, Han X, Zhao X, Dong L, Li Y, Song Z, Yan Z, Zheng N, Shi C, Wang Z, Yang S, Xiong Z, Zhang M, Sun G, Zheng X, Gou M, Ji C, Du J, Zheng H, Doležel J, Deng XW, Stein N, Yang Q, Zhang K, Wang D (2021) A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. *Nat Genet* 53(4):574–584. doi: 10.1038/s41588-021-00808-z
- Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)* 27(21):2987–2993. doi: 10.1093/bioinformatics/btr509
- Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 25(16):2078–2079. doi: 10.1093/bioinformatics/btp352
- Li H, Ralph P (2019) Local PCA Shows How the Effect of Population Structure Differs Along the Genome. *Genetics* 211(1):289–304. doi: 10.1534/genetics.118.301747
- Li Y, Haseneyer G, Schön C-C, Ankerst D, Korzun V, Wilde P, Bauer E (2011) High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. *BMC Plant Biol* 11(1):6. doi: 10.1186/1471-2229-11-6
- Lian X, Zhang S, Huang G, Huang L, Zhang J, Hu F (2021) Confirmation of a Gametophytic Self-Incompatibility in *Oryza longistaminata*. *Front. Plant Sci.* 12:576340. doi: 10.3389/fpls.2021.576340
- Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics (Oxford, England)* 28(18):2397–2399. doi: 10.1093/bioinformatics/bts444
- Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016) Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLoS genetics* 12(2):e1005767. doi: 10.1371/journal.pgen.1005767

## 6. References

---

- Lundqvist A (1956) SELF-INCOMPATIBILITY IN RYE. *Hereditas* 42(3-4):293–348. doi: 10.1111/j.1601-5223.1956.tb03021.x
- Lundqvist A (1960) THE ORIGIN OF SELF-COMPATIBILITY IN RYE. *Hereditas* 46(1-2):1–19. doi: 10.1111/j.1601-5223.1960.tb03076.x
- Lundqvist A (1962) THE NATURE OF THE TWO-LOCI INCOMPATIBILITY SYSTEM IN GRASSES. *Hereditas* 48(1-2):153–168. doi: 10.1111/j.1601-5223.1962.tb01804.x
- Lux T (2020) Structural gene prediction and annotation for *Secale cereale* Lo7. e!DAL - Plant Genomics and Phenomics Research Data Repository (PGP), IPK Gatersleben, Seeland OT Gatersleben, Corrensstraße 3, 06466, Germany
- Lynch M, Ritland K (1999) Estimation of Pairwise Relatedness With Molecular Markers
- Ma R, Yli-Mattila T, 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(3):210–221. doi: 10.1111/j.1601-5223.2004.01844.x
- Malécot G. (1948) Les mathématiques de l'hérédité, Paris: Masson et Cie
- Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C (2012) Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity* 108(3):285–291. doi: 10.1038/hdy.2011.73
- Manzanares C, Barth S, Thorogood D, Byrne SL, Yates S, Czaban A, Asp T, Yang B, Studer B (2016) A Gene Encoding a DUF247 Domain Protein Cosegregates with the S Self-Incompatibility Locus in Perennial Ryegrass. *Molecular biology and evolution* 33(4):870–884. doi: 10.1093/molbev/msv335
- Maraci Ö, Özkan H, Bilgin R (2018) Phylogeny and genetic structure in the genus *Secale*. *PLOS ONE* 13(7):e0200825. doi: 10.1371/journal.pone.0200825
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* 17(1):10. doi: 10.14806/ej.17.1.200
- Martis MM, Zhou R, Haseneyer G, Schmutz T, Vrána J, Kubaláková M, König S, Kugler KG, Scholz U, Hackauf B, Korzun V, Schön C-C, Dolezel J, Bauer E, Mayer KFX, Stein N (2013) Reticulate evolution of the rye genome. *The Plant cell* 25(10):3685–3698. doi: 10.1105/tpc.113.114553
- Mascher M, Wu S, Amand PS, Stein N, Poland J (2013) Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. *PLOS ONE* 8(10):e76925. doi: 10.1371/journal.pone.0076925
- Massicotte P, South A (2024) rnaturalearth: World Map Data from Natural Earth. R package version 1.0.1.9000
- McVean G (2009) A genealogical interpretation of principal components analysis. *PLoS genetics* 5(10):e1000686. doi: 10.1371/journal.pgen.1000686
- Meisner J, Albrechtsen A (2019) Testing for Hardy-Weinberg equilibrium in structured populations using genotype or low-depth next generation sequencing data. *Molecular ecology resources* 19(5):1144–1152. doi: 10.1111/1755-0998.13019
- Melonek J, Korzun V, Hackauf B (2021) Genomics of Self-Incompatibility and Male-Fertility Restoration in Rye. In: Rabanus-Wallace MT, Stein N (eds) *The Rye Genome*. Springer International Publishing, Cham, pp 181–212
- Melz G, Kaczmarek J, Szigat G (1990) Genetical analysis of rye (*Secale cereale* L.). Location of self-fertility genes in different inbred lines. *Genetica Polonica* 31(1)
- Melz G, Winkel A (1987) Genetical analysis of rye (*Secale Cereale* L.) III. Self-fertility of the rye mutant VD - inheritance and gene location. *Genetica Polonica* 28(1-2)
- Miles CC, Wayne M (2008) Quantitative Trait Locus (QTL) Analysis
- Mitra S (2019) Multiple Data Analyses and Statistical Approaches for Analyzing Data from Metagenomic Studies and Clinical Trials. In: Anisimova M (ed) *Evolutionary Genomics*. Springer New York, New York, NY, pp 605–634
- Monteiro F, Vidigal P, Barros AB, Monteiro A, Oliveira HR, Viegas W (2016) Genetic Distinctiveness of Rye *In situ* Accessions from Portugal Unveils a New Hotspot of Unexplored Genetic Resources. *Frontiers in plant science* 7:1334. doi: 10.3389/fpls.2016.01334

## 6. References

---

- Morton NE (2005) Linkage disequilibrium maps and association mapping. *The Journal of clinical investigation* 115(6):1425–1430. doi: 10.1172/JCI25032
- Moussouni S, Pintaud J-C, Vigouroux Y, Bouguedoura N (2017) Diversity of Algerian oases date palm (*Phoenix dactylifera* L., Arecaceae): Heterozygote excess and cryptic structure suggest farmer management had a major impact on diversity. *PloS one* 12(4):e0175232. doi: 10.1371/journal.pone.0175232
- Müller R, Kaj I, Mugal CF (2022) A Nearly Neutral Model of Molecular Signatures of Natural Selection after Change in Population Size. *Genome biology and evolution* 14(5). doi: 10.1093/gbe/evac058
- Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. *The Plant cell* 21(8):2194–2202. doi: 10.1105/tpc.109.068437
- Nei M, Kumar S (2000) Molecular Evolution and Phylogenetics
- Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases.
- Nei M, Tajima F (1980) DNA Polymorphism Detectable by Restriction Endonucleases
- Nelson CW, Hughes AL (2015) Within-host nucleotide diversity of virus populations: insights from next-generation sequencing. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 30:1–7. doi: 10.1016/j.meegid.2014.11.026
- Newell MA, Cook D, Tinker NA, Jannink J-L (2011) Population structure and linkage disequilibrium in oat (*Avena sativa* L.): implications for genome-wide association studies. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* 122(3):623–632. doi: 10.1007/s00122-010-1474-7
- Novembre J, Stephens M (2008) Interpreting principal component analyses of spatial population genetic variation. *Nature genetics* 40(5):646–649. doi: 10.1038/ng.139
- Ochoa A, Storey JD (2021) Estimating FST and kinship for arbitrary population structures. *PLoS genetics* 17(1):e1009241. doi: 10.1371/journal.pgen.1009241
- Ohta T (2017) Nearly Neutral Theory. In: Reference Module in Life Sciences. Elsevier
- Okazaki A, Yamazaki S, Inoue I, Ott J (2021) Population genetics: past, present, and future. *Human genetics* 140(2):231–240. doi: 10.1007/s00439-020-02208-5
- Oppermann M, Weise S, Dittmann C, Knüpffer H (2015) GBIS: the information system of the German Genebank. *Database : the journal of biological databases and curation* 2015:bav021. doi: 10.1093/database/bav021
- Otyama PI, Wilkey A, Kulkarni R, Assefa T, Chu Y, Clevenger J, O'Connor DJ, Wright GC, Dezern SW, MacDonald GE, Anglin NL, Cannon EKS, Ozias-Akins P, Cannon SB (2019) Evaluation of linkage disequilibrium, population structure, and genetic diversity in the U.S. peanut mini core collection. *BMC genomics* 20(1):481. doi: 10.1186/s12864-019-5824-9
- Parat F, Schwertfirm G, Rudolph U, Miedaner T, Korzun V, Bauer E, Schön C-C, Tellier A (2016) Geography and end use drive the diversification of worldwide winter rye populations. *Molecular ecology* 25(2):500–514. doi: 10.1111/mec.13495
- Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, Legge SE, Bishop S, Cameron D, Hamshere ML, Han J, Hubbard L, Lynham A, Mantripragada K, Rees E, MacCabe JH, McCarroll SA, Baune BT, Breen G, Byrne EM, Dannlowski U, Eley TC, Hayward C, Martin NG, McIntosh AM, Plomin R, Porteous DJ, Wray NR, Caballero A, Geschwind DH, Huckins LM, Ruderfer DM, Santiago E, Sklar P, Stahl EA, Won H, Agerbo E, Als TD, Andreassen OA, Bækvad-Hansen M, Mortensen PB, Pedersen CB, Børglum AD, Bybjerg-Grauholt J, Djurovic S, Durmishi N, Pedersen MG, Golimbet V, Grove J, Hougaard DM, Mattheisen M, Molden E, Mors O, Nordentoft M, Pejovic-Milovancevic M, Sigurdsson E, Silagadze T, Hansen CS, Stefansson K, Stefansson H, Steinberg S, Tosato S, Werge T, Collier DA, Rujescu D, Kirov G, Owen MJ, O'Donovan MC, Walters JTR (2018) Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* 50(3):381–389. doi: 10.1038/s41588-018-0059-2

## 6. References

---

- Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS genetics* 2(12):e190. doi: 10.1371/journal.pgen.0020190
- Pereira GS, Garcia AAF, Margarido GRA (2018) A fully automated pipeline for quantitative genotype calling from next generation sequencing data in autopolyploids. *BMC bioinformatics* 19(1):398. doi: 10.1186/s12859-018-2433-6
- Prins P, Smart G, Arends D, Mulligan MK, Williams RW, Jansen RC (2019) Systems Genetics for Evolutionary Studies. *Methods in molecular biology* (Clifton, N.J.) 1910:635–652. doi: 10.1007/978-1-4939-9074-0\_21
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959. doi: 10.1093/genetics/155.2.945
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, Bakker PIW de, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81(3):559–575. doi: 10.1086/519795
- Qu J, Kachman SD, Garrick D, Fernando RL, Cheng H (2020) Exact Distribution of Linkage Disequilibrium in the Presence of Mutation, Selection, or Minor Allele Frequency Filtering. *Frontiers in genetics* 11:362. doi: 10.3389/fgene.2020.00362
- R Core Team (2023) R: A language and environment for statistical computing. <https://www.R-project.org/>
- Rabanus-Wallace MT, Hackauf B, Mascher M, Lux T, Wicker T, Gundlach H, Baez M, Houben A, Mayer KFX, Guo L, Poland J, Pozniak CJ, Walkowiak S, Melonek J, Praz CR, Schreiber M, Budak H, Heuberger M, Steuernagel B, Wulff B, Börner A, Byrns B, Čížková J, Fowler DB, Fritz A, Himmelbach A, Kaithakottil G, Keilwagen J, Keller B, Konkin D, Larsen J, Li Q, Myśkow B, Padmarasu S, Rawat N, Sesiz U, Biyiklioglu-Kaya S, Sharpe A, Šimková H, Small I, Swarbreck D, Toegelová H, Tsvetkova N, Voylokov AV, Vrána J, Bauer E, Bolibok-Bragoszewska H, Doležel J, Hall A, Jia J, Korzun V, Laroche A, Ma X-F, Ordon F, Özkan H, Rakoczy-Trojanowska M, Scholz U, Schulman AH, Siekmann D, Stojąłowski S, Tiwari VK, Spannagl M, Stein N (2021) Chromosome-scale genome assembly provides insights into rye biology, evolution and agronomic potential. *Nature genetics* 53(4):564–573. doi: 10.1038/s41588-021-00807-0
- Rabanus-Wallace MT, Stein N (2021) The Rye Genome. Springer International Publishing, Cham
- Rebong D, Henriquez Inoa S, Moore VM, Reberg-Horton SC, Mirsky S, Murphy JP, Leon RG (2024) Breeding allelopathy in cereal rye for weed suppression. *Weed Sci* 72(1):30–40. doi: 10.1017/wsc.2023.64
- Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. *Genet. Res.* 67(2):175–185. doi: 10.1017/S0016672300033620
- Ritland K, Travis S (2004) Inferences involving individual coefficients of relatedness and inbreeding in natural populations of *Abies*. *Forest Ecology and Management* 197(1-3):171–180. doi: 10.1016/j.foreco.2004.05.012
- Robledo-Arnuncio JJ, Klein EK, Muller-Landau HC, Santamaría L (2014) Space, time and complexity in plant dispersal ecology. *Movement ecology* 2(1):16. doi: 10.1186/s40462-014-0016-3
- Rohner M, Manzanares C, Yates S, Thorogood D, Copetti D, Lübbertstedt T, Asp T, Studer B (2023) Fine-Mapping and Comparative Genomic Analysis Reveal the Gene Composition at the S and Z Self-incompatibility Loci in Grasses. *Molecular biology and evolution* 40(1). doi: 10.1093/molbev/msac259
- Roshevitz RY (1947) A monograph of the wild, weedy and cultivated species of rye
- Schiemann E (1948) Weizen, Roggen, Gerste : Systematik, Geschichte und Verwendung. Gustav Fischer Verlag
- Schittenhelm S, Kraft M, Wittich K-P (2014) Performance of winter cereals grown on field-stored soil moisture only. *European Journal of Agronomy* 52:247–258. doi: 10.1016/j.eja.2013.08.010
- Schlegel RHJ (2014) Rye. *Genetics, breeding, and cultivation. Life science*. CRC Press, Boca Raton, FL
- Schreiber M, Gao Y, Koch N, Fuchs J, Heckmann S, Himmelbach A, Börner A, Özkan H, Maurer A, Stein N, Mascher M, Dreissig S (2022) Recombination landscape divergence between populations is

## 6. References

---

- marked by larger low-recombining regions in domesticated rye. *Molecular biology and evolution* 39(6). doi: 10.1093/molbev/msac131
- Schreiber M, Himmelbach A, Börner A, Mascher M (2019) Genetic diversity and relationship between domesticated rye and its wild relatives as revealed through genotyping-by-sequencing. *Evolutionary applications* 12(1):66–77. doi: 10.1111/eva.12624
- Schreiber M, Özkan H, Komatsuda T, Mascher M (2021) Evolution and Domestication of Rye. In: Rabanus-Wallace MT, Stein N (eds) *The Rye Genome*. Springer International Publishing, Cham, pp 85–100
- Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet* 44(7):825–830. doi: 10.1038/ng.2314
- Sella G, Barton NH (2019) Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. *Annual Review of Genomics and Human Genetics* 20(Volume 20, 2019):461–493. doi: 10.1146/annurev-genom-083115-022316
- Sencer HA (1975) Study of variation in the genus *Secale* L. and on the origin of the cultivated rye
- Sencer HA, Hawkes JG (1980) On the origin of cultivated rye. *Biological Journal of the Linnean Society* 13(4):299–313
- Shang H-Y, Wei Y-M, Wang X-R, Zheng Y-L (2006) Genetic diversity and phylogenetic relationships in the rye genus *Secale* L. (rye) based on *Secale cereale* microsatellite markers. *Genet. Mol. Biol.* 29(4):685–691. doi: 10.1590/s1415-47572006000400018
- Shinozuka H, Cogan NOI, Smith KF, Spangenberg GC, Forster JW (2010) Fine-scale comparative genetic and physical mapping supports map-based cloning strategies for the self-incompatibility loci of perennial ryegrass (*Lolium perenne* L.). *Plant Mol Biol* 72(3):343–355. doi: 10.1007/s11103-009-9574-y
- Shivanna KR, Heslop-Harrison Y, Heslop-Harrison J (1982) THE POLLEN-STIGMA INTERACTION IN THE GRASSES. 3. FEATURES OF THE SELF-INCOMPATIBILITY RESPONSE. *Acta Botanica Neerlandica* 31(4):307–319. doi: 10.1111/j.1438-8677.1982.tb01637.x
- Sidhu JS, Ramakrishnan SM, Ali S, Bernardo A, Bai G, Abdullah S, Ayana G, Sehgal SK (2019) Assessing the genetic diversity and characterizing genomic regions conferring Tan Spot resistance in cultivated rye. *PLOS ONE* 14(3):e0214519. doi: 10.1371/journal.pone.0214519
- Siekmann D, Jansen G, Zaar A, Kilian A, Fromme FJ, Hackauf B (2021) A Genome-Wide Association Study Pinpoints Quantitative Trait Genes for Plant Height, Heading Date, Grain Quality, and Yield in Rye (*Secale cereale* L.). *Frontiers in plant science* 12:718081. doi: 10.3389/fpls.2021.718081
- Singh M, Kumar M, Albertsen MC, Young JK, Cigan AM (2018) Concurrent modifications in the three homeologs of Ms45 gene with CRISPR-Cas9 lead to rapid generation of male sterile bread wheat (*Triticum aestivum* L.). *Plant Mol Biol* 97(4-5):371–383. doi: 10.1007/s11103-018-0749-2
- Skuza L, Szućko I, FILIP E, Strzała T (2019) Genetic diversity and relationship between cultivated, weedy and wild rye species as revealed by chloroplast and mitochondrial DNA non-coding regions analysis. *PLOS ONE* 14(2):e0213023. doi: 10.1371/journal.pone.0213023
- Skuza L, Szućko-Kociuba I, FILIP E, ADAMCZYK A (2018) DNA Barcoding in Selected Species and Subspecies of Rye (*Secale*) Using Three Chloroplast Loci (matK, rbcL, trnH-psbA). *Not Bot Horti Agrobo* 47(1):54–62. doi: 10.15835/nbha47111248
- Stratton D (2008) Inbreeding and Population Structure
- Stutz HC (1957) A Cytogenetic Analysis of the Hybrid *Secale Cereale* L. x *Secale Montanum* Guss. and Its Progeny. *Genetics* 42(3):199–221. doi: 10.1093/genetics/42.3.199
- Stutz HC (1972) On the Origin of Cultivated Rye. *American Journal of Botany*(59, No. 1):59–70
- Sul JH, Martin LS, Eskin E (2018) Population structure in genetic studies: Confounding factors and mixed models. *PLoS genetics* 14(12):e1007309. doi: 10.1371/journal.pgen.1007309
- Sun Y, Guo L, Zhu Q-H, Fan L (2022a) When domestication bottleneck meets weed. *Molecular plant* 15(9):1405–1408. doi: 10.1016/j.molp.2022.08.002

## 6. References

---

- Sun Y, Shen E, Hu Y, Wu D, Feng Y, Lao S, Dong C, Du T, Hua W, Ye C-Y, Zhu J, Zhu Q-H, Cai D, Skuza L, Qiu J, Fan L (2022b) Population genomic analysis reveals domestication of cultivated rye from weedy rye. *Molecular plant* 15(3):552–561. doi: 10.1016/j.molp.2021.12.015
- Tabangin ME, Woo JG, Martin LJ (2009) The effect of minor allele frequency on the likelihood of obtaining false positives. *BMC proceedings* 3 Suppl 7(Suppl 7):S41. doi: 10.1186/1753-6561-3-S7-S41
- Tajima F, Nei M (1982) Biases of the estimates of DNA divergence obtained by the restriction enzyme technique. *Journal of molecular evolution* 18(2):115–120. doi: 10.1007/BF01810830
- Takayama S, Isogai A (2005) Self-incompatibility in plants. *Annual review of plant biology* 56:467–489. doi: 10.1146/annurev.arplant.56.032604.144249
- Takebayashi N, Morrell PL (2001) Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. *American Journal of Botany* 88(7):1143–1150
- Targonska-Karasek M, Boczkowska M, Podyma W, Pasnik M, Niedzielski M, Rucinska A, Nowak-Zyczynska Z, Rakoczy-Trojanowska M (2020) Investigation of obsolete diversity of rye (*Secale cereale* L.) using multiplexed SSR fingerprinting and evaluation of agronomic traits. *Journal of applied genetics* 61(4):513–529. doi: 10.1007/s13353-020-00579-z
- Thorogood D, Hayward MD (1991) The genetic control of self-compatibility in an inbred line of *Lolium perenne* L. *Heredity* 67(2):175–181. doi: 10.1038/hdy.1991.77
- Thorogood D, Yates S, Manzanares C, Skot L, Hegarty M, Blackmore T, Barth S, Studer B (2017) A Novel Multivariate Approach to Phenotyping and Association Mapping of Multi-Locus Gametophytic Self-Incompatibility Reveals S, Z, and Other Loci in a Perennial Ryegrass (Poaceae) Population. *Front. Plant Sci.* 8:1331. doi: 10.3389/fpls.2017.01331
- Tran F, Penniket C, Patel RV, Provart NJ, Laroche A, Rowland O, Robert LS (2013) Developmental transcriptional profiling reveals key insights into Triticeae reproductive development. *The Plant journal : for cell and molecular biology* 74(6):971–988. doi: 10.1111/tpj.12206
- Tucker G, Price AL, Berger B (2014) Improving the power of GWAS and avoiding confounding from population stratification with PC-Select. *Genetics* 197(3):1045–1049. doi: 10.1534/genetics.114.164285
- VanLiere JM, Rosenberg NA (2008) Mathematical properties of the  $r^2$  measure of linkage disequilibrium. *Theoretical population biology* 74(1):130–137. doi: 10.1016/j.tpb.2008.05.006
- Vavilov NI (1917) On the origin of cultivated rye. *Bulletin of Applied Botany*(10(7-10)):561–590
- Vavilov NI (1926) Studies on the Origin of Cultivated Plants. *Bulletin of Applied Botany*(16):1–248
- Vences FJ, Vaquero F, La Prez de Vega M (1987) Phylogenetic relationships in *Secale* (Poaceae): An isozymatic study. *PI Syst Evol* 157(1-2):33–47. doi: 10.1007/BF00939179
- Villanueva B, Fernández A, Saura M, Caballero A, Fernández J, Morales-González E, Toro MA, Pong-Wong R (2021) The value of genomic relationship matrices to estimate levels of inbreeding. *Genetics, selection, evolution : GSE* 53(1):42. doi: 10.1186/s12711-021-00635-0
- Vos PG, Paulo MJ, Voorrips RE, Visser RGF, van Eck HJ, van Eeuwijk FA (2017) Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* 130(1):123–135. doi: 10.1007/s00122-016-2798-8
- Voylokov AV, Fuong FT, Smirnov VG (1993) Genetic studies of self-fertility in rye (*Secale cereale* L.). 1. The identification of genotypes of self-fertile lines for the Sf alleles of self-incompatibility genes. *Theor Appl Genet* 87(5):616–618. doi: 10.1007/BF00221887
- Voylokov AV, Korzun V, Börner A (1998) Mapping of three self-fertility mutations in rye (*Secale cereale* L.) using RFLP, isozyme and morphological markers. *Theor Appl Genet* 97(1-2):147–153. doi: 10.1007/s001220050879
- Wang J (2022) Fast and accurate population admixture inference from genotype data from a few microsatellites to millions of SNPs. *Heredity* 129(2):79–92. doi: 10.1038/s41437-022-00535-z

## 6. References

---

- Wang J, Zhang Z (2021) GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genomics, proteomics & bioinformatics* 19(4):629–640. doi: 10.1016/j.gpb.2021.08.005
- Wang JA (2011) A new likelihood estimator and its comparison with moment estimators of individual genome-wide diversity
- Wang L, Peng H, Ge T, Liu T, Hou X, Li Y (2014a) Identification of differentially accumulating pistil proteins associated with self-incompatibility of non-heading Chinese cabbage. *Acta Botanica Neerlandica* 16(1):49–57. doi: 10.1111/plb.12016
- Wang Q, Tian F, Pan Y, Buckler ES, Zhang Z (2014b) A SUPER powerful method for genome wide association study. *PLOS ONE* 9(9):e107684. doi: 10.1371/journal.pone.0107684
- Wehling P, Hackauf B, Wricke, G. . Inst. fuer Angewandte Genetik (1995) Characterization of the two-factor self-incompatibility system in *Secale cereale* L. *Fortschritte der Pflanzenzuechtung* (Germany) 18
- Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure
- Willing E-M, Dreyer C, van Oosterhout C (2012) Estimates of genetic differentiation measured by F(ST) do not necessarily require large sample sizes when using many SNP markers. *PloS one* 7(8):e42649. doi: 10.1371/journal.pone.0042649
- Wollenweber TE, van Deenen N, Roelfs K-U, Prüfer D, Gronover CS (2021) Microscopic and Transcriptomic Analysis of Pollination Processes in Self-Incompatible *Taraxacum koksaghyz*. *Plants* 10(3). doi: 10.3390/plants10030555
- Wricke G, Wehling P (1985) Linkage between an incompatibility locus and a peroxidase isozyme locus (Prx 7) in rye. *Theor Appl Genet* 71(2):289–291. doi: 10.1007/BF00252069
- Wright S (1921) Systems of Mating. V. General Considerations. *Genetics* 6(2):167–178. doi: 10.1093/genetics/6.2.167
- Wright S (1951) The genetical structure of populations. *Annals of eugenics* 15(4):323–354. doi: 10.1111/j.1469-1809.1949.tb02451.x
- Wright S (1978) Evolution and the genetic of population. Variability within and among Natural Populations, 4th edn., vol 4, Chicago: University of Chicago
- Yang B, Thorogood D, Armstead IP, Franklin FCH, Barth S (2009) Identification of genes expressed during the self-incompatibility response in perennial ryegrass (*Lolium perenne* L.). *Plant Mol Biol* 70(6):709–723. doi: 10.1007/s11103-009-9501-2
- Yuan Y, Zhang Q, Zeng S, Gu L, Si W, Zhang X, Tian D, Yang S, Wang L (2017) Selective sweep with significant positive selection serves as the driving force for the differentiation of japonica and indica rice cultivars. *BMC genomics* 18(1):307. doi: 10.1186/s12864-017-3702-x
- Zeng Z-B (2005) QTL mapping and the genetic basis of adaptation: recent developments. In: Mauricio R (ed) *Genetics of Adaptation*. Springer, Dordrecht, pp 25–37
- Zhao P, Zhang L, Zhao L (2015) Dissection of the style's response to pollination using transcriptome profiling in self-compatible (*Solanum pimpinellifolium*) and self-incompatible (*Solanum chilense*) tomato species. *BMC Plant Biol* 15:119. doi: 10.1186/s12870-015-0492-7
- Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* (Oxford, England) 28(24):3326–3328. doi: 10.1093/bioinformatics/bts606
- Zhu C, Gore M, Buckler ES, Yu J (2008a) Status and Prospects of Association Mapping in Plants. *The Plant Genome* 1(1). doi: 10.3835/plantgenome2008.02.0089
- Zhu X, Li S, Cooper RS, Elston RC (2008b) A unified association analysis approach for family and unrelated samples correcting for stratification. *American journal of human genetics* 82(2):352–365. doi: 10.1016/j.ajhg.2007.10.009
- Zohary D, Hopf M, Weiss E (2012) Domestication of Plants in the Old World. The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin, 4th ed. OUP Oxford, Oxford

## 7. Appendix

---

# 7. Appendix

## A1. The S- & Z-locus

Anchoring the **S-locus** to the rye reference genome of ‘Lo7’. SNP markers originate from the rye 600k array (Bauer et al. 2017). References: 1 - Rabanus-Wallace et al. (2021), 2 - Bian et al. (2004). For surrounding markers, as well as DUF247 positions, see source (source: Melonek et al. 2021)

Marker	Rye gene model	Position (Mbp)	No of SNPs	Description	Refs.
tcos3511, tcos137	SECCE1Rv1G0014520	115.1	5	Bifunctional protein Fold	1
	SECCE1Rv1G0014530	115.1	14	Histidine-tRNA ligase	
	SECCE1Rv1G0014540	115.2	1	Desiccation-related protein PCC13-62	
BCD762, TC76051, tcos163, tcos164	SECCE1Rv1G0014550	115.2	10	40S ribosomal protein S4	1, 2
	SECCE1Rv1G0014560	115.3		Translocator protein-like protein	
	SECCE1Rv1G0014570	115.4	15	BTB/POZ and MATH domain-containing protein 2	
	SECCE1Rv1G0014580	115.4	4	Pre-rRNA-processing protein TSR2	
	SECCE1Rv1G0014590	115.5		Acetylglutamate kinase	
	SECCE1Rv1G0014600	115.6	14	BTB/POZ and MATH domain-containing protein 2	
	SECCE1Rv1G0014610	115.6	14	BTB/POZ and MATH domain-containing protein 2	
	SECCE1Rv1G0014620	115.6	10	BTB/POZ and MATH domain-containing protein 2	
tcos162	SECCE1Rv1G0014630	115.6	5	Rubber elongation factor protein putative	1
	SECCE1Rv1G0014640	115.6	5	Translocase of chloroplast 159, chloroplastic	
	SECCE1Rv1G0014650	115.7	5	Xylosyltransferase 1	
	SECCE1Rv1G0014660	115.7	5	E3 ubiquitin-protein ligase	
	SECCE1Rv1G0014690	116.8		Ubiquitin-conjugating enzyme 23	
	SECCE1Rv1G0014710	117.1	7	cDNA clone:J013058P10, full insert sequence	
	SECCE1Rv1G0014750	117.9	11	Carboxyl-terminal peptidase, putative (DUF239)	
tcos3507, tcos5040	SECCE1Rv1G0014770	118.3		Villin	1

Anchoring the **Z-locus** to the rye reference genome of ‘Lo7’. SNP markers originate from the rye 600k array (Bauer et al. 2017). Based on Rabanus-Wallace et al. (2021). (source: Melonek et al. 2021)

Marker	Gene model	Position (Mbp)	No of SNPs	Description
TC89057	SECCE2Rv1G0130710.1	878.2	6	Glycerol kinase
	SECCE2Rv1G0130770.1	878.4		Transmembrane protein, putative (DUF247)
	SECCE2Rv1G0130780.1	878.5		Transmembrane protein, putative (DUF247)
TC116908	SECCE2Rv1G0130790.1	878.7	4	Ubiquitin carboxyl-terminal hydrolase
	SECCE2Rv1G0130800.1	878.7	5	Short chain dehydrogenase/reductase
	SECCE2Rv1G0130810.1	878.8	1	F-box family protein
	SECCE2Rv1G0130820.1	879.0	7	EamA-like transporter family
	SECCE2Rv1G0130830.1	879.0	20	Tryptophan decarboxylase
	SECCE2Rv1G0130840.1	879.0		Methyltransferase-like protein
	SECCE2Rv1G0130870.1	879.2		VQ motif-containing protein, putative
TC101821	SECCE2Rv1G0130890.1	879.4	3	Organic cation transporter protein
	SECCE2Rv1G0130900.1	879.4	7	Clustered mitochondria protein

## 7. Appendix

---

### A2. Passport data

1.396 genotypes. According to Schreiber et al. (2022), Table edited. Sample ID “130\_PI561806\_1” from original data not considered due to error-causing F value at GWAS.

Sample ID	accession	species	subtaxa	country	panel	seq	country-code	domestication status
379_2004_6_21_3_9_1	2004-6-21-3-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
380_2004_6_21_3_9_2	2004-6-21-3-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
381_2004_6_21_3_9_3	2004-6-21-3-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
382_2004_6_21_3_9_4	2004-6-21-3-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
385_2004_6_21_3_9_1	2004-6-21-3-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
386_2004_6_21_3_9_2	2004-6-21-3-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
383_2004_6_21_4_7_1	2004-6-21-4-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
384_2004_6_21_4_7_2	2004-6-21-4-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
13_2004_6_22_2_4_1	2004-6-22-2-4	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
14_2004_6_22_2_4_2	2004-6-22-2-4	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
390_2004_6_22_2_4_1	2004-6-22-2-4	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
391_2004_6_23_2_7_1	2004-6-23-2-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
392_2004_6_23_2_7_2	2004-6-23-2-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
393_2004_6_23_2_7_3	2004-6-23-2-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
397_2004_6_23_3_5_1	2004-6-23-3-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
398_2004_6_23_3_5_2	2004-6-23-3-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
399_2004_6_23_3_5_3	2004-6-23-3-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
400_2004_6_23_3_5_4	2004-6-23-3-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
387_2004_7_3_11_6_1	2004-7-3-11-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
388_2004_7_3_11_6_2	2004-7-3-11-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
389_2004_7_3_11_6_3	2004-7-3-11-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
394_2005_6_25_11_9_1	2005-6-25-11-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
395_2005_6_25_11_9_2	2005-6-25-11-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
396_2005_6_25_11_9_3	2005-6-25-11-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
343_2005_6_26_2_13_1	2005-6-26-2-13	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
344_2005_6_26_2_13_2	2005-6-26-2-13	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
345_2005_6_26_2_13_3	2005-6-26-2-13	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
346_2005_6_26_2_13_4	2005-6-26-2-13	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
347_2005_6_26_3_8_1	2005-6-26-3-8	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
348_2005_6_26_3_8_2	2005-6-26-3-8	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
349_2005_6_26_3_8_3	2005-6-26-3-8	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
350_2005_6_26_3_8_4	2005-6-26-3-8	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
351_2005_6_26_4_5_1	2005-6-26-4-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
352_2005_6_26_4_5_2	2005-6-26-4-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
353_2005_6_26_4_5_3	2005-6-26-4-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
354_2005_6_26_4_5_4	2005-6-26-4-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
24_2005_6_26_7_6_1	2005-6-26-7-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
25_2005_6_26_7_6_2	2005-6-26-7-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
370_2005_6_26_9_5_1	2005-6-26-9-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
371_2005_6_26_9_5_2	2005-6-26-9-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
372_2005_6_26_9_5_3	2005-6-26-9-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
359_2005_6_27_1_6_1	2005-6-27-1-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
360_2005_6_27_1_6_2	2005-6-27-1-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
361_2005_6_27_1_6_3	2005-6-27-1-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
339_2005_6_27_4_9_1	2005-6-27-4-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
340_2005_6_27_4_9_2	2005-6-27-4-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
341_2005_6_27_4_9_3	2005-6-27-4-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
342_2005_6_27_4_9_4	2005-6-27-4-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
373_2005_6_27_4_9_1	2005-6-27-4-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
374_2005_6_27_4_9_2	2005-6-27-4-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
375_2005_6_27_4_9_3	2005-6-27-4-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
376_2005_6_27_4_9_4	2005-6-27-4-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
377_2005_6_27_4_9_5	2005-6-27-4-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
355_2005_6_27_5_13_1	2005-6-27-5-13	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
356_2005_6_27_5_13_2	2005-6-27-5-13	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
357_2005_6_27_5_13_3	2005-6-27-5-13	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
358_2005_6_27_5_13_4	2005-6-27-5-13	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed

Table continued on next page

## 7. Appendix

---

2_2005_6_27_5_14_1	2005-6-27-5-14	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
3_2005_6_27_5_14_2	2005-6-27-5-14	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
287_2005_6_27_5_14_1	2005-6-27-5-14	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
288_2005_6_27_5_14_2	2005-6-27-5-14	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
378_2005_6_27_5_14_1	2005-6-27-5-14	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
401_2005_6_28_12_7_1	2005-6-28-12-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
402_2005_6_28_12_7_2	2005-6-28-12-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
254_2005_6_28_2_11_1	2005-6-28-2-11	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
255_2005_6_28_2_11_2	2005-6-28-2-11	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
299_2005_6_28_3_5_1	2005-6-28-3-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
300_2005_6_28_3_5_2	2005-6-28-3-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
404_2005_6_28_3_5_1	2005-6-28-3-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
236_2005_6_28_4_5_1	2005-6-28-4-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
237_2005_6_28_4_5_2	2005-6-28-4-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
9_2005_6_28_5_7_1	2005-6-28-5-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
10_2005_6_28_5_7_2	2005-6-28-5-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
262_2005_6_28_5_7_1	2005-6-28-5-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
263_2005_6_28_5_7_2	2005-6-28-5-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
405_2005_6_28_5_7_1	2005-6-28-5-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
216_2005_6_28_8_6_1	2005-6-28-8-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
217_2005_6_28_8_6_2	2005-6-28-8-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
214_2005_6_28_9_8_1	2005-6-28-9-8	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
215_2005_6_28_9_8_2	2005-6-28-9-8	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
334_2005_6_29_7_3_1	2005-6-29-7-3	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
335_2005_6_29_7_3_2	2005-6-29-7-3	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
336_2005_6_29_7_3_3	2005-6-29-7-3	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
368_2005_6_29_7_3_1	2005-6-29-7-3	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
369_2005_6_29_7_3_2	2005-6-29-7-3	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
28_25_6_30_1_12_1	2005-6-30-1-12	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
29_25_6_30_1_12_2	2005-6-30-1-12	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
337_2005_6_30_3_6_1	2005-6-30-3-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
338_2005_6_30_3_6_2	2005-6-30-3-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
365_2005_6_30_5_5_1	2005-6-30-5-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
366_2005_6_30_5_5_2	2005-6-30-5-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
367_2005_6_30_5_5_3	2005-6-30-5-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
240_2005_7_5_7_5_1	2005-7-5-7-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
241_2005_7_5_7_5_2	2005-7-5-7-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
260_2006_6_17_5_1_1	2006-6-17-5-1	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
261_2006_6_17_5_1_2	2006-6-17-5-1	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
43_2006_6_20_10_7_1	2006-6-20-10-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
44_2006_6_20_10_7_2	2006-6-20-10-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
51_2006_6_20_11_12_1	2006-6-20-11-12	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
52_2006_6_20_11_12_2	2006-6-20-11-12	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
252_2006_6_24_1_9_1	2006-6-24-1-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
253_2006_6_24_1_9_2	2006-6-24-1-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
64_2006_6_25_2_6_1	2006-6-25-2-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
65_2006_6_25_2_6_2	2006-6-25-2-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
26_2006_6_25_3_6_1	2006-6-25-3-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
41_2006_6_25_3_7_1	2006-6-25-3-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
42_2006_6_25_3_7_2	2006-6-25-3-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
47_2006_6_25_4_1_1	2006-6-25-4-1	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
48_2006_6_25_4_1_2	2006-6-25-4-1	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
39_2006_6_25_5_4_1	2006-6-25-5-4	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
40_2006_6_25_5_4_2	2006-6-25-5-4	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
289_2006_6_25_7_7_1	2006-6-25-7-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
290_2006_6_25_7_7_2	2006-6-25-7-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
45_2006_6_25_8_8_1	2006-6-25-8-8	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
46_2006_6_25_8_8_2	2006-6-25-8-8	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
301_2006_6_25_9_6_1	2006-6-25-9-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
302_2006_6_25_9_6_2	2006-6-25-9-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
37_2006_6_26_1_5_1	2006-6-26-1-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
38_2006_6_26_1_5_2	2006-6-26-1-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
264_2006_6_26_1_5_1	2006-6-26-1-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
265_2006_6_26_1_5_2	2006-6-26-1-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
291_2006_6_26_2_10_1	2006-6-26-2-10	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
292_2006_6_26_2_10_2	2006-6-26-2-10	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
49_2006_6_26_2_11_1	2006-6-26-2-11	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed

Table continued on next page

## 7. Appendix

---

50_2006_6_26_2_11_2	2006-6-26-2-11	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
27_2006_6_26_2_9_1	2006-6-26-2-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
246_2006_6_26_6_5_1	2006-6-26-6-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
247_2006_6_26_6_5_2	2006-6-26-6-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	domesticated:middle
293_2006_6_27_1_8_1	2006-6-27-1-8	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
294_2006_6_27_1_8_2	2006-6-27-1-8	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
35_2006_6_27_1_9_1	2006-6-27-1-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
36_2006_6_27_1_9_2	2006-6-27-1-9	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
305_2006_6_27_2_7_1	2006-6-27-2-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
306_2006_6_27_2_7_2	2006-6-27-2-7	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
311_2006_6_27_4_6_1	2006-6-27-4-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
312_2006_6_27_4_6_2	2006-6-27-4-6	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
17_As slim_1	As slim	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
18_As slim_2	As slim	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
19_As slim_3	As slim	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
138_CSE1_1	CSE_1	Secale cereale	subsp. cereale	Sweden	Hakan2019	GBS	SWE	domesticated:middle
139_CSE1_2	CSE_1	Secale cereale	subsp. cereale	Sweden	Hakan2019	GBS	SWE	domesticated:middle
154_CSE100_1	CSE_100	Secale cereale	subsp. cereale	Czech Republic	Hakan2019	GBS	CZE	feral:nonbrittle
155_CSE100_2	CSE_100	Secale cereale	subsp. cereale	Czech Republic	Hakan2019	GBS	CZE	feral:nonbrittle
105_CSE37_1	CSE_37	Secale cereale	subsp. cereale	Afghanistan	Hakan2019	GBS	AFG	feral:nonbrittle
106_CSE37_2	CSE_37	Secale cereale	subsp. cereale	Afghanistan	Hakan2019	GBS	AFG	feral:nonbrittle
140_CSE70_1	CSE_70	Secale cereale	subsp. cereale	Germany	Hakan2019	GBS	DEU	domesticated:middle
141_CSE70_2	CSE_70	Secale cereale	subsp. cereale	Germany	Hakan2019	GBS	DEU	domesticated:middle
55_PI168130_1	PI-168130	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
56_PI168130_2	PI-168130	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
276_PI168133_1	PI-168133	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
277_PI168133_2	PI-168133	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
118_PI168136_1	PI-168136	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
119_PI168136_2	PI-168136	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
200_PI168164_1	PI-168164	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
134_PI168168_1	PI-168168	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
135_PI168168_2	PI-168168	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
123_PI168169_1	PI-168169	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
90_PI168176_1	PI-168176	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
211_PI168178_1	PI-168178	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
11_PI168181_1	PI-168181	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
12_PI168181_2	PI-168181	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
209_PI168182_1	PI-168182	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
210_PI168182_2	PI-168182	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
70_PI168184_1	PI-168184	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
71_PI168184_2	PI-168184	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
126_PI168185_1	PI-168185	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
166_PI168186_1	PI-168186	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
198_PI168188_1	PI-168188	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
199_PI168188_2	PI-168188	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
167_PI168194_1	PI-168194	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
168_PI168194_2	PI-168194	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
122_PI168196_1	PI-168196	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
120_PI168199_1	PI-168199	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
132_PI168205_1	PI-168205	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
133_PI168205_2	PI-168205	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
323_PI168209_1	PI-168209	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
324_PI168209_2	PI-168209	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
315_PI168211_1	PI-168211	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
316_PI168211_2	PI-168211	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
33_PI168213_1	PI-168213	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
34_PI168213_2	PI-168213	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
295_PI168218_1	PI-168218	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
296_PI168218_2	PI-168218	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
258_PI168219_1	PI-168219	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
259_PI168219_2	PI-168219	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
205_PI168220_1	PI-168220	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
206_PI168220_2	PI-168220	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
66_PI168222_1	PI-168222	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
317_PI168227_1	PI-168227	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
318_PI168227_2	PI-168227	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
238_PI168229_1	PI-168229	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle

Table continued on next page

## 7. Appendix

---

239_PI168229_2	PI-168229	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
285_PI173587_1	PI-173587	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
286_PI173587_2	PI-173587	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
201_PI173589_1	PI-173589	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
89_PI201992_1	PI-201992	Secale cereale	subsp. cereale	Germany	Hakan2019	GBS	DEU	feral:nonbrittle
270_PI205221_1	PI-205221	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
271_PI205221_2	PI-205221	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
282_PI218110_1	PI-218110	Secale cereale	subsp. cereale	Pakistan	Hakan2019	GBS	PAK	feral:nonbrittle
283_PI218110_2	PI-218110	Secale cereale	subsp. cereale	Pakistan	Hakan2019	GBS	PAK	feral:nonbrittle
101_PI220118_1	PI-220118	Secale cereale	subsp. cereale	Afghanistan	Hakan2019	GBS	AFG	feral:nonbrittle
102_PI220118_2	PI-220118	Secale cereale	subsp. cereale	Afghanistan	Hakan2019	GBS	AFG	feral:nonbrittle
187_PI227870_1	PI-227870	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
188_PI228360_1	PI-228360	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
115_PI234655_1	PI-234655	Secale cereale	subsp. cereale	Kazakhstan	Hakan2019	GBS	KAZ	feral:brittle
116_PI234655_2	PI-234655	Secale cereale	subsp. cereale	Kazakhstan	Hakan2019	GBS	KAZ	feral:nonbrittle
189_PI243741_1	PI-243741	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
190_PI243741_2	PI-243741	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
280_PI250744_1	PI-250744	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
281_PI250744_2	PI-250744	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
178_PI250745_1	PI-250745	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
15_PI250746_1	PI-250746	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
16_PI250746_2	PI-250746	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
194_PI250747_1	PI-250747	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
195_PI250885_1	PI-250885	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
129_PI252002_1	PI-252002	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
81_PI253957_1	PI-253957	Secale cereale	subsp. cereale	Afghanistan	Hakan2019	GBS	AFG	feral:brittle
82_PI253957_2	PI-253957	Secale cereale	subsp. cereale	Afghanistan	Hakan2019	GBS	AFG	feral:brittle
244_PI254811_1	PI-254811	Secale cereale	subsp. cereale	Austria	Hakan2019	GBS	AUT	weed
245_PI254811_2	PI-254811	Secale cereale	subsp. cereale	Austria	Hakan2019	GBS	AUT	domesticated:middle
87_PI263561_1	PI-263561	Secale cereale	subsp. cereale	Switzerland	Hakan2019	GBS	CHE	feral:nonbrittle
88_PI265470_1	PI-265470	Secale cereale	subsp. cereale	Finland	Hakan2019	GBS	FIN	feral:nonbrittle
279_PI267102_1	PI-267102	Secale cereale	subsp. cereale	Azerbaijan	Hakan2019	GBS	AZE	feral:nonbrittle
196_PI268281_1	PI-268281	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
163_PI269397_1	PI-269397	Secale cereale	subsp. cereale	Afghanistan	Hakan2019	GBS	AFG	feral:nonbrittle
164_PI269964_1	PI-269964	Secale cereale	subsp. cereale	Pakistan	Hakan2019	GBS	PAK	feral:nonbrittle
165_PI269964_2	PI-269964	Secale cereale	subsp. cereale	Pakistan	Hakan2019	GBS	PAK	feral:brittle
242_PI272333_1	PI-272333	Secale cereale	subsp. cereale	Hungary	Hakan2019	GBS	HUN	feral:nonbrittle
243_PI272333_2	PI-272333	Secale cereale	subsp. cereale	Hungary	Hakan2019	GBS	HUN	feral:nonbrittle
72_PI272338_1	PI-272338	Secale cereale	subsp. cereale	Hungary	Hakan2019	GBS	HUN	feral:nonbrittle
73_PI272338_2	PI-272338	Secale cereale	subsp. cereale	Hungary	Hakan2019	GBS	HUN	feral:nonbrittle
332_PI284842_1	PI-284842	Secale cereale	subsp. cereale	Hungary	Hakan2019	GBS	HUN	feral:nonbrittle
333_PI284842_2	PI-284842	Secale cereale	subsp. cereale	Hungary	Hakan2019	GBS	HUN	feral:nonbrittle
197_PI289814_1	PI-289814	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
256_PI323450_1	PI-323450	Secale cereale	subsp. cereale	Poland	Hakan2019	GBS	POL	feral:nonbrittle
257_PI323450_2	PI-323450	Secale cereale	subsp. cereale	Poland	Hakan2019	GBS	POL	feral:nonbrittle
327_PI326284_1	PI-326284	Secale cereale	subsp. cereale	Azerbaijan	Hakan2019	GBS	AZE	domesticated:bottom
328_PI326284_2	PI-326284	Secale cereale	subsp. cereale	Azerbaijan	Hakan2019	GBS	AZE	feral:nonbrittle
329_PI326286_1	PI-326286	Secale cereale	subsp. cereale	Kazakhstan	Hakan2019	GBS	KAZ	feral:nonbrittle
179_PI330526_1	PI-330526	Secale cereale	subsp. cereale	England	Hakan2019	GBS	GBR	feral:nonbrittle
180_PI330526_2	PI-330526	Secale cereale	subsp. cereale	England	Hakan2019	GBS	GBR	feral:nonbrittle
58_PI334516_1	PI-334516	Secale cereale	subsp. cereale	England	Hakan2019	GBS	GBR	domesticated:middle
59_PI334516_2	PI-334516	Secale cereale	subsp. cereale	England	Hakan2019	GBS	GBR	domesticated:middle
309_PI344970_1	PI-344970	Secale cereale	subsp. cereale	Bosnia and Herzegovina	Hakan2019	GBS	BIH	feral:nonbrittle
310_PI344970_2	PI-344970	Secale cereale	subsp. cereale	Bosnia and Herzegovina	Hakan2019	GBS	BIH	feral:nonbrittle
176_PI344975_1	PI-344975	Secale cereale	subsp. cereale	Montenegro	Hakan2019	GBS	MNE	feral:nonbrittle
177_PI344975_2	PI-344975	Secale cereale	subsp. cereale	Montenegro	Hakan2019	GBS	MNE	feral:nonbrittle
174_PI344979_1	PI-344979	Secale cereale	subsp. cereale	Montenegro	Hakan2019	GBS	MNE	feral:nonbrittle
175_PI344979_2	PI-344979	Secale cereale	subsp. cereale	Montenegro	Hakan2019	GBS	MNE	feral:nonbrittle
173_PI344987_1	PI-344987	Secale cereale	subsp. cereale	Serbia	Hakan2019	GBS	SRB	feral:nonbrittle
172_PI344990_1	PI-344990	Secale cereale	subsp. cereale	Serbia	Hakan2019	GBS	SRB	feral:nonbrittle
61_PI405812_1	PI-405812	Secale cereale	subsp. cereale	Macedonia	Hakan2019	GBS	MKD	feral:nonbrittle
160_PI410534_1	PI-410534	Secale cereale	subsp. cereale	Pakistan	Hakan2019	GBS	PAK	feral:nonbrittle
161_PI410534_2	PI-410534	Secale cereale	subsp. cereale	Pakistan	Hakan2019	GBS	PAK	feral:nonbrittle
303_PI429372_1	PI-429372	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
304_PI429372_2	PI-429372	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
182_PI429373_1	PI-429373	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
183_PI429373_2	PI-429373	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle

Table continued on next page

## 7. Appendix

---

54_PI429376_1	PI-429376	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
321_PI429378_1	PI-429378	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
322_PI429378_2	PI-429378	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
193_PI429379_1	PI-429379	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	feral:nonbrittle
156_PI430003_1	PI-430003	Secale cereale	subsp. cereale	India	Hakan2019	GBS	IND	domesticated:bottom
157_PI430003_2	PI-430003	Secale cereale	subsp. cereale	India	Hakan2019	GBS	IND	feral:nonbrittle
162_PI430004_1	PI-430004	Secale cereale	subsp. cereale	India	Hakan2019	GBS	IND	feral:nonbrittle
76_PI436165_1	PI-436165	Secale cereale	subsp. cereale	NA	Hakan2019	GBS	NA	feral:nonbrittle
268_PI445880_1	PI-445880	Secale cereale	subsp. cereale	NA	Hakan2019	GBS	NA	feral:nonbrittle
6_PI_4459777_1	PI-4459777	Secale cereale	subsp. cereale	Israel	Hakan2019	GBS	ISR	feral:nonbrittle
7_PI_4459777_2	PI-4459777	Secale cereale	subsp. cereale	Israel	Hakan2019	GBS	ISR	feral:nonbrittle
8_PI_4459777_3	PI-4459777	Secale cereale	subsp. cereale	Israel	Hakan2019	GBS	ISR	feral:nonbrittle
278_PI445980_1	PI-445980	Secale cereale	subsp. cereale	Israel	Hakan2019	GBS	ISR	feral:nonbrittle
30_PI445981_1	PI-445981	Secale cereale	subsp. cereale	Israel	Hakan2019	GBS	ISR	feral:nonbrittle
31_PI445981_2	PI-445981	Secale cereale	subsp. cereale	Israel	Hakan2019	GBS	ISR	feral:nonbrittle
60_PI445996_1	PI-445996	Secale cereale	subsp. cereale	Sweden	Hakan2019	GBS	SWE	feral:nonbrittle
325_PI452132_1	PI-452132	Secale cereale	subsp. cereale	China	Hakan2019	GBS	CHN	feral:nonbrittle
326_PI452132_2	PI-452132	Secale cereale	subsp. cereale	China	Hakan2019	GBS	CHN	feral:nonbrittle
79_PI531829_1	PI-531829	Secale cereale	subsp. cereale	Armeniaá	Hakan2019	GBS	ARM	feral:nonbrittle
80_PI531829_2	PI-531829	Secale cereale	subsp. cereale	Armeniaá	Hakan2019	GBS	ARM	feral:nonbrittle
313_PI535147_1	PI-535147	Secale cereale	subsp. cereale	Yugoslavia	Hakan2019	GBS	SRB	domesticated:middle
314_PI535147_2	PI-535147	Secale cereale	subsp. cereale	Yugoslavia	Hakan2019	GBS	SRB	domesticated:middle
202_PI561793_1	PI-561793	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
203_PI561794_1	PI-561794	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
204_PI561794_2	PI-561794	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
127_PI561795_1	PI-561795	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
128_PI561795_2	PI-561795	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
136_PI561796_1	PI-561796	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
137_PI561796_2	PI-561796	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
124_PI561797_1	PI-561797	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
125_PI561797_2	PI-561797	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
92_PI561798_1	PI-561798	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
171_PI561799_1	PI-561799	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
117_PI561801_1	PI-561801	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
248_PI561802_1	PI-561802	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
249_PI561802_2	PI-561802	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
192_PI561803_1	PI-561803	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
319_PI561804_1	PI-561804	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
320_PI561804_2	PI-561804	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
131_PI561806_2	PI-561806	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
121_PI561807_1	PI-561807	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
4_PI_561809_1	PI-561809	Secale cereale	subsp. cereale	Pakistan	Hakan2019	GBS	PAK	feral:brittle
5_PI_561809_2	PI-561809	Secale cereale	subsp. cereale	Pakistan	Hakan2019	GBS	PAK	feral:brittle
184_PI561810_1	PI-561810	Secale cereale	subsp. cereale	Pakistan	Hakan2019	GBS	PAK	feral:brittle
207_PI568118_1	PI-568118	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
208_PI568118_2	PI-568118	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	weed
212_PI568120_1	PI-568120	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
213_PI568120_2	PI-568120	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
74_PI573648_1	PI-573648	Secale cereale	subsp. cereale	Russian Federation	Hakan2019	GBS	RUS	feral:nonbrittle
75_PI573648_2	PI-573648	Secale cereale	subsp. cereale	Russian Federation	Hakan2019	GBS	RUS	feral:nonbrittle
330_PI573649_1	PI-573649	Secale cereale	subsp. cereale	Afghanistan	Hakan2019	GBS	AFG	feral:nonbrittle
331_PI573649_2	PI-573649	Secale cereale	subsp. cereale	Afghanistan	Hakan2019	GBS	AFG	feral:nonbrittle
185_PI578092_1	PI-578092	Secale cereale	subsp. cereale	Pakistan	Hakan2019	GBS	PAK	feral:nonbrittle
181_PI584781_1	PI-584781	Secale cereale	subsp. cereale	Georgia	Hakan2019	GBS	GEO	feral:nonbrittle
186_PI584782_1	PI-584782	Secale cereale	subsp. cereale	Georgia	Hakan2019	GBS	GEO	feral:nonbrittle
20_PI618666_1	PI-618666	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
21_PI618666_2	PI-618666	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
269_PI618671_1	PI-618671	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	feral:nonbrittle
165	R100	Secale cereale	subsp. cereale	Germany	GBS_Schreiber2018	GBS	DEU	domesticated:middle
166	R100	Secale cereale	subsp. cereale	Germany	GBS_Schreiber2018	GBS	DEU	domesticated:top
167	R100	Secale cereale	subsp. cereale	Germany	GBS_Schreiber2018	GBS	DEU	domesticated:top
168	R100	Secale cereale	subsp. cereale	Germany	GBS_Schreiber2018	GBS	DEU	domesticated:top
169	R100	Secale cereale	subsp. cereale	Germany	GBS_Schreiber2018	GBS	DEU	domesticated:top
170	R100	Secale cereale	subsp. cereale	Germany	GBS_Schreiber2018	GBS	DEU	domesticated:top
201	R1001	Secale cereale	subsp. cereale	Switzerland	GBS_Schreiber2018	GBS	CHE	feral:nonbrittle
202	R1001	Secale cereale	subsp. cereale	Switzerland	GBS_Schreiber2018	GBS	CHE	feral:nonbrittle
203	R1001	Secale cereale	subsp. cereale	Switzerland	GBS_Schreiber2018	GBS	CHE	feral:nonbrittle

Table continued on next page

## 7. Appendix

Table continued on next page

## 7. Appendix

## 7. Appendix

---

213	R831	Secale cereale	subsp. cereale	Slovakia	GBS_Schreiber2018	GBS	SVK	domesticated:top
214	R831	Secale cereale	subsp. cereale	Slovakia	GBS_Schreiber2018	GBS	SVK	domesticated:top
215	R831	Secale cereale	subsp. cereale	Slovakia	GBS_Schreiber2018	GBS	SVK	domesticated:top
216	R831	Secale cereale	subsp. cereale	Slovakia	GBS_Schreiber2018	GBS	SVK	domesticated:middle
217	R831	Secale cereale	subsp. cereale	Slovakia	GBS_Schreiber2018	GBS	SVK	domesticated:middle
218	R831	Secale cereale	subsp. cereale	Slovakia	GBS_Schreiber2018	GBS	SVK	domesticated:top
225	R864	Secale cereale	subsp. cereale	Georgia	GBS_Schreiber2018	GBS	GEO	feral:nonbrittle
226	R864	Secale cereale	subsp. cereale	Georgia	GBS_Schreiber2018	GBS	GEO	feral:nonbrittle
227	R864	Secale cereale	subsp. cereale	Georgia	GBS_Schreiber2018	GBS	GEO	feral:nonbrittle
228	R864	Secale cereale	subsp. cereale	Georgia	GBS_Schreiber2018	GBS	GEO	feral:nonbrittle
229	R864	Secale cereale	subsp. cereale	Georgia	GBS_Schreiber2018	GBS	GEO	feral:nonbrittle
230	R864	Secale cereale	subsp. cereale	Georgia	GBS_Schreiber2018	GBS	GEO	feral:nonbrittle
255	R875	Secale cereale	subsp. cereale	Austria	GBS_Schreiber2018	GBS	AUT	domesticated:middle
256	R875	Secale cereale	subsp. cereale	Austria	GBS_Schreiber2018	GBS	AUT	domesticated:bottom
257	R875	Secale cereale	subsp. cereale	Austria	GBS_Schreiber2018	GBS	AUT	domesticated:bottom
258	R875	Secale cereale	subsp. cereale	Austria	GBS_Schreiber2018	GBS	AUT	domesticated:bottom
259	R875	Secale cereale	subsp. cereale	Austria	GBS_Schreiber2018	GBS	AUT	domesticated:middle
260	R875	Secale cereale	subsp. cereale	Austria	GBS_Schreiber2018	GBS	AUT	domesticated:bottom
115	R91	Secale cereale	subsp. cereale	Poland	GBS_Schreiber2018	GBS	POL	domesticated:middle
116	R91	Secale cereale	subsp. cereale	Poland	GBS_Schreiber2018	GBS	POL	domesticated:top
117	R91	Secale cereale	subsp. cereale	Poland	GBS_Schreiber2018	GBS	POL	domesticated:middle
118	R91	Secale cereale	subsp. cereale	Poland	GBS_Schreiber2018	GBS	POL	domesticated:middle
119	R91	Secale cereale	subsp. cereale	Poland	GBS_Schreiber2018	GBS	POL	domesticated:top
120	R91	Secale cereale	subsp. cereale	Poland	GBS_Schreiber2018	GBS	POL	domesticated:middle
595	R93	Secale cereale	subsp. cereale	Germany	GBS_Schreiber2018	GBS	DEU	domesticated:middle
596	R93	Secale cereale	subsp. cereale	Germany	GBS_Schreiber2018	GBS	DEU	domesticated:top
597	R93	Secale cereale	subsp. cereale	Germany	GBS_Schreiber2018	GBS	DEU	domesticated:middle
598	R93	Secale cereale	subsp. cereale	Germany	GBS_Schreiber2018	GBS	DEU	domesticated:middle
599	R93	Secale cereale	subsp. cereale	Germany	GBS_Schreiber2018	GBS	DEU	domesticated:top
600	R93	Secale cereale	subsp. cereale	Germany	GBS_Schreiber2018	GBS	DEU	domesticated:top
537	R937	Secale cereale	subsp. cereale	Italy	GBS_Schreiber2018	GBS	ITA	domesticated:middle
538	R937	Secale cereale	subsp. cereale	Italy	GBS_Schreiber2018	GBS	ITA	domesticated:middle
539	R937	Secale cereale	subsp. cereale	Italy	GBS_Schreiber2018	GBS	ITA	domesticated:top
540	R937	Secale cereale	subsp. cereale	Italy	GBS_Schreiber2018	GBS	ITA	domesticated:top
541	R937	Secale cereale	subsp. cereale	Italy	GBS_Schreiber2018	GBS	ITA	domesticated:middle
542	R937	Secale cereale	subsp. cereale	Italy	GBS_Schreiber2018	GBS	ITA	domesticated:top
465	R944	Secale cereale	subsp. cereale	Italy	GBS_Schreiber2018	GBS	ITA	domesticated:top
466	R944	Secale cereale	subsp. cereale	Italy	GBS_Schreiber2018	GBS	ITA	domesticated:top
467	R944	Secale cereale	subsp. cereale	Italy	GBS_Schreiber2018	GBS	ITA	domesticated:top
468	R944	Secale cereale	subsp. cereale	Italy	GBS_Schreiber2018	GBS	ITA	domesticated:top
469	R944	Secale cereale	subsp. cereale	Italy	GBS_Schreiber2018	GBS	ITA	domesticated:top
470	R944	Secale cereale	subsp. cereale	Italy	GBS_Schreiber2018	GBS	ITA	domesticated:top
561	R972	Secale cereale	subsp. cereale	Georgia	GBS_Schreiber2018	GBS	GEO	domesticated:middle
562	R972	Secale cereale	subsp. cereale	Georgia	GBS_Schreiber2018	GBS	GEO	domesticated:middle
563	R972	Secale cereale	subsp. cereale	Georgia	GBS_Schreiber2018	GBS	GEO	domesticated:middle
564	R972	Secale cereale	subsp. cereale	Georgia	GBS_Schreiber2018	GBS	GEO	domesticated:middle
565	R972	Secale cereale	subsp. cereale	Georgia	GBS_Schreiber2018	GBS	GEO	domesticated:middle
566	R972	Secale cereale	subsp. cereale	Georgia	GBS_Schreiber2018	GBS	GEO	domesticated:middle
Weining_rye_1	Weining	Secale cereale	subsp. cereale	China	Weining2018	GBS	CHN	feral:nonbrittle
Weining_rye_2	Weining	Secale cereale	subsp. cereale	China	Weining2018	GBS	CHN	domesticated:bottom
Weining_rye_3	Weining	Secale cereale	subsp. cereale	China	Weining2018	GBS	CHN	feral:nonbrittle
Weining_rye_4	Weining	Secale cereale	subsp. cereale	China	Weining2018	GBS	CHN	feral:nonbrittle
Weining_rye_5	Weining	Secale cereale	subsp. cereale	China	Weining2018	GBS	CHN	domesticated:bottom
Weining_rye_6	Weining	Secale cereale	subsp. cereale	China	Weining2018	GBS	CHN	feral:nonbrittle
403_2004_6_29_10_4_1	2004-6-29-10-4	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	
297_2004_6_29_6_5_1	2004-6-29-6-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	
298_2004_6_29_6_5_2	2004-6-29-6-5	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	
57_2006_6_25_1_1	2006-6-25-1-1	Secale cereale	subsp. cereale	Turkey	Hakan2019	GBS	TUR	
250_CISE38_1	Cise_38	Secale cereale	subsp. cereale	USA	Hakan2019	GBS	USA	
251_CISE38_2	Cise_38	Secale cereale	subsp. cereale	USA	Hakan2019	GBS	USA	
364_CISE38_1	Cise_38	Secale cereale	subsp. cereale	USA	Hakan2019	GBS	USA	
109_CSE10_1	CSE_10	Secale cereale	subsp. cereale	USA	Hakan2019	GBS	USA	
110_CSE10_2	CSE_10	Secale cereale	subsp. cereale	USA	Hakan2019	GBS	USA	
113_CSE108_1	CSE_108	Secale cereale	subsp. cereale	Japan	Hakan2019	GBS	JPN	
114_CSE108_2	CSE_108	Secale cereale	subsp. cereale	Japan	Hakan2019	GBS	JPN	
103_CSE110_1	CSE_110	Secale cereale	subsp. cereale	South Korea	Hakan2019	GBS	KOR	
104_CSE110_2	CSE_110	Secale cereale	subsp. cereale	South Korea	Hakan2019	GBS	KOR	

Table continued on next page

## 7. Appendix

---

99_CSE12_1	CSE_12	Secale cereale	subsp. cereale	USA	Hakan2019	GBS	USA	
144_CSE185_1	CSE_185	Secale cereale	subsp. cereale	Canada	Hakan2019	GBS	CAN	
145_CSE185_2	CSE_185	Secale cereale	subsp. cereale	Canada	Hakan2019	GBS	CAN	
111_CSE35_1	CSE_35	Secale cereale	subsp. cereale	USA	Hakan2019	GBS	USA	
112_CSE35_2	CSE_35	Secale cereale	subsp. cereale	USA	Hakan2019	GBS	USA	
158_CSE79_1	CSE_79	Secale cereale	subsp. cereale	Australia	Hakan2019	GBS	AUS	
159_CSE79_2	CSE_79	Secale cereale	subsp. cereale	Australia	Hakan2019	GBS	AUS	
Lo7_1	Lo7	NA	NA	NA	Hakan2019	GBS	NA	
Lo7_2	Lo7	NA	NA	NA	Hakan2019	GBS	NA	
Lo7_3	Lo7	NA	NA	NA	Hakan2019	GBS	NA	
Lo7_4	Lo7	NA	NA	NA	Hakan2019	GBS	NA	
Lo7_5	Lo7	NA	NA	NA	Hakan2019	GBS	NA	
96_PI20522_1	PI-20522	Secale strictum	NA	Turkey	Hakan2019	GBS	TUR	
97_PI20522_2	PI-20522	Secale strictum	NA	Turkey	Hakan2019	GBS	TUR	
169_PI237927_1	PI-237927	Secale cereale	subsp. cereale	Brazil	Hakan2019	GBS	BRA	
170_PI237927_2	PI-237927	Secale cereale	subsp. cereale	Brazil	Hakan2019	GBS	BRA	
95_PI240286_1	PI-240286	Secale strictum	NA	Turkey	Hakan2019	GBS	TUR	
191_PI240675_1	PI-240675	Secale cereale	subsp. cereale	Uruguay	Hakan2019	GBS	URY	
100_PI261400_1	PI-261400	Secale cereale	subsp. cereale	Canada	Hakan2019	GBS	CAN	
307_PI261401_1	PI-261401	Secale cereale	subsp. cereale	Canada	Hakan2019	GBS	CAN	
308_PI261401_2	PI-261401	Secale cereale	subsp. cereale	Canada	Hakan2019	GBS	CAN	
266_PI314964_1	PI-314964	Secale cereale	subsp. cereale	Brazil	Hakan2019	GBS	BRA	
267_PI314964_2	PI-314964	Secale cereale	subsp. cereale	Brazil	Hakan2019	GBS	BRA	
146_PI330407_1	PI-330407	Secale cereale	subsp. cereale	South Africaá	Hakan2019	GBS	ZAF	
147_PI330407_2	PI-330407	Secale cereale	subsp. cereale	South Africaá	Hakan2019	GBS	ZAF	
93_PI345739_1	PI-345739	Secale cereale	subsp. cereale	Australia	Hakan2019	GBS	AUS	
94_PI345739_2	PI-345739	Secale cereale	subsp. cereale	Australia	Hakan2019	GBS	AUS	
152_PI345740_1	PI-345740	Secale cereale	subsp. cereale	Australia	Hakan2019	GBS	AUS	
153_PI345740_2	PI-345740	Secale cereale	subsp. cereale	Australia	Hakan2019	GBS	AUS	
284_PI401401_1	PI-401401	Secale strictum	NA	Iran	Hakan2019	GBS	IRN	
32_PI401402_1	PI-401402	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	
77_PI401404_1	PI-401404	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	
78_PI401404_2	PI-401404	Secale cereale	subsp. cereale	Iran	Hakan2019	GBS	IRN	
1_PI_436165_1	PI-436165	Secale cereale	subsp. cereale	Chile	Hakan2019	GBS	CHL	
85_PI436190_1	PI-436190	Secale cereale	subsp. cereale	Chile	Hakan2019	GBS	CHL	
86_PI436190_2	PI-436190	Secale cereale	subsp. cereale	Chile	Hakan2019	GBS	CHL	
83_PI436192_1	PI-436192	Secale cereale	subsp. cereale	Chile	Hakan2019	GBS	CHL	
84_PI436192_2	PI-436192	Secale cereale	subsp. cereale	Chile	Hakan2019	GBS	CHL	
91_PI445973_1	PI-445973	Secale cereale	subsp. cereale	USA	Hakan2019	GBS	USA	
272_PI445998_1	PI-445998	Secale cereale	subsp. cereale	Canada	Hakan2019	GBS	CAN	
273_PI445998_2	PI-445998	Secale cereale	subsp. cereale	Canada	Hakan2019	GBS	CAN	
62_PI445999_1	PI-445999	Secale cereale	subsp. cereale	Canada	Hakan2019	GBS	CAN	
63_PI445999_2	PI-445999	Secale cereale	subsp. cereale	Canada	Hakan2019	GBS	CAN	
67_PI446022_1	PI-446022	Secale cereale	subsp. cereale	Mexico	Hakan2019	GBS	MEX	
68_PI446022_2	PI-446022	Secale cereale	subsp. cereale	Mexico	Hakan2019	GBS	MEX	
69_PI446022_3	PI-446022	Secale cereale	subsp. cereale	Mexico	Hakan2019	GBS	MEX	
107_PI446023_1	PI-446023	Secale cereale	subsp. cereale	Mexico	Hakan2019	GBS	MEX	
108_PI446023_2	PI-446023	Secale cereale	subsp. cereale	Mexico	Hakan2019	GBS	MEX	
142_PI446027_1	PI-446027	Secale cereale	subsp. cereale	New Zealandá	Hakan2019	GBS	NZL	
143_PI446027_2	PI-446027	Secale cereale	subsp. cereale	New Zealandá	Hakan2019	GBS	NZL	
22_PI535199_1	PI-535199	Secale cereale	subsp. cereale	USA	Hakan2019	GBS	USA	
23_PI535199_2	PI-535199	Secale cereale	subsp. cereale	USA	Hakan2019	GBS	USA	
362_PI535199_1	PI-535199	Secale cereale	subsp. cereale	USA	Hakan2019	GBS	USA	
363_PI535199_2	PI-535199	Secale cereale	subsp. cereale	USA	Hakan2019	GBS	USA	
98_PI543398_1	PI-543398	Secale cereale	subsp. cereale	Argentina	Hakan2019	GBS	ARG	
274_PI568257_1	PI-568257	Secale cereale	subsp. cereale	Russian Federationá	Hakan2019	GBS	RUS	
275_PI568257_2	PI-568257	Secale cereale	subsp. cereale	Russian Federationá	Hakan2019	GBS	RUS	
R1003_10A	R1003	Secale vavilovii	NA	Armenia	GBS_R1003_R1572_additional	GBS	ARM	
R1003_10B	R1003	Secale vavilovii	NA	Armenia	GBS_R1003_R1572_additional	GBS	ARM	
R1003_10C	R1003	Secale vavilovii	NA	Armenia	GBS_R1003_R1572_additional	GBS	ARM	
R1003_10D	R1003	Secale vavilovii	NA	Armenia	GBS_R1003_R1572_additional	GBS	ARM	
R1003_10E	R1003	Secale vavilovii	NA	Armenia	GBS_R1003_R1572_additional	GBS	ARM	
R1003_10F	R1003	Secale vavilovii	NA	Armenia	GBS_R1003_R1572_additional	GBS	ARM	
R1003_10G	R1003	Secale vavilovii	NA	Armenia	GBS_R1003_R1572_additional	GBS	ARM	
R1003_10H	R1003	Secale vavilovii	NA	Armenia	GBS_R1003_R1572_additional	GBS	ARM	
R1003_11A	R1003	Secale vavilovii	NA	Armenia	GBS_R1003_R1572_additional	GBS	ARM	
R1003_11B	R1003	Secale vavilovii	NA	Armenia	GBS_Schreiber2018	GBS	ARM	

Table continued on next page

## 7. Appendix

## 7. Appendix

---

R1063_2	R1063	Secale vavilovii	NA	Poland	GBS_new_samples_2018	GBS	POL	
R1063_3	R1063	Secale vavilovii	NA	Poland	GBS_new_samples_2018	GBS	POL	
R1063_4	R1063	Secale vavilovii	NA	Poland	GBS_new_samples_2018	GBS	POL	
R1063_5	R1063	Secale vavilovii	NA	Poland	GBS_new_samples_2018	GBS	POL	
R1063_6	R1063	Secale vavilovii	NA	Poland	GBS_new_samples_2018	GBS	POL	
R1064_1	R1064	Secale vavilovii	NA	Poland	GBS_new_samples_2018	GBS	POL	
R1064_2	R1064	Secale vavilovii	NA	Poland	GBS_new_samples_2018	GBS	POL	
R1064_3	R1064	Secale vavilovii	NA	Poland	GBS_new_samples_2018	GBS	POL	
R1064_4	R1064	Secale vavilovii	NA	Poland	GBS_new_samples_2018	GBS	POL	
R1064_5	R1064	Secale vavilovii	NA	Poland	GBS_new_samples_2018	GBS	POL	
R1064_6	R1064	Secale vavilovii	NA	Poland	GBS_new_samples_2018	GBS	POL	
R1108_1	R1108	Secale strictum	subsp. kuprijanovii	Kazakhstan	GBS_new_samples_2018	GBS	KAZ	
R1108_2	R1108	Secale strictum	subsp. kuprijanovii	Kazakhstan	GBS_new_samples_2018	GBS	KAZ	
R1108_3	R1108	Secale strictum	subsp. kuprijanovii	Kazakhstan	GBS_new_samples_2018	GBS	KAZ	
R1108_4	R1108	Secale strictum	subsp. kuprijanovii	Kazakhstan	GBS_new_samples_2018	GBS	KAZ	
R1108_5	R1108	Secale strictum	subsp. kuprijanovii	Kazakhstan	GBS_new_samples_2018	GBS	KAZ	
R1108_6	R1108	Secale strictum	subsp. kuprijanovii	Kazakhstan	GBS_new_samples_2018	GBS	KAZ	
R1119_1	R1119	Secale strictum	subsp. anatolicum	Turkey	GBS_new_samples_2018	GBS	TUR	
R1119_2	R1119	Secale strictum	subsp. anatolicum	Turkey	GBS_new_samples_2018	GBS	TUR	
R1119_3	R1119	Secale strictum	subsp. anatolicum	Turkey	GBS_new_samples_2018	GBS	TUR	
R1119_4	R1119	Secale strictum	subsp. anatolicum	Turkey	GBS_new_samples_2018	GBS	TUR	
R1119_5	R1119	Secale strictum	subsp. anatolicum	Turkey	GBS_new_samples_2018	GBS	TUR	
R1119_6	R1119	Secale strictum	subsp. anatolicum	Turkey	GBS_new_samples_2018	GBS	TUR	
R1125_1	R1125	Secale vavilovii	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R1125_2	R1125	Secale vavilovii	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R1125_3	R1125	Secale vavilovii	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R1125_4	R1125	Secale vavilovii	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R1125_5	R1125	Secale vavilovii	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R1125_6	R1125	Secale vavilovii	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R1127_1	R1127	Secale vavilovii	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R1127_2	R1127	Secale vavilovii	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R1127_3	R1127	Secale vavilovii	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R1127_4	R1127	Secale vavilovii	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R1127_5	R1127	Secale vavilovii	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R1127_6	R1127	Secale vavilovii	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R1128_1	R1128	Secale vavilovii	NA	Soviet Union	GBS_new_samples_2018	GBS	RUS	
R1128_2	R1128	Secale vavilovii	NA	Soviet Union	GBS_new_samples_2018	GBS	RUS	
R1128_3	R1128	Secale vavilovii	NA	Soviet Union	GBS_new_samples_2018	GBS	RUS	
R1128_4	R1128	Secale vavilovii	NA	Soviet Union	GBS_new_samples_2018	GBS	RUS	
R1128_5	R1128	Secale vavilovii	NA	Soviet Union	GBS_new_samples_2018	GBS	RUS	
R1128_6	R1128	Secale vavilovii	NA	Soviet Union	GBS_new_samples_2018	GBS	RUS	
567	R1134	Secale strictum	former cereale	Switzerland	GBS_Schreiber2018	GBS	CHE	
568	R1134	Secale strictum	former cereale	Switzerland	GBS_Schreiber2018	GBS	CHE	
569	R1134	Secale strictum	former cereale	Switzerland	GBS_Schreiber2018	GBS	CHE	
570	R1134	Secale strictum	former cereale	Switzerland	GBS_Schreiber2018	GBS	CHE	
571	R1134	Secale strictum	former cereale	Switzerland	GBS_Schreiber2018	GBS	CHE	
572	R1134	Secale strictum	former cereale	Switzerland	GBS_Schreiber2018	GBS	CHE	
233_R1152_1	R1152	Secale vavilovii	NA	NA	Hakan2019	GBS	NA	
R1156_1	R1156	Secale vavilovii	NA	Soviet Union	GBS_new_samples_2018	GBS	RUS	
R1156_2	R1156	Secale vavilovii	NA	Soviet Union	GBS_new_samples_2018	GBS	RUS	
R1156_3	R1156	Secale vavilovii	NA	Soviet Union	GBS_new_samples_2018	GBS	RUS	
R1156_4	R1156	Secale vavilovii	NA	Soviet Union	GBS_new_samples_2018	GBS	RUS	
R1156_5	R1156	Secale vavilovii	NA	Soviet Union	GBS_new_samples_2018	GBS	RUS	
R1156_6	R1156	Secale vavilovii	NA	Soviet Union	GBS_new_samples_2018	GBS	RUS	
R1184_1	R1184	Secale strictum	NA	Argentina	GBS_new_samples_2018	GBS	ARG	
R1184_2	R1184	Secale strictum	NA	Argentina	GBS_new_samples_2018	GBS	ARG	
R1184_3	R1184	Secale strictum	NA	Argentina	GBS_new_samples_2018	GBS	ARG	
R1184_4	R1184	Secale strictum	NA	Argentina	GBS_new_samples_2018	GBS	ARG	
R1184_5	R1184	Secale strictum	NA	Argentina	GBS_new_samples_2018	GBS	ARG	
R1184_6	R1184	Secale strictum	NA	Argentina	GBS_new_samples_2018	GBS	ARG	
R1210_1	R1210	Secale strictum	subsp. africanum	South Africa	GBS_new_samples_2018	GBS	ZAF	
R1210_2	R1210	Secale strictum	subsp. africanum	South Africa	GBS_new_samples_2018	GBS	ZAF	
R1210_3	R1210	Secale strictum	subsp. africanum	South Africa	GBS_new_samples_2018	GBS	ZAF	
R1210_4	R1210	Secale strictum	subsp. africanum	South Africa	GBS_new_samples_2018	GBS	ZAF	
R1210_5	R1210	Secale strictum	subsp. africanum	South Africa	GBS_new_samples_2018	GBS	ZAF	
R1210_6	R1210	Secale strictum	subsp. africanum	South Africa	GBS_new_samples_2018	GBS	ZAF	
R1572_1A	R1572	Secale cereale	subsp. cereale	Germany	GBS_R1003_R1572_additional	GBS	DEU	

Table continued on next page

## 7. Appendix

## 7. Appendix

---

156	R2431	Secale strictum	subsp. strictum	Bulgaria	GBS_Schreiber2018	GBS	BGR	
519	R2432	Secale vavilovii x cereale	NA	Afghanistan	GBS_Schreiber2018	GBS	AFG	
520	R2432	Secale vavilovii x cereale	NA	Afghanistan	GBS_Schreiber2018	GBS	AFG	
521	R2432	Secale vavilovii x cereale	NA	Afghanistan	GBS_Schreiber2018	GBS	AFG	
522	R2432	Secale vavilovii x cereale	NA	Afghanistan	GBS_Schreiber2018	GBS	AFG	
523	R2432	Secale vavilovii x cereale	NA	Afghanistan	GBS_Schreiber2018	GBS	AFG	
524	R2432	Secale vavilovii x cereale	NA	Afghanistan	GBS_Schreiber2018	GBS	AFG	
R2433_1	R2433	Secale vavilovii	NA	NA	GBS_new_samples_2018	GBS	NA	
R2433_2	R2433	Secale vavilovii	NA	NA	GBS_new_samples_2018	GBS	NA	
R2433_3	R2433	Secale vavilovii	NA	NA	GBS_new_samples_2018	GBS	NA	
R2433_4	R2433	Secale vavilovii	NA	NA	GBS_new_samples_2018	GBS	NA	
R2433_5	R2433	Secale vavilovii	NA	NA	GBS_new_samples_2018	GBS	NA	
R2433_6	R2433	Secale vavilovii	NA	NA	GBS_new_samples_2018	GBS	NA	
25	R2434	Secale vavilovii x cereale	NA	Russia	GBS_Schreiber2018	GBS	RUS	
26	R2434	Secale vavilovii x cereale	NA	Russia	GBS_Schreiber2018	GBS	RUS	
27	R2434	Secale vavilovii x cereale	NA	Russia	GBS_Schreiber2018	GBS	RUS	
28	R2434	Secale vavilovii x cereale	NA	Russia	GBS_Schreiber2018	GBS	RUS	
29	R2434	Secale vavilovii x cereale	NA	Russia	GBS_Schreiber2018	GBS	RUS	
30	R2434	Secale vavilovii x cereale	NA	Russia	GBS_Schreiber2018	GBS	RUS	
453	R2446	Secale sylvestre	NA	Ukraine	GBS_Schreiber2018	GBS	UKR	
454	R2446	Secale sylvestre	NA	Ukraine	GBS_Schreiber2018	GBS	UKR	
455	R2446	Secale sylvestre	NA	Ukraine	GBS_Schreiber2018	GBS	UKR	
456	R2446	Secale sylvestre	NA	Ukraine	GBS_Schreiber2018	GBS	UKR	
457	R2446	Secale sylvestre	NA	Ukraine	GBS_Schreiber2018	GBS	UKR	
458	R2446	Secale sylvestre	NA	Ukraine	GBS_Schreiber2018	GBS	UKR	
R2456_1	R2456	Secale strictum	subsp. kuprijanovii	Italy	GBS_new_samples_2018	GBS	ITA	
R2456_2	R2456	Secale strictum	subsp. kuprijanovii	Italy	GBS_new_samples_2018	GBS	ITA	
R2456_3	R2456	Secale strictum	subsp. kuprijanovii	Italy	GBS_new_samples_2018	GBS	ITA	
R2456_4	R2456	Secale strictum	subsp. kuprijanovii	Italy	GBS_new_samples_2018	GBS	ITA	
R2456_5	R2456	Secale strictum	subsp. kuprijanovii	Italy	GBS_new_samples_2018	GBS	ITA	
R2456_6	R2456	Secale strictum	subsp. kuprijanovii	Italy	GBS_new_samples_2018	GBS	ITA	
145	R2456	Secale strictum	subsp. kuprijanovii	Italy	GBS_Schreiber2018	GBS	ITA	
146	R2456	Secale strictum	subsp. kuprijanovii	Italy	GBS_Schreiber2018	GBS	ITA	
147	R2456	Secale strictum	subsp. kuprijanovii	Italy	GBS_Schreiber2018	GBS	ITA	
148	R2456	Secale strictum	subsp. kuprijanovii	Italy	GBS_Schreiber2018	GBS	ITA	
149	R2456	Secale strictum	subsp. kuprijanovii	Italy	GBS_Schreiber2018	GBS	ITA	
150	R2456	Secale strictum	subsp. kuprijanovii	Italy	GBS_Schreiber2018	GBS	ITA	
R247_1	R247	Secale vavilovii	NA	NA	GBS_new_samples_2018	GBS	NA	
R247_2	R247	Secale vavilovii	NA	NA	GBS_new_samples_2018	GBS	NA	
R247_3	R247	Secale vavilovii	NA	NA	GBS_new_samples_2018	GBS	NA	
R247_4	R247	Secale vavilovii	NA	NA	GBS_new_samples_2018	GBS	NA	
R247_5	R247	Secale vavilovii	NA	NA	GBS_new_samples_2018	GBS	NA	
R247_6	R247	Secale vavilovii	NA	NA	GBS_new_samples_2018	GBS	NA	
327	R2596	Secale vavilovii x cereale	NA	Iran	GBS_Schreiber2018	GBS	IRN	
328	R2596	Secale vavilovii x cereale	NA	Iran	GBS_Schreiber2018	GBS	IRN	
329	R2596	Secale vavilovii x cereale	NA	Iran	GBS_Schreiber2018	GBS	IRN	
330	R2596	Secale vavilovii x cereale	NA	Iran	GBS_Schreiber2018	GBS	IRN	
331	R2596	Secale vavilovii x cereale	NA	Iran	GBS_Schreiber2018	GBS	IRN	
332	R2596	Secale vavilovii x cereale	NA	Iran	GBS_Schreiber2018	GBS	IRN	
R26_1	R26	Secale x derzhavini	NA	Germany	GBS_new_samples_2018	GBS	DEU	
R26_2	R26	Secale x derzhavini	NA	Germany	GBS_new_samples_2018	GBS	DEU	
R26_3	R26	Secale x derzhavini	NA	Germany	GBS_new_samples_2018	GBS	DEU	
R26_4	R26	Secale x derzhavini	NA	Germany	GBS_new_samples_2018	GBS	DEU	
R26_5	R26	Secale x derzhavini	NA	Germany	GBS_new_samples_2018	GBS	DEU	
R26_6	R26	Secale x derzhavini	NA	Germany	GBS_new_samples_2018	GBS	DEU	
R264_1	R264	Secale strictum	subsp. africanum	South Africa	GBS_new_samples_2018	GBS	ZAF	
R264_2	R264	Secale strictum	subsp. africanum	South Africa	GBS_new_samples_2018	GBS	ZAF	
R264_3	R264	Secale strictum	subsp. africanum	South Africa	GBS_new_samples_2018	GBS	ZAF	
R264_4	R264	Secale strictum	subsp. africanum	South Africa	GBS_new_samples_2018	GBS	ZAF	
R264_5	R264	Secale strictum	subsp. africanum	South Africa	GBS_new_samples_2018	GBS	ZAF	
R264_6	R264	Secale strictum	subsp. africanum	South Africa	GBS_new_samples_2018	GBS	ZAF	
R278_1	R278	Secale strictum	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R278_2	R278	Secale strictum	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R278_3	R278	Secale strictum	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R278_4	R278	Secale strictum	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R278_5	R278	Secale strictum	NA	Turkey	GBS_new_samples_2018	GBS	TUR	
R278_6	R278	Secale strictum	NA	Turkey	GBS_new_samples_2018	GBS	TUR	

Table continued on next page

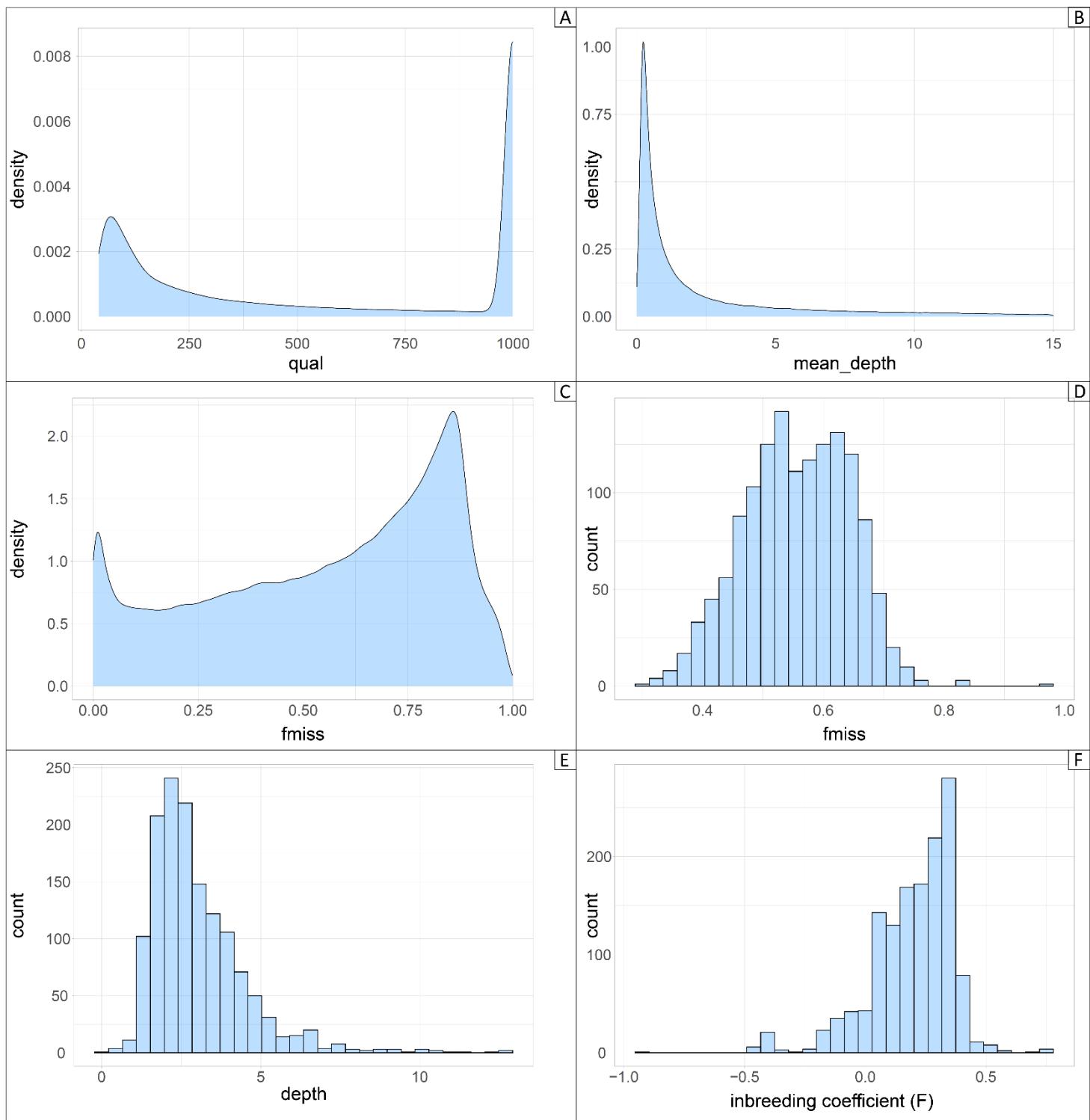
## 7. Appendix

## 7. Appendix

## 7. Appendix

### A3. Unfiltered single nucleotide variant data

**Quality control** for filtering vcf data. Calculated with vcftools (Danecek et al. 2011), visualized in R (version 4.2.3; R Core Team 2023). **A: Variant quality.** (Phred encoded) site quality. Mean depth for each of the variants. Calculated with --site-mean-depth. **B: Variant mean depth.** Calculated with --site-mean-depth. **C: Variant missingness.** Measure of how many individuals lack a genotype at a call site. Calculated with --missing-site. **D: Proportion of missing data per individual.** Calculated with --missing-indv. **E: Mean depth per individual.** Calculated with --depth. **F: Heterozygosity and inbreeding coefficient per individual.** Calculated with --het.



## 7. Appendix

---

### A4. Principal component analysis (PCA) – Variance Proportion

Explanation of variance for the first 32 eigenvectors (EV) for different *Secale* taxa. Variance proportion (all *Secale* taxa) include *S. cereale* subsp. *cereale*, *S. cereale* subsp. *vavilovii*, *S. strictum*, *S. sylvestre*. Variance proportion (*S. cereale*) include *S. cereale* subsp. *cereale*, *S. cereale* subsp. *vavilovii*.

<b>EV</b>	<b>Variance proportion (all <i>Secale</i> taxa)</b>
1	15.46%
2	8.78%
3	2.47%
4	1.76%
5	1.56%
6	1.21%
7	1.09%
8	0.87%
9	0.65%
10	0.59%
11	0.51%
12	0.43%
13	0.41%
14	0.39%
15	0.34%
16	0.33%
17	0.33%
18	0.31%
19	0.31%
20	0.29%
21	0.29%
22	0.28%
23	0.26%
24	0.26%
25	0.26%
26	0.25%
27	0.25%
28	0.24%
29	0.23%
30	0.23%
31	0.22%
32	0.21%

<b>EV</b>	<b>Variance proportion (<i>S. cereale</i>)</b>
1	4.40%
2	2.45%
3	1.17%
4	1.03%
5	0.80%
6	0.77%
7	0.71%
8	0.67%
9	0.61%
10	0.53%
11	0.49%
12	0.45%
13	0.45%
14	0.36%
15	0.35%
16	0.34%
17	0.32%
18	0.32%
19	0.31%
20	0.31%
21	0.30%
22	0.30%
23	0.29%
24	0.28%
25	0.28%
26	0.27%
27	0.27%
28	0.27%
29	0.27%
30	0.26%
31	0.26%
32	0.25%

## 7. Appendix

---

### A5. Calculation of Cross-entropy criterion

Rounded values. Difference = cross-entropy criterion (K) - cross-entropy criterion (K+1). According to Fritchot et al. (2014), the third digit of the cross-entropy criterion can be significant for data sets including 1000 individuals genotyped at over 20,000 SNPs. Significant value marked blue.

numbers of ancestral populations (K)	cross-entropy criterion	Difference
1	0.50500	0.04081
2	0.46419	0.01148
3	0.45271	0.01616
4	0.43655	0.00501
5	0.43154	0.00402
6	0.42752	0.00237
7	0.42515	0.00339
8	0.42176	0.00294
9	0.41882	0.00220
10	0.41662	0.00118
11	0.41544	0.00153
12	0.41391	0.00219
13	0.41172	0.00138
14	0.41034	0.00176
15	0.40858	0.00113
16	0.40745	0.00143
17	0.40603	0.00166
18	0.40437	0.00213
19	0.40223	0.00037
20	0.40186	0.40186

## 7. Appendix

---

### A6. LD decay – average Pair-wise correlation ( $r^2$ ) values

Pair-wise correlation ( $r^2$ ) between marker pairs of all 29,616 SNPs were calculated for an area of 200 Mb for all chromosomes. Average  $r^2$  values of 0.1 kb intervals for 10 kb were calculated, as well as 100 kb intervals for the first 1 Mb (first table). Second table (next page) shows  $r^2$  values of 1 Mb intervals for the whole area of 200 Mb.

start [kb]	end [kb]	R.2									
0.0	0.1	0.184	2.8	2.9	0.059	5.6	5.7	0.034	8.3	8.4	0.019
0.1	0.2	0.127	2.9	3.0	0.041	5.7	5.8	0.088	8.4	8.5	0.051
0.2	0.3	0.087	3.0	3.1	0.043	5.8	5.9	0.115	8.5	8.6	0.036
0.3	0.4	0.097	3.1	3.2	0.085	5.9	6.0	0.061	8.6	8.7	0.092
0.4	0.5	0.079	3.2	3.3	0.112	6.0	6.1	0.023	8.7	8.8	0.021
0.5	0.6	0.087	3.3	3.4	0.028	6.1	6.2	0.017	8.8	8.9	0.034
0.6	0.7	0.064	3.4	3.5	0.019	6.2	6.3	0.046	8.9	9.0	0.093
0.7	0.8	0.056	3.5	3.6	0.071	6.3	6.4	0.023	9.0	9.1	0.021
0.8	0.9	0.061	3.6	3.7	0.051	6.4	6.5	0.074	9.1	9.2	0.045
0.9	1.0	0.073	3.7	3.8	0.051	6.5	6.6	0.132	9.2	9.3	0.021
1.0	1.1	0.090	3.8	3.9	0.069	6.6	6.7	0.087	9.3	9.4	0.035
1.1	1.2	0.095	3.9	4.0	0.074	6.7	6.8	0.030	9.4	9.5	0.089
1.2	1.3	0.090	4.0	4.1	0.059	6.8	6.9	0.072	9.5	9.6	0.035
1.3	1.4	0.065	4.1	4.2	0.127	6.9	7.0	0.032	9.6	9.7	0.032
1.4	1.5	0.095	4.2	4.3	0.064	7.0	7.1	0.128	9.7	9.8	0.033
1.5	1.6	0.106	4.3	4.4	0.051	7.1	7.2	0.085	9.8	9.9	0.028
1.6	1.7	0.059	4.4	4.5	0.038	7.2	7.3	0.035	9.9	10.0	0.005
1.7	1.8	0.055	4.5	4.6	0.039	7.3	7.4	0.068	0.0	100.0	0.137
1.8	1.9	0.045	4.6	4.7	0.060	7.4	7.5	0.057	100.0	200.0	0.033
1.9	2.0	0.088	4.7	4.8	0.057	7.5	7.6	0.032	200.0	300.0	0.035
2.0	2.1	0.063	4.8	4.9	0.061	7.6	7.7	0.034	300.0	400.0	0.034
2.1	2.2	0.042	4.9	5.0	0.026	7.7	7.8	0.096	400.0	500.0	0.032
2.2	2.3	0.068	5.0	5.1	0.055	7.8	7.9	0.024	500.0	600.0	0.033
2.3	2.4	0.052	5.1	5.2	0.053	7.9	8.0	0.024	600.0	700.0	0.030
2.4	2.5	0.040	5.2	5.3	0.034	8.0	8.1	0.011	700.0	800.0	0.031
2.5	2.6	0.048	5.3	5.4	0.070	8.1	8.2	0.014	800.0	900.0	0.029
2.6	2.7	0.054	5.4	5.5	0.063	8.2	8.3	0.033	900.0	1,000.0	0.029
2.7	2.8	0.115	5.5	5.6	0.063						

## 7. Appendix

start [Mb]	end [Mb]	R.2									
1	2	0.028	51	52	0.025	101	102	0.025	151	152	0.024
2	3	0.027	52	53	0.024	102	103	0.024	152	153	0.025
3	4	0.026	53	54	0.024	103	104	0.024	153	154	0.025
4	5	0.026	54	55	0.023	104	105	0.025	154	155	0.025
5	6	0.026	55	56	0.024	105	106	0.025	155	156	0.024
6	7	0.026	56	57	0.024	106	107	0.025	156	157	0.024
7	8	0.026	57	58	0.024	107	108	0.025	157	158	0.025
8	9	0.024	58	59	0.024	108	109	0.024	158	159	0.026
9	10	0.025	59	60	0.024	109	110	0.025	159	160	0.026
10	11	0.024	60	61	0.024	110	111	0.025	160	161	0.027
11	12	0.025	61	62	0.023	111	112	0.025	161	162	0.024
12	13	0.025	62	63	0.025	112	113	0.024	162	163	0.025
13	14	0.024	63	64	0.023	113	114	0.025	163	164	0.025
14	15	0.025	64	65	0.024	114	115	0.026	164	165	0.025
15	16	0.024	65	66	0.025	115	116	0.026	165	166	0.026
16	17	0.025	66	67	0.024	116	117	0.025	166	167	0.025
17	18	0.025	67	68	0.024	117	118	0.025	167	168	0.026
18	19	0.024	68	69	0.024	118	119	0.025	168	169	0.026
19	20	0.024	69	70	0.026	119	120	0.025	169	170	0.026
20	21	0.025	70	71	0.025	120	121	0.025	170	171	0.026
21	22	0.024	71	72	0.024	121	122	0.025	171	172	0.026
22	23	0.023	72	73	0.024	122	123	0.025	172	173	0.024
23	24	0.023	73	74	0.023	123	124	0.026	173	174	0.025
24	25	0.023	74	75	0.024	124	125	0.024	174	175	0.025
25	26	0.024	75	76	0.024	125	126	0.025	175	176	0.027
26	27	0.024	76	77	0.024	126	127	0.025	176	177	0.025
27	28	0.025	77	78	0.024	127	128	0.024	177	178	0.025
28	29	0.024	78	79	0.024	128	129	0.025	178	179	0.025
29	30	0.023	79	80	0.024	129	130	0.024	179	180	0.026
30	31	0.024	80	81	0.024	130	131	0.025	180	181	0.024
31	32	0.024	81	82	0.025	131	132	0.027	181	182	0.025
32	33	0.023	82	83	0.025	132	133	0.025	182	183	0.026
33	34	0.024	83	84	0.023	133	134	0.026	183	184	0.026
34	35	0.024	84	85	0.024	134	135	0.024	184	185	0.028
35	36	0.023	85	86	0.024	135	136	0.025	185	186	0.025
36	37	0.025	86	87	0.025	136	137	0.024	186	187	0.025
37	38	0.023	87	88	0.025	137	138	0.027	187	188	0.025
38	39	0.023	88	89	0.024	138	139	0.025	188	189	0.027
39	40	0.024	89	90	0.026	139	140	0.025	189	190	0.025
40	41	0.024	90	91	0.025	140	141	0.024	190	191	0.025
41	42	0.024	91	92	0.026	141	142	0.025	191	192	0.027
42	43	0.024	92	93	0.026	142	143	0.023	192	193	0.027
43	44	0.023	93	94	0.024	143	144	0.026	193	194	0.025
44	45	0.024	94	95	0.024	144	145	0.025	194	195	0.026
45	46	0.024	95	96	0.024	145	146	0.024	195	196	0.027
46	47	0.023	96	97	0.025	146	147	0.026	196	197	0.027
47	48	0.024	97	98	0.025	147	148	0.023	197	198	0.027
48	49	0.024	98	99	0.024	148	149	0.024	198	199	0.027
49	50	0.023	99	100	0.024	149	150	0.025	199	200	0.026
50	51	0.024	100	101	0.025	150	151	0.026			

## 7. Appendix

---

### A7. LD decay at SNPs – average Pair-wise correlation ( $r^2$ ) values

Pair-wise correlation ( $r^2$ ) between marker pairs of SNPs detected in GWAS and markers in a 20 Mb upstream and downstream range of SNP. Average  $r^2$  values of 1 Mb intervals were calculated.  $r^2$  values, which are higher than the threshold of  $r^2 = 0.1$  are marked blue. First table = SNP results of GWAS with all taxa. Second table = SNP results of GWAS without *S. strictum*.

Chr	start [Mb]	end [Mb]	R.2	SNP	Chr	start [Mb]	end [Mb]	R.2	SNP	Chr	start [Mb]	end [Mb]	R.2	SNP
1R	-20	-19	0.017	SNP_116		14	15	0.003	SNP_3568		-4	-3	0.034	
	-19	-18	0.015			15	16	0.007			-1	0	0.098	
	-16	-15	0.049			18	19	0.010			0	1	0.072	
	-15	-14	0.020			19	20	0.000			1	2	0.032	
	-14	-13	0.007			-20	-19	0.003			2	3	0.010	
	-11	-10	0.027			-17	-16	0.044			4	5	0.015	
	-8	-7	0.023			-15	-14	0.003			6	7	0.016	
	-7	-6	0.017			-14	-13	0.011			8	9	0.015	
	-5	-4	0.005			-9	-8	0.006			9	10	0.045	
	-4	-3	0.040			-8	-7	0.025			11	12	0.013	
	-2	-1	0.022			-7	-6	0.071			12	13	0.041	
	-1	0	0.058			-5	-4	0.003			14	15	0.002	
	0	1	0.018			-4	-3	0.033			16	17	0.024	
	3	4	0.023	SNP_3467		-3	-2	0.014			17	18	0.064	
	4	5	0.008			-2	-1	0.002			18	19	0.017	
	12	13	0.026			-1	0	0.079			19	20	0.029	
	14	15	0.017			0	1	0.009			-20	-19	0.004	
	18	19	0.011			1	2	0.008			-16	-15	0.006	
	19	20	0.014			2	3	0.025			-14	-13	0.006	
	-20	-19	0.004			5	6	0.022	SNP_4667		-13	-12	0.009	
	-19	-18	0.013			6	7	0.042			-12	-11	0.004	
	-18	-17	0.002			7	8	0.016			-11	-10	0.002	
	-16	-15	0.006			8	9	0.005			-10	-9	0.002	
	-15	-14	0.002			9	10	0.002			-9	-8	0.003	
	-14	-13	0.002			10	11	0.001			-8	-7	0.008	
	-12	-11	0.009			11	12	0.023			-7	-6	0.006	
	-10	-9	0.003			13	14	0.056			-6	-5	0.005	
	-9	-8	0.005			14	15	0.018			-5	-4	0.012	
	-8	-7	0.002			15	16	0.031			-3	-2	0.004	
	-7	-6	0.003			16	17	0.026			-2	-1	0.005	
	-4	-3	0.006			17	18	0.042			-1	0	0.082	
	-2	-1	0.007			18	19	0.022			0	1	0.024	
	0	1	0.068			19	20	0.019			1	2	0.005	
	4	5	0.004	2R		-20	-19	0.047	SNP_4224		4	5	0.003	
	5	6	0.004			-19	-18	0.052			5	6	0.004	
	6	7	0.007			-18	-17	0.017			6	7	0.009	
	8	9	0.007			-15	-14	0.023			9	10	0.006	
	9	10	0.001			-14	-13	0.001			10	11	0.008	
	10	11	0.003			-11	-10	0.013			12	13	0.007	
	11	12	0.003			-6	-5	0.039			14	15	0.003	
	13	14	0.011			-5	-4	0.013			15	16	0.004	

Table continued on next page

## 7. Appendix

	16	17	0.003			1	2	0.001			-3	-2	0.019
	17	18	0.008			2	3	0.022			-1	0	0.027
	18	19	0.003			3	4	0.022			0	1	0.081
	19	20	0.005			5	6	0.014			1	2	0.014
	-20	-19	0.016			6	7	0.007			4	5	0.009
	-19	-18	0.018			7	8	0.052			5	6	0.024
	-17	-16	0.010			8	9	0.031			6	7	0.015
	-16	-15	0.015			9	10	0.008			7	8	0.040
	-15	-14	0.010			11	12	0.014			9	10	0.009
	-14	-13	0.006			12	13	0.004			10	11	0.016
	-12	-11	0.009			14	15	0.003			11	12	0.013
	-11	-10	0.028			15	16	0.008			12	13	0.031
	-10	-9	0.016			16	17	0.007			13	14	0.006
	-9	-8	0.022			17	18	0.014			14	15	0.004
	-8	-7	0.021			18	19	0.004			15	16	0.009
	-7	-6	0.015			19	20	0.029			16	17	0.003
	-2	-1	0.018			-19	-18	0.059			17	18	0.009
	-1	0	0.031			-18	-17	0.017			19	20	0.018
	0	1	0.022	SNP_8211		-17	-16	0.003			-16	-15	0.008
	1	2	0.011			-13	-12	0.016			-15	-14	0.010
	2	3	0.015			-12	-11	0.074			-14	-13	0.005
	3	4	0.010			-11	-10	0.010			-13	-12	0.017
	4	5	0.019			-10	-9	0.028			-12	-11	0.008
	5	6	0.009			-9	-8	0.007			-11	-10	0.004
	7	8	0.032			-7	-6	0.037			-10	-9	0.016
	8	9	0.021			-6	-5	0.040			-9	-8	0.005
	9	10	0.022			-3	-2	0.044			-8	-7	0.007
	10	11	0.023			-2	-1	0.022			-7	-6	0.009
	11	12	0.013			-1	0	0.115			-6	-5	0.012
	14	15	0.010			0	1	0.014			-5	-4	0.002
	15	16	0.018			8	9	0.031			-4	-3	0.002
	16	17	0.009			10	11	0.011			-2	-1	0.009
	18	19	0.025			13	14	0.004			-1	0	0.011
	19	20	0.013			15	16	0.025			0	1	0.014
	-19	-18	0.011			16	17	0.038			1	2	0.015
	-18	-17	0.006			17	18	0.041			3	4	0.011
	-16	-15	0.003			18	19	0.008			5	6	0.008
	-15	-14	0.015			19	20	0.006			7	8	0.004
	-14	-13	0.002			-20	-19	0.005			8	9	0.005
	-13	-12	0.014			-19	-18	0.006			11	12	0.006
	-12	-11	0.017			-16	-15	0.005			15	16	0.007
3R	-11	-10	0.015	SNP_11215		-14	-13	0.011			16	17	0.008
	-10	-9	0.015			-13	-12	0.006			-19	-18	0.003
	-9	-8	0.005			-11	-10	0.008			-18	-17	0.001
	-8	-7	0.006			-9	-8	0.014			-15	-14	0.001
	-7	-6	0.010			-8	-7	0.001			-12	-11	0.003
	-5	-4	0.008			-7	-6	0.010			-11	-10	0.001
	-2	-1	0.004			-6	-5	0.015			-10	-9	0.001
	-1	0	0.062			-5	-4	0.021			-8	-7	0.001
	0	1	0.026			-4	-3	0.003			-7	-6	0.002

Table continued on next page

## 7. Appendix

-6	-5	0.003		-18	-17	0.022		0	1	0.044			
-5	-4	0.005		-17	-16	0.000		1	2	0.067			
-4	-3	0.002		-16	-15	0.022		2	3	0.005			
-3	-2	0.003		-14	-13	0.016		3	4	0.033			
-2	-1	0.001		-11	-10	0.040		4	5	0.083			
-1	0	0.003		-10	-9	0.133		9	10	0.019			
0	1	0.013		-9	-8	0.053		11	12	0.002			
1	2	0.002		-8	-7	0.045		14	15	0.044			
2	3	0.002		-7	-6	0.002		16	17	0.017			
5	6	0.003		-6	-5	0.033		17	18	0.006			
6	7	0.004		-3	-2	0.126		19	20	0.004			
7	8	0.002		-2	-1	0.059		-17	-16	0.001			
10	11	0.001		-1	0	0.151		-16	-15	0.073			
11	12	0.002		0	1	0.129		-14	-13	0.001			
12	13	0.002		2	3	0.184		-13	-12	0.001			
14	15	0.002		3	4	0.107		-6	-5	0.001			
15	16	0.002		4	5	0.067		-2	-1	0.010			
17	18	0.003		5	6	0.027		-1	0	0.016			
18	19	0.003		6	7	0.032		0	1	0.001			
19	20	0.004		8	9	0.066		1	2	0.005			
6R	-9	-8	0.003		12	13	0.214		5	6	0.078	SNP_27431	
	-8	-7	0.004		14	15	0.087		6	7	0.006		
	-6	-5	0.002		15	16	0.097		9	10	0.041		
	-5	-4	0.002		16	17	0.090		10	11	0.009		
	-1	0	0.172		18	19	0.018		11	12	0.001		
	0	1	0.072		-16	-15	0.000		13	14	0.002		
	1	2	0.014		-15	-14	0.018		14	15	0.001		
	3	4	0.002		-14	-13	0.009		16	17	0.014		
	4	5	0.011		-7	-6	0.012		18	19	0.026		
	5	6	0.003		-5	-4	0.003		-20	-19	0.011		
	8	9	0.003		-4	-3	0.009		-19	-18	0.008		
	10	11	0.006		-3	-2	0.018		-17	-16	0.057		
	12	13	0.004		-2	-1	0.011		-16	-15	0.019		
	14	15	0.001		-1	0	0.030		-15	-14	0.016		
	15	16	0.003		0	1	0.021		-14	-13	0.005		
	18	19	0.002		1	2	0.010		-12	-11	0.005		
	19	20	0.002		2	3	0.006		-11	-10	0.013		
	-17	-16	0.165		4	5	0.013		-9	-8	0.013		
	-16	-15	0.142		5	6	0.008		-7	-6	0.005		
	-15	-14	0.165		6	7	0.008		-6	-5	0.017	SNP_28497	
	-12	-11	0.169		7	8	0.069		-5	-4	0.005		
	-6	-5	0.105		10	11	0.017		-4	-3	0.007		
	-4	-3	0.006		12	13	0.009		-2	-1	0.015		
	-1	0	0.615		13	14	0.011		-1	0	0.011		
	0	1	0.325		15	16	0.027		0	1	0.213		
	3	4	0.052		7R	-9	-8	0.038		3	4	0.007	
	10	11	0.089			-6	-5	0.008		4	5	0.006	
	11	12	0.030			-4	-3	0.004		6	7	0.007	
	-20	-19	0.036			-2	-1	0.048		9	10	0.002	
	-19	-18	0.037			-1	0	0.987		10	11	0.005	

Table continued on next page

## 7. Appendix

---

	14	15	0.014	
	15	16	0.001	
	16	17	0.008	
	18	19	0.000	

Chr	start [Mb]	end [Mb]	R.2	SNP
4R	-18	-17	0.003	SNP_16769
	-16	-15	0.006	
	-12	-11	0.006	
	-9	-8	0.000	
	-6	-5	0.002	
	-5	-4	0.003	
	-3	-2	0.008	
	-2	-1	0.002	
	-1	0	0.108	
	0	1	0.006	
	3	4	0.014	
	5	6	0.030	
	6	7	0.002	
	7	8	0.003	
	8	9	0.010	

Chr	start [Mb]	end [Mb]	R.2	SNP
7R	-19	-18	0.001	SNP_25651
	-17	-16	0.001	
	-14	-13	0.001	
	-13	-12	0.002	
	-12	-11	0.002	
	-11	-10	0.002	
	-7	-6	0.001	
	-3	-2	0.001	
	-2	-1	0.001	
	-1	0	0.045	
	0	1	0.001	
	1	2	0.002	
	2	3	0.001	
	3	4	0.009	
	5	6	0.002	
	6	7	0.002	
	8	9	0.000	
	9	10	0.001	
	10	11	0.000	
	13	14	0.002	
	16	17	0.001	
	17	18	0.002	
	18	19	0.003	

## 8. Indices

### 8.1 List of abbreviations

bp	Base pairs
chr	Chromosome
DArT	Diversity Array Technology
ELMO	Engulfment and cell motility
EV	Eigenvector
$F$	Inbreeding coefficient
FarmCPU	Fixed and random model Circulating Probability Unification
FDR	false discovery rate
FEM	Fixed Effect Model
$F_{ST}$	Fixation Index
GAPIT	Genomic Association and Prediction Integrated Tool
GBIS	Genebank information system of the German Federal ex situ Genebank at IPK Gatersleben
CaM	Calmodulin
GRIN	Germplasm Resource Information Network
GWAS	Genome-wide association studies
HC	High confidence
HW	Hardy-Weinberg
HWE	Hardy-Weinberg-Equilibrium
kb	Kilo bases
LC	Low confidence
LD	Linkage Disequilibrium
maf	Minor allele frequency
Mb	mega bases
MLM	Mixed Linear Model
MLMM	Multiple Loci Linear Mixed Model

## 8. Indices

---

MTA	Marker trait associations
NA	Not Available
PC	Principal component
PCA	Principal component analysis
PCO	Principal coordinate
pers. commun.	Personal communication
Pos	Position
QTN	Quantitative trait nucleotide
REM	Random Effect Model
SNP	Single nucleotide polymorphisms
SI	Self-incompatibility
SC	Self-compatibility
SF	Self-fertility
SUPER	Settlement of MLM Under Progressively Exclusive Relationship
XTH	Xyloglucan endotransglucosylase/hydrolase
$\pi$	Nucleotide diversity

## 8. Indices

---

### 8.2 List of figures

Figure 1 Model for SI in grasses.....	7
Figure 2 Geographical origin.....	13
Figure 3 Principal component analysis (PCA) .....	24
Figure 4 Cross-entropy criterion.....	26
Figure 5 Ancestry coefficient analysis.....	27
Figure 6 Nucleotide diversity ( $\pi$ ) .....	29
Figure 7 Inbreeding coefficient ( $F$ ).....	30
Figure 8 Genome-Wide Association Study (GWAS) for all <i>Secale</i> taxa.....	31
Figure 9 Genome-Wide Association Study (GWAS) for <i>Secale</i> taxa without <i>S. strictum</i> .....	33
Figure 10 Linkagen disequilibrium (LD) decay for all rye chromosomes .....	35
Figure 11 Linkagen disequilibrium (LD) decay for QTLs.....	38

### 8.3 List of tables

Table 1 Taxonomic relationships within the genus <i>Secale</i> . (Daskalova and Spetsov 2020).....	2
Table 2 The species and subspecies of <i>Secale</i> L. (Feldman and Levy 2023) .....	3
Table 3 Fixation Indices ( $F_{ST}$ ).....	28
Table 4 List of all QTLs for all <i>Secale</i> taxa.....	32
Table 5 List of all QTLs for <i>Secale</i> taxa without <i>S. strictum</i> .....	34
Table 6 Gene Annotation.....	39

## 9. Acknowledgements

I would like to express my sincere gratitude to Dr. Steven Dreißig for accepting me into his research group on Plant Reproductive Genetics (Emmy Noether Research Group), which is hosted by the Chair of Plant Breeding (Prof. Dr. Klaus Pillen) at the Martin-Luther-University Halle-Wittenberg (MLU), and for allowing me to participate in his research.

I would like to express my special thanks for providing the topic, but above all for the supervision and the opportunity to discuss technical questions at any time. The supervision enabled me to combine theoretical knowledge of population genetic analyses with practical implementation using bioinformatic methods.

Furthermore, I would like to thank Christina Waesch (PhD student) for her advice and suggestions on how to improve my analyses. I would also like to thank the entire research group team for making me feel welcome.

Finally, I would like to thank Prof. Dr. Klaus Pillen for being the first reviewer of my thesis.

## 10. Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich diese Masterarbeit selbständig und ohne fremde Hilfe verfasst und andere als die angegebenen Quellen und Hilfsmittel nicht benutzt habe.

Halle (Saale), 04.04.2024