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

Nikolaj Vendelbo (nive@nordicseed.com)
Aarhus University / Nordic Seed A/S
Khalid Mahmood
Nordic Seed A/S
Pernille Sarup
Nordic Seed A/S
Peter Kristensen
Nordic Seed A/S
Jihad Orabi
Nordic Seed A/S
Ahmed Jahoor
Nordic Seed A/S

Research Article

Keywords: Restoration of male fertility (Rf), Cytoplasmic male sterility (CMS), Pentatricopeptide repeat pro-tein (PPR), Mitochondrial transcription termination factor (mTERF), 600K SNP array, Genome-wide association study (GWAS), RNAseq, chi-square, Linkage disequilibrium

Posted Date: August 31st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-275908/v2

Version of Record: A version of this preprint was published at International Journal of Molecular Sciences on August 27th, 2021. See the published version at https://doi.org/10.3390/ijms22179277.

Genetic architecture of male fertility restoration in a hybrid breeding system of rye (*Secale cereale* L.)

Nikolaj M. Vendelbo*[‡1,2], Khalid Mahmood [1‡], Pernille Sarup [1], Peter S. Kristensen [1], Jihad Orabi [1], Ahmed Jahoor [1,3]

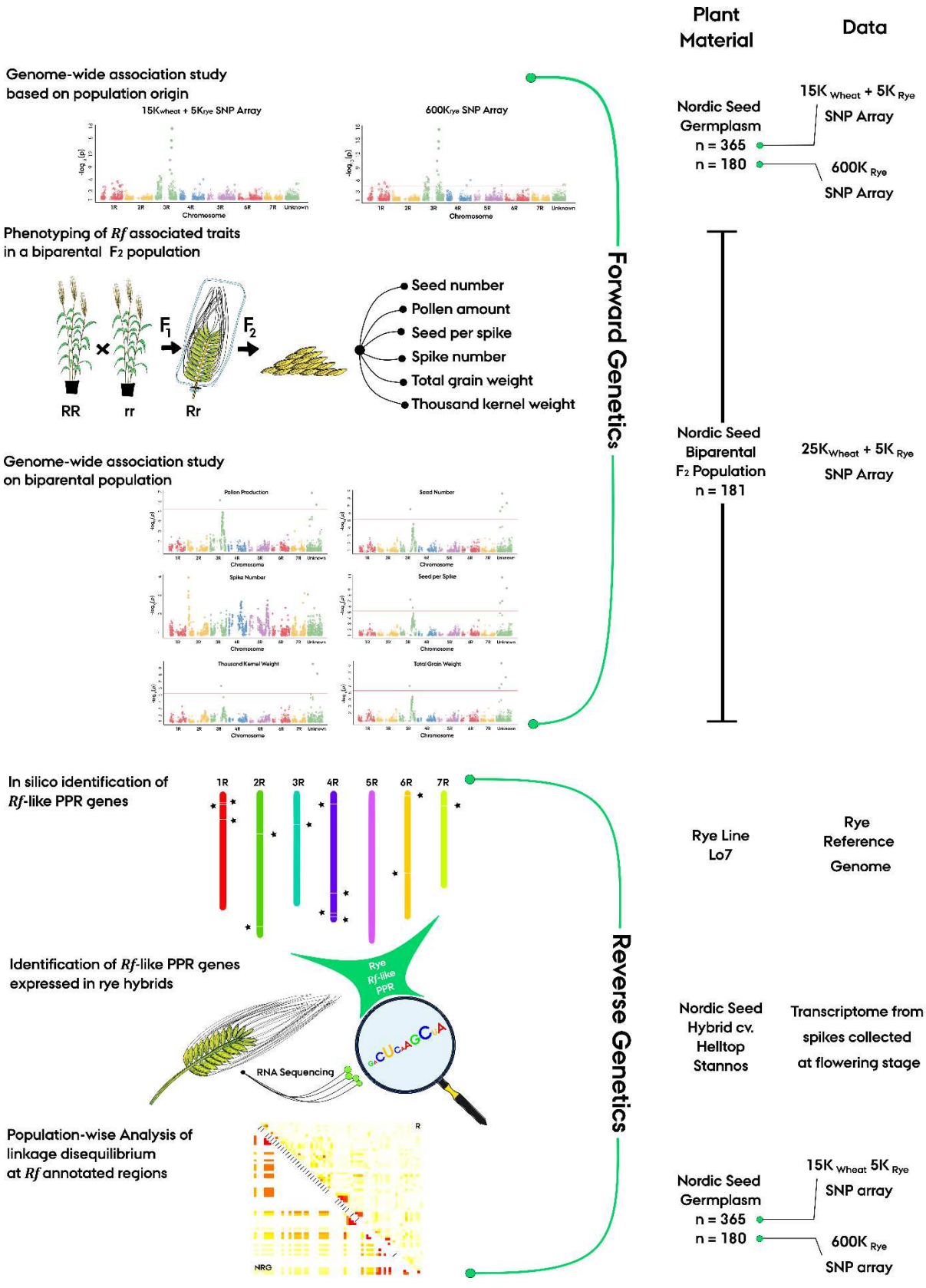
[1] Nordic Seed A/S, Grindsnabevej 25, Odder, 8300, Denmark
[2] Department of Agroecology, Faculty of Technology, Aarhus University, Forsøgsvej 1, Flakkebjerg, Slagelse, 4200, Denmark
[3] Department of Plant Breeding, The Swedish University of Agricultural Sciences, Alnarp, 23053, Sweden
[*] For correspondence (e-mail nive@nordicseed.com)
[‡] These authors contributed equally to this work

Abstract

The 'Gülzow' (G) type cytoplasmic male sterility (CMS) system in hybrid rye (*Secale cereale* L.) breeding exhibits a strong and environmentally stable restoration of male fertility (*Rf*). While having received little scientific attention, three G-type *Rf* genes had been identified on 4RL (*Rfg1*) and two minor genes on 3R (*Rfg2*) and 6R (*Rfg3*) chromosome. Here, we report a comprehensive investigation of the genetics underlying restoration of male fertility in a large G-type CMS breeding system using a palette of complementing forward and reverse genetic analysis. This includes (i) genome wide association studies (GWAS) on a G-type germplasm, (ii) GWAS on a biparental mapping population, (iii) *in silico* identification of *Rf*-like pentatricopeptide repeat (RFL-PPR) genes and their expressed in G-type rye hybrids, and (iv) mining patterns in linkage disequilibrium. Our findings provide compelling evidence of a novel major G-type non-PPR *Rf* gene on the 3RL chromosome. In the *in silico* analysis, we identified 22 RFL-PPR of which 15 were expressed in the transcriptome of G-type hybrids. Our findings provides a novel insight into the underlying genetics of male fertility restoration in a G-type CMS system in rye. The discovery made in this study is distinct to known P- and C-type systems in rye in addition to known CMS systems in barley and wheat. This study constitutes a steppingstone towards understanding the restoration of male fertility in G-type CMS system and a potential resources for addressing the inherent issues of the P-type system.

Keywords: Gülzow gene pool, restoration of male fertility (*Rf*), cytoplasmic male sterility (CMS), Pentotricopeptide repeat (PPR), Genome wide association study (GWAS), RNAseq, linkage disequilibrium, chi-square

Graphical abstract



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 rely on the existence of cytoplasmic male-sterility (CMS) and restoration of male-fertility (*Rf*) genes that resides in genetically distinct parental populations [3,4]. This system efficiently enables control of the parental crossing in the field as a prerequisite for large scale hybrid seed production [5].

In hybrid rye numerous CMS systems exists 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) chromosome, and three minor genes on 3RL, 4RL and 5R chromosome in 'Pampa' (P) type cytoplasm [7-10]. Less prevalent CMS systems include 'Gülzow' (G) type originating from the Austrian population of rye variety 'Schlägler alt' [11], R-type originating from a Russian population, [12], C- [13] and S- [14] type originating from an old Polish cultivar 'Smolickie'. In the G-type CMS system one major gene have been identified on 4RL (*Rfg1*) and two minor on 3R (*Rfg2*) and 6R (*Rfg3*) chromosome [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 partial and highly environmental unstable [9,19-21]. In addition to a potential loss in grain yield, impartial pollination renders the cultivar susceptible to fungal infection of 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 of which the predominance exhibits

74 unsatisfactory restoration [19,20]. In 1991, several non-adapted Argentinian and Iranian rye populations
75 with high frequency of restorer gametes were identified [25]. Crossing of an elite maternal line with one
76 of these non-adapted exotics led to observations of significantly higher restoration levels and
77 environmental stability [26,27]. In order to steer the introgression of novel superior exotic *Rf* genes
78 through marker assisted selection, molecular markers were developed for *Rfp1*, *Rfp2*, *Rfp3* [8,9].
79 Hybrids carrying an exotic *Rf* gene were, however, found to exhibit a significant reduction in grain
80 yield by 4.4% to 9.4% caused by linkage drag effects or epistatic interactions associated with the
81 exotic *Rf* gene [28]. Despite these deleterious effects, hybrid cultivars carrying the exotic *Rfp1* have
82 been introduced to the North European market by a patented brand PollenPlus® [29]. In contrary,
83 hybrids derived from the less prevalent G-type cytoplasm is characterized by a complete and
84 environmental stable restoration of male fertility [30]. Having received little scientific attention the
85 underlying genetics of the G-type CMS system, however, remains largely unexplored [15].
86 The male sterility factors in CMS lines are encoded by mitochondrial genes that cause a defect in the
87 production of viable pollen [31]. Male fertility can be restored by nuclear *Rf* genes encoding proteins
88 that bind specifically to the CMS conferring transcript, preventing the expression of the mitochondrial
89 factor [32,33]. Majority of the proteins encoded by known *Rf* genes belong to the pentatricopeptide repeat
90 (PPR) superfamily known to house more than 450 members in most plant species [34,35]. PPR proteins
91 target the mitochondrial or chloroplast mRNA, participating in a range of post transcriptional
92 processes (RNA editing, splicing, cleavage and translation) with profound effects on organelle
93 biogenesis and function [36-38]. Proteins of the PPR superfamily are characterized by up to 30x tandem
94 repeats of a canonical 35-amino-acid motif, forming an α-solenoid structure [39]. Based on the
95 organization of their motifs, PPR proteins can be divided in two subclasses, i) PLS class containing
96 characteristic triplets of P, L ('long', ≈ 36 amino acids) and S ('short', ≈ 31 amino acids) motifs, and
97 ii) P class solely containing the canonical motif [36,40]. The *Rf*-like (RFL) genes predominantly belong

to the P-class subfamily of PPR proteins [41,42]. In *Poaceae* species, *Rf*-like (RFL) PPR genes have been reported to comprise ≈ 10 % of the PPR gene compliment with 26 genes identified in barley (*Hordeum vulgare* L.) and 25 in perennial ryegrass (*Lolium perenne* L.)[43,44]. While *Rf* genes have been characterized as RFL-PPR in both barley (*Hordeum vulgare* L.) Rfm1; [45], sorghum (*Sorghum bicolore* L.) Rf1; [46], maize (*Zea mays* L.) Rf5; [47], and rice (*Oryza sativa* L.) Rf4; [48,]Rf5; [49,]Rf6; [50] there is little available information on RFL-PPRs in rye.

In this paper we report a comprehensive study of the genetics underlying male-fertility restoration in G-type CMS based hybrid rye breeding system. The objective of this study was to identify major and minor G-type *Rf* genes. This was approached through (i) Genome wide association studies (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, (iii) *in silico* identification of RFL-PPR genes expressed in rye hybrids, (iv) Studying patterns of linkage disequilibrium on *Rf*-annotated single-nucleotide polymorphism (SNP) markers in the breeding and the mapping population. This knowledge will serve as a steppingstone towards developing novel hybrid cultivars exhibiting superior and environmentally stable restoration of male-fertility to maximize grain yield and enhance ergot resistance.

Materials & Methods

Plant material

In total 365 Nordic Seed Germany GmbH inbred hybrid rye (*Secale cereale* L.) elite breeding component lines were selected for this study, comprising 242 restorer, 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 Austrian population of rye variety 'Schlägler alt' [11,15]. Genetic structure of the germplasm has been thoroughly characterized in a recent study by Vendelbo,

et al. [51]. A biparental mapping population was developed from a hybrid rye cv. Stannos, deriving from the cross of a cytoplasmic male-sterile (CMS) line msG214135 and a restorer line R3966.

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 and associated traits to 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 hour 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 hours of light at 8°C for a week and hereafter 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 repeated to generate a F_2 biparental mapping population. To quantify the degree of male-fertility restoration in the F_2 population, in total 181 F_2 plants were rigorously phenotyped for pollen production using a customized visual 1-9 scale (1: no pollen, 9: large quantity of pollen) at four timepoints. At harvest, 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 get a comprehensive phenotypic dataset on the inheritance of male-fertility restoration in the population. 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 restoration of male-fertility (*Rf*) genes using a χ^2 test [52]. An F_2 plant was considered 'sterile', if it either yielded less than 20 seeds or scored ≤ 2 in pollen production.

Molecular markers

All rye lines included in this investigation were genotyped using a custom Illumina Infinium $15K_{wheat}$ [53] and $5K_{Rye}$ [54,55] 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. [55]. The F_2 biparental mapping population was genotyped on a custom Illumina Infinium $25K_{wheat}$ and $5K_{Rye}$ SNP array, denoted 30K, enriched with additional 10K wheat markers deriving from the 90K wheat SNP array by Wang, et al. [53] compared to the 20K array. Mapping position of SNP markers derived from the 90K wheat were identified by blasting the marker sequences to the rye reference genome at a significance threshold of the e-value at 10^{-5}, selecting the physical position of the top hit [56,57].

Data analysis

Genetic analysis of SNP marker data was done in R studio (v. 1.1.463) interface in R statistical software (v. 3.6.3) by application of various predesigned packages [58,59].

Genome wide association study

Discovery of *Rf* associated SNP markers was done by genome-wide association study (GWAS) using genomic association and prediction integration tool (GAPIT) (v.3) package in R [60]. 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.

Identification of pentatricopeptide repeats (PPR) and restoration fertility-like PPR genes in rye

For identification of members of the PPR protein family in the draft genome assembly of rye (*Secale cereale* L.) by Bauer, et al. [55], all known PPR domain sequences in plants were obtained from the Pfam database [61]. These PPR domains were then blasted against protein sequences of the rye genome

using NCBI BLASTP tool [56]. The PPR domains were furthermore used to develop a PPR profile matrix using the 'hhmbuild' program in the HMMER package [62]. This matrix was then utilized to identify PPR genes amongst the 27784 coding sequences reported in the rye draft genome. Identified PPR genes were, hereafter, studied and predictive information on protein functions and conserved sequence elements obtained through a customized InterProScan (v. 5) pipeline by scanning on the PANTHER, PROSITE, Pfam, and SUPERFAMILY databases [63]. The InterProScan was implemented in OmicsBox (v. 1.2.4) [64]. PPR gene sequences were then aligned, using the NCBI BLAST platform to known *Rf*-genes from barley (*Hordeum vulgare* L.), rice (*Oryza sativa* L.), maize (*Zea mays* L.), and stiff brome (*Brachypodium distachyon* L.) [43,44,56]. Hits with a minimum 50% identity and 50% query coverage were collected, and PPR genes present in at least three of the species were considered a candidate restorer of fertility like (RFL) PPR gene. Coding- and protein sequences of all above mentioned species were downloaded from the Ensemble Plants database [65]. Physical location of the RFL PPR genes was identified by mapping to the recently published reference genome by Rabanus-Wallace, et al. [57]. Mapping was conducted using the NCBI BLASTN tool at a significance threshold of the expected value (e-value) at 10^{-5} selecting the position of the top hit [56].

RNA-seq data analysis of RFL-PPRs in G-type hybrids of rye

Nordic Seed hybrid rye cv. Helltop and cv. Stannos belong to G-type hybrid breeding system of rye. For the identification of causative *Rf* genes in the G-type breeding system, expression of the candidate RFL-PPR genes were investigated in *de novo* transcriptome assemblies of these two hybrids. The transcript data obtained from the spikes of G-type hybrids at the time of flowering. The raw reads from this library has been deposited in sequence read archive (SRA) with submission ID "PRJNA612415 and can be accessed here (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA612415). We investigated the RFL-PPR transcripts

192 in the individual plants of these two G-type hybrids. High quality *de novo* transcriptome assembly of
193 these two hybrids has recently been published by Mahmood et al., (2020).
194 In order to identify RFL-PPR genes expressed in G-type hybrids, the assembled transcripts were
195 translated to coding protein sequences and blasted against the protein sequences of the RFL-PPR
196 genes using the NCBI BLASTP tool at a significance threshold of the e-value at 10^{-5} [56]. With no
197 available gene annotation for the rye draft genome, annotation was conducted through assigning the
198 Gene Ontology terms using OmicsBox program (v. 1.2.4) [64]. Functions of sequences were predicted
199 at a significance threshold of the e-value at 10^{-3}, annotation cutoff of 55 and evidence code set to 0.8
200 for the different categories as implemented in OmicsBox. The annotation of expressed transcripts was
201 provided only when they shared similarities with known RFL-PPRs at a significance threshold of the
202 e-value at 10^{-10}.

203 **Linkage disequilibrium**

204 Analysis of linkage disequilibrium was conducted to investigate whether the populations portrayed
205 evidence of conservation in regions harboring annotated restorer of male fertility markers.
206 Comparative analysis of pairwise linkage disequilibrium (LD) between parental populations was
207 done using SnpStats (v. 1.36.0) R package with depth set to 10 and LD estimated as the coefficient
208 of determination (r^2) [66]. Heatmap of the pairwise LD was constructed using LDheatmap (v. 0.99-7)
209 R package [67]. The analysis was conducted on the entire germplasm using the 20K genotype data and
210 a subset population using the 600K gene data. Analysis of pairwise LD was, furthermore, also utilized
211 to determine the position of a subset of highly *Rf* associated wheat derived SNP markers that could
212 not be mapped to the rye reference genome. The analysis was conducted by calculating the pairwise
213 LD amongst the individual wheat markers and the entire entity of mapped informative markers.
214
215

Results

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 markers 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) (Table 1). All SNP arrays portrayed a uniform distribution of markers across the rye genome (Table 1). In total, 4419 informative markers were identified in the 20K array on the entire germplasm as 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 (n = 181), 3493 informative markers were identified out of which 1088 derived from the 5K rye array, 808 from the 600K rye array and 1597 from the 90K wheat array.

Table 1. Chromosomal distribution and polymorphism information content (PIC) of quality filtered single nucleotide polymorphism (SNP) markers deriving from three genotyping platforms on Nordic Seed hybrid rye (*Secale cereale* L.) elite breeding lines (**20K, 600K)** and F_2 biparental mapping population (**30K**).

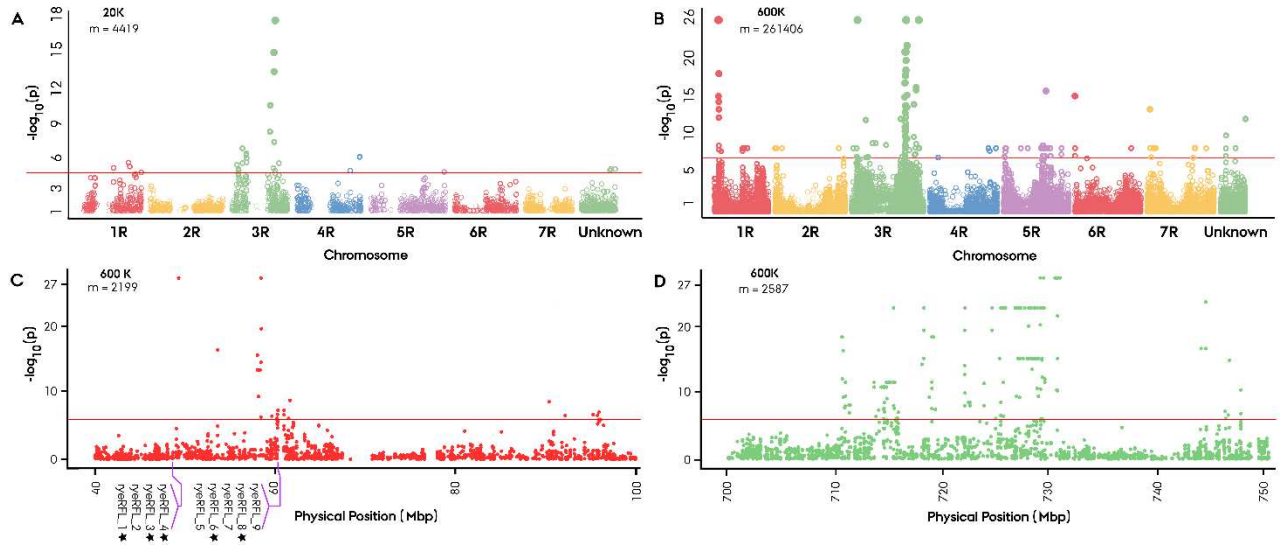
Genotyping platform	**20K**	**30K**	**600K**
lines	**365**	**181**	**180**
Chromosome	**Markers**	**Markers**	**Markers**
1R	590	332	33854
2R	711	395	33698
3R	631	386	31493
4R	515	299	32555
5R	669	374	37073
6R	516	322	36872
7R	528	340	38918
Unknown	259	1045	16943
Total	4419	3493	261406
PIC	0.26	0.34	0.23

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 (Fig. 1A, Supplementary material 1). 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 729-730 Mbp (Fig. 1B, D, Supplementary material 2). 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 (Fig. 1C). On 4RL, a less significantly associated marker ($-\log_{10}(p)$ = 5.4) was identified at 885 Mbp by the 20K GWAS. Solitary markers in the 600K GWAS analysis were disregarded from further investigation.

Fig. 1 Manhattan plot for genome wide association study (GWAS) on population origin (case controle) on Nordic Seed elite hybrid rye 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 including position of identified restoration of male-fertility (*Rf*) like pentatricopeptide repeat (RFL-PPR) genes of which RFL-PPRs expressed in G-type hybrids are marked with an asterix, **D**) Manhattan plot of the 600K SNP array 3RL region.



Biparental population

A biparental F_2 population consisting of 181 individuals was developed from the hybrid cv. Stannos. The population was phenotyped for six restoration of male fertility as well as related traits to

restoration in order to get a comprehensive dataset on the inheritance of 'Gülzow' (G) type *Rf* genes. Seed number and pollen production were found, on basis of our observations, to be the most representative *Rf* associated traits (Fig. 2A-B).

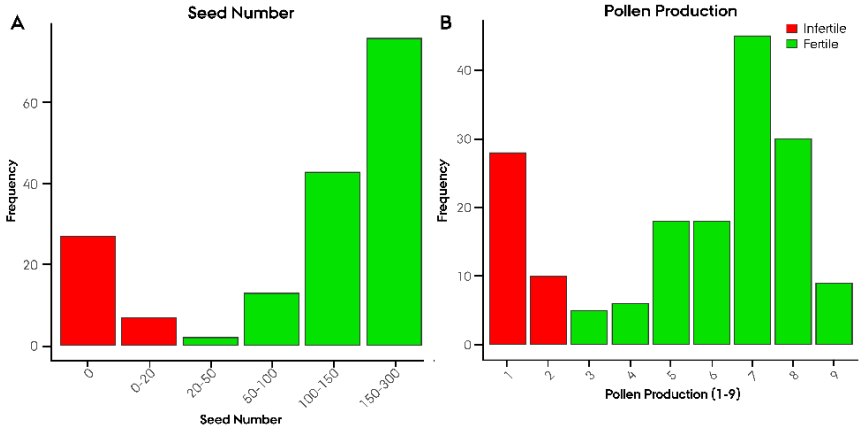


Fig. 2 Phenotypic distribution of restoration of male fertility related traits, **A**) Seed number, and **B**) Pollen Production in 181 F_2 plants derived from a hybrid rye cv. Stannos

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 a χ^2 test. Intriguingly, the observed segregation ratios were in accordance with a monogenic dominant inheritance of male fertility restoration with χ^2 (1, $n_{infertile}$ = 38, $n_{fertile}$ = 143) = 2.26, p = .13 for seed number and χ^2 (1, $n_{infertile}$ = 43, $n_{fertile}$ = 138) = 1.11, p = .29 for pollen production (Supplementary material 3). GWAS led to the identification of 16 *Rf* associated SNP markers of which, 5 markers showed a significant association with $-\log_{10}(p) > 5.2$ (Fig. 3, Supplementary material 3).

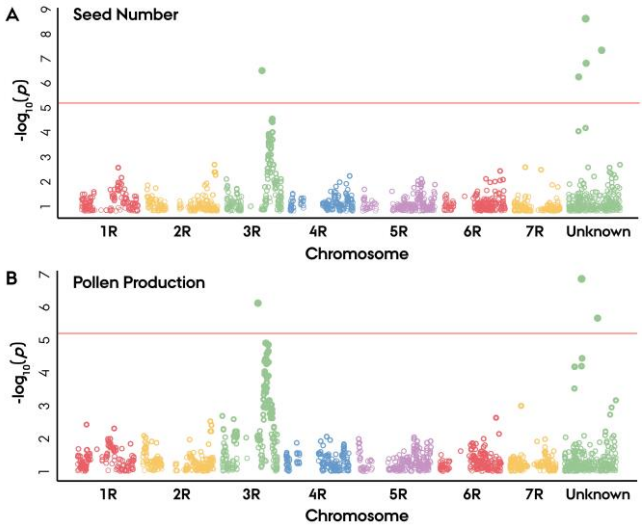
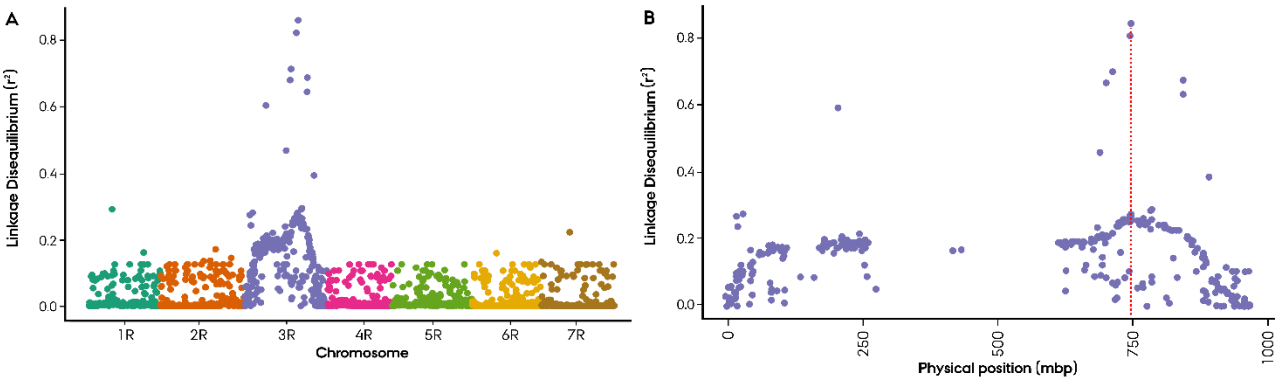


Fig. 3 Manhattan plot for genome wide association study (GWAS) on two restoration of male fertility related phenotypic scores, **A)** Seed Number and **B**) Pollen Production (1-9) in a F_2 biparental population (n = 181) derived from the hybrid rye cv. Stannos. Significant association was identified using criterion of $-\log_{10}(p) > 5.2$ depicted as a red line.

On 3RL, a twin-peak was identified spanning from 627 to 809 Mbp with its highest associated marker ($-\log_{10}(p)$ = 6.66) localized at 627 Mbp. The remaining four significantly associated SNP markers derived from the 90K wheat SNP array. None of the four markers could successfully be mapped to the 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 (Supplementary material 4). 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 (Supplementary Fig. 2). The top *Rf* associated wheat marker AX_158558079 exhibited a max LD of 0.85 at 747 Mbp (Fig. 4A-B).

Fig. 4 Genome wide pairwise linkage disequilibrium (LD) between a highly *Rf* associated wheat derived SNP marker, AX_158558079 and 3493 informative SNP markers in 181 rye (*Secale cereale* L.) F_2 plants. **A**) Chromosome-wise distribution of LD, **B**) LD distribution on 3R chromosome.



Identification of restoration of fertility-like pentatricopeptide repeat (RFL-PPR) genes in rye

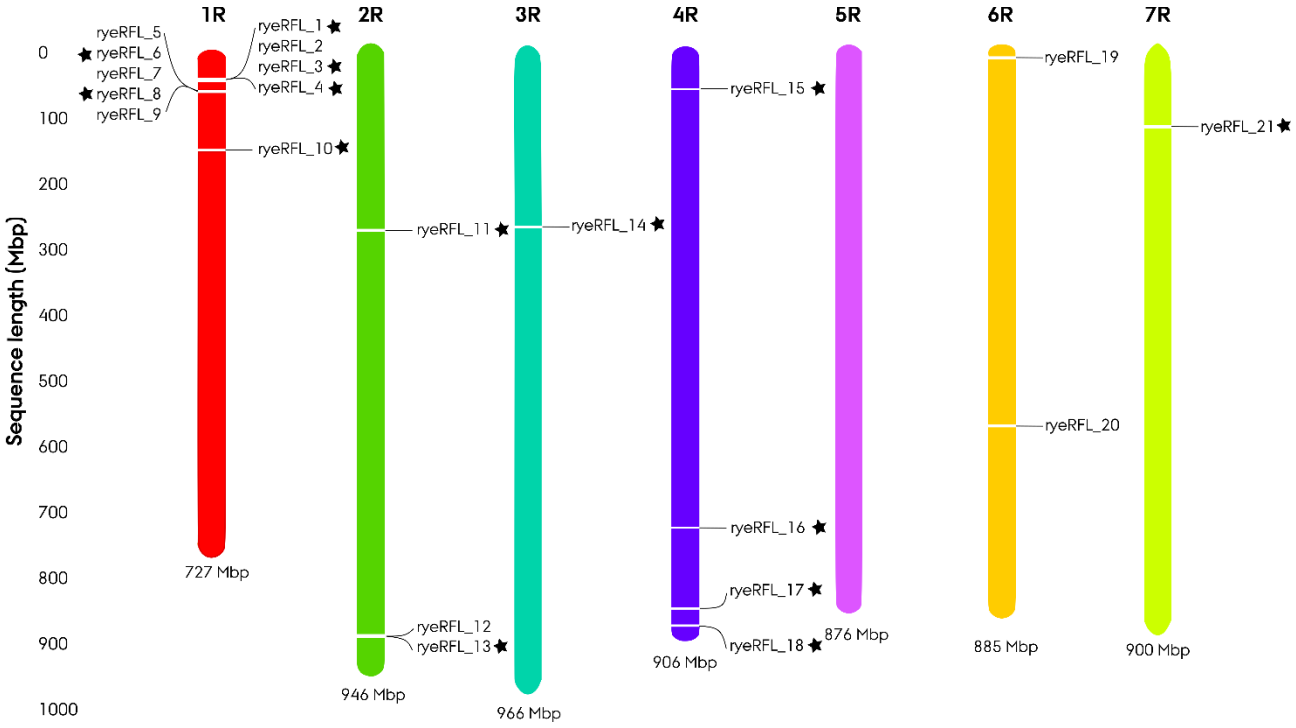
For identification of PPR genes within the 27784 coding sequences of the draft rye refence genome, a PPR gene profile matrix was developed using available sequence information. Scanning of the

285 reference genome led to the identification of 232 PPR genes (Supplementary material 5). Out of
286 these, 22 RFL-PPR genes were identified on basis on homology to known *Rf* genes in grass species
287 (Supplementary material 6). Genomic location of these in the rye reference genome can be seen in
288 Fig. 5. On average, the RFL-PPR genes portrayed 14 PPR domains and had an average sequence
289 length of 3355 bp.

290 **Expression of RFL-PPRs in the spikes of G-type hybrids at the time of flowering**

291 To investigate, whether the 22 identified rye RFL-PPR genes were expressed in 'G' type based
292 breeding germplasm, the transcriptome data from two hybrids cv. Helltop and cv. Stannos were
293 analyzed. As the RNA-seq data was obtained from spikes at flowering, it should be expected that *Rf*
294 genes are actively expressed in fertile hybrids. In the cv. Helltop transcriptome assembly, 18 RFL
295 PPR genes were identified, whereas in the cv. Stannos assembly, 16 were expressed at the time of
296 flowering. Among the expressed RFL-PPR, 15 were shared in both hybrids. The RFL-PPR exclusive
297 to either hybrid might involve in hybrid specific function, but not the restoration function as both
298 hybrids belong to same G-type breeding system. Hence, we focused on the RFL-PPR common in
299 both hybrids. Among these RFL-PPR, five were located on 1R, four on 4R, two on 2R, one on 3R
300 and one on 7R chromosome (Fig 5). One expressed RFL-PPR mapped to the sequences that are not
301 placed to any chromosome in reference genome. The exact location of expressed as well as non-
302 expressed RFL-PPR on respective chromosome were depicted in graphics (Fig. 5).
303
304
305
306

Fig. 5 Genomic location of 22 *in silico* annotated restoration of male fertility-like (RFL) pentotricopeptide repeats (PPR) genes in the rye reference genome depicted in physical distance (Megabase pair, Mbp). RFL-PPRs identified in both spike transcriptome assemblies of two rye hybrids, cv. Stannos and cv. Helltop are marked with an asterix.



Linkage disequilibrium and fixation indices at regions harboring restoration of male fertility annotated 600K SNP markers

With restoration of male fertility presumed to constitute a population defining trait in the germplasm, one of the objectives of this study was to identify genetic fingerprints of conservation at 63 *Rf* annotated marker locations (Supplementary material 7). This was addressed by estimation of pairwise linkage disequilibrium (LD) and marker fixation indices (F_{st}) in the entire germplasm using both the 20K and 600K SNP array genotype data (Table 2A-B, Supplementary material 8). The NRG&CMS population was found to exhibited a large proportion of monomorphic markers at the *Rf* annotated regions, 68.3% and 70.5% on the two sites at 3RL and 34.0% and 48.3% on the two sites at 4RL (Table 2). In contrast, the restorer population portrayed a level of monomorphism below 1% at all sites. Calculation of fixation indices led to the finding of a unique population-wise differentiation on

3R, exhibiting a intrachromosomal F_{st} of 0.34 in comparative to the overall genomewise F_{st} of 0.23 (Fig. 6). Fixation indices in the *Rf* annotated regions ranged from 0.29 to 0.37 for the two sites on 3RL and 0.18 to 0.28 on the two sites on 4RL (Table 2A-B). LD was consistently higher for both populations on 3R with a mean LD of 0.65 for the NRG and 0.21 for the R population in comparison to 0.44 and 0.19 on 4RL respectively (Table 2A-B). Heatmaps of the pairwise LD were constructed for both populations using the 600K SNP array genotype data as a comparative tool to visualize the LD landscape at the *Rf* annotated regions (Fig. 7).

Table 2. Mean pairwise linkage disequilibrium (LD) and fixation indices (F_{st}) at restoration of male fertility (*Rf*) annotated regions on 3R and 4R chromosome in Nordic Seed hybrid rye elite breeding germplasm using **A**) 20K and **B**) 600K SNP genotype data.

A			**Chromosome**	**3R**		**4R**	
			Markers	631		515	
			F_{st}	0.30		0.21	
	Intra Chromosomal	**Intrachromosomal LD**	**NRG&CMS (n = 123)**	0.70		0.38	
			Restorer (n = 242)	0.20		0.13	
20K			**Region (Mbp)**	800-810	865-875	570-600	805-900
			Markers	30	43	44	89
			***Rf* annotated markers**	4	4	6	53
			F_{st}	0.29	0.37	0.18	0.28
	Region	**LD**	**NRG&CMS (n = 123)**	0.80	0.39	0.30	0.53
		Monomorphic Markers		19	18	18	46
		LD	**Restorer (n = 242)**	0.14	0.12	0.14	0.08
		Monomorphic Markers		0	0	0	1
B			**Chromosome**	**3R**		**4R**	
			Markers	31493		32555	
			F_{st}	0.34		0.20	
	Intra Chromosomal	**Intrachromosomal LD**	**NRG&CMS (n = 88)**	0.71		0.50	
			Restorer (n = 92)	0.36		0.29	
600K			**Region (Mbp)**	800-810	865-875	570-600	805-910
			Markers	460	553	1428	5655
			***Rf* annotated markers**	4	4	6	53
			F_{st}	0.37	0.32	0.20	0.23
	Region	**LD**	**NRG&CMS (n = 88)**	0.71	0.68	0.42	0.51
		Monomorphic Markers		314	390	486	2732
		LD	**Restorer (n = 92)**	0.33	0.25	0.27	0.25
		Monomorphic Markers		0	1	11	10

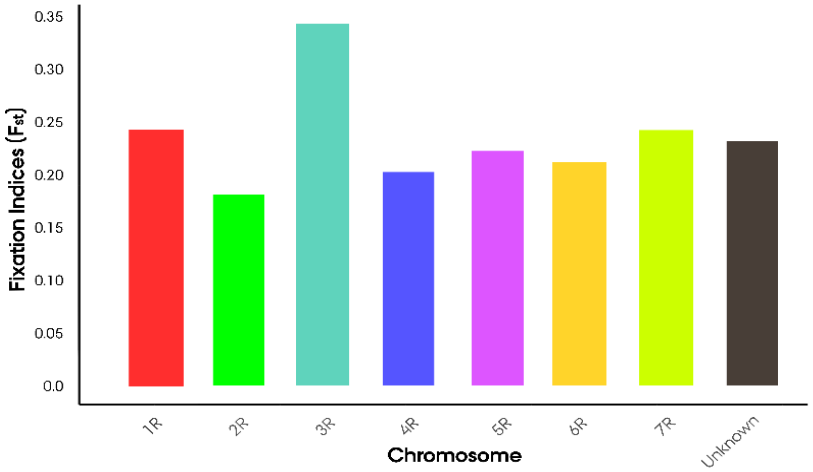


Fig. 6 Intrachromosomal fixation indices (F_{st}) of Nordic Seed hybrid rye elite breeding germplasm using 600K rye SNP array genotype data. Populations comprised of 92 restorer and 88 non-restorer germplasm lines.

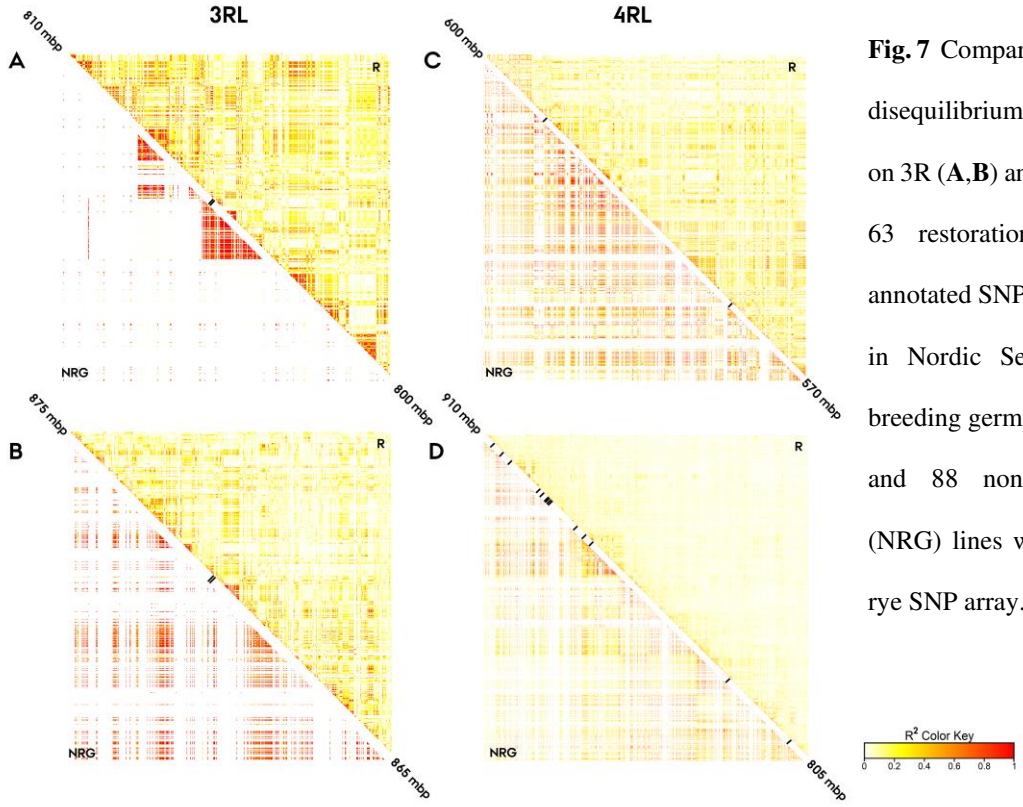


Fig. 7 Comparative pairwise linkage disequilibrium (R^2) of four regions on 3R (**A**,**B**) and 4R (**C**,**D**) harboring 63 restoration of male fertility annotated SNP markers (black lines) in Nordic Seed hybrid rye elite breeding germplasm. 92 restorer (R) and 88 non-restorer germplasm (NRG) lines were genotyped 600K rye SNP array.

Discussion

While the less common 'Gülzow' (G) type based systems demonstrate superior restoration of male fertility, it has received little scientific attention in the past. This is the first study since Melz and Adolf [15] 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 (*Rfg*1) and two modifying genes on 3R (*Rfg2*) and 6R (*Rfg3*) [15,68]. Recent years technological and scientific advances have progressively accelerated the genomic resources available in rye with the latest additions being the 600K high-density SNP array by Bauer, et al. [55] and chromosomal scale rye reference genome 'Lo7' by Rabanus-Wallace, et al. [57]. Using a comprehensive palette of complementing forward- and reverse genetics approaches, 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 modifying gene on 3RL chromosome.

Indications of a major restoration of male-fertility like pentatricopeptide repeat (RFL-PPR) gene on 1RS

While case control genome wide association study (GWAS) is a useful tool for providing an insight in the genetics differentiating of the parental gene pools, it has several limitations. Utilizing population origin of lines as 'phenotypic' input, statistically associated markers in case control GWAS, can either be a population defining trait such as a *Rf* QTL, or a product of population structure. In a recent population study by Vendelbo, et al. [51] 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 introducing a risk of false positives [69,70]. To moderate the effect of these confounding factors, Genomic Association and Prediction Integrated Tool (GAPIT) used to conduct the GWAS therefore utilizes a compressed mixed linear model [60,71]. Large LD blocks on the other hand introduces an uncorrectable confounding factor. Long distance LD complicates the disentanglement of actual causal variants from linked neutral markers, which can in term lead to spurious associations [70].

In the 600K case control GWAS, a unique strong peak was identified on 1RS (Fig. 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, in the genome scan 9 out of the 22 identified RFL-PPRs were found to reside in this region situated in two clusters at 46.3-47.1 Mbp and 61.5-62.0 Mbp on 1RS. In the RNA-seq data, five of these were found to be expressed at flowering stage in both G-type hybrids (Fig. 5). A similar hotspot for RFL-PPRs was identified in the rye reference genome 'Lo7' on the proximal region of 1RS [57]. In the barley reference genome 'Morex' Melonek, et al. [44] discovered a similar enrichment with 10 out of 26 RFL-PPRs situated on 1HS, a region highly syntenic to rye 1RS [72]. Functional annotation of the 10 RFL-PPRs on 1RS led to the finding that six were annotated as '*Rf1* mitochondrial like' (Supplementary material 6). In rice, *Rf1* has been found to restore fertility by cleavage of an *atp6-oRf79* dicistronic gene, impeding the accumulation of the cytotoxic peptide ORF79 and hence leading to the recovery of pollen potency [73].

Consistent with our findings, Miedaner, et al. [7] reported a major P-type *Rf* gene on 1RS in a German inbred rye line 'L18'. In wheat, two major *Rf* genes have likewise been identified on 1AS (*Rf1*) and 1BS (*Rf3*) chromosome, syntenic to rye 1RS [74-76]. 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 (Fig. 3, Fig. 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 a *Rf* gene. Being that the region coincides with a large cluster of RFL-PPRs, of which five are co-expressed in both assayed hybrid cultivars, supports the initial explanation of an additional major G-type RFL-PPR gene on 1 RS in the germplasm.

Modifying G-type *Rf* genes

In order to identify the complement of major and minor *Rf* genes in the G-type CMS system, GWAS was complemented with estimations of LD and F_{st} at sites reported to harbor P-type *Rf* genes. At present a single minor *Rf* gene have been identified on 3R in both G-type Rfg3; [15] and P-type [7] CMS

systems. The modifying P-type *Rf* gene has been mapped to 3RL with subsequent association of eight 600K SNP markers by Bauer, et al. [55] at 806.1 and 869.5 Mbp (Supplementary material 7). In the mapping population GWAS, five markers were found to be associated with a *Rf* QTL at 807.1 – 808.7 Mbp with a mean LOD of 4.7 (Fig. 3, Supplementary material 3). This region was furthermore characterized by an enrichment in LD in both parental populations with the NRG&CMS population portraying a large proportion of monomorphic markers indicative of a strict conservation (Table 2, Fig. 7A). Similar to the region surrounding the site believed to harbor the major Rf gene on 3RL at ≈ 747 Mbp, the region 800-820 Mbp exhibited a highly conserved haplotype for fertile and sterile F_2 plants with a χ^2 test (1, $n_{infertile} = 44$, $n_{fertile} = 134$) = 0.01, $p = .93$. (Supplementary material 9). These findings in conjunction suggest the additional presence of a minor *Rf* gene on 3RL, which would further accentuate the unique role of 3RL in G-type CMS system. We can however not exclude the possible confounding effect of long distance LD creating a spurious association between the site at 807 Mbp and the proximal major *Rf* gene.

On 4RL, a marker was found to be associated to a population differentiating site in the 20K case control GWAS (Supplementary material 1). Intriguingly, the marker located at 885 Mbp was annotated as *Rfp3*-associated and furthermore co-localized with an RFL-PPR gene at 889 Mbp expressed in both of the assayed G-type hybrids (Fig. 5, Supplementary material 6) [8]. Neither the 600K case control nor mapping population GWAS portrayed any evidence of a *Rf* gene on 4RL (Fig. 1, Fig. 3). Mining of patterns in LD at the region housing the *Rf* annotated markers likewise did not show any evidence of a *Rf* gene under selection in the germplasm (Fig. 7C,D). Contrary to the C- and P-type CMS systems, this suggests a negligible role of 4R in the G-type systems. Furthermore, in addition to a major *Rf* gene on 3RL exclusive to the G-type CMS system, our findings suggest the potential presence of an additional minor gene on the distal region of 3RL.

Decisive role of 3R in the G-type CMS breeding system

In our GWAS study, we found a strong coinciding peak on 3RL in both the 20K and 600K case control suggesting that 3R houses a population differentiating trait (Fig. 1A-B). This finding was consistent with the discovery of a distinctly higher interchromosomal fixation indices (F_{st}) on 3R (Fig. 6). Furthermore, Vendelbo, et al. [51] reported a singular enrichment of interchromosomal LD for both parental populations on 3R. In conjunction, these findings accentuate the pivotal role of the 3R chromosome in the assayed G-type hybrid rye elite breeding germplasm.

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 (Supplementary Fig. 1). Segregation ratio of *Rf* associated traits in the mapping population was found to be in accordance with a monogenic dominant inheritance of a *Rf* gene by χ^2 test consistent with the singular peak identified in the case control GWAS (Fig. 1A-B). GWAS on the phenotypic dataset from the mapping population confirmed that the major *Rf* gene co-localized with the region identified in the case control GWAS on 3RL (Fig. 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 (*Triticum aestivum* L.) array, could not be mapped to the rye reference genome 'Lo7' [53,57]. This was resolved by a chromosome-wise LD mapping of each of the wheat markers, with the highest associated marker mapping to 747 Mbp (Fig. 7, Supplementary Fig. 2).

Novel major *Rf* gene unique to the G-type CMS breeding system

While no major *Rf* gene on 3RL has to our knowledge been identified in neither the G-type or P-type CMS system in rye, Melz and Adolf [15] reported a minor G-type *Rf* gene on 3R (*Rfg2*) and Miedaner, et al. [7] a minor P-type on 3RL. In the case of *Rfg2,* inconclusive segregation ratios of primary trisomics of rye 3R led to the assumption that 3R likely housed a minor gene. While Melz and Adolf

436 [15] cautiously interpreted this anomaly as a product of a modifying gene their findings suggest that
437 something of significance is occurring on 3R in the G-type CMS system. It therefore remains open
438 whether the identified major G-type gene is *Rfg2* previously misclassified as a minor gene, or an
439 unreported *Rf* gene on 3RL.
440 In P-type CMS systems, the primary site of interest in term of male fertility restoration is on 4RL
441 housing three major *Rf* genes *Rfp1, Rfp2, Rfp*3 [8,9,28]. Identified by Miedaner, et al. [7] the minor P-type
442 *Rf* gene on 3R has since received no scientific attention, suggesting that 3R constitutes a minor role
443 in male fertility restoration in the P-type CMS system.
444 To our knowledge no *Rf* gene have been reported on chromosome segments orthologous to rye 3RL
445 in any of the domesticated species residing within the botanical tribe *Triticeae*. In wheat, major *Rf*
446 genes have been identified on 1AS (*Rf1*), 1BS (*Rf3*), 6AS (*Rf9*), 6BS (*Rf4, Rf6*), 6D (*Rf5*) and 7D
447 (*Rf2*) chromosome [74,75,77-80]. In barley (*Hordeum vulgare* L.), two major *Rf* genes have been identified
448 on 6HS (*Rfm1*, *Rfm3*) [81,82]. Intriguingly, Martis, et al. [72] discovered that the distal region of 3RL and
449 4RL were conserved syntenic segments of an ancestral *Triticea* chromosome a6. In a comparative
450 analysis, they found that the segment on 3RL portrayed distinctly less collinearity than all other
451 syntenic segments suggesting a differential evolution of 3RL during rye speciation. In contrary, the
452 syntenic segment on 4RL was found to be highly conserved in *Brachypodium distachyon* L., rice
453 (*Oryza sativa* L.), sorghum (*Sorghum bicolore* L.) and barley. Rye 4RL, a region housing three major
454 P-type *Rf* genes, was found to be syntenic to barley 6HS [72]. These results are consistent with the
455 findings of Hackauf, et al. [83], who reported that the segment housing *Rfp1* on 4RL exhibits an ortholog
456 on wheat 6DS and barley 6H. In a subsequent study, Hackauf, et al. [8] furthermore, proposed that *Rfp3*
457 on 4RL likely maps to an orthologous segment housing *Rf6* on wheat 6BS and *Rfm1* on barley 6HS
458 [79]. These findings provide further evidence of a conserved synteny between rye 4RL and wheat 6AS-
459 6BS-6DS, explained by the later divergence of these two species from an ancestral *Triticeae*

progenitor than barley [84]. Consistent with Börner, et al. [68], this suggests a likely conservation of genes controlling CMS restoration across *Triticeae* species. In these CMS systems, 4RL-6HS-6AS/BS/DS chromosomes house the predominance of identified major *Rf* genes [83]. Furthermore, this accentuates the novelty of a major *Rf* gene on 3RL in the G-type CMS system, with no known *Rf* genes on ortholog chromosome segments in other *Triticeae* species.

Non-Pentatricopeptide repeat *Rf* gene on 3RL

In recent years, 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 *i.a.* barley *Rfm1* [45], sorghum *Rf1* [46], maize *Rf5* [47], and rice *Rf4*, *Rf5*, *Rf6* [48-50]. The results of genome scan and RNA-seq data analysis for expression of RFL-PPR clearly point towards the existence of non-RFL-PPR on 3R to cause restoration in G-type breeding system. In the *in silico* analysis, we successfully annotated 232 PPRs in the draft rye reference genome 'Lo7' out of which 22 were identified as RFL-PPRs by stringent comparison to existing sequence information on characterized RFL-PPRs (Supplementary material 5, Supplementary material 6). Our findings are consistent with the observations made by Melonek, et al. [44], who identified 26 RFL-PPRs in barley, and Sykes, et al. [43] who identified 25 in perennial ryegrass (*Lolium perenne* L.).

None of the identified RFL-PPRs were however found to co-localize with the major *Rf* gene on 3RL (Fig. 5). There are several possible scenarios that might explain this observation. Firstly, the identified major *Rf* gene is unique, not resembling any of the presently characterized RFL-PPRs used to design the capture profile matrix. Secondly, the developed pipeline employed too stringent criteria for discrimination of RFL-PPRs. Thirdly, the major *Rf* gene on 3RL is indeed a non-PPR *Rf* gene. While it is not feasible to further address the first scenario at present, two PPRs were indeed found to co-localized with the identified region on 3RL at 743.7 Mbp (ryePPR_3R_25) and 751.6 Mbp (ryePPR_3R_26) (Fig. 3, Fig. 4, Supplementary material 5). *In silico* annotation of these revealed

that the former encodes a P-class PPR gene composed of 7 motifs furthermore containing motifs belonging to the acetamidase-formamidase family, while the latter a PLS-class PPR gene composed of only 5 motifs. While both P- and PLS-class PPR genes have been characterized as *Rf*-like we concluded that these PPRs were correctly annotated as non *Rf*-like on basis of both motif number and complementary functions not known to be related to *Rf* [45-47]. Furthermore, in the RNA-seq data analysis, neither of the PPRs were found to be expressed in either of the rye hybrid cv. Stannos and cv. Helltop at flowering. In conjunction, this suggest that these PPRs are not functional *Rf* genes in our material supporting the notion about the presence of a non-PPR *Rf* gene on 3RL.

Intriguingly, a growing body of *Rf* genes are 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; [85], acyl-carrier protein synthase Rf17, in rice; [86], aldehyde dehydrogenase Rf2, in maize; [87], bHLH transcription factor Rf4, in maize; [88], and *Rf1*, a peptidase, in Sugar beet (*Beta vulgaris* L.) [89]. In rye, Hackauf, et al. [8] furthermore, reported a close linkage between *Rfp1* and *Rfp3* on 4RL to mitochondrial transcription termination factors (mTERF) genes. In the recent rye reference genome 'Lo7', 131 mTERF genes were identified of which four were closely situated around the *Rfp1* [57,90]. A similar observation was made by Bernhard, et al. [82] who identified two mTERF genes closely linked to *Rfm3* on barley 6HS, syntenic to rye 4RL. Pan, et al. [91] successively observed a role of mTERF genes in kernel development in maize connecting the gene family to the reproductive system of plants. These recent additions caused Kubo, et al. [92] to propose a revision of the existing PPR and non-PPR classification of *Rf* genes into three groups on basis of characteristic features. These are, I) association with a post-transcriptional mechanism for regulating mitochondrial gene expression, II) R gene-like copy number variant at the locus, III) lack of a direct link with a mitochondrial ORF associated with CMS. With no RFL-PPRs on 3RL it is reasonable to presume that the causative *Rf* gene belongs to this expanding non-PPR class.

CMS Systems in Hybrid Rye Breeding

On basis of male fertility restoration requirements and genetic similarity of sterilizing cytoplasms G, C, and R-type CMS systems have been proposed to belong to the larger Vavilovii (V) type [6,93,94]. 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 descend 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 neither 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.

Conclusion

In this study, we demonstrated the strength of complimenting forward and reverse genetics approaches providing a comprehensive tool set for dissecting 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 ortholog in neither barley nor wheat. With no co-localizing RFL-PPRs on 3RL, this suggests that the *Rf* gene belongs to the expanding non-PPR class. 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.

Acknowledgement

We would like to thank laboratory technician Hanne Svenstrup at Nordic Seed A/S (Dyngby, Denmark) for contribution to the genotypic data collection and Anette Deterding and Marlene Walbrodt at Nordic Seed Germany GmbH (Nienstädt, Germany) for phenotyping the F_2 population. The plant material included in this study was provided by Nordic Seed Germany GmbH (Nienstädt,

Germany). The 20K SNP genotyping was performed by Trait Genetics (Gatersleben, Germany). The 600K SNP genotyping was performed by Eurofins Genomics (Skejby, Denmark).

Financial statement

The research was funded by Innovation Fund Denmark (grant no. 8053-00085B, 7039-00016B) and Pajbjerg Foundation.

Data availability

The data that support the findings of this study are presented in the supplementary materials and methods and/or available from the corresponding authors upon request

Author contributions

All authors were involved in the study design. N.M.V. performed the bioinformatic analysis, visual output and wrote the manuscript. K.M. conducted the *in silico* study for the identification of PPR and RFL-PPR genes in rye draft genome and annotation of markers identified in GWAS study. K.M. performed the analysis for identification of RFL-PPRs expressed in 'G' type hybrids. P.S.K. oversaw the biparental mapping population experiment and the phenotypic evaluation of *Rf* associated traits. J.O. was responsible for all communication with Trait Genetics and Eurofins Genomics conducting the SNP genotyping. K.M., P.S., J.O. and A.H. was involved in the intellectual input for the study including interpretation of results. All authors were involved in the conceptualization of the study and revision of the manuscript.

Competing Interests

All authors are employees in the plant breeding company Nordic Seed A/S. The employment does not alter the authors' adherence to all of Nature Plants policies.

References

1 Miedaner, T. & Hübner, M. Quality demands for different uses of hybrid rye. *Tagungder Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs* **61**, 45-49 (2011).

2 Laidig, F. *et al.* 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. *Theoretical and Applied Genetics* **130**, 981-998, doi:10.1007/s00122-017-2865-9 (2017).

3 Geiger, H. H. & Miedaner, T. Hybrid rye and heterosis. *Genetics and Exploitation of Heterosis in crops*, 439-450 (1999).

4 Chen, L. & Liu, Y. G. Male sterility and fertility restoration in crops. *Annual review of plant biology* **65**, 579-606, doi:10.1146/annurev-arplant-050213-040119 (2014).

5 Kim, Y. J. & Zhang, D. Molecular Control of Male Fertility for Crop Hybrid Breeding. *Trends Plant Science* **23**, 53-65, doi:10.1016/j.tplants.2017.10.001 (2018).

6 Lapinski, M. & Stojalowski, S. Occurence and genetic identity of male sterility-inducing cytoplasm in rye. *Plant Breeding and Seed Science* **48** (2003).

7 Miedaner, T. *et al.* Mapping of genes for male-fertility restoration in 'Pampa' CMS winter rye (Secale cereale L.). *Theoretical and Applied Genetics* **101**, 1226-1233 (2000).

8 Hackauf, B., Bauer, E., Korzun, V. & Miedaner, T. Fine mapping of the restorer gene Rfp3 from an Iranian primitive rye (Secale cereale L.). *Theoretical and Applied Genetics* **130**, 1179-1189, doi:10.1007/s00122-017-2879-3 (2017).

9 Stracke, S. *et al.* Development of PCR-based markers linked to dominant genes for male-fertility restoration in Pampa CMS of rye (Secale cereale L.). *Theoretical and Applied Genetics* **106**, 1184-1190, doi:10.1007/s00122-002-1153-4 (2003).

10 Niedziela, A., Brukwinski, W. & Bednarek, P. T. Genetic mapping of pollen fertility restoration QTLs in rye (Secale cereale L.) with CMS Pampa. *Journal of Applied Genetics*, 1-14, doi:10.1007/s13353-020-00599-9 (2021).

11 Adolf, K. A new source of spontaneous sterility in winter rye-preliminary results. *Proceedings of Eucarpia Meeting of the Cereal Section on Rye* **1985**, 11-13 (1986).

12 Kobylianskii, V. D. Production of sterile analogues of winter rye varieties, sterile mantainers and fertile restorers. *Trudy po Prikladnoi Botanike Genetike i Selektsii* (1971).

13 Lapinski, M. Cytoplasmic-genic type of male sterility in Secale montanum Guss. *Wheat Information Service* **35** (1972).

14 Madej, L. Research on male sterility in rye. *Hodowla Rosl Aklim Nasienn* (1975).

15 Melz, G. & Adolf, K. Genetic analysis of rye (Secale cereale L.) Genetics of male sterility of the G-type. *Theoretical and Applied Genetics* **82**, 761-764, doi:10.1007/BF00227322 (1991).

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. *Journal of Applied Genetics* **57**, 439-451, doi:10.1007/s13353-016-0347-4 (2016).

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 Breeding* **123**, 428-433 (2004).

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. *Journal of Applied Genetics* **46**, 371-373 (2005).

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. *Fortschritte der Pflanzenzuechtung* (1995).

21 Kodisch, A. *et al.* Ergot infection in winter rye hybrids shows differential contribution of male and female genotypes and environment. *Euphytica* **216**, doi:10.1007/s10681-020-02600-2 (2020).

22 Klotz, J. L. Activities and Effects of Ergot Alkaloids on Livestock Physiology and Production. *Toxins (Basel)* **7**, 2801-2821, doi:10.3390/toxins7082801 (2015).

23 Blaney, B. J., Molloy, J. B. & Brock, I. J. Alkaloids in Australian rye ergot (Claviceps purpurea) sclerotia: implication for food and stockfeed regulations. *Animal Production Science* **49**, 975-982 (2009).

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 Breeding* **129**, 58-62, doi:10.1111/j.1439-0523.2009.01646.x (2010).

25 Geiger, H. H. & Miedaner, T. Genetic basis and phenotypic stability of male-fertility restoration in rye. *Vorträge für Pflanzenzüchtung* (1996).

26 Miedaner, T., Wilde, P. & Wortmann, H. Combining ability of non-adapted sources for male-fertility restoration in Pampa CMS of hybrid rye. *Plant Breeding* **124**, 39-43 (2004).

27 Falke, K. C., Wilde, P. & Miedaner, T. Rye introgression lines as source of alleles for pollen-fertility restoration in Pampa CMS. *Plant Breeding* **128**, 528-531, doi:10.1111/j.1439-0523.2008.01589.x (2009).

28 Miedaner, T. *et al.* Correlated effects of exotic pollen-fertility restorer genes on agronomic and quality traits of hybrid rye. *Plant Breeding* **136**, 224-229, doi:10.1111/pbr.12456 (2017).

29 KWS. *PollenPlus*, <https://www.kws.com/gb/en/products/cereals/hybrid-rye/pollenplus-kws-files-for-ergot-patent-in-hybrid-rye/> (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 Breeding and Seed Science* **47**, 47-55 (2003).

31 Gaborieau, L., Brown, G. G. & Mireau, H. The Propensity of Pentatricopeptide Repeat Genes to Evolve into Restorers of Cytoplasmic Male Sterility. *Frontiers in Plant Science* **7**, 1816, doi:10.3389/fpls.2016.01816 (2016).

32 Dellanoy, E., Stanley, W. A., Bond, C. S. & Small, I. D. Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. *Biochemical Society Transactions* **35**, 1643-1647 (2007).

33 Barkan, A. *et al.* A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genetic* **8**, e1002910, doi:10.1371/journal.pgen.1002910 (2012).

34 O'Toole, N. *et al.* On the expansion of the pentatricopeptide repeat gene family in plants. *Molecular Biology and Evolution* **25**, 1120-1128, doi:10.1093/molbev/msn057 (2008).

35 Lurin, C. *et al.* Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16**, 2089-2103, doi:10.1105/tpc.104.022236 (2004).

36 Schmitz-Linneweber, C. & Small, I. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends in Plant Science* **13**, 663-670, doi:10.1016/j.tplants.2008.10.001 (2008).

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* **205**, 239-249, doi:10.1534/genetics.116.196105 (2017).

38 Liu, Y. J. *et al.* A Plastid-Localized Pentatricopeptide Repeat Protein is Required for Both Pollen Development and Plant Growth in Rice. *Scientific reports* **7**, 1-12, doi:10.1038/s41598-017-10727-x (2017).

39 Gully, B. S. *et al.* The solution structure of the pentatricopeptide repeat protein PPR10 upon binding atpH RNA. *Nucleic Acids Research* **43**, 1918-1926, doi:10.1093/nar/gkv027 (2015).

40 Ban, T. *et al.* Structure of a PLS-class pentatricopeptide repeat protein provides insights into mechanism of RNA recognition. *Journal of Biological Chemistry* **288**, 31540-31548, doi:10.1074/jbc.M113.496828 (2013).

41 Fujii, S. & Small, I. The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytologist* **191**, 37-47, doi:10.1111/j.1469-8137.2011.03746.x (2011).

42 Uyttewaal, M. *et al.* Characterization of Raphanus sativus pentatricopeptide repeat proteins encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. *Plant Cell* **20**, 3331-3345, doi:10.1105/tpc.107.057208 (2008).

43 Sykes, T. *et al.* In Silico Identification of Candidate Genes for Fertility Restoration in Cytoplasmic Male Sterile Perennial Ryegrass (Lolium perenne L.). *Genome biology and evolution* **9**, 351-362, doi:10.1093/gbe/evw047 (2017).

44 Melonek, J. *et al.* High intraspecific diversity of Restorer-of-fertility-like genes in barley. *The Plant Journal* **97**, 281-295 (2019).

45 Rabanus-Wallace, M. T. *et al.* Chromosome-scale genome assembly provides insights into rye biology and evolution. *Nature Genetics* **(In press)**, doi:10.1101/2019.12.11.869693 (2021).

46 Wright, Stephen I., Ness, Rob W., Foxe, John P. & Barrett, Spencer C. H. Genomic Consequences of Outcrossing and Selfing in Plants. *International Journal of Plant Sciences* **169**, 105-118, doi:10.1086/523366 (2008).

47 Rizzolatti, C. *et al.* Map-based cloning of the fertility restoration locus Rfm1 in cultivated barley (Hordeum vulgare). *Euphytica* **213**, doi:10.1007/s10681-017-2056-4 (2017).

48 Klein, R. R. *et al.* 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. *Theoretical and Applied Genetics* **111**, 994-1012, doi:10.1007/s00122-005-2011-y (2005).

49 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. *Molecular and Cellular Biology* **28**, 5337-5547, doi:10.1128/MCB.00563-08 (2008).

50 Kazama, T. & Toriyama, K. A fertility restorer gene, Rf4, widely used for hybrid rice breeding encodes a pentatricopeptide repeat protein. *Rice* **7**, 1-5 (2014).

51 Hu, J. *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* **24**, 109-122, doi:10.1105/tpc.111.093211 (2012).

52 Huang, W. *et al.* Pentatricopeptide-repeat family protein RF6 functions with hexokinase 6 to rescue rice cytoplasmic male sterility. *Proceedings of the National Academy of Sciences* **112**, 14984-14989, doi:10.1073/pnas.1511748112 (2015).

53 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* **15**, e0239541, doi:10.1371/journal.pone.0239541 (2020).

54 McHugh, M. L. The chi-square test of independence. *Biochemia medica* **23**, 143-149, doi:10.11613/bm.2013.018 (2013).

55 Wang, S. *et al.* Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol Journal* **12**, 787-796, doi:10.1111/pbi.12183 (2014).

56 Haseneyer, G. *et al.* From RNA-seq to large-scale genotyping - genomics resources for rye (Secale cereale L.). *BMC Plant Biol* **11**, 131, doi:10.1186/1471-2229-11-131 (2011).

57 Bauer, E. *et al.* Towards a whole-genome sequence for rye (Secale cereale L.). *The Plant Journal* **89**, 853-869, doi:10.1111/tpj.13436 (2017).

58 NCBI. *National Center for Biotechnology Information.* , <https://www.ncbi.nlm.nih.gov> (2020).

59 Rstudio Team. Rstudio: Integrated Development for R. RStudio, Inc., Boston. http://www.rstudio.com (2015).

60 R Core Team. R a language and environment for statistical computing. R foundation for statistical Computing, Vienna, Austria. https://www.R-project.org/. (2020).

61 Lipka, A. E. *et al.* GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**, 2397-2399, doi:10.1093/bioinformatics/bts444 (2012).

62 El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Research* **47**, D427-D432, doi:10.1093/nar/gky995 (2019).

63 HMMER. *HMMER: biosequence analysis using profile hidden Markov models*, <http://hmmer.org> (2020).

64 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240 (2014).

65 Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676 (2005).

66 Howe, K. L. *et al.* Ensembl Genomes 2020-enabling non-vertebrate genomic research. *Nucleic Acids Research* **48**, D689-D695, doi:10.1093/nar/gkz890 (2020).

67 Clayton, D. snpStats: SnpMatrix and XSnpMatrix classes and methods. *R package version 1.36.0* (2019).

68 Shin, J. H., Blay, S., McNey, B. & Graham, J. LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *Journal of Statistical Software* **16**, 1-10, doi:10.18637/jss.v000.i00 (2006).

69 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.). *Theoretical and Applied Genetics* **97**, 99-102 (1998).

70 Vilhjalmsson, B. J. & Nordborg, M. The nature of confounding in genome-wide association studies. *Nature Reviews Genetics* **14**, 1-2, doi:10.1038/nrg3382 (2013).

71 Korte, A. & Farlow, A. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* **9**, 29 (2013).

72 Zhang, Z. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nature genetics* **42**, 355-360, doi:10.1038/ng.546 (2010).

73 Martis, M. M. *et al.* Reticulate evolution of the rye genome. *Plant Cell* **25**, 3685-3698, doi:10.1105/tpc.113.114553 (2013).

74 Kazama, T., Nakamura, T., Watanabe, M., Sugita, M. & Toriyama, K. Suppression mechanism of mitochondrial ORF79 accumulation by Rf1 protein in BT-type cytoplasmic male sterile rice. *The Plant Journal* **55**, 619-628, doi:10.1111/j.1365-313X.2008.03529.x (2008).

75 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. *Molecular Genetics and Genomics* **293**, 451-462, doi:10.1007/s00438-017-1396-z (2018).

76 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. *Theoretical and Applied Genetics* **130**, 1253-1266, doi:10.1007/s00122-017-2885-5 (2017).

77 Devos, K. M. *et al.* Chromosomal rearrangements in the rye genome relative to what of wheat. *Theoretical and Applied Genetics* **85**, 673-680 (1993).

78 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 Science* **24**, 17-20 (1984).

79 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. *Frontiers in Plant Science* **11**, doi:10.3389/fpls.2020.577475 (2020).

80 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. *Molecular and General Genetics MGG* **247**, 351-357, doi:10.1007/BF00293203 (1995).

81 Du, H., Maan, S. S. & Hammond, J. J. Genetic Analysis of Male-Fertility Restoration in Wheat: III. Effects of Aneuploidy. *Crop science* **31**, 319-322 (1991).

82 Ui, H. *et al.* High-resolution genetic mapping and physical map construction for the fertility restorer Rfm1 locus in barley. *Theoretical and Applied Genetics* **128**, 283-290, doi:10.1007/s00122-014-2428-2 (2015).

83 Bernhard, T., Koch, M., Snowdon, R. J., Friedt, W. & Wittkop, B. Undesired fertility restoration in msm1 barley associates with two mTERF genes. *Theoretical and Applied Genetics* **132**, 1335-1350, doi:10.1007/s00122-019-03281-9 (2019).

84 Hackauf, B., Korzun, V., Wortmann, H., Wilde, P. & Wehling, P. Development of conserved ortholog set markers linked to the restorer gene Rfp1 in rye. *Molecular Breeding* **30**, 1507-1518, doi:10.1007/s11032-012-9736-5 (2012).

85 Huang, S. *et al.* Phylogenetic analysis of the acetyl-CoA carboxylase and 3-phosphoglycerate kinase loci in wheat and other grasses. *Plant molecular biology* **48**, 805-820 (2002).

86 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. *The Plant Journal* **65**, 359-367, doi:10.1111/j.1365-313X.2010.04427.x (2011).

87 Fujii, S. & Toriyama, K. Suppressed expression of RETROGRADE-REGULATED MALE STERILITY restores pollen fertility in cytoplasmic male sterile rice plants. *Proceedings in the National Academy of Sciences* **106**, 9513-9518 (2009).
88 Liu, F. & Schnable, P. S. Functional specialization of maize mitochondrial aldehyde dehydrogenases. *Plant Physiology* **130**, 1657-1674, doi:10.1104/pp.012336 (2002).
89 Jaqueth, J. S. *et al.* Fertility restoration of maize CMS-C altered by a single amino acid substitution within the Rf4 bHLH transcription factor. *The Plant Journal* **101**, 101-111, doi:10.1111/tpj.14521 (2020).
90 Kitazaki, K. *et al.* Post-translational mechanisms are associated with fertility restoration of cytoplasmic male sterility in sugar beet (Beta vulgaris). *The Plant Journal* **83**, 290-299, doi:10.1111/tpj.12888 (2015).
91 Wilde, P. *et al.* Restorer Plants. KWS SAAT SE. (U.S. Patent App. 16/064304). (2019).
92 Pan, Z. *et al.* A Mitochondrial Transcription Termination Factor, ZmSmk3, Is Required for nad1 Intron4 and nad4 Intron1 Splicing and Kernel Development in Maize. *G3: Genes, Genomes, Genetics* **9**, 2677-2686, doi:10.1534/g3.119.400265 (2019).
93 Kubo, T., Arakawa, T., Honma, Y. & Kitazaki, K. What Does the Molecular Genetics of Different Types of Restorer-of-Fertility Genes Imply? *Plants* **9**, doi:10.3390/plants9030361 (2020).
94 Lapinski, M. & Stojalowski, S. The C-source of sterility-inducing cytoplasm in rye: Origin, identity and occurrence. *Vorträge für Pflanzenzüchtung* **35**, 51-60 (1996).
95 Warsecha, R. & Salak-Warzecha, K. Comparative studies on CMS sources in rye. *Vorträge für Pflanzenzüchtung* **35**, 39-49 (1996).

Nikolaj M. Vendelbo
Khalid Mahmood
Pernille Sarup
Peter S. Kristensen
Jihad Orabi
Corresponding author(s): Ahmed Jahoor

Last updated by author(s): 25/2/2021

Editorial Policy Checklist

This form is used to ensure compliance with Nature Research editorial policies related to research ethics and reproducibility. For further information, please see our editorial policies site. All relevant questions on the form must be answered.

Competing interests

Policy information about competing interests

Competing interests declaration

In the interest of transparency and to help readers form their own judgements of potential bias, Nature Research journals require authors to declare any competing financial and/or non-financial interest in relation to the work described in the submitted manuscript.

☒ We declare that none of the authors have competing financial or non-financial interests as defined by Nature Research.

☐ We declare that one or more of the authors have a competing interest as defined by Nature Research.

Authorship

Policy information about authorship

Prior to submission all listed authors must agree to all manuscript contents, the author list and its order and the author contribution statements. Any changes to the author list after submission must be approved by all authors.

☒ We have read the Nature Research Authorship Policy and confirm that this manuscript complies.

Data availability

Policy information about availability of data

Data availability statement

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

☒ We have provided a full data availability statement in the manuscript.

Mandated accession codes (where applicable)

Confirm that all relevant data are deposited into a public repository and that accession codes are provided.

☒ All relevant accession codes are provided ☐ Accession codes will be available before publication ☐ No data with mandated deposition

Code availability

Policy information about availability of computer code

Code availability statement

For all studies using custom code or mathematical algorithm that is deemed central to the conclusions, the manuscript must include a statement under the heading "Code availability" describing how readers can access the code, including any access restrictions. Code availability statements should be provided as a separate section after the data availability statement but before the References.

☒ We have provided a full code availability statement in the manuscript

Data presentation

For all data presented in a plot, chart or other visual representation confirm that:

n/a	Confirmed
☐	☒ Individual data points are shown when possible, and always for $n \leq 10$
☐	☒ The format shows data distribution clearly (e.g. dot plots, box-and-whisker plots)
☐	☒ Box-plot elements are defined (e.g. center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers)
☐	☒ Clearly defined error bars are present and what they represent (SD, SE, CI) is noted

Image integrity

Policy information about image integrity

☒ We have read Nature Research's image integrity policy and all images comply.

Unprocessed data must be provided upon request. Please double-check figure assembly to ensure that all panels are accurate (e.g. all labels are correct, no inadvertent duplications have occurred during preparation, etc.).

Where blots and gels are presented, please take particular care to ensure that lanes have not been spliced together, that loading controls are run on the same blot, and that unprocessed scans match the corresponding figures.

Additional policy considerations

Some types of research require additional policy disclosures. Please indicate whether each of these apply to your study. If you are not certain, please read the appropriate section before selecting a response.

Does not apply	Involved in the study
☐	☐ Macromolecular structural data
☐	☒ Unique biological materials
☐	☐ Research animals and/or animal-derived materials that require ethical approval
☐	☐ Human embryos, gametes and/or stem cells
☐	☐ Human research participants
☐	☐ Clinical data

Macromolecular structural data

Policy information about special considerations for specific types of data

Validation report

☐ We have provided an official validation report from wwPDB for all macromolecular structures studied.

Biological materials

Policy information about availability of materials

Obtaining biological materials: All rye lines assayed in the study belong to an elite hybrid rye breeding germplasm and require an MTA for transfer to another party.

☐ We have described these restrictions in the manuscript. ☒ We have described how to obtain all materials in the manuscript.

Research animals

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Ethical compliance

☐ We have complied with all relevant ethical regulations and include a statement affirming this in the manuscript.

Ethics committee

☐ We have disclosed the name(s) of the board and institution that approved the study protocol in the manuscript.

Human embryos, gametes and stem cells

Policy information about studies involving human embryos, gametes and stem cells

Manuscripts involving the use of human embryos, gametes or stem cells must include an ethics statement that provides the following information:
- The institutional and/or licensing committee(s) that approved the study protocol
- Confirmation that informed consent was obtained from all recipients and/or donors of cells or tissues
- The conditions for donating materials for the research

☐ We have read the Nature Research policy on human embryos, gametes and stem cells and have complied with policy requirements.

Human research participants

Policy information about studies involving human research participants

Ethical compliance

☐ We have complied with all relevant ethical regulations and include a statement affirming this in the manuscript.

Ethics committee

Confirm that the manuscript states the name(s) of the board and/or institution that:

☐ Approved the study protocol -OR- ☐ Provided guidelines for study procedures (if protocol approval is not required)

Informed consent

☐ We have obtained informed consent from all participants and this is noted in the manuscript.

Identifiable images

For publication of identifiable images of research participants, confirm that consent to publish was obtained and is noted in the Methods.

Authors must ensure that consent meets the conditions set out in the Nature Research participant release form.

☐ Yes ☐ No identifiable images of human research participants

Clinical studies

Policy information about clinical studies

Clinical trial registration

☐ We have provided the trial registration number from ClinicalTrials.gov or an equivalent agency in the manuscript.

Phase 2 and 3 randomized controlled trials

We have provided the CONSORT checklist with your submission.

☐ Yes ☐ No ☐ Not a phase 2/3 randomized controlled trial

Tumor marker prognostic studies

We have followed the REMARK reporting guidelines.

☐ Yes ☐ No ☐ Not a tumor marker prognostic study

I certify that all the above information is complete and correct.

Typed signature Nikolaj Meisner Vendelbo Date 25/2/2021

Figures

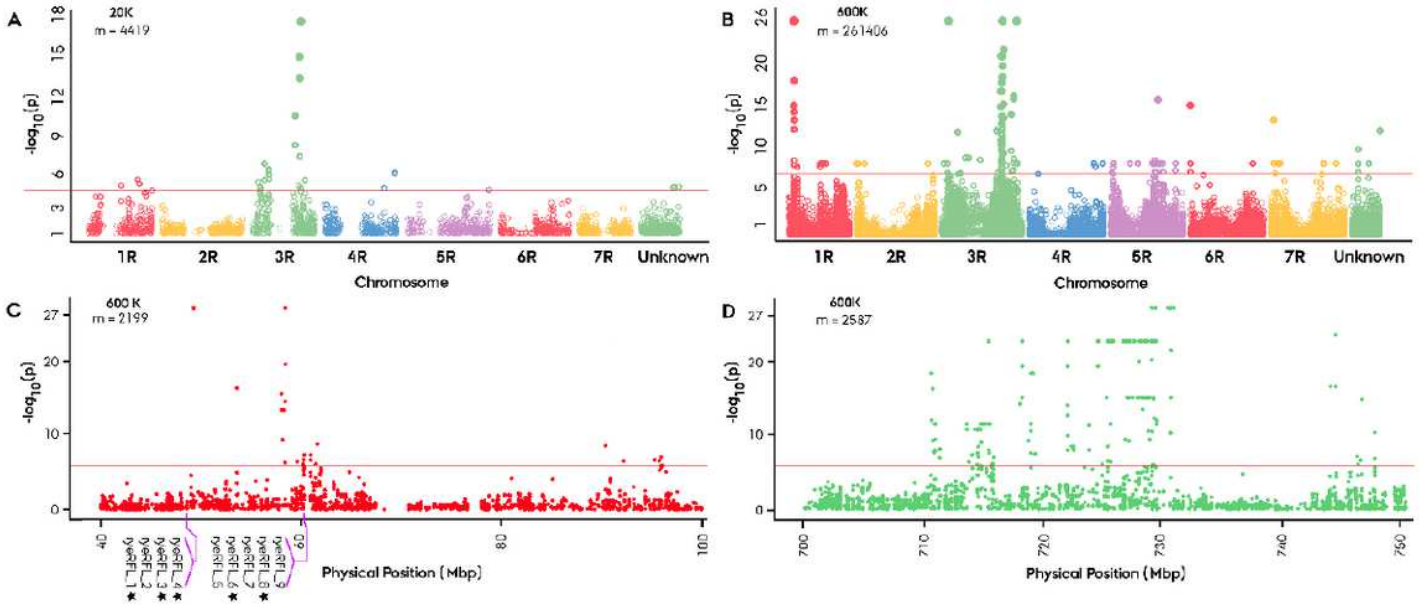


Figure 1

Manhattan plot for genome wide association study (GWAS) on population origin (case controle) on Nordic Seed elite hybrid rye 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 1RSregion including position of identified restoration of male-fertility (Rf) like pentatricopeptide repeat (RFL-PPR) genes of which RFL-PPRs expressed in G-type hybrids are marked with an asterix, D) Manhattan plot of the 600K SNP array 3RL region.

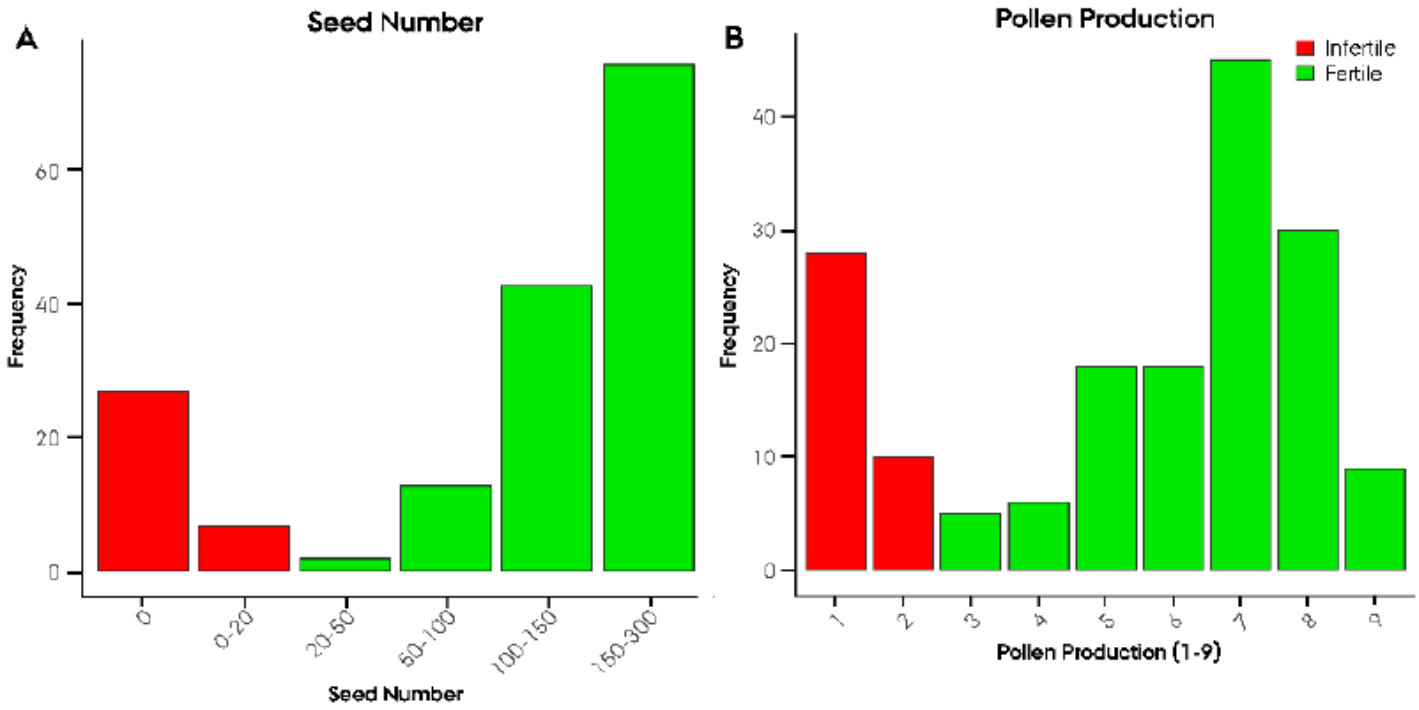


Figure 2

Phenotypic distribution of restoration of male fertility related traits, A) Seed number, and B) Pollen Production in 181 F2 plants derived from a hybrid rye cv. Stannos

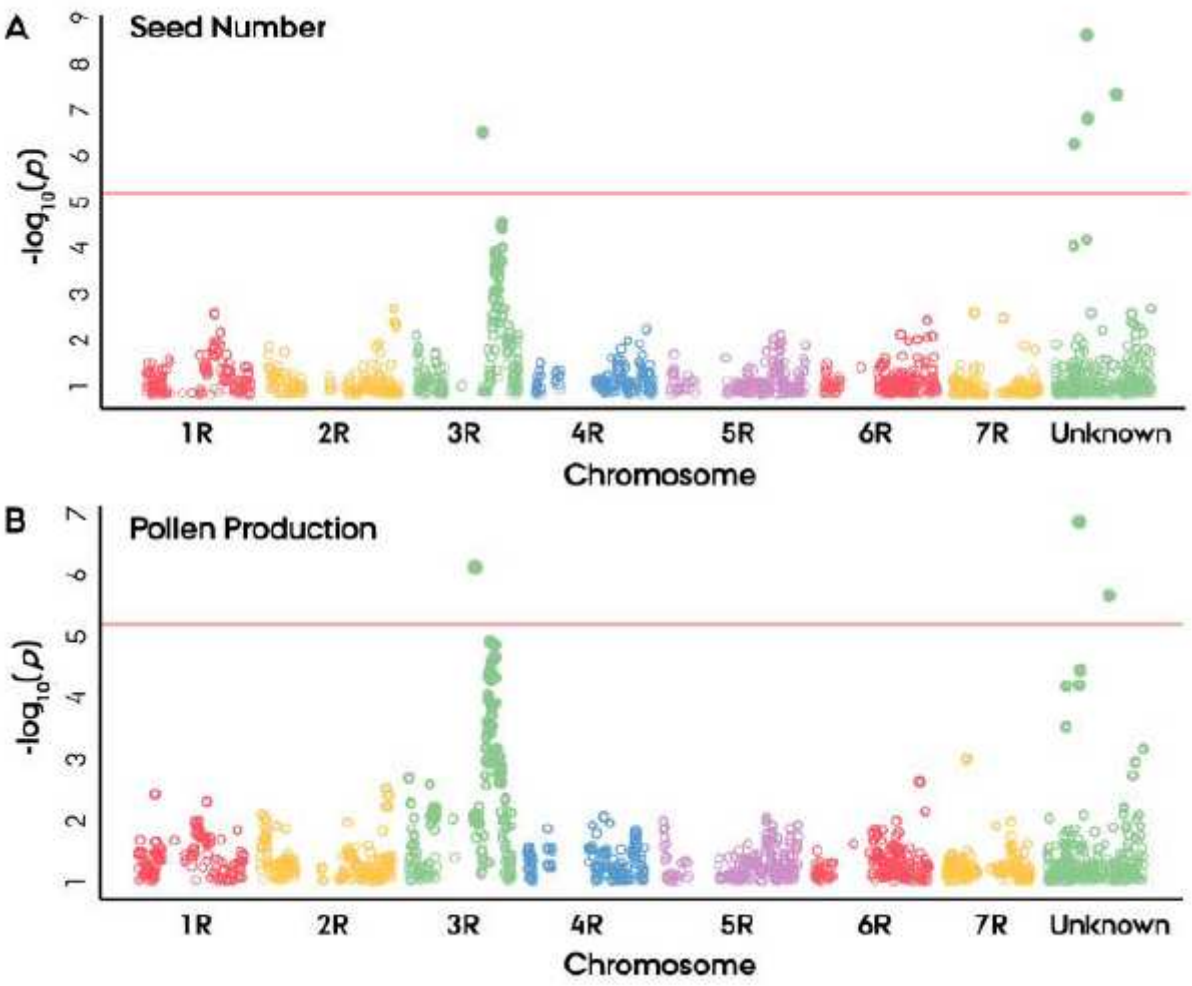


Figure 3

Manhattan plot for genome wide association study (GWAS) on two restoration of male fertility related phenotypic scores, A) Seed Number and B) Pollen Production (1-9) in a F2 biparental population (n = 181) derived from the hybrid rye cv. Stannos. Significant association was identified using criterion of -log10(p) > 5.2 depicted as a red line.

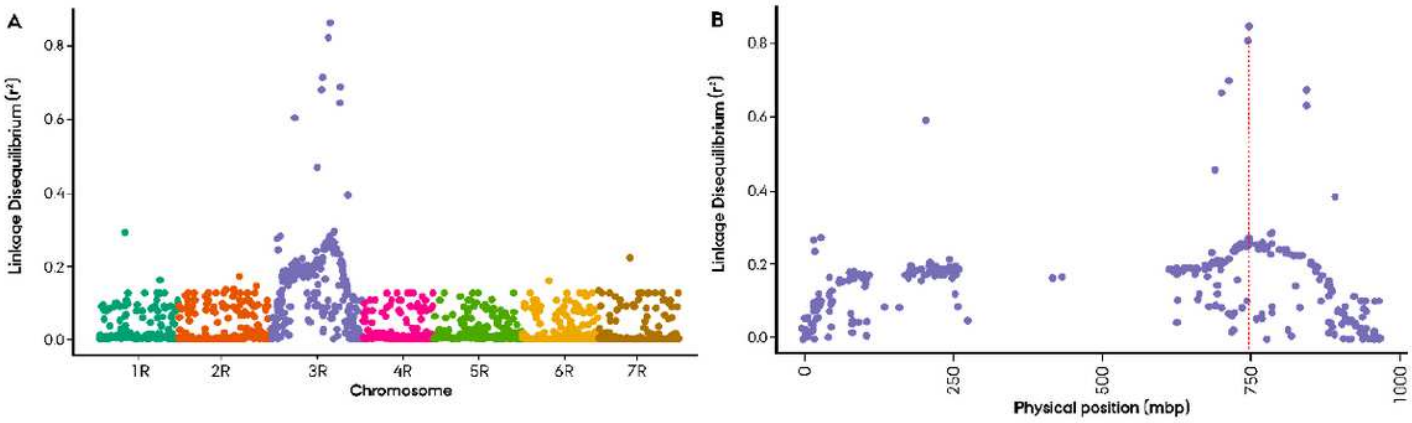


Figure 4

Genome wide pairwise linkage disequilibrium (LD) between a highly Rf associated wheat derived SNP marker, AX_158558079 and 3493 informative SNP markers in 181 rye (Secale cereale L.) F2 plants. A) Chromosome-wise distribution of LD, B) LD distribution on 3R chromosome.

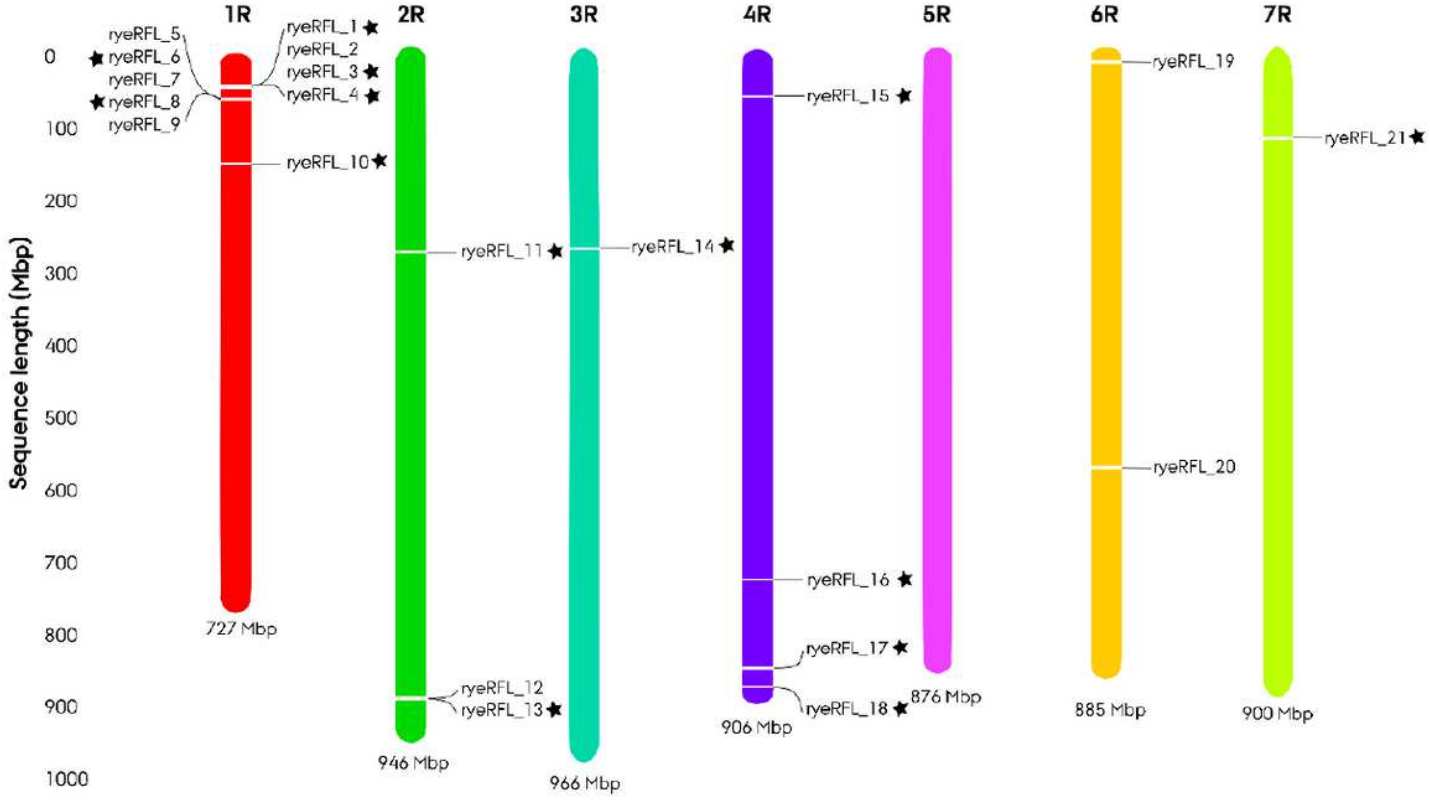


Figure 5

Genomic location of 22 in silico annotated restoration of male fertility-like (RFL) pentotricopeptide repeats (PPR) genes in the rye reference genome depicted in physical distance (Megabase pair, Mbp). RFL-PPRs identified in both spike transcriptome assemblies of two rye hybrids, cv. Stannos and cv. Helltop are marked with an asterix.

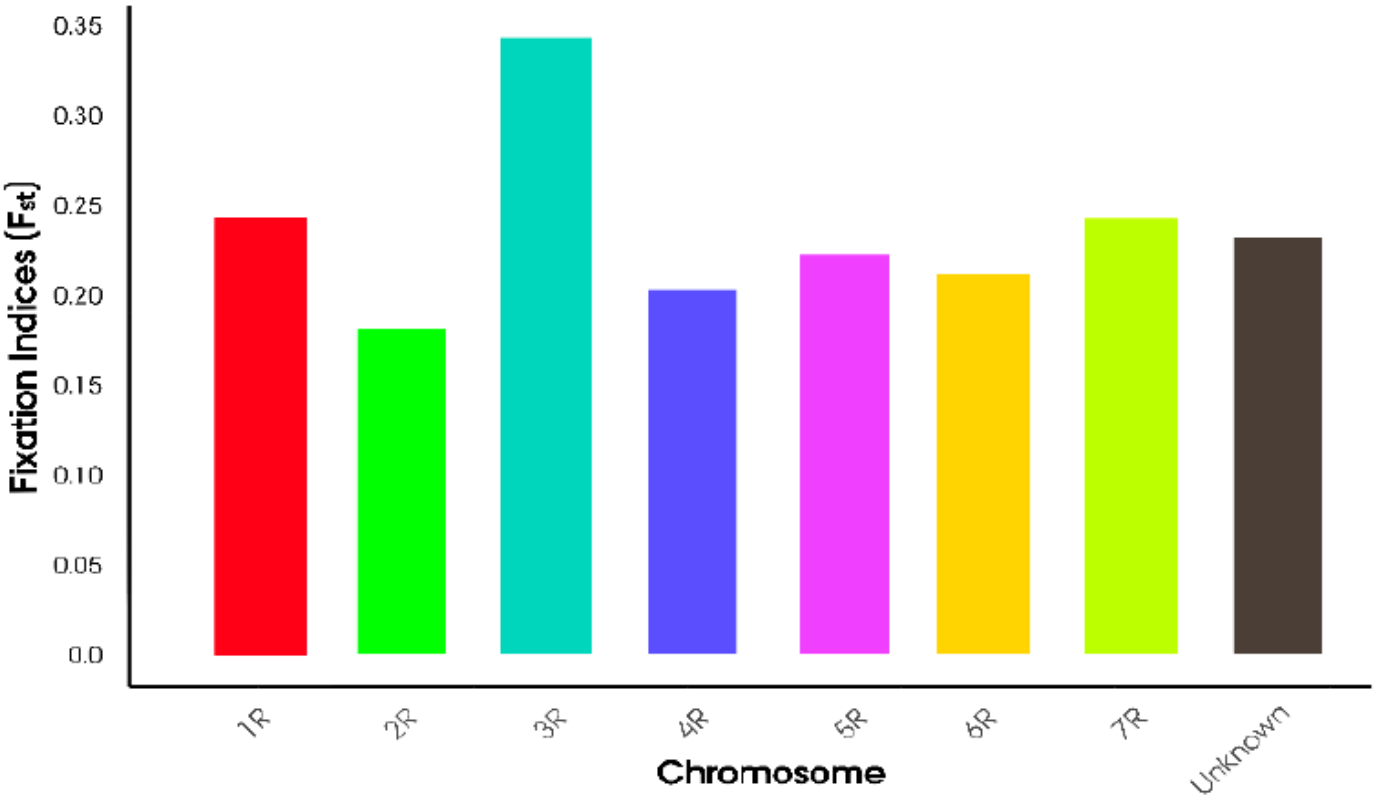


Figure 6

Intrachromosomal fixation indices (Fst) of Nordic Seed hybrid rye elite breeding germplasm using 600K rye SNP array genotype data. Populations comprised of 92 restorer and 88 non-restorer germplasm lines.

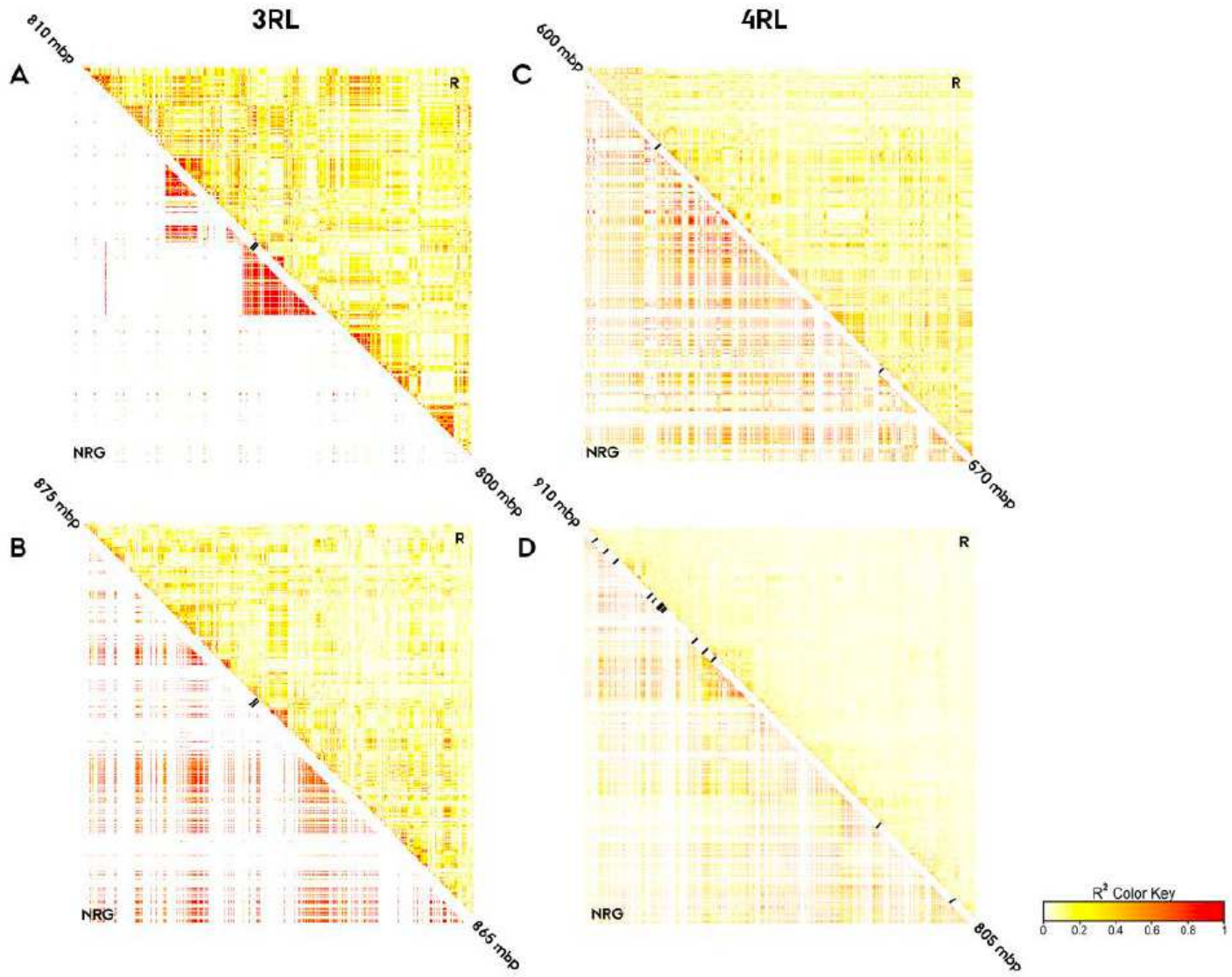


Figure 7

Comparative pairwise linkage disequilibrium (R2) of four regions on 3R (A,B) and 4R (C,D) harboring 63 restoration of male fertility annotated SNP markers (black lines) in Nordic Seed hybrid rye elite breeding germplasm. 92 restorer (R) and 88 non-restorer germplasm (NRG) lines were genotyped 600K rye SNP array.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.zip