Article

Genomic Scan of Male Fertility Restoration Genes in a 'Gülzow' Type Hybrid Breeding System of Rye (*Secale cereale* L.)

Nikolaj Meisner Vendelbo [1,2,*,†], Khalid Mahmood [1,†], Pernille Sarup [1], Peter Skov Kristensen [1], Jihad Orabi [1] and Ahmed Jahoor [1,3]

[1] Nordic Seed A/S, Grindsnabevej 25, 8300 Odder, Denmark; khma@nordicseed.com (K.M.); pesa@nordicseed.com (P.S.); pskr@nordicseed.com (P.S.K.); jior@nordicseed.com (J.O.); ahja@nordicseed.com (A.J.)
[2] Department of Agroecology, Faculty of Technology, Aarhus University, Forsøgsvej 1, Flakkebjerg, 4200 Slagelse, Denmark
[3] Department of Plant Breeding, The Swedish University of Agricultural Sciences, 23053 Alnarp, Sweden
* Correspondence: nive@nordicseed.com
† These authors contributed equally to this work.

Abstract: Efficient and stable restoration of male fertility (Rf) is a prerequisite for large-scale hybrid seed production but remains an inherent issue in the predominant fertility control system of rye (*Secale cereale* L.). The 'Gülzow' (G)-type cytoplasmic male sterility (CMS) system in hybrid rye breeding exhibits a superior Rf. While having received little scientific attention, one major G-type *Rf* gene has been identified on 4RL (*Rfg1*) and two minor genes on 3R (*Rfg2*) and 6R (*Rfg3*) chromosomes. Here, we report a comprehensive investigation of the genetics underlying restoration of male fertility in a large G-type CMS breeding system using recent advents in rye genomic resources. This includes: (I) genome-wide association studies (GWAS) on G-type germplasm; (II) GWAS on a biparental mapping population; and (III) an RNA sequence study to investigate the expression of genes residing in Rf-associated regions in G-type rye hybrids. Our findings provide compelling evidence of a novel major G-type non-PPR *Rf* gene on the 3RL chromosome belonging to the mitochondrial transcription termination factor gene family. We provisionally denote the identified novel *Rf* gene on 3RL *RfNOS1*. The discovery made in this study is distinct from known P- and C-type systems in rye as well as recognized CMS systems in barley *(Hordeum vulgare* L.) and wheat (*Triticum aestivum* L.). We believe this study constitutes a stepping stone towards understanding the restoration of male fertility in the G-type CMS system and potential resources for addressing the inherent issues of the P-type system.

Keywords: restoration of male fertility (*Rf*); cytoplasmic male sterility (CMS); pentatricopeptide repeat protein (PPR); mitochondrial transcription termination factor (mTERF); 600K SNP array; genome-wide association study (GWAS); RNAseq; chi-square; linkage disequilibrium

1. Introduction

In recent years, hybrids have become the predominant class of cultivated winter rye (*Secale cereale* L.) in Northern Europe [1]. Outperforming population-based cultivars, hybrids in rye demonstrate strong heterotic effects on all developmental and yield characteristics [2,3]. Breeding of hybrids relies on the existence of cytoplasmic male sterility (CMS) and restoration of male fertility (Rf) genes that reside in genetically distinct parental populations [3,4]. This system efficiently enables control of parental crossing in the field, which is a prerequisite for large-scale hybrid seed production [5]. In hybrid rye, numerous CMS systems exist, of which the most predominant is the Pampa (P) type [6]. In this system, five major P-type *Rf* genes have been identified on 1RS, 4RL (*Rfp1, Rfp2, Rfp3*), and 6R (dominant modifier) chromosomes, and three minor genes on 3RL, 4RL, and 5R chromosomes [7–10]. The less prevalent CMS systems include the 'Gülzow' (G) type

originating from the population of rye variety 'Schlägler alt' [11], R-type originating from a Russian population [12], and C- [13] and S- [14] types originating from the old Polish population of rye variety 'Smolickie'. In the G-type CMS system, one major gene has been identified on 4RL (*Rfg1*) and two minor genes on 3R (*Rfg2*) and 6R (*Rfg3*) chromosomes [15]. In the C-type CMS system, two major *Rf* genes have been identified on 4RL (*Rfc1*) and 6RS (*Rfc2*) [16,17]. Intriguingly, Stojalowski et al. [18] observed a linkage between major *Rf* genes on 4RL for all three CMS systems, C-type (*Rfc1*), G-type (*Rfg1*), and P-type (*Rfp1*, *Rfp2*, *Rfp3*) to the same marker loci. This finding accentuates the pivotal importance of 4RL across CMS systems in hybrid rye breeding.

Restoration of male fertility in hybrids derived from the predominant P-type cytoplasm is frequently incomplete and highly environmentally unstable [9,19–21]. In addition to a potential loss in grain yield, partial pollination renders the cultivar susceptible to fungal infection by ergot (*Claviceps purpurea* (Fr.) Tul.) which can contaminate the rye grains with toxic sclerotia [22–24]. The P-type system is inherently shaped by the low frequency of restorer gametes in European populations in which the predominance exhibits unsatisfactory restoration [19,20]. In 1991, several non-adapted Argentinian and Iranian rye populations with a high frequency of restorer gametes were identified [25]. Crossing an elite maternal line with one of these non-adapted exotics led to observations of significantly higher restoration levels and environmental stability [26,27]. In order to steer the introgression of novel superior exotic *Rf* genes through marker-assisted selection, molecular markers were developed for *Rfp1*, *Rfp2*, and *Rfp3* [8,9]. Hybrids carrying an exotic *Rf* gene were, however, found to exhibit a significant reduction in grain yield by 4.4% to 9.4% caused by linkage drag effects or epistatic interactions associated with the exotic *Rf* gene [28]. Despite these deleterious effects, hybrid cultivars carrying the exotic *Rfp1* have been introduced to the Northern European market by a patented brand, PollenPlus® [29]. In contrast, hybrids derived from the less prevalent G-type cytoplasm are characterized by a complete and environmental stable restoration of male fertility [30]. However, having received little scientific attention, the underlying genetics of the G-type CMS system remain largely unexplored [15].

The male sterility factors in CMS lines are encoded by mitochondrial genes that cause a defect in the production of viable pollen [31]. While multiple gene families have been linked with male fertility restoration, a distinct clade of the pentatricopeptide repeat (PPR) RNA-binding factor family referred to as *Rf*-like PPR (RFL-PPR) constitutes the predominant class of isolated *Rf* genes [5]. Proteins of the PPR superfamily are characterized by up to 30 tandem repeats of a canonical 35-amino-acid motif [32]. Based on motif composition, PPR genes are divided into two subclasses: the P class solely containing the canonical motif and the PLS class containing triplets of P, L ('long', ≈36 aa), and S ('short', ≈31 aa) motifs [33,34]. While the P class has predominantly been associated with RFL-PPR genes, instances of the PLS class have also been identified [35]. In rye, 591 PPR genes have been identified, out of which 83 belong to the RFL-PPR clade [36] (Table S8a). PPR proteins target mitochondrial or chloroplast mRNA, participating in a range of post-transcriptional processes (RNA editing, splicing, cleavage, and translation) with profound effects on organelle biogenesis and function [33,37,38]. In wheat, the RFL-PPR genes *Rf1* and *Rf3* have been shown to bind to a mitochondrial *orf279* transcript, induce cleavage, and prevent the expression of the sterilizing factor [39]. Within grasses, several isolated *Rf* genes have been characterized as RFL-PPRs, including *Rfm1* in barley [35], *Rf1* in sorghum (*Sorghum bicolore* L.) [40], *Rf5* in maize [41], and *Rf4*, *Rf5*, and *Rf6* in rice (*Oryza sativa* L.) [42–44]. In the C-type CMS system of rye, the *Rfc1* locus has been found to reside in close proximity to a cluster of RFL-PPR genes on 4RL [36].

Another gene family less prevalently associated with Rf is the mitochondrial transcription termination factors (mTERF) [45]. In rye, 131 mTERF genes have been identified [36] (Table S8b). Similar to the PPR, mTERF genes encode helical repeat proteins that target mitochondrial DNA, regulating the expression of mitochondrial genes [46]. Within grasses, mTERF genes have been associated with restoration of male fertility in barley *Rfm3* on

the 6HS chromosome [47] and wheat *Rf9* on the 6BS chromosome [48]. In the P-type CMS system of rye, the *Rfp1* locus has been found to reside in close proximity to four mTERF genes [49,50]. In consistence, hotspots of RFL-PPR and mTERF genes have been identified in regions harboring known *Rf* genes in rye [36].

In this paper, we report an investigation of the genetics underlying male fertility restoration in G-type CMS-based hybrid rye breeding systems. The objective of this study was to identify major and minor G-type *Rf* genes. This was approached through: (I) genome -wide association study (GWAS) on a G-type CMS hybrid rye breeding germplasm; (II) GWAS on a biparental mapping population for studying the inheritance of male fertility restoration; and (III) gene expression analysis of PPR, RFL-PPR, and mTERF genes residing in Rf-associated blocks in two G-type hybrid cultivars for the identification of causative genes. This knowledge will serve as a stepping stone towards developing novel hybrid cultivars exhibiting superior and environmentally stable restoration of male fertility to maximize grain yield and enhance ergot resistance.

2. Results

2.1. Analysis of Genotyping Data

Prior to bioinformatic analysis using the single nucleotide polymorphism (SNP) array genotype data, a quality filtration was conducted to remove monomorphic, non-informative markers. Polymorphism information content (PIC) was calculated as a measure of the identified marker's informativeness, with a mean PIC of 0.26 for the 20K platform ($n = 365$), 0.34 for the 30K platform ($n = 181$), and 0.23 for the 600K platform ($n = 180$). All SNP arrays portrayed a uniform distribution of markers across the rye genome (Table S1). In total, 4419 informative markers were identified in the 20K array on the entire germplasm and can be found thoroughly characterized in a recent study by Vendelbo et al. (2020). A subset of this germplasm was genotyped on the recent rye 600K array, yielding 261406 informative markers. In the F_2 mapping population composed of 181 plants, 3493 informative markers were identified, out of which 1088 were derived from the 5K rye array, 808 from the 600K rye array, and 1597 from the 90K wheat array.

2.2. Genome-Wide Association Study—Case Control

Genome-wide association study (GWAS) was conducted using population origin as phenotypic input in a case control analysis for an initial 'crude' identification of potential restoration of male fertility (Rf) genes in the germplasm. The 20K GWAS analysis produced a distinct peak in the Manhattan plot at 724 to 745 Mbp on 3RL, with the highest associated marker ($-\log_{10}(p) = 19.1$) located at 745 Mbp (Figure 1A, Table S2). In the successive 600K GWAS analysis, a similar peak was identified at 710–747 Mbp with the highest associated markers ($-\log_{10}(p) = 27.07$) located between 729 and 730 Mbp (Figure 1B–D, Table S3).

The identified Rf-associated region harbored five PLS-class PPR genes and one mTERF gene in the 'Lo7' reference genome (Figure 1D, Table S7a). In addition, a unique peak portraying a similarly strong association was found on 1RS at 49.3–58.5 Mbp in the 600K GWAS analysis (Figure 1C). The region harbored 22 RFL-PPR genes organized in four clusters, 2 P-class PPR genes, and 1 mTERF gene (Figure 1C, Table S7b). Out of 20 significant associated markers residing at the site, 18 mapped to a narrow peak from 58.02 to 58.47 Mbp on 1 RS (Table S3).

2.3. Biparental Population

A biparental F_2 population consisting of 181 plants was developed from the 'Gülzow' (G)-type hybrid cv. Stannos. The population was phenotyped for six Rf-associated traits as well as traits related to restoration in order to obtain a comprehensive dataset on the inheritance of G-type *Rf* genes. Seed number and pollen production were found, on the basis of our observations, to be the most representative Rf-associated traits (Figure 2A,B).
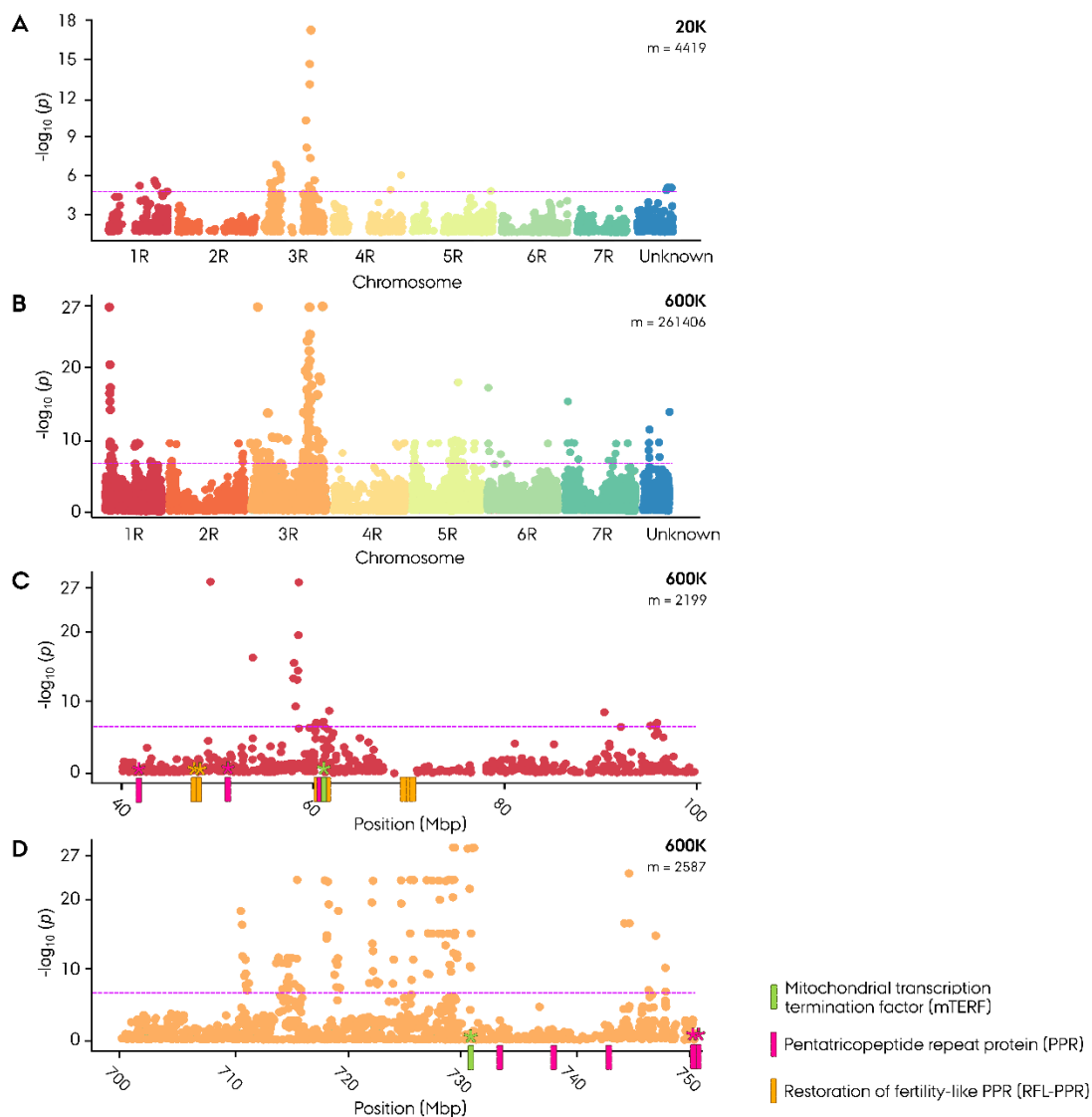
Figure 1. Manhattan plot for genome-wide association study (GWAS) on population origin ('case control') in Nordic Seed elite hybrid rye (*Secale cereale* L.) breeding germplasm. (**A**) Genome-wise Manhattan plot of 20K SNP array GWAS (*n* = 365). (**B**) Genome-wise Manhattan plot of 600K SNP array GWAS (*n* = 180). (**C**) Manhattan plot of the 600K SNP array 1RS region. (**D**) Manhattan plot of the 600K SNP array 3RL region. Position of major restoration of male fertility-associated genes has been included with genes expressed in G-type hybrids marked with an asterisk. Significant association was identified using criterion of $-\log_{10}(p) > 4.95$ in 20K and >6.72 in 600K depicted as a magenta line.

The observed segregation ratio of sterile and fertile F$_2$ plants was tested for goodness of fit to the expected Mendelian ratio at the scenario of one, two, and three major *Rf* genes using an χ^2 test. Intriguingly, the observed segregation ratios were in accordance with a monogenic dominant inheritance of male fertility restoration with χ^2 (1, $n_{\text{,infertile}} = 38$, $n_{\text{,fertile}} = 143$) = 2.26, $p = 0.13$ for seed number and χ^2 (1, $n_{\text{infertile}} = 43$, $n_{\text{fertile}} = 138$) = 1.11, $p = 0.29$ for pollen production (Table S4). GWAS led to the identification of 16 Rf-associated SNP markers, of which 5 markers showed a significant association with $-\log_{10}(p) > 5.2$ (Figure 3). On 3RL, a twin peak was identified in the GWAS. The first peak spanning from 627 to 769 Mbp with the highest associated marker derived from the 90K wheat array ($-\log_{10}(p) = 6.66$) was localized at 627 Mbp. The majority of the Rf-associated markers were located around 745 Mbp (Figure 3, Table S4). The second peak, comprising four markers, spanned from 807.1 to 808.7 Mbp with its highest associated marker ($-\log_{10}(p) = 4.68$) just below the significance threshold of 4.85.
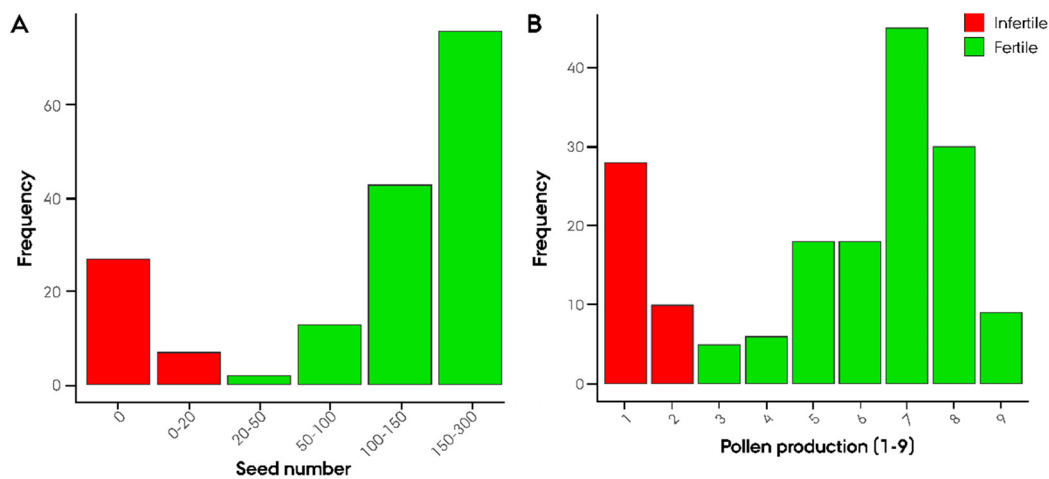
Figure 2. Phenotypic distribution of restoration of male-fertility-related traits. (**A**) Seed number and (**B**) pollen production in 181 F_2 plants derived from a 'Gülzow' type hybrid rye (*Secale cereale* L.) cv. Stannos.
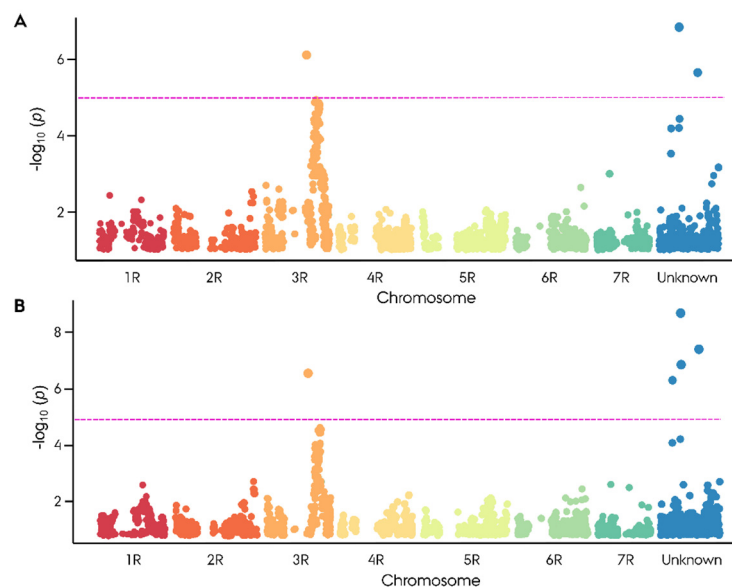

Figure 3. Manhattan plot for genome-wide association study on the restoration of male-fertility-related phenotypic scores. (**A**) Seed number and (**B**) pollen production (1–9) in an F_2 biparental population composed of 181 plants derived from the hybrid rye cv. Stannos. In total, 3494 informative SNP markers were used. Significant association was identified using criterion of $-\log_{10}(p) > 4.85$ depicted as a magenta line.

The remaining four significantly associated SNP markers were derived from the 90K wheat SNP array. None of the four markers could successfully be mapped to the 'Lo7' rye reference genome. Two of these markers mapped to the short arm of the wheat 3B chromosome, including the highest Rf-associated ($-\log_{10}(p) = 9.12$) wheat marker AX_158558079. One of the remaining moderately Rf-associated markers mapped to the short arm of the wheat 1A chromosome, while the last had no available mapping position in wheat. With no mapping position, genome-wide pairwise linkage disequilibrium was calculated for each of the four highly Rf-associated wheat-derived SNP markers to determine their position based on linkage to mapped markers (Table S5). All four wheat markers exhibited a singular peak on 3RL with a top LD ranging from 0.43 to 0.97 in the region spanning 701 to 747 Mbp (Figure S2). The top Rf-associated wheat marker AX_158558079 exhibited a max LD of 0.85 at 747 Mbp (Figure 4A,B). None of the

Rf-associated wheat-derived markers showed linkage towards the Rf-associated marker cluster at 807.1–808.7 Mbp (Table S9).
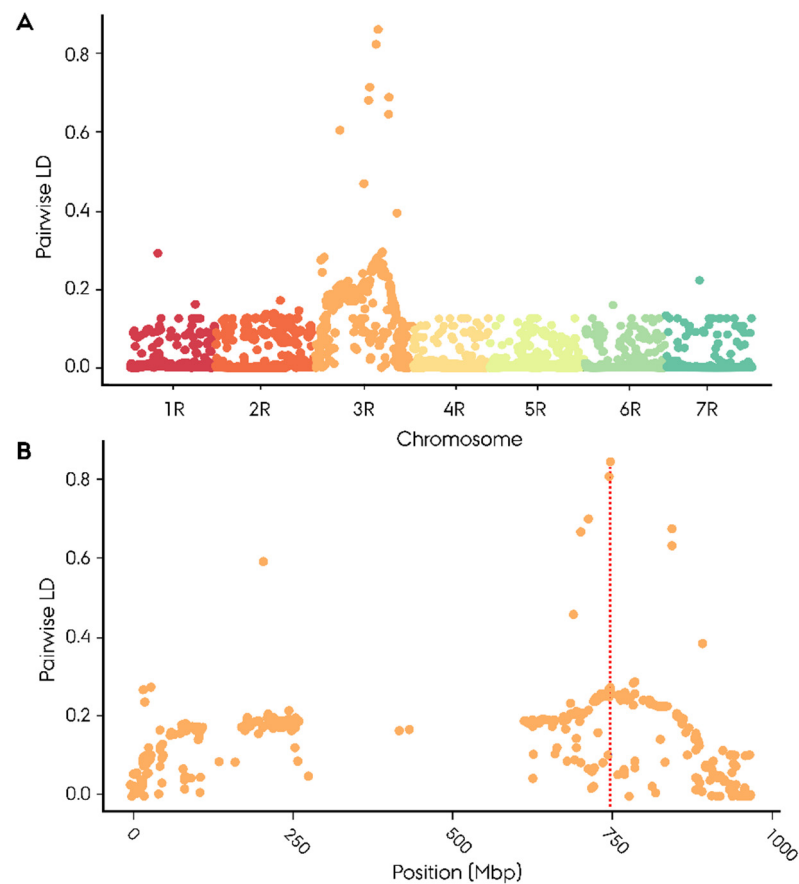

Figure 4. Mapping of an SNP marker derived from 90K wheat array highly associated with restoration of male fertility in 181 (*Secale cereale* L.) F_2 plants by genome-wide pairwise linkage disequilibrium (LD) towards 2448 informative SNP markers with accurate mapping position on 'Lo7' rye reference genome. (**A**) Genome-wide distribution of LD, (**B**) LD distribution on 3R chromosome.

2.4. Genomic Scan and Expression of Genes Residing within Restoration of Male-Fertility-Associated Regions of G-Type Hybrids

To identify a likely candidate *Rf* gene, a genomic scan of identified Rf-associated regions was performed and gene expression was investigated in two G-type hybrids. As the RNA-seq data were obtained from spikes at flowering, active expression of *Rf* genes would be expected in the fertile hybrids. The Rf-associated region spanning 710 to 760 Mbp was found to harbor 448 genes in the 'Lo7' reference genome, out of which 272 were expressed in cv. Helltop and 266 in cv. Stannis (Table S7a). Among the expressed genes, 251 were co-expressed in both hybrids. Amongst the panel of PPR, RFL-PPR and mTERF genes residing in the Rf-associated regions, and two out of five PLS-class PPR genes, at 752.1 and 759.1 Mbp, were co-expressed in both G-type hybrids (Figure 1D, Table 1). The site also harbored a single mTERF gene at 731.7 Mbp, likewise co-expressed in both G-type hybrids. The mTERF gene co-localized with the top associated marker in the 600K case control GWAS with less than 1 Mbp distance (Figure 1D, Table S3).

Table 1. Pentatricopeptide repeat protein (PPR) and mitochondrial transcription termination factor (mTERF) genes in restoration of male-fertility-associated regions identified on 1RS (40–70 Mbp) and 3RL (710–760 Mbp) chromosome in 'Lo7' rye (*Secale cereale* L.) reference genome. Genes expressed in 'Gülzow' type hybrid cultivars at flowering are marked as green, and genes not expressed as red.

Chromosome	Position (Mbp)	Gene Length (bp)	Annotation	Expression	
				cv. Helltop	cv. Stannos
1R	42.9	822	PPR (P Type)	green	green
	46.2	2091	RFL-PPR (P Type)	green	red
	46.2	2613	RFL-PPR (P Type)	green	green
	46.3	2601	RFL-PPR (P Type)	red	red
	46.9	2487	RFL-PPR (P Type)	red	red
	47.1	2490	RFL-PPR (P Type)	green	green
	47.1	1833	RFL-PPR (P Type)	green	green
	51.9	1659	PPR (P Type)	green	green
	61.1	417	RFL-PPR (P Type)	red	red
	61.3	287	PPR (P Type)	red	red
	61.3	213	PPR (P Type)	red	red
	61.5	2499	RFL-PPR (P Type)	red	red
	61.5	1866	mTERF	green	green
	61.8	2472	RFL-PPR (P Type)	red	green
	61.9	2448	RFL-PPR (P Type)	red	red
	61.9	1460	RFL-PPR (P Type)	red	red
	61.9	2451	RFL-PPR (P Type)	red	red
	62.0	2499	RFL-PPR (P Type)	red	red
	70.8	510	mTERF-like	red	red
	70.8	540	mTERF-like	red	green
3R	731.7	462	mTERF	green	green
	733.4	1986	PPR (PLS Type)	green	red
	738.4	1743	PPR (PLS Type)	red	green
	743.7	1461	PPR (PLS Type)	red	green
	751.6	1356	PPR (PLS Type)	red	red
	752.1	1548	PPR (PLS Type)	green	green
	759.7	825	PPR (PLS Type)	green	green

The Rf-associated site on 1RS spanning from 40 to 70 Mbp was found to harbor 380 genes in the 'Lo7' reference genome, out of which 166 were expressed in cv. Helltop and 160 in cv. (Table S7b) Stannos. Among the expressed genes, 144 were co-expressed in both G-type hybrids. Out of 13 RFL-PPR genes residing in the region, three situated at 46.2–47.1 Mbp were co-expressed in both G-type hybrids (Figure 1C, Table 1). The site also harbored two P-class PPR genes, at 42.9 and 51.9 Mbp, and a single mTERF gene, at 61.5 Mbp, likewise co-expressed in both G-type hybrids. The mTERF gene resided 3.2 Mbp from the top associated marker on 1RS.

3. Discussion

While the less common 'Gülzow' (G)-type system demonstrates superior restoration of male fertility, it has received little scientific attention in the past. This is the first study since Melz et al. [30] in 2003 to investigate the genetics underlying male fertility restoration in G-type CMS hybrid rye breeding systems. Until now, only three G-type restoration of male fertility (Rf) genes have been reported, a major gene located on 4RL (*Rfg1*) and two modifying genes on 3R (*Rfg2*) and 6R (*Rfg3*) [15,51]. By exploiting recent advents in rye genomic resources, we succeeded in identifying a novel major G-type *Rf* gene on 3RL, in addition to further evidence of a major gene on 1RS and a modifying gene on 3RL

chromosome. Our findings provide a novel insight into the differentiation of the G-type fertilization control system from the predominant P-type.

3.1. Indications of a Major Restoration of Male-Fertility-Like Pentatricopeptide Repeat Gene on 1RS

While case control genome-wide association study (GWAS) is a useful tool for providing an insight into the genetics differentiating the parental gene pools, it has several limitations. Using population origin of lines as 'phenotypic' input, statistically associated markers in case control GWAS, can either be a population-defining trait such as an *Rf* QTL, or a product of population structure.

In a recent population study by Vendelbo et al. [52] on the entire G-type hybrid rye elite breeding germplasm, the maternal NRG & CMS population was found to exhibit considerable population structure and vast LD blocks. Unequal relatedness among individuals and population structure introduces a confounding effect that might cause spurious marker associations and introduce a risk of false positives [53,54]. To moderate the effect of these confounding factors, the GAPIT software used to conduct the GWAS therefore utilizes a compressed mixed linear model [55,56]. Large LD blocks, on the other hand, introduce an uncorrectable confounding factor. Long-distance LD complicates the disentanglement of actual causal variants from linked neutral markers, which can in turn lead to spurious associations [54].

In the 600K case control GWAS, a unique strong peak was identified on 1RS (Figure 1B,C). While the evidential significance of case control GWAS is insufficient to draw definitive conclusions, it provides an insight into pivotal genomic sites differentiating the parental populations. Intriguingly, the region was found to harbor 22 out of the 83 annotated *Rf*-like PPR (RFL-PPR) genes in the 'Lo7' reference genome situated in three large clusters (Figure 1C) [36]. In wheat, the syntenic segment has been reported to house two major *Rf* genes, *Rf1* (1AS) and *Rf3* (1BS), both belonging to the RFL-PPR gene family [39,57,58]. In rye, a major P-type *Rf* gene has likewise been identified on 1RS in a German inbred rye line 'L18' [7]. This *Rf* gene has, however, received little attention and it remains unknown whether the underlying *Rf* gene belongs to the RFL-PPR family.

Examining the RNA-seq data, none of the eight RFL-PPR genes residing in close proximity to the top associated marker peak were, however, found to be co-expressed at the flowering stage in the G-type hybrids (Figure 1C, Table S7b). Instead, a mitochondrial transcription termination factor (mTERF) gene residing 3.2 Mbp from the peak was found to be co-expressed.

While these findings suggest that the germplasm houses an additional major G-type *Rf* gene on 1RS, we did not observe any Rf-associated QTLs on 1RS in the mapping population GWAS (Figures 3 and 4). This can either be due to the absence of the major *Rf* gene on 1RS in the pollen father of cv. Stannos, or that the region harbors a population-defining trait other than an *Rf* gene. Given that the locus aligns with a region containing numerous known Rf genes in both wheat and rye, as well as an mTERF gene co-expressed in both G-type hybrids, it is not unlikely that a large G-type Rf gene occurs on 1RS in the germplasm.

3.2. Modifying G-Type Restoration of Male Fertility Genes

While incapable of restoring male fertility, minor *Rf* genes are believed to enhance ('modify') the effect of major *Rf* genes [7]. At present, two minor *Rf* genes have been identified in the G-type CMS hybrid rye breeding systems on the 3R (*Rfg2*) and 6R (*Rfg3*) chromosomes [15]. In the mapping population, we identified an *Rf* QTL at 807.1–808.7 Mbp, indicative of a minor *Rf* gene (Table S4). While no positional information exists on the *Rfg2* gene, a minor P-type *Rf* gene has been identified in the German inbred line 'L18' with associated markers mapping to 806.1 and 869.5 Mbp [7]. Calculation of pair-wise LD ruled out the presence of spurious association between the potential minor *Rf* gene and the downstream major *Rf* gene, confirming an independent QTL (Table S9). It remains uncertain whether the potential minor *Rf* gene represents a unique gene or the previously described *Rfg2*.

3.3. Decisive Role of 3R in the G-Type CMS Breeding System

The initial genome scan for population-differentiating traits using case control GWAS indicated that the 3R chromosome played a unique role in the G-type CMS system (Figure 1A,B). This is consistent with previous population studies on the assayed germplasm, identifying a singular enrichment of interchromosomal LD for both parental populations on the 3R chromosome, suggesting a conservation of a population-defining trait or traits [52].

To investigate whether the population-differentiating region on 3RL harbored a G-type *Rf* gene, a biparental mapping population was developed. In contrast to the case control GWAS, the biparental mapping population is not subject to confounding issues related to population structure. The segregation ratio of Rf-associated traits in the mapping population was found to be in accordance with a monogenic dominant inheritance of an *Rf* gene by χ^2 test consistent with the singular peak identified in the case control GWAS (Figure 1A,B). Since minor *Rf* genes are incapable of restoring male fertility on their own, they do not influence the segregation ratio of infertile/fertile F_2 progeny and are hence not 'caught' in the χ^2 test. Utilizing the phenotypic dataset on Rf-associated traits from the mapping population, we identified an Rf-associated region on 3RL consistent with findings in the initial case control GWAS (Figure 3A,B). The precise position of the causative *Rf* gene was, however, initially obscured by the finding that the four most associated SNP markers, deriving from the 90K wheat array, could not be accurately mapped to the rye reference genome 'Lo7' [36,59]. Instead, chromosome-wide LD mapping of each Rf-associated wheat marker was conducted, with the top associated marker mapping to 747 Mbp (Figure 4, Table S5). Intriguingly, while the 4R chromosome plays a pivotal role in the P- and C-type CMS systems in rye, housing *Rfp1*, *Rfp2*, *Rfp3*, and *Rfc1*, we found no evidence of G-type genes on 4R in either case control or mapping population GWAS (Figures 1 and 3) [8,9,28,60]. Instead, our findings suggest that the 3R chromosome plays a unique role in the G-type CMS system, possibly harboring both a minor and major *Rf* gene.

3.4. Novel Major Restoration of Male Fertility Gene Unique to the G-Type CMS Breeding System

While no major *Rf* gene on the 3RL chromosome has to our knowledge been identified in either of the known CMS systems in rye, minor genes have been identified in both the G-type and P-type [7,15]. In the case of *Rfg2*, ambiguous segregation ratios of primary trisomics of rye 3R led to the assumption that 3R likely housed a minor G-type *Rf* gene. While Melz and Adolf [15] cautiously interpreted this anomaly as a product of a modifying gene, their findings suggest that something of significance is occurring on 3R in the G-type CMS system. It therefore remains open whether the identified major G-type gene is *Rfg2*, previously misclassified as a minor gene, or an unreported *Rf* gene on 3RL. We propose to denote the major G-type *Rf* gene on 3RL *RfNOS1*.

To our knowledge, no *Rf* gene has been reported on chromosome segments orthologous to rye 3RL in any of the domesticated species residing within the botanical tribe Triticeae. In wheat, major *Rf* genes have been identified on 1AS (*Rf1*), 1BS (*Rf3*), 6AS (*Rf9*), 6BS (*Rf4*, *Rf6*), 6D (*Rf5*), and 7D (*Rf2*) chromosomes [48,57,58,61–63]. In barley (*Hordeum vulgare* L.), two major *Rf* genes have been identified on 6HS (*Rfm1*, *Rfm3*) [47,64]. Intriguingly, Martis et al. [65] discovered that the distal regions of 3RL and 4RL conserved syntenic segments of an ancestral Triticeae chromosome a6. In a comparative analysis, they found that the segment on 3RL portrayed distinctly less collinearity than all other syntenic segments, suggesting a differential evolution of 3RL during rye speciation. On the contrary, the syntenic segment on 4RL was found to be highly conserved in *Brachypodium distachyon* L., rice (*Oryza sativa* L.), sorghum (*Sorghum bicolore* L.), and barley. Rye 4RL, a region housing three major P-type and one C-type *Rf* genes, was found to be syntenic to barley 6HS [60,65]. These results are consistent with the findings of Hackauf et al. [50], who reported that the segment housing *Rfp1* on 4RL exhibits an ortholog on wheat 6DS and barley 6H. In a subsequent study, Hackauf et al. [8], furthermore, proposed that *Rfp3* on 4RL likely maps to an orthologous segment housing *Rf6* on wheat 6BS and

Rfm1 on barley 6HS [62]. These findings provide further evidence of a conserved synteny between these chromosomal segments and the conservation of *Rf* genes pivotal to the fertilization control systems across rye (P- and C-type), barley, and wheat [51]. Together, this accentuates the novelty of our discovery of a major *Rf* gene, *RfNOS1*, on 3RL in the G-type CMS system in rye, with no known *Rf* genes on orthologous chromosome segments in other Triticeae species.

3.5. Non-Pentatricopeptide Repeat Restoration of Male Fertility Gene on 3RL

The majority of characterized *Rf* genes have been assigned to the PPR superfamily, denoted as *Rf*-like PPR genes or RFL-PPRs. In domesticated Poaceae species, this includes wheat *Rf1* and *Rf3* [39], barley *Rfm1* [35], sorghum *Rf1* [40], maize *Rf5* [41], and rice *Rf4*, *Rf5*, and *Rf6* [42–44]. A genome scan for Rf-associated genes at the site on 3RL harboring the major G-type *Rf* gene *RfNOS1* in the 'Lo7' reference genome revealed no RFL-PPR genes. Instead, the region harbored five PLS-class PPR genes either misclassified, not resembling known PLS-type RFL-PPR genes, or pointing towards a non-PPR *Rf* gene on 3RL (Figure 1D, Table S8a). Out of 83 RFL-PPRs annotated in the 'Lo7' reference genome, none of them belong to the PLS class. Whereas a less prevalent group, PLS-type RFL-PPR genes exists, as observed in the case of *Rfm1* in barley [35]. It, therefore, cannot be ruled out that the RFL-PPR annotation of 'Lo7' is incomplete, inclined to annotate the prevalent P-class as RFL-PPR genes, while disregarding the less well-represented PLS class. However, both of the PLS-class PPR genes residing in the Rf-associated region on 3RL that were co-expressed in the two G-type hybrids resided more than 20 Mbp from the top associated peak in the 600K case control GWAS (Figure 1D, Table S8b). In conjunction, this suggests that the major *Rf* gene on 3RL belongs to the non-PPR family.

A growing body of *Rf* genes is now being characterized as non-PPR *Rf* genes, adding to the complexity of male fertility restoration. Until now, this includes glycine-rich proteins (Rf2, in rice; [66]), acyl-carrier protein synthase (Rf17, in rice; [67]), aldehyde dehydrogenase (Rf2, in maize; [68]), bHLH transcription factor (Rf4, in maize; [69]), and *Rf1*, a peptidase, in sugar beet (*Beta vulgaris* L.) [70]. The region on 3RL harbored several genes related to plant fertility that were co-expressed in both G-type hybrids at anthesis (Table S7a). Amongst these were two KATANIN genes located at 730.3 and 759.9 Mbp. In *Arabidopsis*, KATANIN is required for fertility, embryo development, and seed production [71]. Genes involved in flower development, including MADS-box transcription factor [72–74], BZIP transcription factor [75], and basic helix-loop-helix (bHLH) DNA-binding superfamily protein [69] were likewise found co-expressed within the region.

Intriguingly, the region was found to harbor a single mTERF gene located at 731 Mbp, coinciding with the top-associated peak in the 600K case control GWAS (Figure 1D, Table S8b). In rye, Hackauf et al. [8] reported a close linkage between *Rfp1* and *Rfp3* on 4RL to mTERF genes. A similar observation was made by Bernhard et al. [47], who identified two mTERF genes closely linked to *Rfm3* on barley 6HS, syntenic to rye 4RL. Pan et al. [76] successively observed the role of mTERF genes in kernel development in maize, connecting the gene family to the reproductive system of plants. This newly identified mTERF gene, which is expressed at flowering in both G-type hybrids, is a potential candidate for the novel major G-type *Rf* gene on 3RL. We denote the gene as *RfNOS1*, belonging to the expanding class of non-PPR *Rf* genes.

3.6. CMS Systems in Hybrid Rye Breeding

On the basis of male fertility restoration requirements and genetic similarity of steriliz-ing cytoplasms, G-, C-, and R-type CMS systems have been proposed to belong to the larger Vavilovii (V) type [6,77,78]. In a comprehensive study by Lapinski and Stojalowski [6] on 50 rye populations from 23 countries, the vast majority of male sterility sources were found to belong to the V-type. Populations with European descent were predominantly found to carry the V-type sterility-inducing cytoplasm, while the P-type was exclusively observed in lines descending from South America. Nonetheless, with no previous report of a major

Rf gene on 3RL in either R, S, or C-type CMS system, such a unilateral grouping as V-type seems premature. From our observations, the G-type CMS system distinguishes itself from the other CMS systems by a less pivotal role of *Rf* genes on 4RL.

4. Methods

4.1. Plant Material

In total, 365 Nordic Seed Germany GmbH inbred hybrid rye (*Secale cereale* L.) elite breeding lines were selected for this study, comprising 242 restorers, 116 non-restorer germplasm (NRG), and 7 cytoplasmic male sterile (CMS) lines. The CMS male sterility is based on the 'Gülzow' (G)-type cytoplasm originating from the population of the rye variety 'Schlägler alt' [11,15]. Population structure and information on genetic characteristics of the germplasm are presented in a recent study by Vendelbo et al. [52]. A biparental mapping population was developed from a hybrid rye cv. Stannos. DNA extraction was performed using an adapted SDS-based method according to USDA [79] after Pallotta et al. [80] on an equivalent of 75 mg of plant material collected from the primary leaves of two seven-day-old, seedlings per line. DNA concentration and 260/280 nm ratio of samples were measured using an Epoch^{TM} microplate spectrophotometer (Biotek$^{®}$ Vermont, Winooski, VT, USA) and evidence of fragmentation by size visualization on a 1.2% agarose gel.

4.2. Biparental Mapping Population

To investigate the inheritance of male fertility restoration in the G-type CMS-based Nordic Seed breeding system, a biparental mapping population was developed. The population was phenotyped for restoration of male fertility, including traits associated with restoration. Seeds of the hybrid cv. Stannos (F_1) were sown in pots containing a course-grain sphagnum substrate at Nordic Seed Germany GmbH greenhouse facilities. The seedlings were cultivated under a 16 h light regime with night temperatures of 14–16 °C and day temperatures of 18–24 °C. Seven days after sowing, at the 2-leaf stage, seedlings were set to vernalize in a climate chamber under 16 h of light at 8 °C for a week and then 3 °C for the following seven weeks. After vernalization, the pots were transferred to the greenhouse. Prior to anther-protrusion, cellophane bags were put on the spikes to prevent cross-fertilization. At maturity, seeds of a single F_1 plant were harvested and the procedure was repeated to generate an F_2 biparental mapping population. At four timepoints, a total of 181 F2 plants were phenotyped for pollen production using a customized visual 1–9 scale (1: no pollen, 9: large amount of pollen). The plants were, furthermore, scored for number of spikes per plant, total seed number, seeds per spike, total grain weight, and thousand kernel weight in order to obtain a comprehensive phenotypic dataset on the inheritance of male fertility restoration in the population. The segregation ratio of infertile and fertile F_2 plants was tested for goodness of fit to the expected Mendelian ratio at the scenario of one, two, and three major restorations of male fertility (*Rf*) genes using an χ^2 test [81]. An F_2 plant was considered 'sterile' if it either yielded less than 20 seeds or scored ≤ 2 in pollen production.

4.3. Molecular Markers

All rye lines included in this investigation were genotyped using a custom Illumina Infinium 15K$_{wheat}$ [59] and 5K$_{Rye}$ [82,83] single nucleotide polymorphism (SNP) array, denoted 20K, as described by Vendelbo et al. (2020). In addition, 180 lines comprising 88 NRG and 92 restorer lines were also genotyped using the state-of-the-art 600K high-density rye array by Bauer et al. [83]. The F_2 biparental mapping population was genotyped on a custom Illumina Infinium 25K$_{wheat}$ and 5K$_{Rye}$ SNP array, denoted 30K, enriched with an additional 10K wheat markers, compared to the previous 20K array, deriving from the 90K wheat SNP array by Wang et al. [59]. The mapping position of SNP markers derived from the 90K wheat array was found by mapping the marker sequences to the 'Lo7' reference genome using the NCBI blastn (v. 2.9.0+, ML, USA) function at a significance threshold of the e-value at 10^{-5}, selecting the physical position of the top hit [36,84]. The

position of *Rf*-associated markers without accurate mapping position in the 'Lo7' reference genome was determined by calculation of pairwise LD across the entire mapped marker panel. LD was calculated using snpStats (v. 1.36.0, Cambridge, UK) R package with LD set as the coefficient of determination (r^2) [85].

4.4. Data Analysis

Genetic analysis of SNP marker data was performed in R studio (v. 1.1.463, Boston, MA, USA) interface in R statistical software (v. 3.6.3) by application of various predesigned packages [86,87]. Prior to analysis, markers were filtered for marker allele frequency $\geq$0.005, missing individual score $\leq$ 0.2 and missing marker score $\leq$ 0.1 to identify informative markers. The polymorphism information content (PIC) was calculated as an estimate of marker informativeness using SnpReady (v. 0.9.6, Laguna, Philippines) [88].

4.5. Genome-Wide Association Study

The discovery of *Rf*-associated regions was made by a genome-wide association study (GWAS) using the genomic association and prediction integration tool (GAPIT) (v.3) package in R [55]. Phenotypic input for GWAS included all recordings of the biparental F_2 population, and a binary case control for the entire population relative to their population origin using the 20K SNP array and 600K high-density SNP array, respectively. A standard Bonferroni-corrected threshold of α = 0.05 was used as the significance threshold

4.6. RNA-Seq Data Expression Analysis of PPR and mTERF Genes Residing in Rf-Associated Region in G-Type Hybrids of Rye

Annotated major *Rf*-associated genes, pentatricopeptide repeat proteins (PPR), Rf-like PPR, and mitochondrial transcription termination factors (mTERF), residing in regions associated with Rf were identified in the 'Lo7' reference genome [36]. To ascertain the potential causative Rf gene or genes amongst the identified panel, gene expression was investigated in two G-type Nordic Seed hybrid cv. Helltop and cv. Stannos de novo transcriptome assemblies. The transcript data were obtained from spikes of the G-type hybrids at flowering. High-quality de novo transcriptome assembly of the two hybrids has recently been published by Mahmood et al. [89] and raw reads from the transcript library have been made accessible in a sequence read archive repository at (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA612415, accessed on 12 March 2021). Coding and protein sequences of the Rf-associated gene panel were extracted from the 'Lo7' reference genome repository. The de novo assembled transcriptome assemblies were likewise translated to coding protein sequences with TransDecoder (v. 5.5.0, Valencia, Spain). The NCBI blastx and blastp (v. 2.9.0+, Maryland, USA) functions were used to blast coding and protein sequences of Rf-associated genes in the panel towards the generated protein transcriptome database at a significance threshold of the e-value at 10^{-5} to ascertain genes expressed in the two G-type hybrid cv. at flowering stage [84].

4.7. Graphical Editing

Graphs and figures were outputted from R in .svg format and manually curated using Inkscape (v. 1.1) program (https://inkscape.org/, accessed on 7 May 2021).

5. Conclusions

In this study, we exploited recent advents in rye genomic resources to dissect the genetics underlying restoration of male fertility in a G-type CMS system. Our findings provide compelling evidence of a novel major G-type *Rf* gene on 3RL with no known orthologues in either barley or wheat. Gene mining of the Rf-associated region on 3RL led to the identification of an mTERF gene co-expressed in two G-type hybrids as a candidate gene for restoration. We propose to denote the novel G-type *Rf* gene as *RfNOS1*. Conclusively, our investigation provides a novel insight into the genetics of male fertility restoration in

a G-type CMS system and its differentiation to rye P- and C-types in addition to known CMS systems in barley and wheat.

References

1. Miedaner, T.; Hübner, M. Quality demands for different uses of hybrid rye. *Tag. Ver. Pflanz. Saatgutkaufleute Osterr.* **2011**, *61*, 45–49.
2. Laidig, F.; Piepho, H.P.; Rentel, D.; Drobek, T.; Meyer, U.; Huesken, A. Breeding progress, variation, and correlation of grain and quality traits in winter rye hybrid and population varieties and national on-farm progress in Germany over 26 years. *Theor. Appl. Genet.* **2017**, *130*, 981–998. [CrossRef] [PubMed]
3. Geiger, H.H.; Miedaner, T. Hybrid rye and heterosis. In *Genetics and Exploitation of Heterosis in Crops*; American Society of Agronomy-Crop Science Society of America: Madison, WI, USA, 1999; pp. 439–450.
4. Chen, L.; Liu, Y.G. Male sterility and fertility restoration in crops. *Annu. Rev. Plant Biol.* **2014**, *65*, 579–606. [CrossRef] [PubMed]
5. Kim, Y.J.; Zhang, D. Molecular control of male fertility for crop hybrid breeding. *Trends Plant Sci.* **2018**, *23*, 53–65. [CrossRef] [PubMed]
6. Lapinski, M.; Stojalowski, S. Occurence and genetic identity of male sterility-inducing cytoplasm in rye. *Plant Breed. Seed Sci.* **2003**, *48*, 7–23.
7. Miedaner, T.; Glass, C.; Dreyer, F.; Wilde, P.; Wortmann, H.; Geiger, H.H. Mapping of genes for male-fertility restoration in 'Pampa' CMS winter rye (*Secale cereale* L.). *Theor. Appl. Genet.* **2000**, *101*, 1226–1233. [CrossRef]
8. Hackauf, B.; Bauer, E.; Korzun, V.; Miedaner, T. Fine mapping of the restorer gene Rfp3 from an Iranian primitive rye (*Secale cereale* L.). *Theor. Appl. Genet.* **2017**, *130*, 1179–1189. [CrossRef]
9. Stracke, S.; Schilling, A.G.; Forster, J.; Weiss, C.; Glass, C.; Miedaner, T.; Geiger, H.H. Development of PCR-based markers linked to dominant genes for male-fertility restoration in Pampa CMS of rye (*Secale cereale* L.). *Theor. Appl. Genet.* **2003**, *106*, 1184–1190. [CrossRef]
10. Niedziela, A.; Brukwinski, W.; Bednarek, P.T. Genetic mapping of pollen fertility restoration QTLs in rye (*Secale cereale* L.) with CMS Pampa. *J. Appl. Genet.* **2021**, *62*, 1–14. [CrossRef]
11. Adolf, K. A new source of spontaneous sterility in winter rye-preliminary results. *Proc. EUCARPIA Meet. Cereal Sect. Rye* **1986**, *1985*, 11–13.
12. Kobylianskii, V.D. Production of sterile analogues of winter rye varieties, sterile mantainers and fertile restorers. *Trudy Po Prikladnoi Botanike Genetike Selektsii* **1971**, 76–85.
13. Lapinski, M. Cytoplasmic-genic type of male sterility in Secale montanum Guss. *Wheat Inf. Serv.* **1972**, *35*, 25–28.
14. Madej, L. Research on male sterility in rye. *Hodowla Rosl Aklim Nasienn* **1975**, *10*, 590–593.
15. Melz, G.; Adolf, K. Genetic analysis of rye (*Secale cereale* L.) genetics of male sterility of the G-type. *Theor. Appl. Genet.* **1991**, *82*, 761–764. [CrossRef]

16. Milczarski, P.; Hanek, M.; Tyrka, M.; Stojalowski, S. The application of GBS markers for extending the dense genetic map of rye (*Secale cereale* L.) and the localization of the Rfc1 gene restoring male fertility in plants with the C source of sterility-inducing cytoplasm. *J. Appl. Genet.* **2016**, *57*, 439–451. [CrossRef] [PubMed]

17. Stojalowski, S.; Apińaski, M.; Masojć, P. RAPD markers linked with restorer genes for the C-sources of cytoplasmic male sterility in rye (*Secale cereale* L.). *Plant Breed.* **2004**, *123*, 428–433. [CrossRef]

18. Stojalowski, S.; Jaciubek, M.; Masojć, P. Rye SCAR markers for male fertility restoration in the P cytoplasm are also applicable to marker-assisted selection in C cytoplasn. *J. Appl. Genet.* **2005**, *46*, 371–373. [PubMed]

19. Yuan, Y. *Umweltstabilität der Cytoplasmitsch-Genisch Vererbten Männlichen Sterilität (CMS) bei Roggen (Secale cereale L.)*; Verlag UE Grauer: Stuttgart, Germany, 1995.

20. Geiger, H.H.; Yuan, Y.; Miedaner, T.; Wilde, P. Environmental sensitivity of cytoplasmic genic male sterility (CMS) in *Secale cereale* L. *Fortschr. Pflanz.* **1995**, *18*, 7–18.

21. Kodisch, A.; Wilde, P.; Schmiedchen, B.; Fromme, F.-J.; Rodemann, B.; Tratwal, A.; Oberforster, M.; Wieser, F.; Schiemann, A.; Jørgensen, L.N.; et al. Ergot infection in winter rye hybrids shows differential contribution of male and female genotypes and environment. *Euphytica* **2020**, *216*, 1–14. [CrossRef]

22. Klotz, J.L. Activities and effects of ergot alkaloids on livestock physiology and production. *Toxins* **2015**, *7*, 2801–2821. [CrossRef] [PubMed]

23. Blaney, B.J.; Molloy, J.B.; Brock, I.J. Alkaloids in Australian rye ergot (Claviceps purpurea) sclerotia: Implication for food and stockfeed regulations. *Anim. Prod. Sci.* **2009**, *49*, 975–982. [CrossRef]

24. Miedaner, T.; Mirdita, V.; Rodemann, B.; Drobeck, T.; Rentel, D. Genetic variation of winter rye cultivars for their ergot (Claviceps purpurea) reaction tested in a field design with minimized interplot interference. *Plant Breed.* **2010**, *129*, 58–62. [CrossRef]

25. Geiger, H.H.; Miedaner, T. Genetic basis and phenotypic stability of male-fertility restoration in rye. *Vorträge Pflanz.* **1996**, 27–38.

26. Miedaner, T.; Wilde, P.; Wortmann, H. Combining ability of non-adapted sources for male-fertility restoration in Pampa CMS of hybrid rye. *Plant Breed.* **2004**, *124*, 39–43. [CrossRef]

27. Falke, K.C.; Wilde, P.; Miedaner, T. Rye introgression lines as source of alleles for pollen-fertility restoration in pampa cms. *Plant Breed.* **2009**, *128*, 528–531. [CrossRef]

28. Miedaner, T.; Herter, C.P.; Goßlau, H.; Wilde, P.; Hackauf, B.; Pillen, K. Correlated effects of exotic pollen-fertility restorer genes on agronomic and quality traits of hybrid rye. *Plant Breed.* **2017**, *136*, 224–229. [CrossRef]

29. KWS. PollenPlus. Available online: https://www.kws.com/gb/en/products/cereals/hybrid-rye/pollenplus-kws-files-for-ergot-patent-in-hybrid-rye/ (accessed on 30 June 2021).

30. Melz, G.; Melz, G.; Hartman, F. Genetics of a male-sterile rye of 'G-type' with results of the first F1-hybrids. *Plant Breed. Seed Sci.* **2003**, *47*, 47–55.

31. Gaborieau, L.; Brown, G.G.; Mireau, H. The propensity of pentatricopeptide repeat genes to evolve into restorers of cytoplasmic male sterility. *Front. Plant Sci.* **2016**, *7*, 1816. [CrossRef]

32. Gully, B.S.; Cowieson, N.; Stanley, W.A.; Shearston, K.; Small, I.D.; Barkan, A.; Bond, C.S. The solution structure of the pentatricopeptide repeat protein PPR10 upon binding atpH RNA. *Nucleic Acids Res.* **2015**, *43*, 1918–1926. [CrossRef]

33. Schmitz-Linneweber, C.; Small, I. Pentatricopeptide repeat proteins: A socket set for organelle gene expression. *Trends Plant Sci.* **2008**, *13*, 663–670. [CrossRef]

34. Ban, T.; Ke, J.; Chen, R.; Gu, X.; Tan, M.H.; Zhou, X.E.; Kang, Y.; Melcher, K.; Zhu, J.K.; Xu, H.E. Structure of a PLS-class pentatricopeptide repeat protein provides insights into mechanism of RNA recognition. *J. Biol. Chem.* **2013**, *288*, 31540–31548. [CrossRef] [PubMed]

35. Rizzolatti, C.; Bury, P.; Tatara, E.; Pin, P.A.; Rodde, N.; Bergès, H.; Budar, F.; Mireau, H.; Gielen, J.J.L. Map-based cloning of the fertility restoration locus Rfm1 in cultivated barley (Hordeum vulgare). *Euphytica* **2017**, *213*. [CrossRef]

36. Rabanus-Wallace, M.T.; Hackauf, B.; Mascher, M.; Lux, T.; Wicker, T.; Gundlach, H.; Báez, M.; Houben, A.; Mayer, K.F.X.; Guo, L.; et al. Chromosome-scale genome assembly provides insights into rye biology, evolution and agronomic potential. *Nat. Genet.* **2021**, *53*, 564–573. [CrossRef] [PubMed]

37. Qi, W.; Yang, Y.; Feng, X.; Zhang, M.; Song, R. Mitochondrial function and maize kernel development requires Dek2, a pentatricopeptide repeat protein involved in nad1 mRNA splicing. *Genetics* **2017**, *205*, 239–249. [CrossRef] [PubMed]

38. Liu, Y.J.; Liu, X.; Chen, H.; Zheng, P.; Wang, W.; Wang, L.; Zhang, J.; Tu, J. A plastid-localized pentatricopeptide repeat protein is required for both pollen development and plant growth in rice. *Sci. Rep.* **2017**, *7*, 1–12. [CrossRef]

39. Melonek, J.; Duarte, J.; Martin, J.; Beuf, L.; Murigneux, A.; Varenne, P.; Comadran, J.; Specel, S.; Levadoux, S.; Bernath-Levin, K.; et al. The genetic basis of cytoplasmic male sterility and fertility restoration in wheat. *Nat. Commun.* **2021**, *12*, 1036. [CrossRef]

40. Klein, R.R.; Klein, P.E.; Mullet, J.E.; Minx, P.; Rooney, W.L.; Schertz, K.F. Fertility restorer locus Rf1 [corrected] of sorghum (Sorghum bicolor L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. *Theor. Appl. Genet.* **2005**, *111*, 994–1012. [CrossRef]

41. Beick, S.; Schmitz-Linneweber, C.; Williams-Carrier, R.; Jensen, B.; Barkan, A. The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. *Mol. Cell. Biol.* **2008**, *28*, 5337–5547. [CrossRef] [PubMed]

42. Kazama, T.; Toriyama, K. A fertility restorer gene, Rf4, widely used for hybrid rice breeding encodes a pentatricopeptide repeat protein. *Rice* **2014**, *7*, 1–5. [CrossRef]

43. Hu, J.; Wang, K.; Huang, W.; Liu, G.; Gao, Y.; Wang, J.; Huang, Q.; Ji, Y.; Qin, X.; Wan, L.; et al. The rice pentatricopeptide repeat protein RF5 restores fertility in Hong-Lian cytoplasmic male-sterile lines via a complex with the glycine-rich protein GRP162. *Plant Cell* **2012**, *24*, 109–122. [CrossRef]

44. Huang, W.; Yu, C.; Hu, J.; Wang, L.; Dan, Z.; Zhou, W.; He, C.; Zeng, Y.; Yao, G.; Qi, J.; et al. Pentatricopeptide-repeat family protein RF6 functions with hexokinase 6 to rescue rice cytoplasmic male sterility. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 14984–14989. [CrossRef]

45. Barkan, A.; Small, I. Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* **2014**, *65*, 415–442. [CrossRef]

46. Robles, P.; Quesada, V. Research progress in the molecular functions of plant mTERF proteins. *Cells* **2021**, *10*, 205. [CrossRef]

47. Bernhard, T.; Koch, M.; Snowdon, R.J.; Friedt, W.; Wittkop, B. Undesired fertility restoration in msm1 barley associates with two mTERF genes. *Theor. Appl. Genet.* **2019**, *132*, 1335–1350. [CrossRef]

48. Shahinnia, F.; Geyer, M.; Block, A.; Mohler, V.; Hartl, L. Identification of Rf9, a gene contributing to the genetic complexity of fertility restoration in hybrid wheat. *Front. Plant Sci.* **2020**, *11*, 1720. [CrossRef] [PubMed]

49. Wilde, P.; Korzun, V.; Menzel, J.; Ruonan, Z.; Stein, N.; Hackauf, B. Restorer Plants. KWS SAAT SE. U.S. Patent App. 16/064304, 9 May 2019.

50. Hackauf, B.; Korzun, V.; Wortmann, H.; Wilde, P.; Wehling, P. Development of conserved ortholog set markers linked to the restorer gene Rfp1 in rye. *Mol. Breed.* **2012**, *30*, 1507–1518. [CrossRef]

51. Börner, A.; Korzun, V.; Polley, A.; Malyshev, S.; Melz, G. Genetics and molecular mapping of a male fertility restoration locus (Rfg1) in rye (*Secale cereale* L.). *Theor. Appl. Genet.* **1998**, *97*, 99–102. [CrossRef]

52. Vendelbo, N.M.; Sarup, P.; Orabi, J.; Kristensen, P.S.; Jahoor, A. Genetic structure of a germplasm for hybrid breeding in rye (*Secale cereale* L.). *PLoS ONE* **2020**, *15*, e0239541. [CrossRef]

53. Vilhjalmsson, B.J.; Nordborg, M. The nature of confounding in genome-wide association studies. *Nat. Rev. Genet.* **2013**, *14*, 1–2. [CrossRef]

54. Korte, A.; Farlow, A. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* **2013**, *9*, 1–9. [CrossRef]

55. Lipka, A.E.; Tian, F.; Wang, Q.; Peiffer, J.; Li, M.; Bradbury, P.J.; Gore, M.A.; Buckler, E.S.; Zhang, Z. GAPIT: Genome association and prediction integrated tool. *Bioinformatics* **2012**, *28*, 2397–2399. [CrossRef]

56. Zhang, Z.; Ersoz, E.; Lai, C.Q.; Todhunter, R.J.; Tiwari, H.K.; Gore, M.A.; Bradbury, P.J.; Yu, J.; Arnett, D.K.; Ordovas, J.M.; et al. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **2010**, *42*, 355–360. [CrossRef] [PubMed]

57. Geyer, M.; Albrecht, T.; Hartl, L.; Mohler, V. Exploring the genetics of fertility restoration controlled by Rf1 in common wheat (Triticum aestivum L.) using high-density linkage maps. *Mol. Genet. Genom.* **2018**, *293*, 451–462. [CrossRef]

58. Wurschum, T.; Leiser, W.L.; Weissmann, S.; Maurer, H.P. Genetic architecture of male fertility restoration of Triticum timopheevii cytoplasm and fine-mapping of the major restorer locus Rf3 on chromosome 1B. *Theor. Appl. Genet.* **2017**, *130*, 1253–1266. [CrossRef] [PubMed]

59. Wang, S.; Wong, D.; Forrest, K.; Allen, A.; Chao, S.; Huang, B.E.; Maccaferri, M.; Salvi, S.; Milner, S.G.; Cattivelli, L.; et al. Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol. J.* **2014**, *12*, 787–796. [CrossRef] [PubMed]

60. Stojalowski, S.A.; Milczarski, P.; Hanek, M.; Bolibok-Bragoszewska, H.; Myskow, B.; Kilian, A.; Rakoczy-Trojanowska, M. DArT markers tightly linked with the Rfc1 gene controlling restoration of male fertility in the CMS-C system in cultivated rye (*Secale cereale* L.). *J. Appl. Genet.* **2011**, *52*, 313–318. [CrossRef]

61. Maan, S.S.; Luchen, K.A.; Bravo, J.M. Genetic analyses of male-fertility restoration in wheat. I. Chromosomal location of Rf genes 1. *Crop. Sci.* **1984**, *24*, 17–20. [CrossRef]

62. Ma, Z.Q.; Zhao, Y.H.; Sorrells, M.E. Inheritance and chromosomal locations of male fertility restoring gene transferred from Aegilops umbellulata Zhuk. to *Triticum aestivum* L. *Mol. Gen. Genet. MGG* **1995**, *247*, 351–357. [CrossRef]

63. Du, H.; Maan, S.S.; Hammond, J.J. Genetic analysis of male-fertility restoration in wheat: III. Effects of aneuploidy. *Crop. Sci.* **1991**, *31*, 319–322. [CrossRef]

64. Ui, H.; Sameri, M.; Pourkheirandish, M.; Chang, M.C.; Shimada, H.; Stein, N.; Komatsuda, T.; Handa, H. High-resolution genetic mapping and physical map construction for the fertility restorer Rfm1 locus in barley. *Theor. Appl. Genet.* **2015**, *128*, 283–290. [CrossRef]

65. Martis, M.M.; Zhou, R.; Haseneyer, G.; Schmutzer, T.; Vrana, J.; Kubalakova, M.; Konig, S.; Kugler, K.G.; Scholz, U.; Hackauf, B.; et al. Reticulate evolution of the rye genome. *Plant Cell* **2013**, *25*, 3685–3698. [CrossRef]

66. Itabashi, E.; Iwata, N.; Fujii, S.; Kazama, T.; Toriyama, K. The fertility restorer gene, Rf2, for lead rice-type cytoplasmic male sterility of rice encodes a mitochondrial glycine-rich protein. *Plant J.* **2011**, *65*, 359–367. [CrossRef]

67. Fujii, S.; Toriyama, K. Suppressed expression of retrogrede-regulated male sterility restores pollen fertility in cytoplasmic male sterile rice plants. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9513–9518. [CrossRef]

68. Liu, F.; Schnable, P.S. Functional specialization of maize mitochondrial aldehyde dehydrogenases. *Plant Physiol.* **2002**, *130*, 1657–1674. [CrossRef]

69. Jaqueth, J.S.; Hou, Z.; Zheng, P.; Ren, R.; Nagel, B.A.; Cutter, G.; Niu, X.; Vollbrecht, E.; Greene, T.W.; Kumpatla, S.P. Fertility restoration of maize CMS-C altered by a single amino acid substitution within the Rf4 bHLH transcription factor. *Plant J.* **2020**, *101*, 101–111. [CrossRef]

70. Kitazaki, K.; Arakawa, T.; Matsunaga, M.; Yui-Kurino, R.; Matsuhira, H.; Mikami, T.; Kubo, T. Post-translational mechanisms are associated with fertility restoration of cytoplasmic male sterility in sugar beet (*Beta vulgaris*). *Plant J.* **2015**, *83*, 290–299. [CrossRef] [PubMed]

71. Luptovčiak, I.; Samakovli, D.; Komis, G.; Šamaj, J. KATANIN1 is essential for embryogenesis and seed formation in Arabidopsis. *Front. Plant Sci.* **2017**, *8*, 728. [CrossRef] [PubMed]

72. Schreiber, D.N.; Bantin, J.; Dresselhaus, T. The MADS box transcription factor ZmMADS2 is required for anther and pollen maturation in maize and accumulates in apoptotic bodies during anther dehiscence. *Plant Physiol.* **2004**, *134*, 1069–1079. [CrossRef] [PubMed]

73. Zhou, L. A MADS-box transcription factor related to fertility conversion in male sterile wheat lines. *Acta Agron. Sin.* **2008**, *34*, 598. [CrossRef]

74. Linke, B.; Nothnagel, T.; Börner, T. Flower development in carrot CMS plants: Mitochondria affect the expression of MADS box genes homologous to GLOBOSA and DEFICIENS. *Plant J.* **2003**, *34*, 27–37. [CrossRef]

75. Zou, M.; Guan, Y.; Ren, H.; Zhang, F.; Chen, F. A bZIP transcription factor, OsABI5, is involved in rice fertility and stress tolerance. *Plant Mol. Biol.* **2008**, *66*, 675–683. [CrossRef]

76. Pan, Z.; Ren, X.; Zhao, H.; Liu, L.; Tan, Z.; Qiu, F. A mitochondrial transcription termination factor, ZmSmk3, is required for nad1 intron4 and nad4 intron1 splicing and kernel development in maize. *G3 Genes Genomes Genet.* **2019**, *9*, 2677–2686. [CrossRef]

77. Lapinski, M.; Stojalowski, S. The C-source of sterility-inducing cytoplasm in rye: Origin, identity and occurrence. *Vorträge Pflanz.* **1996**, *35*, 51–60.

78. Warsecha, R.; Salak-Warzecha, K. Comparative studies on CMS sources in rye. *Vorträge Pflanz.* **1996**, *35*, 39–49.

79. USDA. United States Department of Agriculture: Wheat and Barley DNA Extraction Protocol (96-Well Plate Format). Available online: https://www.ars.usda.gov/ARSUserFiles/60701500/SmallGrainsGenotypingLaboratory/Protocols/wheat%20and%20barleyDNA%20extraction_original.pdf (accessed on 7 July 2021).

80. Pallotta, M.A.; Warner, P.; Fox, R.L.; Kuchel, H.; Jefferies, S.J.; Langridge, P. Marker assisted wheat breeding in the southern region of Australia. *Proc. Tenth Int. Wheat Genet. Symp.* **2003**, 789–791.

81. McHugh, M.L. The chi-square test of independence. *Biochem. Med.* **2013**, *23*, 143–149. [CrossRef]

82. Haseneyer, G.; Schmutzer, T.; Seidel, M.; Zhou, R.; Mascher, M.; Schon, C.C.; Taudien, S.; Scholz, U.; Stein, N.; Mayer, K.F.; et al. From RNA-seq to large-scale genotyping-genomics resources for rye (*Secale cereale* L.). *BMC Plant Biol.* **2011**, *11*, 131. [CrossRef] [PubMed]

83. Bauer, E.; Schmutzer, T.; Barilar, I.; Mascher, M.; Gundlach, H.; Martis, M.M.; Twardziok, S.O.; Hackauf, B.; Gordillo, A.; Wilde, P.; et al. Towards a whole-genome sequence for rye (*Secale cereale* L.). *Plant J.* **2017**, *89*, 853–869. [CrossRef]

84. NCBI. National Center for Biotechnology Information. Available online: https://www.ncbi.nlm.nih.gov (accessed on 13 June 2021).

85. Clayton, D. snpStats: SnpMatrix and XSnpMatrix Classes and Methods. R Package Version 1.36.0. 2019. Available online: https://rdrr.io/bioc/snpStats/ (accessed on 7 July 2021).

86. Rstudio Team. Rstudio: Integrated development for R. RStudio, Inc., Boston. 2015. Available online: http://www.rstudio.com (accessed on 15 June 2021).

87. R Core Team. *R a Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021; Available online: https://www.R-project.org/ (accessed on 13 June 2021).

88. Granato, I.S.C.; Galli, G.; de Oliveira Couto, E.G.; e Souza, M.B.; Mendonça, L.F.; Fritsche-Neto, R. snpReady: A tool to assist breeders in genomic analysis. *Mol. Breed.* **2018**, *38*, 102. [CrossRef]

89. Mahmood, K.; Orabi, J.; Kristensen, P.S.; Sarup, P.; Jorgensen, L.N.; Jahoor, A. De novo transcriptome assembly, functional annotation, and expression profiling of rye (*Secale cereale* L.) hybrids inoculated with ergot (*Claviceps purpurea*). *Sci. Rep.* **2020**, *10*, 13475. [CrossRef] [PubMed]