Technische Universität München

Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt

Lehrstuhl für Pflanzenzüchtung

Genetic studies towards a refined understanding of the fertility-restoring loci *Rf1* and *Rf3* in G-type cytoplasmic male-sterile wheat

Manuel Geyer

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzende:		Prof. Dr. Caroline Gutjahr
Prüfer der Dissertation:	1.	apl. Prof. Dr. Volker Mohler
	2.	Prof. Dr. Chris-Carolin Schön

Die Dissertation wurde am 08.05.2019 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 09.09.2019 angenommen.

Contents

List of figures

List of tables

List of abbreviations

BAC	Bacterial artificial chromosome
BC_1	Backcross generation one
BLAST	Basic local alignment search tool
bp	Base pair
BT	Bora-Taichung
CAPS	Cleaved amplified polymorphic sequence
Cas9	CRISPR-associated 9
CHA	Chemical hybridising agent
CIM	Composite interval mapping
cM	Centimorgan
CMS	Cytoplasmic male sterility (Cytoplasmic male-sterile)
CRISPR	Clustered regularly inter-spaced short palindromic repeats
DArT	Diversity array technology
EST	Expressed sequence tag
FDR	False-discovery rate
F_2	Filial generation two
Gbp	Giga base pairs
GCA	General combining ability
GS	Genome-wide selection
HL	Hong-Lian
LOD	Logarithm of the odds
MABC	Marker-assisted backcrossing
MAF	Minor allele frequency
MARS	Marker-assisted recurrent selection
MAS	Marker-assisted selection
Mbp	Mega base pairs
mtDNA	Mitochondrial DNA
NB-LRR	Nucleotide-binding and leucine-rich repeat
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PPR	Pentatricopeptide repeat
QTL	Quantitative trait locus
Rf	Restorer-of-fertility
RFL-PPR	Rf-like PPR
RFLP	Restriction fragment length polymorphism
SCA	Specific combining ability
SNP	Single-nucleotide polymorphism
SPT	Seed production technology
SSR	Simple sequence repeat
T_a	Annealing temperature
U	Enzyme unit
WA	Wild-abortive

1 Introduction

1.1 The importance of wheat breeding

Wheats are annual grasses belonging to the genus *Triticum* in the botanical tribe Triticeae, which also contains the crop species barley (*Hordeum vulgare*) and rye (*Secale cereale*). The genus comprises several diploid, tetraploid and hexaploid species with a common basic chromosome number of seven (Snape and Pánková 2007). Wheat was one of the first crops that were domesticated as part of the Neolithic Revolution, which marked the beginning of agriculture. The earliest cultivated wheat species were diploid einkorn (*Triticum monococcum*; genome A^mA^m) and allotetraploid emmer (*Triticum turgidum*; genome BBA^uA^u), of which the latter arose from the hybridisation of *Triticum urartu* (genome A^uA^u) and an unknown *Aegilops* species (genome BB) closely related to *Aegilops speltoides* (Salse et al. 2008). Hybridisation between domesticated emmer (*Triticum turgidum* ssp. *dicoccum*; genome BBA^uA^u) and *Aegilops tauschii* (genome DD) led to the formation of hexaploid Asian spelt (*Triticum aestivum* ssp. *spelta*) and its descendant common wheat (*Triticum aestivum* ssp. *aestivum*; both genome BBA^uA^uDD). Hexaploid European spelt (*Triticum aestivum* ssp. *spelta*) originated from the hybridisation of domesticated emmer and common wheat and comprises the same three subgenomes as Asian spelt and common wheat (Faris 2014). Archaeological records suggested that the domestication of wheat began about 8,000 B.C. in the core area of the Fertile Crescent in present-day south-eastern Turkey and northern Syria, from where it started spreading to Europe, Africa and Asia about 6,000 B.C. Wheat was introduced to America in 1529 and was taken to Australia in 1788 (Lev-Yadun et al. 2000; Feldman 2001). Driven by the evolution under cultivation, wheat has adapted to a wide range of environmental conditions, from 67° N in Scandinavia and Russia to 45° S in Argentina, including elevated regions in the tropics and subtropics (Feldman 2001; Shewry 2009). Owing to its adaptation to different environments, wheat is the most widely cultivated crop worldwide occupying 16% of arable land area. It is grown in more than 120 countries and contributes 11% to total crop production. The countries with the highest wheat production are China, India and Russia (Langridge 2017; FAOSTAT 2019). About 95% of the wheat grown worldwide is common wheat, with most of the remaining 5% being tetraploid durum wheat

(*Triticum turgidum* ssp. *durum*). Other wheats, such as spelt, emmer and einkorn, exist as niche products that are cultivated in regions including Spain, Turkey, the Balkans, the Indian subcontinent and Alpine regions (Fossati and Ingold 2001; Shewry 2009). Wheat has become one of the major staple crops and is of critical importance for global food security. As starch accounts for 65–75% of its grain, wheat is a considerable source of carbohydrates for a large part of the world population. Despite the low protein content of 7–17%, wheat contributes about 20% to the daily *per capita* protein intake worldwide compared to 12% for rice (*Oryza* spp.) and 4% for maize (*Zea mays*). Additionally, wheat provides minerals, vitamins, essential amino acids and dietary fibre (Shewry 2009; Ndolo and Beta 2017; FAOSTAT 2019). The end use of wheat is mainly determined by grain hardness, the content and composition of proteins and dough strength. Whereas wheat with low protein content and soft grains is processed to pastry, higher protein contents and harder grains are necessary for bread and pasta (Batey 2017; Ndolo and Beta 2017). Additionally, wheat has several other end uses, including animal feed and industrial starch production (Curtis and Halford 2014).

From 1961 to 2017, global wheat production increased by 247%, whereas the land area used for wheat cultivation expanded only by 7% (FAOSTAT 2019). This yield increase can be mainly explained by the breeding progress initiated by the Green Revolution and growth in the use of inputs, such as irrigation and fertilisers (Evenson and Gollin 2003; Langridge 2017). At the same time, the demand for wheat has increased as a result of population growth, an increasing *per capita* consumption of animal-derived products and the use of wheat starch for non-food applications (Curtis and Halford 2014). As the world population is predicted to rise from its current level of 7.7 billion to 9.8 billion by 2050 and 11.2 billion by 2100, the demand for wheat will further increase (UN DESA 2019). Tilman et al. (2011) estimated that the global crop production demand will increase by 100–110% from 2005 to 2050. Previous studies suggested that boosting crop yields, rather than clearing new land area, is the most sustainable approach to meet these demands (Godfray et al. 2010; Foley et al. 2011). However, at the current yield gain of 0.9% per year, wheat yields will only be increased by ~38% till 2050 (Ray et al. 2013). Future strategies to lift yields must also consider the detrimental effects of climate change as global wheat yields are predicted to decline by 4.1–6.4% with a

1 °C increase of global temperature (Liu et al. 2016a). The most promising strategies to meet this challenge are innovations in wheat breeding, together with an enhanced deployment of existing wheat varieties and improved crop management (Foley et al. 2011). To accelerate the rate of genetic gain, an intensified use of genetic and genomic resources and the application of advanced breeding techniques must be considered.

1.2 Genomic resources and marker-assisted selection in wheat

Recent years have witnessed the development of extensive genomic resources, which have the potential to accelerate future progress in wheat breeding. The availability of genomic resources for common wheat has been impeded by the large size of its genome, which comprises 17 giga base pairs (Gbp), the high level of homoeology between its three subgenomes and its high content of repetitive sequences (Eversole et al. 2017). To overcome obstacles associated with such complex genomes, several target enrichment strategies have been developed that enable sequencing of specific genomic regions. These strategies include target enrichment based on polymerase chain reaction (PCR), hybridisation-based enrichment, restriction enzyme-based enrichment and enrichment of expressed gene sequences (Cronn et al. 2012). In wheat, the first initiatives to develop genomic resources focused on expressed sequence tags (ESTs). Physical mapping of ESTs enabled first comparative analyses of common wheat and rice genomes (Sorrells et al. 2003; La Rota and Sorrells 2004). Currently available transcriptomic sequences of common wheat are organised in 178,464 UniGene clusters (NCBI 2019). The first wheat sequence information beyond expressed genes was generated by sequencing individual bacterial artificial chromosomes (BACs; Devos et al. 2005) and BAC-end sequencing (Paux et al. 2006), which was followed by survey sequencing of chromosome 3B (Paux et al. 2008; Choulet et al. 2010). A low-coverage whole-genome survey sequence of the common wheat cultivar Chinese Spring was enabled by the advent of second generation sequencing technology (Brenchley et al. 2012). For the same cultivar, the International Wheat Genome Sequencing Consortium (IWGSC) developed a chromosome-based survey sequence (IWGSC CSS) by sequencing individual flow-sorted chromosome arms to a depth of at least 30-fold coverage (IWGSC 2014). The first common wheat reference sequence (RefSeq v1.0) was obtained from Chinese Spring by integrating the

chromosome-based survey sequence with sequences of the BAC minimal tiling path for each chromosome, a whole-genome sequence assembly (IWGSC WGA v0.4) and other genomic resources (IWGSC 2018). Besides the efforts to develop genomic sequences for common wheat, whole-genome sequences were also generated for the relatives *T. urartu*, *Ae. tauschii*, *Ae. speltoides* and *Triticum sharonensis* (Jia et al. 2013; Ling et al. 2013; Marcussen et al. 2014; Luo et al. 2017). The genomic and transcriptomic sequences of wheat are a valuable resource for comparative and phylogenetic studies and to analyse the distribution, expression and function of genes. In addition, sequencing of different wheat accessions has enabled the detection of sequence variants. These variants have been deployed to develop marker arrays, which are useful tools to study ancestral relationships, genomic patterns of diversity and marker-trait associations in mapping experiments (Akhunov et al. 2009; Wang et al. 2014; Winfield et al. 2016).

Over recent decades, genetic gain in wheat breeding has been mainly achieved using phenotypic selection in the context of breeding methods such as pedigree selection, single seed descent, doubled haploids, bulk selection and backcross breeding (Bentley and Mackay 2017). Indirect selection methods based on genetically associated morphological or biochemical traits were early discovered as alternatives to the direct phenotypic selection for the trait of interest. With the advent of the first molecular markers (restriction fragment length polymorphisms; RFLPs) in the 1980's, the traditional phenotypic selection could be extended by marker-assisted selection (MAS), which exploits the genetic linkage between a marker locus and a trait locus (Beckmann and Soller 1986). The development of RFLP markers was followed by marker technologies including simple sequence repeat (SSR) markers, diversity array technology (DArT) markers and single-nucleotide polymorphism (SNP) markers, which allowed a denser coverage of the wheat genome. A major advantage of MAS compared to phenotypic selection is its ability to select on the basis of the marker genotype of a single plant, a factor which can substantially reduce the costs of selection. Molecular markers also enable the concurrent selection for alleles at multiple quantitative trait loci (QTL). Furthermore, MAS facilitates the maintenance and fixation of alleles that are not amenable to phenotypic selection (Koebner and Summers 2003). The major prerequisite for most forms of MAS is the availability of suitable molecular markers, which can be identified using an experimental population with known phenotypes and

marker genotypes. Markers closely linked to a monogenic trait are usually identified by linkage mapping of the discrete phenotype in a biparental population. Markers for the selection of QTL can be detected by interval mapping (Lander and Botstein 1989) and its extensions composite interval mapping (Zeng 1994) and multiple interval mapping (Kao et al. 1999). Whereas these approaches were designed for QTL analysis within a single segregating population, several methods were developed to detect QTL and estimate their effects across populations (Muranty 1996; Xie et al. 1998; Xu 1998). Interval mapping methods are based on the calculation of a logarithm of the odds (LOD) statistic for the presence of a putative QTL at pre-defined positions in each marker interval on a given genetic map. A popular alternative to interval mapping is association mapping, which exploits the linkage disequilibrium between trait alleles and marker alleles without the necessity of a controlled population of related individuals and a linkage map (Thornsberry et al. 2001). Depending on the number of loci controlling a trait, different strategies of MAS should be considered. For monogenic and oligogenic traits, marker-assisted backcrossing (MABC) and forward selection are appropriate strategies to introgress and pyramide favourable alleles in elite germplasm (Bernardo 2008). In wheat breeding programs, MABC was successfully applied to transfer resistance genes for stripe rust (Randhawa et al. 2009), stem rust (Yadav et al. 2015) and leaf rust (Yadawad et al. 2017) into the genetic background of adapted varieties. A backcrossing strategy using molecular markers was also effectively deployed for the introgression of quality-affecting alleles encoding for traits such as pre-harvest sprouting tolerance (Kumar et al. 2010) and grain protein content (Vishwakarma et al. 2016). These MAS strategies are not effective in the case of many QTL with small effects because estimated effects of minor QTL are often inconsistent and the combination of favourable alleles at all QTL becomes more and more difficult as the number of QTL increases. An approach that effectively deploys MAS for many QTL is F_2 enrichment, which increases the frequencies of the favourable alleles by culling F_2 individuals which are homozygous for the non-favourable alleles and subsequent inbreeding of the selected individuals (Bernardo 2008). Bonnett et al. (2005) and Wang et al. (2007) demonstrated that F_2 enrichment is a suitable approach for wheat breeding programs to select for multiple QTL while the population size required to recover the target genotype is greatly reduced. An extension of the F_2 enrichment strategy is marker-assisted recurrent selection (MARS), which involves the prediction of a marker score for each

F_2 individual based on marker effects estimated in QTL analyses. On the basis of the predicted marker score, individuals can be selected and the process of prediction and selection is repeated over several cycles (Bernardo 2008). The application of MARS in wheat breeding programs was illustrated by Charmet et al. (1999) using the example of baking quality. A major drawback of the aforementioned selection methods is the fact that undetected minor-effect QTL are not considered. This shortcoming can be overcome by genome-wide selection (GS; Meuwissen et al. 2001). GS is a form of MAS which considers all available markers for the simultaneous estimation of marker effects in a calibration set with known phenotypes and marker genotypes. The selection of individuals from a prediction set is then based on breeding values, which are estimated based on the available marker genotypes and the determined marker effects. GS was successfully applied in wheat populations for quantitative traits such as grain quality (Heffner et al. 2011), grain yield (Poland et al. 2012), quantitative stem rust resistance (Rutkoski et al. 2014) and fusarium head blight resistance (Jiang et al. 2017b). Recent studies validating the practicality of GS for grain yield and quality parameters over breeding cycles suggested the implementation of GS in future wheat breeding programs (Battenfield et al. 2016; Song et al. 2017; Michel et al. 2019).

1.3 Hybrid wheat

1.3.1 Exploiting heterosis in wheat

A promising strategy to increase the rate of genetic gain is hybrid breeding, which is state of the art in several crop species, including maize and rice (Crow 1998; Cheng et al. 2007). In wheat, only 1% of the total area is cultivated with hybrid varieties, with France, Germany, China and India as the most important countries (Longin et al. 2012). Although hybrid wheat has played a minor role for wheat production, a potential shift from line breeding to hybrid breeding is one of the most debated issues in the wheat breeder community. The crucial motivation for hybrid breeding is heterosis, which is the superiority of a crossbred variety compared to its parental inbred lines (Shull 1952). Heterosis can be explained as a result of (1) the complementation of recessive, unfavourable alleles at multiple loci (partial or complete dominance), (2) the superiority

of heterozygosity compared to homozygosity (overdominance) and (3) the favourable interaction between alleles at different loci (epistasis). Most likely, the contribution of each of these components depends on several factors including the species and the trait of interest (Fu et al. 2015). For grain yield heterosis in wheat, dominance effects play a less prominent role compared to epistatic effects (Jiang et al. 2017a). Heterosis in wheat was first observed for heading date, plant height and leaf width (Freeman 1919). Subsequent studies also reported heterosis for important traits such as grain yield, the number of spikes per plant, the number of kernels per spike, thousand kernel weight and baking quality parameters (Briggle et al. 1964; Halloran 1975; Perenzin et al. 1998; Corbellini et al. 2002). Estimates of heterosis varied between these studies, which may be due to the limited number of tested hybrids. To precisely estimate the heterosis potential in wheat, Longin et al. (2013) examined 1,604 test hybrids together with their parental lines in multiple environments. In these experiments, the hybrids were superior to the mean of their parents for grain yield (10.7%), frost susceptibility (−7.2%), leaf rust (−8.4%) and septoria tritici blotch (−9.3%). However, the gain in plant height of the hybrids (9.2%) was accompanied with an increased lodging susceptibility (21.7%). Comparing the yields of the best hybrid and the best line variety, they observed a maximum commercial heterosis of 13.2%. The superiority of hybrids was also demonstrated by Gowda et al. (2012), who evaluated the grain yield performance of 940 hybrids in multiple environments and observed a maximum commercial heterosis of 22.4%. Although the estimated heterosis is sufficient to compensate the higher seed production costs associated with hybrid varieties, a far better heterosis was observed in allogamous species such as maize (Flint-Garcia et al. 2009). The inferior heterosis in wheat can be explained by a lower degree of dominance, the lack of heterotic groups, epistatic effects and favourable interactions between homoeologous loci, a phenomenon termed fixed heterosis (Abel et al. 2005; Longin et al. 2012). A frequently claimed advantage of hybrid crops is the higher grain yield stability of hybrids compared to line varieties. The numerous studies comparing the grain yield stability of hybrid wheat and inbred lines led to contrasting results (Borghi et al. 1988; Bruns and Peterson 1998; Oury et al. 2000; Koemel et al. 2004). This may be due to the circumstance that a high number of environments is necessary to obtain precise estimates for the grain yield stability of individual genotypes (Becker 1987; Piepho 1998). Mühleisen et al. (2014) addressed this issue by comparing groups of hybrids to groups of inbred lines

and found a superior grain yield stability for hybrids.

1.3.2 Hybridising wheat using G-type CMS

Wheats are almost completely autogamous species with an outcrossing rate of usually less than 1% (Lawrie et al. 2006). Therefore, a hybridising system that efficiently forces outcrossing between parental inbred lines is a prerequisite for a long-term success of hybrid wheat. Several fertility-controlling systems have been explored in wheat, of which cytoplasmic male sterility (CMS) and chemical hybridising agents (CHAs) have been the most widely discussed approaches (Whitford et al. 2013). Although CHAs are used for the major part of global commercial hybrid wheat seed production, they are associated with several disadvantages (Longin et al. 2012). A successful use of CHAs strongly depends on the treated female inbred line, the composition and concentration of the CHA and the growth stage and the environmental condition at the time of application. Deviations from optimal conditions can either lead to sterility of both male and female reproductive organs or to incomplete male sterility, which results in an undesired mixture of hybrid and selfed seeds. In addition, the availability of CHAs depends on the regulation of the responsible authority. As a consequence of these drawbacks, genetic-based hybridising systems such as CMS are promising alternatives.

CMS in plants is based on the incompatibility between the nuclear and the mitochondrial genome, which leads to a disruption of pollen production. Genes causing male sterility have been associated with chimeric open reading frames in the mitochondrial genome. The sterility-inducing effect of these genes can be neutralised by *restorer-of-fertility* (*Rf*) genes in the nuclear genome (Caruso et al. 2012). As the mitochondrial genome is maternally inherited, a CMS line can be obtained by recurrent backcrossing of a cytoplasm donor with any genotype not carrying fertility-restoring alleles. The propagation of CMS lines is enabled by male-fertile maintainer lines, which carry an identical nuclear genome combined with the original cytoplasm. CMS hybrids can be produced by crossing the CMS line with a restorer line, which is homozygous for at least one nuclear fertility-restoring gene (Whitford et al. 2013). In wheat, the first CMS line was developed by Kihara (1951), who combined the nucleus of common wheat with the cytoplasm of *Aegilops caudata*. Few years later, the cytoplasm of *Aegilops ovata*

(Fukasawa 1953) and *Triticum timopheevii* (Wilson and Ross 1962) were also found to cause male sterility (Yen et al. 1969). Although the cytoplasms of as many as 35 species were evaluated for hybrid wheat production, the CMS system based on the *T. timopheevii* cytoplasm, which is also known as G-type CMS, gained widespread use because of the adverse effects of other cytoplasms, or because no advantage existed over G-type CMS (Hayward 1975; Virmani and Edwards 1983; Ahmed et al. 2001; Adugna et al. 2004; Koekemoer et al. 2011). CMS lines with this cytoplasm develop shrivelled, sickle-shaped anthers, which do not dehisce to shed pollen grains. The tapetal cells of CMS lines form less starch and persist longer compared to tapetal cells of euplasmic wheat (Joppa et al. 1966). The development of pollen grains in CMS lines is probably normal until tetrad stage and when the exine is formed. However, the mature pollen grains of these plants are fewer in number, are smaller, have a thinner intine, contain little or no starch and do not always form three nuclei (Joppa et al. 1966; de Vries and Ie 1970; Odenbach 1970). The aberrations in the anther and pollen grain development may be explained by the poorly developed vascular bundles in the anthers of CMS lines (Joppa et al. 1966), but the exact physiological process underlying CMS is still unclear. A possible candidate gene that may cause the sterility-inducing effect of the G-type cytoplasm is the open reading frame *orf256*. The *orf256* sequence is a chimaera of the first 33 nucleotides of the *coxI* gene and an unknown sequence (Rathburn and Hedgcoth 1991). Whereas *orf256* was detected in the mitochondrial DNA (mtDNA) of several species including *T. timopheevii*, it is absent in the mtDNA of common wheat (Rathburn and Hedgcoth 1991; Hedgcoth et al. 2002). The *orf256* sequence is expressed as a protein only in CMS lines but not in *T. timopheevii* and lines carrying the G-type cytoplasm and nuclear fertility-restorer genes (Song and Hedgcoth 1994). Further candidates for the sterility-inducing effect of the G-type cytoplasm are the mitochondrial genes *atp6* and *orf25*, which were shown to be polymorphic between the cytoplasms of common wheat and *T. timopheevii* (Mohr et al. 1993). A possible target of such novel or rearranged mitochondrial expression products could be TaF_Ad, which is located in the nuclear genome and encodes for the F_Ad subunit of the ATP synthase. It was shown that the expression of TaF_Ad is under mitochondrial retrograde regulation in the anthers of a G-type CMS line (Xu et al. 2008). Despite the associations between the expression of these candidate genes and the CMS phenomenon, the exact molecular mechanism behind the sterility-inducing effect of the *T. timopheevii*

cytoplasm has yet to be discovered.

1.3.3 Restoration of male fertility in G-type CMS wheat

Restoration of male fertility in CMS wheat was simultaneously discovered by Schmidt et al. (1962) and Wilson (1962) by crossing CMS lines and *T. timopheevii* introgression lines carrying nuclear fertility-restoring genes. Subsequently, restoration capacity was also observed in common wheat (Oehler and Ingold 1966; Zeven 1967; Sinha et al. 2013), European spelt (Kihara and Tsunewaki 1967), rye (Curtis and Lukaszewski 1993) and *Aegilops umbellulata* (Ma et al. 1995). Fertility-restoring loci detected in the nuclear genome of these species were numbered from *Rf1* (Livers 1964) to *Rf8* (Sinha et al. 2013), with the exception of *Rfc3* and *Rfc4* (Curtis and Lukaszewski 1993). As these designations were often used to refer to different loci, the nomenclature is not definite without further information in the cases of *Rf2*, *Rf3*, *Rf4* and *Rf6* (Tahir and Tsunewaki 1969; Yen et al. 1969; Maan 1985; Ma et al. 1995; Kojima et al. 1997). Designations used for detected restorer loci are listed in Table 1. Whereas the molecular mechanism behind fertility restoration in wheat with G-type cytoplasm is still unclear, studies in several species revealed similarities which may be transferable to the CMS system in wheat. In most of the studied CMS systems, the fertility-restoring gene in the nuclear genome encodes for a pentatricopeptide repeat (PPR) protein. PPR proteins are a large family of proteins found in all eukaryotic lineages, especially in plants. They are known to play important roles in transcription, RNA processing, stability, editing and translation, primarily in organelle genomes but also in the nuclear genome. PPR proteins contain tandemly repeated PPR motifs, of which each motif comprises 35 amino acids and forms a helix-turn-helix structure. The repeated PPR motif produces a superhelix with a central groove that is able to bind RNA (Manna 2015). Identified fertility-restoring PPR (Rf-PPR) proteins and the related Rf-like PPR (RFL-PPR) proteins form a clade within the P subclass of PPR proteins (Fujii et al. 2011). Several studies showed that Rf-PPR proteins are targeted to mitochondria and act by suppressing the expression of the sterility-inducing mitochondrial transcripts (Dahan and Mireau 2013). The first identified *Rf* gene was *Rf-PPR592* from petunia (*Petunia hybrida*). The product of this gene is a mitochondrially targeted protein comprising 14 repeats of the PPR motif. It was shown that *Rf-PPR592* can restore fertility of CMS

petunia plants by decreasing the amount of the PCF protein, which is a chimeric protein encoded in the mitochondrial genome (Nivison and Hanson 1989; Bentolila et al. 2002). Rf-PPR592 is associated with the inner membrane of mitochondria in a large protein complex that binds mRNA of the *pcf* gene (Gillman et al. 2007). PPR proteins inhibiting the expression of CMS-inducing proteins by post-transcriptional regulation were also found to cause fertility restoration in rice (Komori et al. 2004; Kazama et al. 2008; Tang et al. 2014; Igarashi et al. 2016), rapeseed (*Brassica napus*; Menassa et al. 1999; Liu et al. 2016b) and radish (*Raphanus sativus*; Koizuka et al. 2003; Desloire et al. 2003; Uyttewaal et al. 2008). Besides PPR proteins, fertility-restoring genes were also found to encode for other protein groups which mostly act at a metabolic level rather than inhibiting the expression of CMS-conferring transcripts (Gaborieau et al. 2016). These protein groups comprise an aldehyde dehydrogenase in maize (Liu et al. 2001), an acyl carrier protein and glycine-rich proteins in rice (Fujii and Toriyama 2009; Itabashi et al. 2011; Hu et al. 2012) and a peptidase in sugar beet (Kitazaki et al. 2015). Sequences of the identified *Rf* genes, especially the *Rf-PPR* sequences, are a valuable resource to identify and validate candidate genes for the fertility-restoring loci in wheat.

Although initial experiments revealed restoration capacity in several species, none of the identified restorer accessions led to a complete and stable fertility restoration in hybrids with the G-type cytoplasm (Keydel 1973; Hayward 1975). Incomplete fertility restoration eventually manifests in a reduced seed set. Whereas the degrees of fertility can vary from a single seed per plant to an almost complete seed set, it usually decreases from the base to the apex of a spike (Sage 1972). The reduced seed set counteracts the effect of heterosis on grain yield, which is a major motivation for hybrid wheat breeding. Additionally, unpollinated florets of male-sterile wheat lines have shown high susceptibility to infections with ergot (*Claviceps purpurea*; Mantle and Swan 1995). Thus, incomplete fertility restoration might also impair the quality of hybrid wheat products. It was demonstrated that environmental factors strongly affect fertility restoration, but optimal conditions have yet to be determined (Johnson et al. 1967; Jošt 1982). Furthermore, fertility restoration depends on epistatic interactions between restorer and modifier genes, implicating that the genetic components of both restorer and CMS lines influence the extent of fertility restoration (Jošt 1982; Maan

Table 1 Designations of previously detected loci for the restoration of male fertility in G-type cytoplasm. Detected loci without designation are not included

Locus	Chromosome	Germplasm resource	Reference
Rf1	1A	R3 (*T. timopheevii* introgression line)	Robertson and Curtis (1967)
Rf1	1A	R4 (*T. timopheevii* introgression line)	Yen et al. (1969)
Rf1	1A	R1 (*T. timopheevii* introgression line)	Odenbach (1970)
Rf1	1A	R113 (common wheat)	Maan (1985)
Rf1	1A	L19, L22 (common wheat)	Du et al. (1991)
Rf2	7D	R3 (*T. timopheevii* introgression line)	Tahir and Tsunewaki (1969)
Rf2	6B	*T. timopheevii* introgression line	Yen et al. (1969)
Rf3	1B	European spelt (var. *duhamelianum*)	Tahir and Tsunewaki (1969)
Rf3	6D	*T. timopheevii* introgression line	Yen et al. (1969)
Rf3	1B	Primepi (common wheat)	Ma and Sorrells (1995)
Rf4	7D	R4 (*T. timopheevii* introgression line)	Yen et al. (1969)
Rf4	6B	R113 (common wheat)	Maan (1985)
Rf4	6B	L11, L13 (common wheat)	Du et al. (1991)
Rf5	6D	Not documented	Kojima et al. (1997)
Rf6	6U	*Ae. umbellulata* translocation line	Ma et al. (1995)
Rf6	5D	Not documented	Kojima et al. (1997)
Rf7	7B	Not documented	Kojima et al. (1997)
Rf8	2D	PWR4099 (common wheat)	Sinha et al. (2013)
Rfc3	6R	Wheat-rye addition line	Curtis and Lukaszewski (1993)
Rfc4	4R	Wheat-rye addition line	Curtis and Lukaszewski (1993)

et al. 1984). Since it was demonstrated that hybrids of crosses between different restorer lines in sterile cytoplasm showed a superior seed set compared to hybrids derived from crosses between restorer and CMS lines, stacking of major restorer and modifier genes in elite restorer lines has been a potential solution to overcome partial sterility in CMS hybrids (Johnson and Patterson 1977).

Two promising genetic components for this approach are the fertility-restoring loci *Rf1* and *Rf3*. *Rf1* was first detected on chromosome 1A of the *T. timopheevii* introgression line R3 (Livers 1964; Robertson and Curtis 1967; Tahir and Tsunewaki 1969). A restorer locus on this chromosome was also located in other lines with introgressions from *T. timopheevii* and *Triticum zhukovskyi* but also in the spring wheat accession R113

and its descendants, for which no introgressions from wild relatives are documented. It has been assumed that the restoration capacity of chromosome 1A is caused by the same locus, namely *Rf1*, in these accessions (Yen et al. 1969; Odenbach 1970; Bahl and Maan 1973; Maan et al. 1984; Maan 1985; Du et al. 1991). The restorer locus *Rf3* was discovered on chromosome 1B of European spelt (var. *duhamelianum*; Tahir and Tsunewaki 1969). A restorer locus on the same chromosome was detected in the common wheat cultivars Primepi, Prof. Marchal and Chinese Spring, and it has been speculated that the restoration capacity of these cultivars can also be explained by *Rf3* (Bahl and Maan 1973; Kučera 1982; Ma and Sorrells 1995). A prerequisite to exploit *Rf1* and *Rf3* in practical hybrid breeding programs is a method that enables a time- and cost-effective selection for these loci. In the process of breeding restorer lines, phenotypic selection for restorer genes is time consuming because each selection step requires testcrosses with CMS lines. Furthermore, phenotypic selection is not effective because observed phenotypic values do not only depend on the presence of restorer genes but also on effects of epistasis and genotype-by-environment interactions (Jošt 1982; Maan et al. 1984). Therefore, molecular markers closely linked to restorer loci could be a valuable tool in hybrid wheat breeding programs. Although several studies have been conducted to locate *Rf1* and *Rf3*, molecular markers suitable for the selection of these loci have been unavailable. *Rf* genes identified in several other crop species provide a valuable resource for a better understanding of cytonuclear interactions and efficiently exploit CMS in plant breeding. Despite this previous knowledge, no studies have been undertaken to investigate candidates for *Rf* genes in wheat. Another essential prerequisite for the success of CMS hybrid wheat is the knowledge of the distribution of fertility-restoring genes in wheat species, especially in common wheat. Findings about the distribution of *Rf* genes could facilitate the development and maintenance of heterotic groups. As the knowledge of the origin of *Rf1* is based solely on pedigree data, and there is one source for *Rf1* which cannot be traced back to *T. timopheevii*, it has been unclear if the restoring *Rf1* allele is exclusively derived from *T. timopheevii*, or if it may also be present in common wheat chromatin. Although *Rf3* was reported in European spelt and the mentioned common wheat cultivars, it has yet to be evaluated how *Rf3* alleles are distributed in these taxa.

1.4 Objectives of the present study

The general objective of the present study was the genetic characterisation of the restorer loci *Rf1* and *Rf3*. Both loci were genetically mapped in several biparental populations and identified molecular markers were used for the analysis of diverse wheat accessions. The specific objectives were to

(1) investigate the identity of *Rf1* and *Rf3* across diverse restorer accessions,

(2) identify molecular markers suitable for marker-assisted selection for these loci,

(3) estimate the effects of *Rf1* and *Rf3* across diverse genetic backgrounds,

(4) identify candidates of the *RFL-PPR* gene family by linkage mapping,

(5) reveal whether the restoring *Rf1* allele is exclusively derived from *T. timopheevii* and

(6) evaluate the distribution of the restoring *Rf3* allele in common and spelt wheat considering population structure.

The findings of this study should contribute to a refined understanding of *Rf1* and *Rf3*, thereby facilitating an efficient exploitation of these loci in CMS hybrid wheat breeding programs.

2 Materials and methods

2.1 Plant material

Three backcross populations (BC_1) were developed to genetically map the *restorer-of-fertility* (*Rf*) locus *Rf1*. For this, each of the *Rf1* donor accessions R3, R113 and L19 was used to pollinate the cytoplasmic male-sterile (CMS) common wheat line CMS-Sperber (*Triticum aestivum* ssp. *aestivum*). Male sterility of each CMS-Sperber plant was validated by analysing one control spike. BC_1 descendants were obtained by backcrossing the resulting F_1 plants with the maintainer line Sperber. To determine the success of the crosses between CMS-Sperber and the *Rf1* donor accessions, one spike per F_1 plant was evaluated for fertility restoration. For successful crosses, restored fertility on these control spikes was expected. The resulting mapping populations derived from Sperber and the *Rf1* donors R3, R113 and L19 comprised 197, 201 and 230 individuals, respectively. R3 descends from the recurrent backcross *Triticum timopheevii*/3*Marquis and was supposed to carry the restorer loci *Rf1* on chromosome 1A and *Rf2* on chromosome 7D (Livers 1964; Tahir and Tsunewaki 1969; Bahl and Maan 1973). The restorer line R113 originates from the common wheat accessions Era (male-sterile), MNII-54-30 (male-sterile), WI271 and an unknown F_2 plant. It was expected to harbour *Rf1* and the restorer locus *Rf4* on chromosome 6B, but no introgressions from wild species are documented (Maan et al. 1984). L19 is derived from the backcross Chris (mono-6B)/R113//Chris and was supposed to carry only *Rf1* (Maan 1985; Du et al. 1991).

Seven populations (BC_1 and F_2) were developed to genetically map the restorer locus *Rf3*. The restorer accessions used for these populations comprised the common wheat lines Primepi, PR143 and Palmaress and the European spelt (*Triticum aestivum* ssp. *spelta*) cultivars Badenkrone, Badenstern, Holstenkorn and Schwabenspelz. Whereas the restorer accession Badenstern was combined with the male-sterile line CMS-609-73, the other restorer lines were crossed with CMS-Sperber. The F_1 plants derived from Primepi, PR143, Badenkrone, Badenstern, Schwabenspelz and Palmaress were backcrossed with the respective maintainer line to obtain BC_1 populations. A F_2 population was developed for the cross involving Holstenkorn by isolating the spikes of F_1

Table 2 Description of the populations for genetic mapping of the restorer loci *Rf1* and *Rf3*

Population	Type	*N*	Environment	Segregating *Rf* loci and chromosomes (expected)
PopR3	BC_1	197	Greenhouse 2016	*Rf1*, *Rf2* (Livers 1964)
PopR113	BC_1	201	Greenhouse 2016	*Rf1*, *Rf4* (Maan et al. 1984)
PopL19	BC_1	230	Greenhouse 2016	*Rf1* (Du et al. 1991)
PopPrimepi	BC_1	193	Greenhouse 2014	*Rf3* (Bahl and Maan 1973)
PopPR143	BC_1	221	Greenhouse 2015	*Rf1*, *Rf2*, *Rf3* (Patterson et al. 1996)
PopPalmaress	BC_1	91	Greenhouse 2017	4A, 5B, 7D (Kučera 1982)
PopBadenkrone	BC_1	290	Greenhouse 2014	*Rf3* (Tahir and Tsunewaki 1969)
PopBadenstern	BC_1	220	Greenhouse 2014	*Rf3* (Tahir and Tsunewaki 1969)
PopHolstenkorn	F_2	101	Field 2014	*Rf3* (Tahir and Tsunewaki 1969)
PopSchwabenspelz	BC_1	288	Greenhouse 2014	*Rf3* (Tahir and Tsunewaki 1969)

plants using glassine bags. Male sterility of the CMS lines and the success of the crosses between CMS lines and the *Rf3* donor accessions were validated as described in the previous paragraph. The mapping populations derived from Primepi, PR143, Badenkrone, Badenstern, Holstenkorn, Schwabenspelz and Palmaress comprised 193, 221, 290, 220, 101, 288 and 91 plants, respectively. Primepi is a French winter wheat cultivar, which was supposed to carry *Rf3* and possibly a second restorer gene on chromosome 5D (Bahl and Maan 1973). PR143 has the parentage *T. timopheevii*/3*Marquis//CMS-Monon/Primepi and was expected to harbour a combination of *Rf1*, *Rf2*, *Rf3* and maybe a restorer gene on chromosome 5D (Patterson et al. 1996). For European spelt, no other restorer genes than *Rf3* have been reported yet (Tahir and Tsunewaki 1969; Kojima et al. 1997). The French winter wheat cultivar Palmaress was supposed to carry restorer genes on chromosomes 4A, 5B and 7D (Panayotov et al. 1975; Kučera 1982). A mapping population derived from this cultivar was analysed to validate the results from Panayotov et al. (1975). Experimental populations for linkage mapping of *Rf1* and *Rf3* are summarised in Table 2. From hereafter, all mapping populations are designated by "Pop" followed by the name of the *Rf* donor line. The genetic relationship between PopR3, PopR113, PopL19 and PopPrimepi was analysed by principal coordinate analysis (PCoA) of whole-genome marker data as described in section *2.6 Distribution of marker haplotypes in the genomic target regions across wheat species.*

The diagnostic ability of markers linked to the *Rf3* locus was evaluated in a diversity panel comprising common wheat and European spelt accessions. Common wheat accessions served as positive or negative controls, depending on literature references and mapping results of the present study. For a sample of these common wheat accessions, the previous knowledge of *Rf3* was validated by testcrosses with male-sterile line CMS-Sperber and subsequent evaluation of fertility restoration on spikes of F_1 descendants. To augment the diversity panel, the restoration capacity of 30 European spelt cultivars was evaluated. Spelt accessions were used as positive controls for *Rf3* if they belonged to the variety *duhamelianum* and if they were able to restore fertility in testcrosses with CMS-Sperber. For this, it was assumed that *Rf3* is the only restorer locus explaining restoration capacity in the *duhamelianum* variety (Tahir and Tsunewaki 1969; Kojima et al. 1997). Spelt accessions that may not belong to this taxon but that were shown to carry *Rf3* by linkage mapping were also employed as positive controls. Spelt accessions served as negative controls if they did not restore fertility in testcrosses with CMS-Sperber. Fertility-restoring spelt cultivars which may not belong to the *duhamelianum* variety and which were not shown to carry *Rf3* by linkage mapping were not included in the diversity panel. The diversity panel comprised 29 common wheat accessions and 25 European spelt cultivars.

The distribution of *Rf1* was analysed in all hitherto known *Rf1* donor accessions (R1, R2, R3, R4, R5, R113 and L19), 17 unique *T. timopheevii* accessions and a diverse set of 524 common wheat breeding lines. The same set of breeding lines and the same *T. timopheevii* accessions were analysed to evaluate the distribution of *Rf3*. In addition, the distribution of *Rf3* was analysed in Primepi and 30 European spelt cultivars, which were also analysed for the diversity panel. The relationship between *Rf3* genotypes of spelt cultivars and their genetic similarity to common wheat was studied using a set of 368 German common wheat cultivars as a reference.

The restorer accessions R1, R2, R3, R4, R5, R113 and L19 were kindly provided by Finn Borum (Sejet Plant Breeding, Denmark). *T. timopheevii* accessions with a TRI designation and the winter wheat cultivars Minister, Palmaress, Primepi and Prof. Marchal were obtained from the gene bank of the IPK Gatersleben, Germany. The winter wheat accessions PR143 and PR189 were received from the National Small Grain Collection, USA. The maintainer lines Navojoa and Vorobey were released by CIMMYT, Mexico.

The 524 common wheat breeding lines which were used to evaluate the distribution of *Rf1* and *Rf3* were provided by the German breeding companies Saatzucht Bauer GmbH & Co. KG, Saatzucht Josef Breun GmbH & Co. KG, Limagrain GmbH, SECOBRA Saatzucht GmbH, Saatzucht Streng-Engelen GmbH & Co. KG and Lantmännen SW Seed GmbH. All other germplasm used in the present study was in stock at the Bavarian State Research Center for Agriculture. The male-sterile lines CMS-Sperber and CMS-609-73, which were used for the development of mapping populations, were selected from eight available CMS lines on the basis of genetic homogeneity. To compare the homogeneity of the CMS lines, for each of these lines and the corresponding maintainer line, a sample was analysed using eight simple sequence repeat (SSR) markers and the chloroplast-specific marker *TaCM15*, which was used to differentiate euplasmic plants from plants with G-type cytoplasm (Tomar et al. 2014). As Finn Borum (Sejet Plant Breeding, Denmark) reported that the *Rf1* donor accessions R1, R3 and R5 may be heterogenous, for all *Rf1* donor accessions used in this study, a sample of plants was phenotyped for plant height, time of ear emergence and awn length. The sample of the L19 accession was additionally genotyped using seven SSR markers. For each of these restorer accessions, a single plant was selected that exhibited the most common characteristics within the sample. Only the selected plants were used for the genetic analysis in the present study. The genetic uniqueness of the 17 *T. timopheevii* accessions was validated using single-nucleotide polymorphism (SNP) markers.

2.2 Assessment of fertility restoration and segregation analysis

The seedlings of the BC_1 mapping populations and test hybrids were vernalised at 5 °C for two months. The F_2 population PopHolstenkorn was planted and cultivated in the field. All other mapping populations and the test hybrids were grown in 13-cm pots (two plants per pot) under metal-halide lamps in the greenhouse at a target temperature of 15–18 °C. Whereas PopPrimepi was split into three subpopulations, with each being cultivated in a separate greenhouse chamber at a different season of the year, all other populations were grown in single environments. As a measure for the restoration of male fertility, the seed set of isolated spikes was determined. Before the beginning of anthesis, one to four spikes of each plant were isolated using glassine

bags. After ripening, the bagged spikes were harvested and the number of kernels and spikelets was recorded. The seed set of a plant was defined as the number of kernels divided by the number of spikelets, averaged over its isolated spikes. Plants were considered male-sterile if they contained no seeds and male-fertile if at least one seed was observed. The mean seed set of the three subsets of PopPrimepi was analysed using the one-way analysis of variance and Tukey's HSD test. Observed ratios of fertile and sterile plants of the mapping populations were compared to expected segregation patterns using the χ^2 goodness-of-fit test. The modes of the seed set distributions of the fertile descendants were analysed using Hartigan's dip test. The significance level was set at $\alpha = 0.05$ for all tests. Statistical analysis of the seed set was performed using R version 3.3.2 (R Core Team 2016).

2.3 Marker development, genotyping and marker selection

Total genomic DNA was extracted from freeze-dried, young leaf tissue according to the protocol reported by Plaschke et al. (1995). The mapping populations involving the *Rf1* donor accessions R3, R113 and L19 were genotyped using a 15k + 5k Infinium® iSelect® array containing 17,267 biallelic SNP markers combined from the 90k iSelect array and the 820k Axiom® array developed by Wang et al. (2014) and Winfield et al. (2016), respectively. Raw marker genotypes from this SNP array were obtained from TraitGenetics GmbH, Germany. For each of the three populations, SNP markers were discarded if the parental SNP genotypes were unavailable, monomorphic or heterozygous. Furthermore, SNP markers were discarded if the minor allele frequency (MAF) was smaller than 10% or the number of missing genotypes exceeded 10%. For groups of SNP markers that exhibited identical genotypes across the descendants of a population, the marker with the fewest missing values was kept, whereas the other markers of the group were not used for the construction of linkage maps. To enable an optimal comparison between the linkage maps of the populations derived from R3, R113 and L19, the duplicate markers in each of these populations were only discarded if they were not kept for linkage mapping in one of the other two populations at this selection step.

Rf3 was initially mapped in PopPrimepi. All individuals of the population were genotyped with SSR markers that are located on the short arm of chromosome 1B (Somers et al. 2004) and that were polymorphic between Sperber and Primepi. Polymerase chain reaction (PCR) mix components and the PCR temperature profile for SSR markers are listed in Table 3 and 4, respectively. Primer sequences and annealing temperatures (T_a) were obtained from GrainGenes (2016). PCR products were separated on polyacrylamide gels (Table 5) in vertical electrophoresis chambers at 50 V. Individuals that exhibited a recombination for SSR markers flanking *Rf3* were genotyped using a 15k Infinium iSelect SNP array, which contained 13,006 biallelic SNP markers that were also included on the aforementioned 15k + 5k SNP array. SNP genotypes of this array were provided by TraitGenetics GmbH, Germany. SNP marker pruning in this population was based on the parental marker genotypes, the MAF and the number of missing marker genotypes as described in the previous paragraph. For SNPs linked to *Rf3* in this population, cleaved amplified polymorphic sequence (CAPS) markers were developed. To increase the size of CAPS fragments, the SNP-flanking DNA sequences retrieved from The Triticeae Toolbox (2016) were searched against the sequence databases CerealsDB (2016), GrainGenes (2016), KOMUGI (2016) and URGI (2016) using the Basic Local Alignment Search Tool (BLAST). Sequences obtained from the BLAST searches were aligned using Clustal Omega (2016). Consensus sequences were the basis for the design of PCR primers using the Primer3 program (Untergasser et al. 2012). The components for the PCR mix and the PCR temperature profile for CAPS markers are listed in Table 3 and 4, respectively. Optimal annealing temperatures for CAPS markers were determined by gradient PCRs from 55 to 64 °C. Restriction enzymes were chosen using the analysis tool SNP2CAPS (IPK Gatersleben, Germany). For the digestion of amplified DNA, 5 µl of the PCR product were incubated together with 2 U restriction enzyme in a reaction volume of 10 µl using enzyme-specific buffer solutions. Digested PCR products were resolved on polyacrylamide gels as described for SSR markers. Individuals of PopPrimepi that exhibited a recombination in the genomic target area were used to validate CAPS markers by comparing the CAPS marker genotypes to the SNP array-derived genotypes. CAPS marker genotypes were compared to genotypes of flanking SSR markers for non-recombinant descendants of this population. For the validation of a closely linked CAPS marker, subsets of the populations derived from PR143, Palmaress, Badenkrone, Badenstern, Holstenkorn and

Schwabenspelz were genotyped. These subsets comprised 170, 91, 284, 87, 85 and 87 individuals, respectively.

Table 3 Components for polymerase chain reaction (PCR) mixes for simple sequence repeat (SSR) and cleaved amplified polymorphic sequence (CAPS) markers

Component	Final concentration	
	SSR	CAPS
DNA	5.3 ng/µl	4.0 ng/µl
Tris-HCl	75.00 mM	75.00 mM
$(NH_4)_2SO_4$	20.00 mM	20.00 mM
Tween 20	0.10‰ (v/v)	0.10‰ (v/v)
dNTPs	0.80 mM	1.00 mM
$MgCl_2$	2.00 mM	1.50 mM
Primer mix (labeled)	0.6 µM	0.5 µM
Taq polymerase	25.33 U/ml	25.00 U/ml

Table 4 PCR temperature profile applied for SSR and CAPS markers. The annealing temperature is denoted by T_a

Step	Temperature	Duration
Initial denaturation	94 °C	3 min
35 cycles	94 °C	10 s
	T_a	30 s
	72 °C	50 s
Final extension	72 °C	15 min

To analyse the distribution of *Rf1* and *Rf3*, the set of 524 breeding lines was genotyped using the aforementioned 15k Infinium iSelect array. The *Rf1* donor accessions and the *T. timopheevii* accessions were genotyped with the 15k + 5k Infinium iSelect array. SNP marker genotypes were provided by TraitGenetics GmbH. Markers with more than 10% missing values or a MAF smaller than 1% were discarded for this analysis. The relationship between the *Rf3* genotype of the 30 spelt cultivars and their genetic similarity to common wheat was evaluated by genotyping the spelt cultivars together

Table 5 Components used for the preparation of polyacrylamide gels

Component	Final concentration
Urea	7.47 M
Tris	0.10 M
Boric acid	0.10 M
EDTA	1.00 mM
Acrylamide	0.67 M
Bisacrylamide	16.16 mM
APS	1.14 mM
TEMED	69.50 mM

with 368 German common wheat cultivars using 33 SSR markers evenly distributed across the wheat genome.

The development of marker assays for candidate genes of the *pentatricopeptide repeat* (*PPR*) gene family was based on target enrichment sequencing, which was conducted by the IPK Gatersleben, Germany (Zhou et al. 2017). The target enrichment sequencing approach was previously described for sequencing of resistance genes of the *nucleotide-binding and leucine-rich repeat* (*NB-LRR*) gene family and was adopted by the IPK Gatersleben for targeted resequencing of *PPR* genes (Jupe et al. 2013; Andolfo et al. 2014; Zhou et al. 2017). RNA baits for target enrichment were designed based on *PPR* gene sequences that were predicted and annotated in genomic and transcriptomic sequences of *T. aestivum*, *Triticum urartu*, *Triticum turgidum*, *Aegilops tauschii* and *Aegilops sharonensis* (Jia et al. 2013; Krasileva et al. 2013; Ling et al. 2013; IWGSC 2014; Chapman et al. 2015; Zhou et al. 2017). The RNA baits were used to capture *PPR*-specific DNA fragments of the maintainer line Sperber and the restorer accessions R3, R113, L19 and Primepi by hybridisation. Reads obtained from capture sequencing were assembled and aligned to the reference sequence of Chinese Spring. Zhou et al. (2017) identified 79 *Rf-like* (*RFL*)-*PPR* genes, which shared high homology with previously identified *Rf* genes of the *PPR* gene family (*Rf-PPR* genes; Fujii et al. 2011). The *RFL-PPR* genes were distributed across chromosomes 1A, 1B, 1D, 2A, 2B, 2D, 5B, 6A, 6B, 6D, 7B and 7D, but most of them were located on the chromosomes of homoeologous group 1. The distribution of *RFL-PPR* genes across the wheat genome

is illustrated in Figure 1. On the short arms of the chromosomes 1A and 1B, 10 and 21 candidate genes for *Rf1* and *Rf3* were detected, respectively. Based on the aligned reads from capture sequencing, Zhou et al. (2017) extracted *RFL-PPR* gene regions to find SNPs between the gene sequences of Sperber and the restorer accessions R3, R113, L19 and Primepi. Between Sperber and the *Rf1* donor accessions, 193 unique SNPs were detected within *RFL-PPR* genes on the short arm of chromosome 1A. These SNPs were distributed across nine of ten candidate genes on this chromosome (*RFL01–RFL10*). *RFL05* did not exhibit any SNP. Between Sperber and Primepi, 87 SNPs were detected within 13 of 21 candidate genes on chromosome 1BS (*RFL11–RFL31*). No SNPs were found in *RFL19*, *RFL24* and *RFL26–RFL31*.

The 280 SNPs detected by Zhou et al. (2017) were provided to the Bavarian State Research Center for Agriculture for marker development and linkage mapping. For each of the two restorer loci, 48 SNPs were selected for the development of competitive allele-specific PCR assays. The selection of SNPs for marker development was based on sequencing depth and the distribution across the candidate genes. Furthermore, SNPs were not used for marker development if non-target SNPs were located in a 40 bp region flanking the target SNP. To improve the design of PCR primers, non-target SNPs identified by multiple sequence alignment (Clustal Omega 2016) were integrated in the SNP-flanking sequences (200 bp). Sequences containing many non-target SNPs were extended up to 400 bp. The consensus sequences of these alignments were the basis for primer design, which was conducted by Fluidigm® GmbH, Germany. One set of primers comprised two allele-specific forward primers and a common reverse primer. The developed 96 SNP markers were used to genotype the populations derived from the restorer lines R3, R113, L19, Primepi and Badenkrone using the Fluidigm EP1 genotyping platform. SNP genotypes were called using the Fluidigm SNP Genotyping Analysis software. Genotypes of all markers within a candidate gene were merged to construct a consensus genotype for the candidate gene.

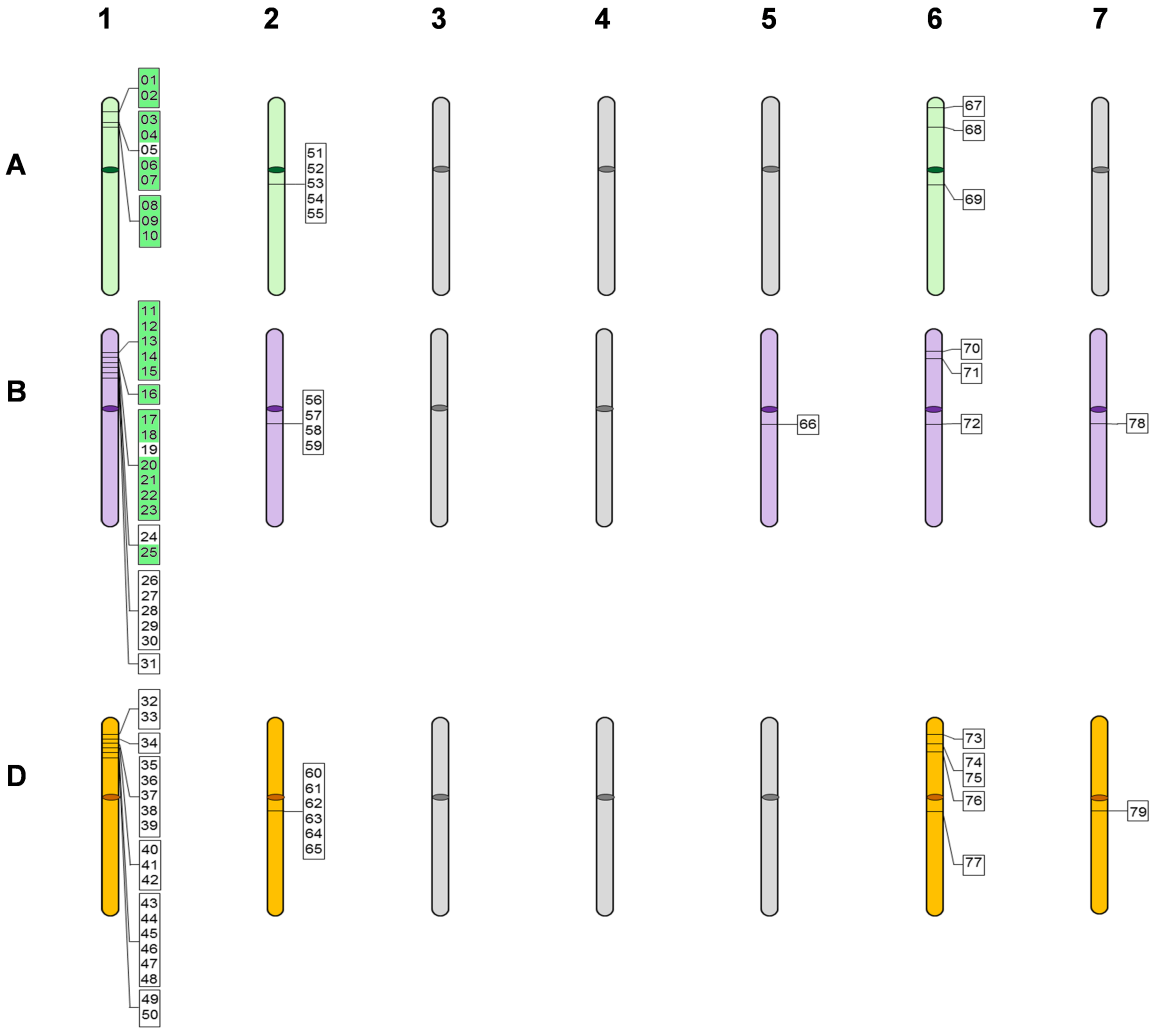

Figure 1 Distribution of 79 *RFL-PPR* genes across the wheat genome (Zhou et al. 2017). Genes are highlighted in green if at least on single-nucleotide polymorphism (SNP) was detected between the parental lines

2.4 Linkage map construction

For the mapping populations derived from *Rf1* donor accessions, linkage maps were constructed based on whole-genome SNP marker genotypes. These linkage maps enabled interval mapping of *Rf1* and other segregating loci that affected fertility restoration. In contrast, in the populations involving the *Rf3* donor lines, *Rf3* was located together with molecular markers by linkage mapping using the binary phenotype. Linkage maps were constructed using JoinMap® version 4.0 (van Ooijen 2006). Linkage groups were established using independence logarithm of the odds (LOD) values larger than 3.0. For the populations derived from *Rf1* donor lines, linkage groups were discarded if they contained less than three markers or spanned less than 10 centimorgans (cM). Linkage mapping was performed using Kosambi's mapping function (Kosambi 1944) and a regression-based procedure developed by Stam (1993). Maps were constructed by adding loci one by one, starting with the two most informative loci. The optimal position of an added locus was determined by comparing a goodness-of-fit statistic for all possible positions. This statistic compared direct recombination frequencies to map-derived recombination frequencies. A marker was removed again if the goodness-of-fit statistic changed by more than 5.0 or if the locus gave rise to negative map distances. After a locus was added to the map, the goodness-of-fit for all possible permutations within a moving window of three adjacent loci was determined, and the best-fitting order was kept. Only pairwise data with recombination frequencies below 0.4 and a LOD value larger than 1.0 were considered for map construction. If loci were removed in this procedure, an attempt was made to integrate them in a second round. All previously removed loci were added to the map in a third round of regression mapping if one or more markers were removed in the second round and if a linkage group harboured loci affecting fertility restoration. For these linkage groups, a marker was manually removed if its mean contribution to the goodness-of-fit statistic exceeded 5.0. A marker of these linkage groups was also discarded if the sum of the absolute differences between the pairwise distances and the map-based distances to the nearest informative neighbour loci was larger than 10 cM. Markers of these linkage groups were manually removed until the results of regression mapping met the abovementioned criteria. The marker order of all linkage groups was validated using linkage maps developed by Wang et al. (2014), Maccaferri et al. (2015a), Maccaferri

et al. (2015b) and Allen et al. (2017). Segregation distortion was analysed using the χ^2 goodness-of-fit test. P values of the χ^2 test were adjusted for multiple testing using a false-discovery rate (FDR) of 0.05 for each population. Consensus maps for quantitative trait locus (QTL) analysis across populations were constructed using the map integration function of JoinMap by following the same procedure described for maps derived from single biparental crosses. Graphical presentation of linkage maps was performed using MapChart version 2.1 (Voorrips 2002).

2.5 QTL analysis

In the populations derived from the *Rf1* donor accessions R3, R113 and L19, loci affecting fertility restoration were detected by QTL analyses. The initial QTL detection and estimation of QTL parameters in these populations was performed using the R package R/qtl (Broman et al. 2003; R Core Team 2016). The standard approach for QTL detection, maximum-likelihood estimation, makes use of the assumption that the residual phenotypic variation is normally distributed (Lander and Botstein 1989). However, this assumption could not be made for the populations of the present study, for which a substantial proportion of the descendants was expected to be completely sterile. Although the standard approach under a normal model can work reasonably well if the fraction of individuals with an extreme phenotypic value is small, this method can lead to the detection of false QTL if many individuals exhibit an extreme phenotypic value (Broman 2003). To avoid this, in the present study, the initial QTL detection was performed by simple interval mapping using the two-part model, which combines two analyses: one for the binary phenotype of all individuals and the other for the quantitative phenotype of the fertile individuals. Conditional genotype probabilities were calculated at a grid of 1-cM intervals and at all marker loci assuming a genotyping error rate of 0.0001. For each position of a putative QTL, maximum-likelihood estimates for the parameters $\theta = (\pi_1, \pi_2, \mu_1, \mu_2, \sigma)$ were obtained by maximising the likelihood function

$$L(\theta) = \prod_i \sum_j p_{ij}(1-\pi_j)^{1-z_i}\{\pi_j f(y_i;\mu_j,\sigma)\}^{z_i},$$

where π_j was the probability that individual i is fertile given QTL genotype j,

μ_j was the mean phenotypic value of individuals with QTL genotype j,

σ was the common phenotypic standard deviation,

p_{ij} was the probability that the QTL genotype of individual i is j,

z_i was the binary phenotype of individual i with $z_i \in \{0, 1\}$ and

$f(y_i; \mu_j, \sigma)$ was the density function for a normal distribution with mean μ and standard deviation σ (Broman 2003).

For each position on the map, the evidence of QTL was summarized by a LOD score, which is the common logarithm of the ratio of the maximum likelihood to the likelihood under the null hypothesis. Genome-wide LOD thresholds were calculated by 10,000 permutations of the phenotypic data (Churchill and Doerge 1994). A genome-wide LOD threshold was defined as the 0.95 quantile of the maximum LOD scores obtained from these permutations. The significance of a QTL was judged based on an empirical P value of the two-part model ($P_{\pi\mu}$), which was defined as the proportion of the maximum LOD scores obtained by permutation that met or exceeded the particular, observed LOD score. As suggested by Broman and Sen (2009), a LOD support interval was defined as the interval in which the LOD score was within 1.5 units of its maximum. Physical positions of markers linked to QTL were obtained from a search of the SNP-flanking sequence against the Chinese Spring RefSeq v1.0 using BLAST (IWGSC 2018).

To obtain more precise estimates for the location of QTL that were detected using the two-part model, a multi-cross QTL analysis was conducted using the iQTLm algorithm implemented in the MCQTL software (Charcosset et al. 2001; Jourjon et al. 2005). The multi-cross QTL analysis was based on a consensus linkage map, which comprised only chromosomes carrying previously detected QTL. The iQTLm algorithm was also applied for QTL analysis within populations to verify previously undetected QTL that were discovered only in the multi-cross analysis. Conditional genotype probabilities were calculated at all marker loci. If a marker interval exceeded 1 cM, genotype probabilities were also calculated at loci in between, allowing a maximal distance of 1 cM. In the process of QTL detection by the iQTLm algorithm, for each chromosome, loci as-

sociated with the phenotype were selected as cofactors by forward selection and used to detect QTL by composite interval mapping (CIM; Zeng 1994). Cofactors were searched along the whole chromosome excluding loci within a 5-cM window of cofactors already included in the linear model. QTL positions were then refined by using the detected QTL as cofactors in another round of CIM. This procedure was repeated until QTL positions converged (Larièpe et al. 2012; Mangin et al. 2014). The iQTLm algorithm was applied for QTL analyses within populations using the following linear model:

$$\mathbf{y} = \mathbf{1}_N\mu + \mathbf{X}_q\mathbf{b}_q + \sum_{c\neq q}(\mathbf{X}_c\mathbf{b}_c) + \mathbf{e},$$

where $\mathbf{y}$ was a $N \times 1$ vector of phenotypic values of N individuals of a population,

$\mathbf{1}_N$ was a $N \times 1$ vector of ones,

μ was the mean phenotypic value of a population,

$\mathbf{X}_q$ and $\mathbf{X}_c$ were $N \times 2$ matrices containing the genotype probabilities of parental alleles at QTL q and cofactor c,

$\mathbf{b}_q$ and $\mathbf{b}_c$ were 2×1 vectors containing the additive effects of parental alleles at QTL q and cofactor c and

$\mathbf{e}$ was a $N \times 1$ vector of residuals with $\mathbf{e} \sim \mathrm{N}(0, \sigma_e^2\mathbf{I})$.

QTL analysis across the three populations derived from R3, R113 and L19 was performed using the following model:

$$\mathbf{y}^* = \mathbf{J}\boldsymbol{\mu}^* + \mathbf{X}_q^*\mathbf{b}_q^* + \sum_{c\neq q}(\mathbf{X}_c^*\mathbf{b}_c^*) + \mathbf{e}^*,$$

where $\mathbf{y}^*$ was a $N^* \times 1$ vector of phenotypic values of N^* individuals of three populations,

$\mathbf{J}$ was a $N^* \times 3$ matrix, whose elements were 0 or 1 depending whether or not individual i^* belonged to one of three populations,

$\boldsymbol{\mu}^*$ was a 3×1 vector containing population-specific phenotypic means,

$\mathbf{X}_q^*$ and $\mathbf{X}_c^*$ were $N^* \times 4$ matrices containing the genotype probabilities of parental alleles at QTL q and cofactor c,

$\mathbf{b}_q^*$ and $\mathbf{b}_c^*$ were 4×1 vectors containing the additive effects of parental alleles at QTL q and cofactor c and

$\mathbf{e}^*$ was a $N^* \times 1$ vector of residuals with $\mathbf{e}^* \sim \mathrm{N}(0, \sigma_{e^*}^2 \mathbf{I})$.

The matrix notation for the two linear models was adopted from Bardol et al. (2013). It was assumed that each parental line carried a different QTL allele and that the effect of the allele of the common parent Sperber was the same in all populations. The additive effects at a QTL were estimated so that they summed up to zero (Jourjon et al. 2005; Bardol et al. 2013). The significance of a locus was determined by comparing stacked models with and without the respective locus using the F test and calculating the $-\log_{10}(P)$ value. Empirical, genome-wide thresholds were calculated by 10,000 permutations at a significance level of $\alpha = 0.05$ (Churchill and Doerge 1994). An approximate LOD score was calculated from the F statistic: $\mathrm{LOD} = 0.217 \mathrm{df}_1 F$, where df_1 was the numerator degree of freedom from the F test (Mangin et al. 2014). The support interval for a QTL was defined as the interval for which the LOD score was within 1.5 units of its maximum. The phenotypic variance explained by a model (R^2) was defined as the ratio of residual variances of a model including no QTL and the full model. The explained phenotypic variance of a single QTL was defined as the ratio of residual variances of two stacked models with and without the respective QTL (Mangin et al. 2014).

In order to estimate the effects of QTL detected by the two-part model or iQTLm in the presence of epistasis, the seed set was regressed on the conditional genotype probabilities considering QTL × QTL interactions within each of the three populations (Haley and Knott 1992). The following linear model was applied:

$$\mathbf{y} = \mathbf{1}_N \mu + \mathbf{X}\mathbf{b} + \mathbf{e},$$

where $\mathbf{y}$ was a $N \times 1$ vector of phenotypic values of N individuals of a population,

$\mathbf{1}_N$ was a $N \times 1$ vector of ones,

μ was the mean phenotypic value of a population,

$\mathbf{X}$ was an $N \times K$ matrix with conditional genotype probabilities of N individuals and K QTL and QTL-by-QTL interactions,

$\mathbf{b}$ was a $K \times 1$ vector with additive effects of K QTL and QTL-by-QTL interactions and

$\mathbf{e}$ was a $N \times 1$ vector of residuals with $\mathbf{e} \sim \mathrm{N}(0, \sigma_e^2 \mathbf{I})$.

Homozygous and heterozygous QTL genotypes were coded as -0.5 and 0.5, respectively (Broman and Sen 2009). The percentage of the phenotypic variance explained by a linear model (R^2) was estimated by the ratio of the explained sum of squares of the full model to the total sum of squares. The estimated ratio of phenotypic variance explained by a QTL or a QTL-by-QTL interaction was defined by the difference between residual sum of squares of models with and without the particular variable, divided by the total sum of squares (Broman and Sen 2009). Statistical significance of epistatic interactions was determined using a heteroscedasticity-consistent Wald test, which compared full and reduced models (Zeileis and Hothorn 2002; Zeileis 2004). Individuals of a population were grouped with respect to their QTL genotypes to analyse the penetrance and expressivity of QTL. According to Griffiths et al. (2000), penetrance was defined as the percentage of individuals with a given genotype that exhibit the associated discrete phenotype. In contrast, expressivity was defined as the quantitative extent to which a given genotype is phenotypically expressed. To define discrete QTL genotypes for this analysis, it was assumed that the true genotype is the one with the higher conditional probability. Ratios of fertile and sterile plants among QTL genotype groups were analysed using Fisher's exact test to determine whether a QTL affected the penetrance of a second QTL and to compare the penetrance of two QTL within a population. For more than one comparison within a population, FDR-adjusted P values were used. To assess whether a QTL affected the expressivity of another QTL and if the expressivities of two QTL differed, the mean seed set of QTL genotype groups was compared using one-way analysis of variance and Tukey's HSD test. Significance levels for statistical tests were set at $\alpha = 0.05$.

2.6 Distribution of marker haplotypes in the genomic target regions across wheat species

The distribution of mapped loci was evaluated based on haplotypes constructed from closely linked molecular markers. Pairwise genetic dissimilarities between marker haplotypes were calculated using modified Roger's distance:

$$d_W = \frac{1}{\sqrt{2m}} \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n_j} (p_{ij} - q_{ij})^2},$$

where m was the number of markers and p_{ij} and q_{ij} were the frequencies of the jth allele at the ith locus (Wright 1978; Reif et al. 2005). The matrix containing the pairwise genetic distances was the basis for hierarchical cluster analysis of marker haplotypes using the complete linkage method implemented in the R base package (R Core Team 2016). Phylograms were drawn using APE (Paradis et al. 2004). Mean genetic distances between haplotypes and groups of haplotypes were compared using Welch's t test, thereby assuming unequal variances ($\alpha = 0.05$). The distribution of *Rf3* in common wheat and spelt was also evaluated in the context of population structure. The population structure of the common wheat breeding lines was analysed by PCoA using the R package APE. Basis for PCoA were modified Roger's distances calculated from whole-genome SNP data. To verify whether *Rf3* was homogeneously distributed within the common wheat population, Pearson's correlation coefficients were calculated between the binary *Rf3* marker genotype information and each of the the top ten eigenvectors determined by PCoA. A one-sample t test was used to determine whether or not the coefficients were different from zero. P values of these tests were adjusted for multiple testing using a FDR of 0.05. The distribution of *Rf3* was additionally estimated in 30 European spelt cultivars using a closely linked marker and testcrosses with CMS-Sperber. The relationship between *Rf3* and spelt population structure was evaluated as described for the common wheat lines. To examine whether or not the *Rf3* genotype of spelt depends on the genetic similarity to common wheat, the population structure of the spelt cultivars was analysed together with 368 common wheat cultivars by PCoA. Mean genetic distances between common wheat and spelt cultivars with and without the restoring *Rf3* allele were compared using Welch's t test ($\alpha = 0.05$).

3 Results

3.1 Segregation of fertility restoration

For each plant of the ten mapping populations, an average of 3.50 spikes was analysed to determine phenotypic values of individual plants (Table 6). Fertility restoration in these populations ranged from complete cytoplasmic male sterility (CMS) to partially and fully restored fertility. The seed set generally increased from the base to the apex of a spike. In all analysed populations, the majority of plants exhibited either sterile or fertile spikes. The ratio of plants that exhibited both sterile and fertile spikes ranged between 0.5% (PopPrimepi) and 19.1% (PopBadenstern), excluding PopHolstenkorn, for which only one spike per plant was evaluated. Focusing on the fertile descendants, the lowest and highest mean seed set was observed in the BC_1 population PopBadenstern and the F_2 population PopHolstenkorn, respectively. The seed set distributions of the fertile plants are summarised in Figure 2. The fertile fractions of the three subsets of PopPrimepi, which were cultivated in different environments, exhibited mean seed sets ($\pm$ standard deviation) of 1.77 ± 0.25, 2.09 ± 0.37 and 1.03 ± 0.38 seeds per spikelet (Figure A1). A one-way analysis of variance and Tukey's HSD test revealed that the means of the three populations differed significantly, indicating that the environment affected the seed set ($P < 0.001$).

For PopR3, more sterile than fertile individuals were observed. This result contradicts the assumption that the restoration capacity of R3 is controlled independently by two *restorer-of-fertility* (*Rf*) loci. Analysing the segregation into fertile and sterile plants using the χ^2-goodness-of-fit test revealed that it also significantly deviated from a possible 1:1 ratio, which would have been expected for a monogenic inheritance of fertility restoration. This indicates incomplete penetrance of the involved restorer locus. The population involving R113 exhibited a ratio of fertile and sterile descendants that did not differ significantly from a 3:1 segregation pattern. This result is consistent with the expected digenic control of restoration capacity of R113. The number of fertile and sterile plants in all other BC_1 populations exhibited no significant deviation from a 1:1 ratio, indicating a monogenic inheritance of fertility restoration. This was also observed for all three subsets of PopPrimepi ($P \geq 0.35$). In the F_2 population

Table 6 Average number of analysed spikes per plant ($\overline{N}_{\text{spikes}}$), mean seed set of fertile individuals ± standard deviation (seeds per spikelet) and segregation of fertility restoration. Ratios of fertile (N_{fertile}) and sterile (N_{sterile}) individuals were compared with theoretical ratios using the χ^2-goodness-of-fit test

Population	Type	$\overline{N}_{\text{spikes}}$	$\text{Mean}_{\text{fertile}}$	N_{fertile}:N_{sterile}	Tested ratio	P
PopR3	BC_1	3.47	0.61 ± 0.36	74:123	1:1	< 0.01
PopR113	BC_1	3.46	0.88 ± 0.54	159:42	3:1	0.18
PopL19	BC_1	3.47	0.98 ± 0.55	102:128	1:1	0.09
PopPrimepi	BC_1	3.48	1.57 ± 0.57	104:89	1:1	0.28
PopPR143	BC_1	4.00	0.91 ± 0.37	117:104	1:1	0.38
PopPalmaress	BC_1	2.02	1.47 ± 0.51	50:41	1:1	0.35
PopBadenkrone	BC_1	3.87	1.02 ± 0.37	132:158	1:1	0.13
PopBadenstern	BC_1	3.95	0.54 ± 0.34	104:116	1:1	0.42
PopHolstenkorn	F_2	1.00	1.94 ± 0.69	70:31	3:1	0.19
PopSchwabenspelz	BC_1	3.85	0.82 ± 0.35	151:137	1:1	0.41

PopHolstenkorn, the segregation of fertility restoration did not differ significantly from a 3:1 ratio of fertile and sterile plants, which was assumed for monogenic inheritance of fertility restoration. Analysing the seed set of the fertile fractions with Hartigan's dip test did not suggest a multimodal distribution in any of the ten populations. The smallest P value of the dip test was observed in PopR113 ($P = 0.10$), for which a digenic control of fertility restoration was expected.

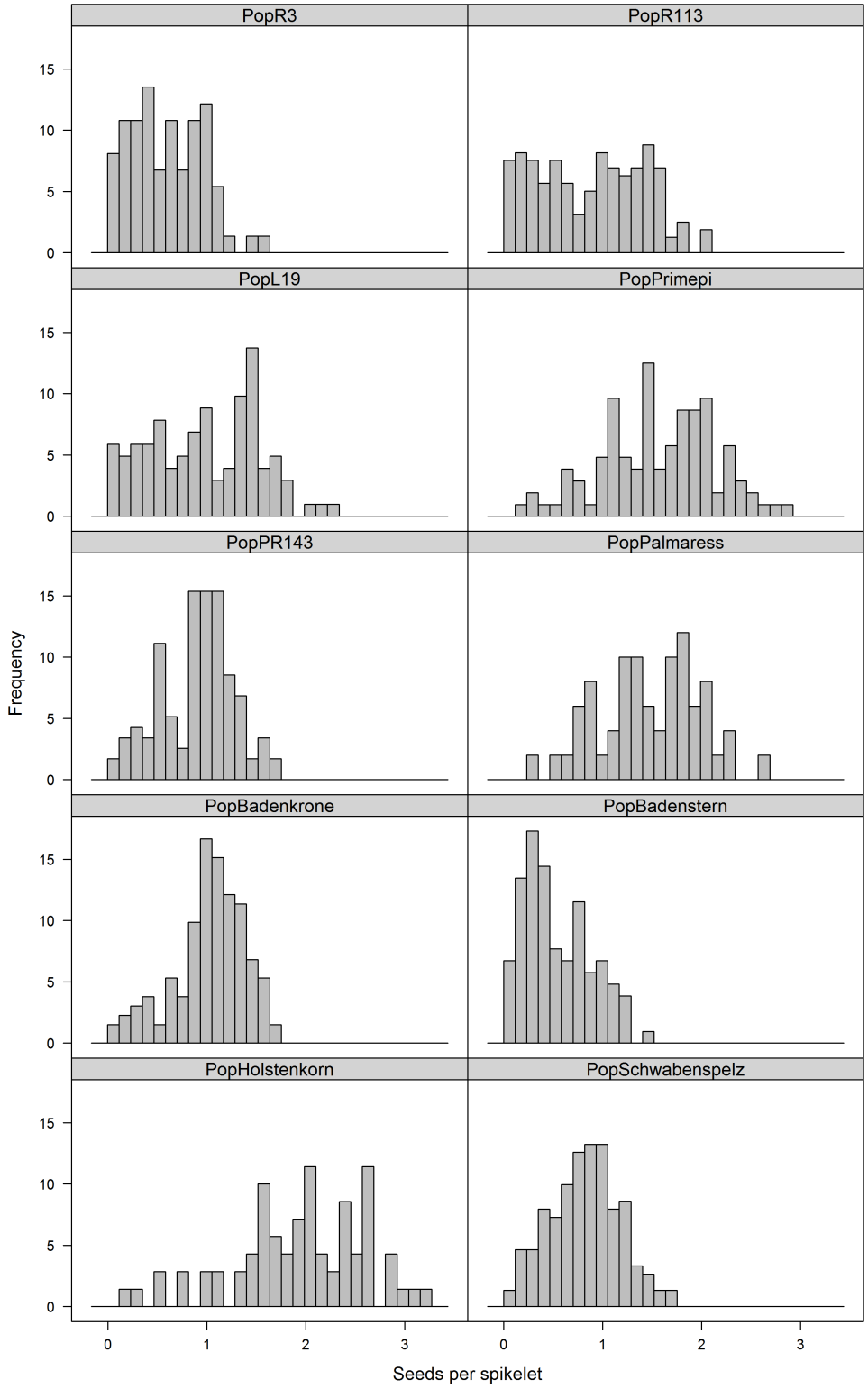

Figure 2 Seed set distributions for the fertile fractions of the ten analysed mapping populations

3.2 Genetic mapping of *Rf1*

3.2.1 Marker selection and linkage mapping

In the populations derived from the *Rf1* donor lines R3, R113 and L19, a total of 2351, 2484 and 2545 polymorphic markers from the 20k single-nucleotide polymorphism (SNP) array, respectively, were selected for the construction of linkage maps. The number of markers discarded at individual selection steps is summarised in Table A1. A principal coordinate analysis (PCoA) of genetic distances between the individuals of the populations did not indicate any within-population structure (Figure A2). Keeping markers that were monomorphic within populations and performing a PCoA across the parental lines and across populations revealed that the genetic distance between R113 and L19 was smaller than the genetic distance between R3 and R113 and between R3 and L19. The genetic distance to Sperber was similar for all *Rf1* donor accessions. An analogous relationship was observed between the populations derived from these accessions (Figure A3).

The linkage map for PopR3 comprised 1992 SNP markers, which were distributed over 31 linkage groups and spanned 2,957.5 centimorgans (cM; Figure 3a). A deviation from the expected segregation ratio was observed for 46 markers on 16 linkage groups ($P \leq 0.047$). Segregation distortion affected mostly the chromosomes 1D, 2D, 3D, 5A and 6A (Figure 4a). The linkage map for PopR113 contained 26 linkage groups with 2066 SNP markers, which spanned a map distance of 2994.4 cM (Figure 3b). In PopR113, 132 markers exhibited segregation distortion ($P \leq 0.042$). Distributed across 11 linkage groups, most deviations from the expected ratio were observed on chromosome 6B. On this chromosome, only the most distal part of the long arm was not affected by segregation distortion. Furthermore, low recombination rates on this chromosome resulted in a map length of only 77.1 cM. Several deviations from the theoretical pattern were also observed on chromosome 2B and 3B (Figure 4b). The PopL19-derived linkage map comprised 2100 SNP markers, which were distributed across 27 linkage groups (Figure 3c). The linkage map of this population spanned a mapping distance of 2971.4 cM. Only nine of the mapped markers exhibited a deviation from the expected segregation ratio ($P \leq 0.015$). These deviating markers were distributed across six linkage groups (Figure 4c). A comparison between the marker

positions of one of the reference linkage maps (Maccaferri et al. 2015b) and the linkage maps derived from PopR3, PopR113 and PopL19 is illustrated in Figures A4, A5 and A6, respectively. Based on the developed linkage maps of the three populations, a consensus map was constructed. For this, only chromosomes that were found to carry quantitative trait loci (QTL) were integrated.

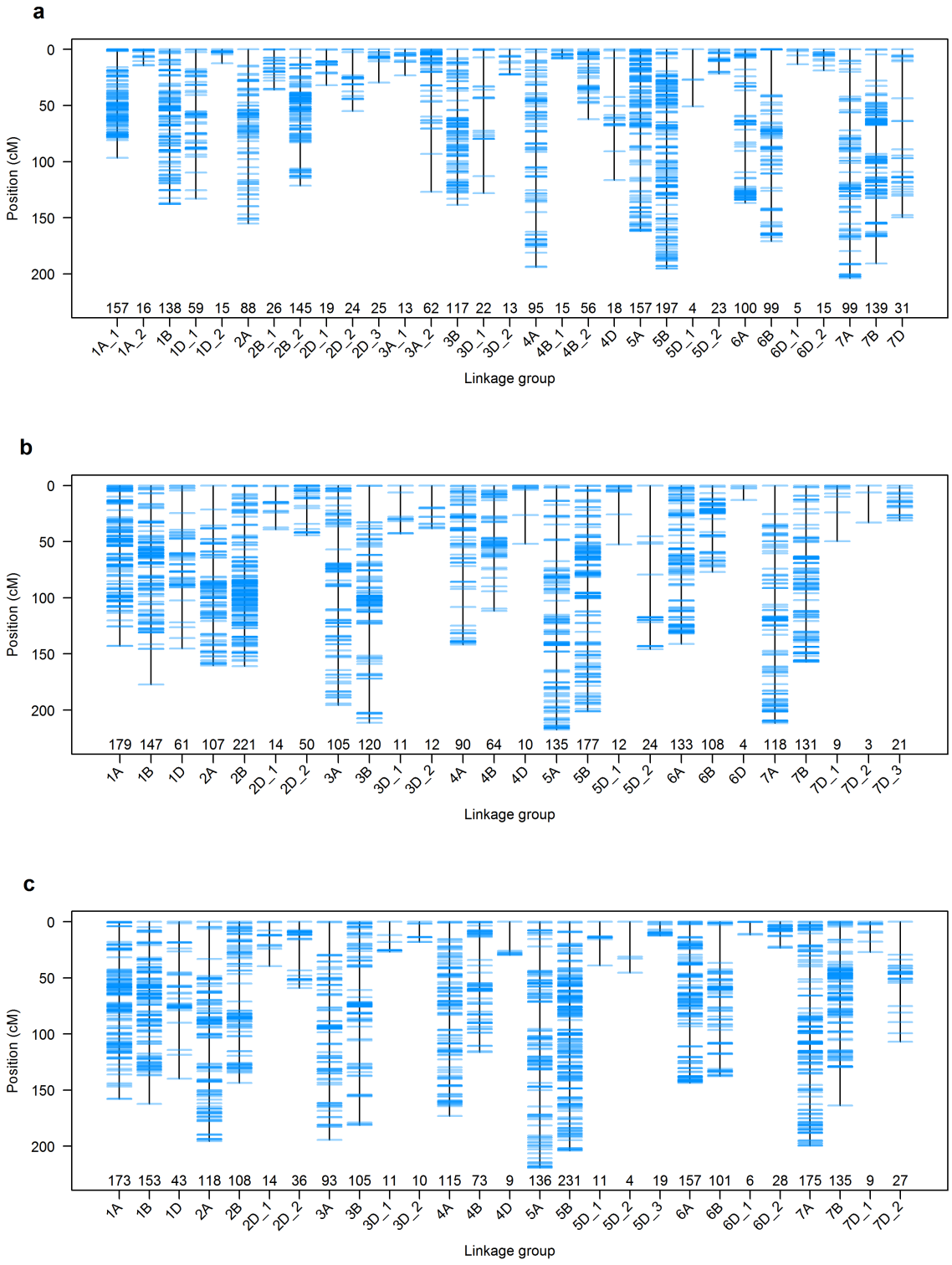

Figure 3 Distribution of single-nucleotide polymorphism (SNP) markers across the linkage groups of PopR3 (**a**), PopR113 (**b**) and PopL19 (**c**). The number of mapped markers per linkage group is depicted at the bottom of each subfigure

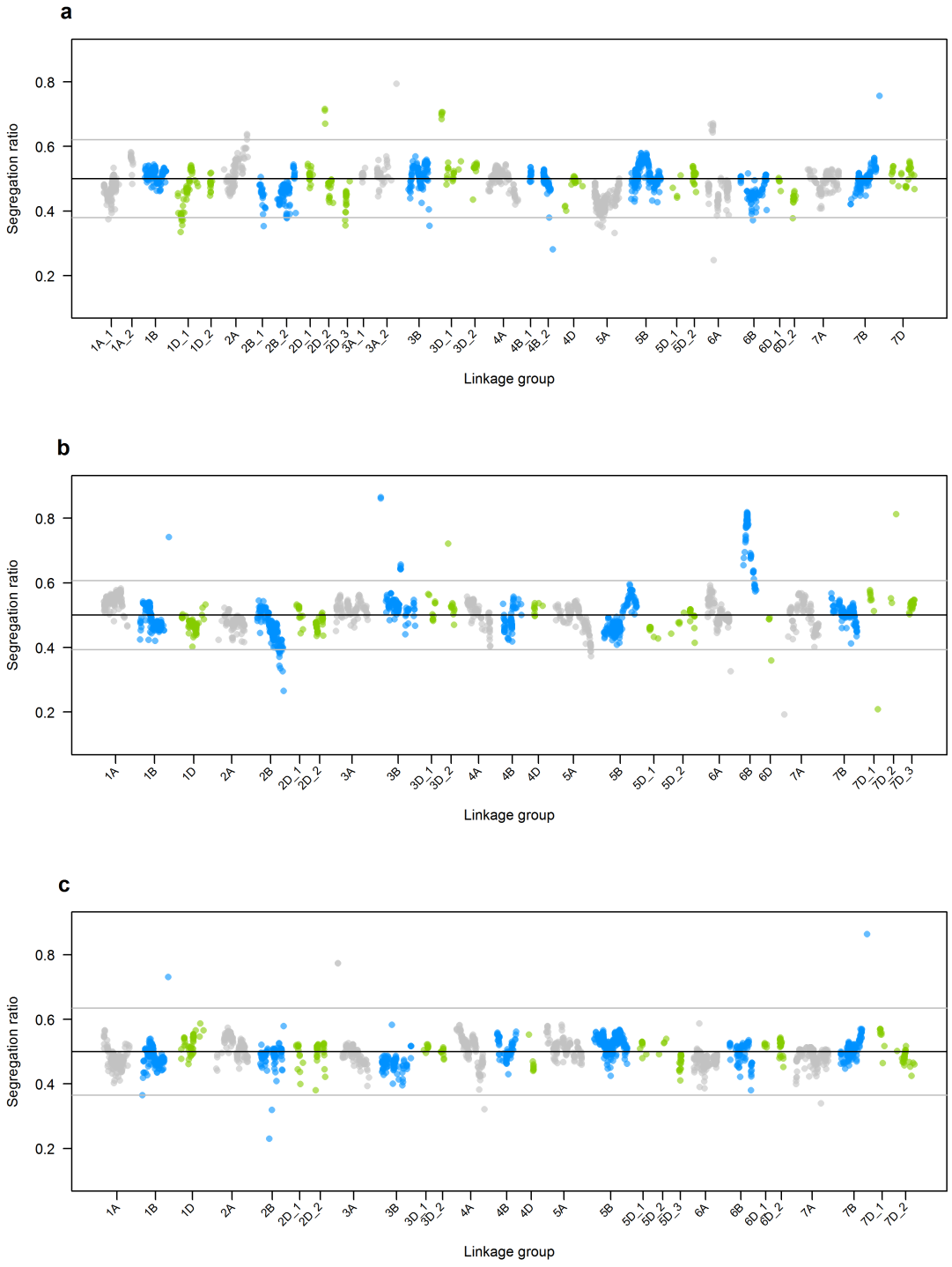

Figure 4 Segregation ratio of SNP markers across the linkage groups of PopR3 (**a**), PopR113 (**b**) and PopL19 (**c**). The segregation ratio was defined as the number of heterozygous genotypes divided by the total number of genotypes. Markers on the subgenomes A, B and D are represented by grey, blue and green dots, respectively. Thresholds for genome-wide false-discovery rates of 0.05 are illustrated by grey lines

3.2.2 QTL analysis

In PopR3, a genome-wide scan using the two-part model revealed one QTL on chromosome 1AS (*Rf1*) and a second QTL on chromosome 1BS, which is hereafter referred to as *QRf.lfl-1BS* (Figure 5a). Genetic map positions of the two loci are illustrated in Figure 6a, Figure 7a and Table 7. Markers within the logarithm of the odds (LOD) support intervals of *Rf1* and *QRf.lfl-1BS* did not deviate from the expected segregation ratios. Whereas the restorer locus *Rf1* affected the binary phenotype, the quantitative phenotype of the fertile plants remained unaffected. In contrast, *QRf.lfl-1BS* affected only the quantitative phenotype of fertile individuals, indicating that it may be a modifier locus.

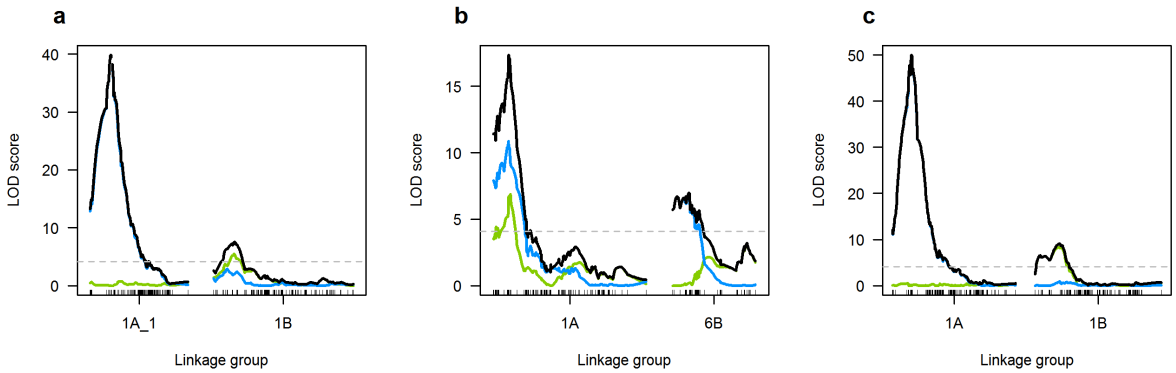

Figure 5 Logarithm of the odds (LOD) score curves for quantitative trait loci (QTL) detected in PopR3 (**a**), PopR113 (**b**) and PopL19 (**c**) using the two-part model. LOD scores for the two-part model, the model for the binary phenotype and the model for the quantitative phenotype of the fertile fraction are illustrated by black, blue and green lines, respectively. The dashed, grey lines represent the thresholds for QTL detection

A genome-wide scan in PopR113 detected one QTL on chromosome 1AS and a second QTL on chromosome 6BS, presumably representing the restorer loci *Rf1* and *Rf4*, respectively (Figure 5b). The genetic map positions of *Rf1* and *Rf4* in this population are displayed in Figure 6b, Figure A10b and Table 7. A possible third QTL on chromosome 6A did not reach significance ($P_{\pi\mu} = 0.09$) and was not analysed further. Segregation patterns of markers within the support interval of the restorer locus on chromosome 1AS did not deviate from the expected ratios. In contrast, all markers on chromosome 6B exhibited the heterozygous genotype more often than was expected. *Rf1* affected both the binary phenotype and the quantitative phenotype of the fertile descendants.

Rf4 influenced the binary phenotype, but no effect was observed on the quantitative phenotype of fertile individuals. In PopL19, two QTL were detected: one on chromosome 1AS, which probably corresponds to *Rf1*, and the other one on chromosome 1BS, which is hereafter called *QRf.lfl-1BS* (Figure 5c). Genetic map positions of *Rf1* and *QRf.lfl-1BS* are illustrated in Figure 6c, Figure 7b and Table 7. No segregation distortion was observed for markers within the support intervals of the two detected loci. Similar as observed in PopR3, *Rf1* influenced the binary phenotype, whereas the QTL on chromosome 1B had only an effect on the quantitative phenotype of the fertile fraction of PopL19, indicating that it is a modifier locus. Genome-wide LOD scores of the two-part model for PopR3, PopR113 and PopL19 are depicted in Figures A7, A8 and A9, respectively.

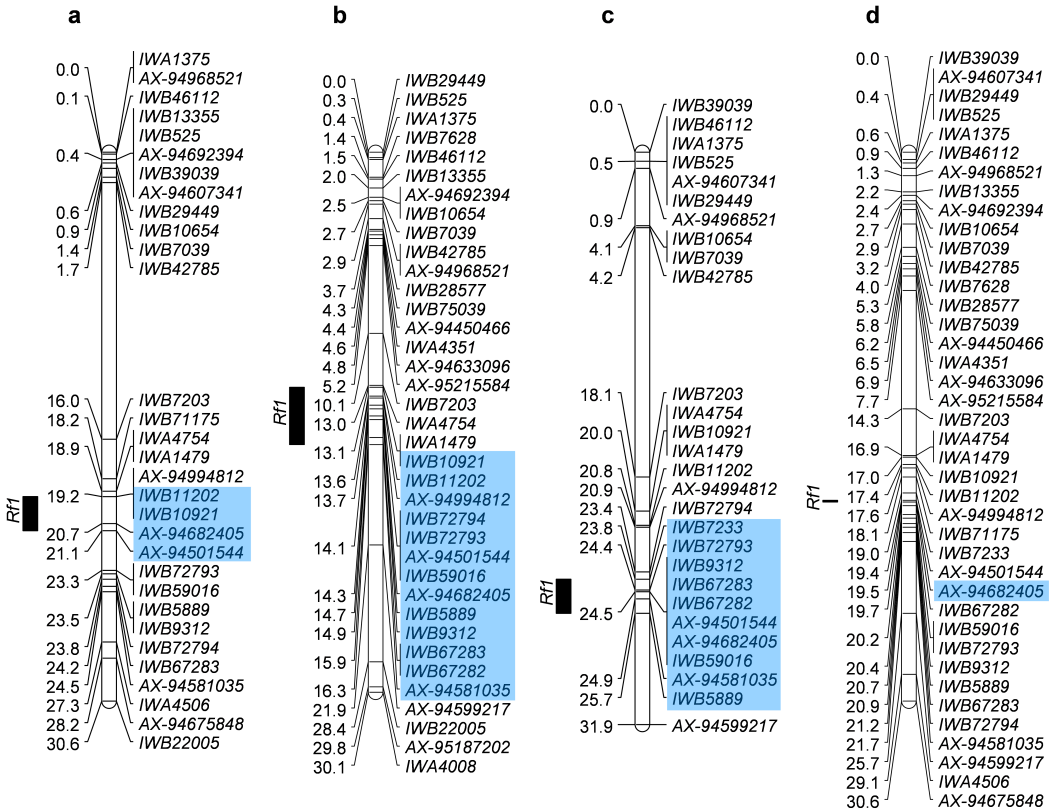

Figure 6 Position of *Rf1* on chromosome-1A linkage maps derived from PopR3 (**a**), PopR113 (**b**) and PopL19 (**c**) and on the consensus map (**d**). Black bars and blue boxes depict 1.5-LOD support intervals for *Rf1*, which were estimated using the two-part model within populations (**a-c**) and iQTLm across populations (**d**). Map positions are given in centimorgan (cM)

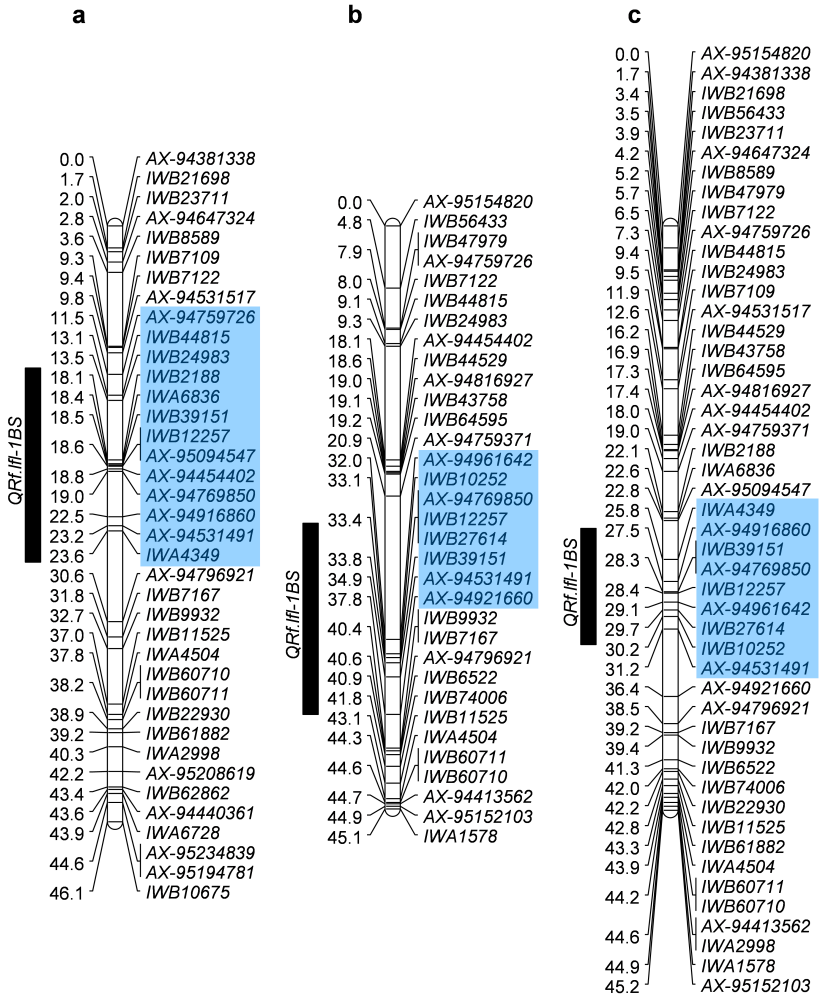

Figure 7 Position of *QRf.lfl-1BS* on chromosome-1B linkage maps derived from PopR3 (**a**) and PopL19 (**b**) and on the consensus map (**c**). Black bars and blue boxes depict 1.5-LOD support intervals for *QRf.lfl-1BS*, which were estimated using the two-part model within populations (**a**, **b**) and iQTLm across populations (**c**). Map positions are given in cM

Table 7 Chromosomes, genetic map positions, 1.5-LOD support intervals (cM), peak markers and their physical positions (mega base pairs; Mbp), maximum LOD scores, and P values for quantitative trait loci (QTL) detected using the two-part model. $P_{\pi\mu}$, P_{π} and P_{μ} are the P values of the two-part model, the model for the binary phenotype and the model for the quantitative phenotype considering only fertile individuals, respectively

Population	Locus	Chromosome	Map position	Peak marker	LOD	$P_{\pi\mu}$	P_{π}	P_{μ}
PopR3	*Rf1*	1AS	20.7 (19.2–21.1)	*AX-94682405* (14.14)	39.9	< 0.001	< 0.001	1
	QRf.lfl-1BS	1BS	20.0 (11.0–26.0)	*AX-94769850* (16.85)	7.6	< 0.001	0.56	< 0.001
PopR113	*Rf1*	1AS	14.3 (13.1–16.3)	*AX-94682405* (14.14)	17.4	< 0.001	< 0.001	< 0.001
	Rf4	6BS	15.0 (0.0–25.0)	*IWA52* (32.33)	7.0	< 0.001	< 0.001	1
PopL19	*Rf1*	1AS	24.5 (23.8–25.7)	*AX-94682405* (14.14)	50.0	< 0.001	< 0.001	1
	QRf.lfl-1BS	1BS	31.0 (23.0–37.8)	*AX-94961642* (22.02)	9.1	< 0.001	1	< 0.001

To obtain more precise estimates of the location of the detected restorer and modifier loci, a consensus map comprising the target chromosomes 1A, 1B and 6B was employed for QTL detection using the iQTLm algorithm within and across the populations PopR3, PopR113 and PopL19. Due to the strong segregation distortion on chromosome 6B in PopR113, linkage maps could not be integrated for this chromosome. Instead, the PopR113-derived map was used for the iQTLm analysis of chromosome 6B. The iQTLm method within PopR3 confirmed the previously detected loci *Rf1* and *QRf.lfl-1BS*. In addition, a further QTL was located on the long arm of chromosome 1B (Figure 8a). This QTL is hereafter referred to as *QRf.lfl-1BL*. In PopR113, the iQTLm algorithm corroborated the previously detected restorer loci *Rf1* and *Rf4*, and, in addition, a third QTL on chromosome 6BL (*QRf.lfl-6BL*) was found (Figure 8b). In the population derived from L19, the iQTLm algorithm discovered the same two QTL that were observed using the two-part model, namely *Rf1* and *QRf.lfl-1BS* (Figure 8c). Combining the phenotypic and genotypic data from all three populations to conduct QTL detection across populations confirmed four of the loci that were detected within populations using iQTLm, namely *Rf1*, *QRf.lfl-1BS*, *Rf4* and *QRf.lfl-6BL*. The QTL on chromosome 1BL (*QRf.lfl-1BL*) did not reach significance in the across-population iQTLm analysis ($P = 0.062$; Figure 9). The 1.5-LOD support interval for *Rf1* was reduced to 0.1 cM on the consensus map using iQTLm across populations compared to 1.9–3.2 cM on the initial linkage maps using the two-part model. Analogously, the support interval for *QRf.lfl-1BS* was narrowed down from 14.8–15.0 cM to 9.0 cM. The location of *Rf4* was delimited to an interval of 16.8 cM compared to 25.0 cM on the PopR113-derived map. The positions of the detected loci on the consensus map are illustrated in Figure 6d, Figure 7c, Figure A10 and Table 8.

Table 8 Chromosomes, positions on the consensus linkage map, 1.5-LOD support intervals (cM), peak markers and their physical positions (Mbp), maximum LOD scores, explained phenotypic variances (R^2) and regression coefficients (β; seeds per spikelet) for QTL detected using the iQTLm algorithm. All genetic factors were significant with P values < 0.001

Population	Genetic factor	Chromosome	Map position	Peak marker	LOD	R^2	$\beta_{Sperber}$	β_{R3}	β_{R113}	β_{L19}
PopR3	*Rf1*	1AS	19.5 (19.4–19.6)	*AX-94682405* (14.14)	47.0	0.53	−0.25	0.25		
	QRf.lfl-1BS	1BS	28.4 (14.4–28.7)	*IWB12257* (17.73)	9.2	0.18	−0.11	0.11		
	QRf.lfl-1BL	1BL	105.4 (92.0–113.4)	*IWB44606* (658.71)	2.5	0.06	−0.06	0.06		
	Full model					0.59				
PopR113	*Rf1*	1AS	19.5 (19.4–21.4)	*AX-94682405* (14.14)	22.8	0.35	−0.33		0.33	
	Rf4	6BS	12.7 (0.0–19.1)	*AX-95173841* (25.60)	4.0	0.09	−0.16		0.16	
	QRf.lfl-6BL	6BL	69.4 (30.3–77.1)	*IWA2431* (712.35)	3.3	0.07	0.13		−0.13	
	Full model					0.41				
PopL19	*Rf1*	1AS	20.4 (19.3–22.8)	*IWB9312* (16.39)	56.6	0.53	−0.42			0.42
	QRf.lfl-1BS	1BS	29.1 (22.0–36.1)	*AX-94961642* (22.02)	10.2	0.17	−0.18			0.18
	Full model					0.59				
Multi-cross	*Rf1*	1AS	19.5 (19.4–19.5)	*AX-94682405* (14.14)	114.0	0.46	−0.51	−0.01	0.17	0.35
	QRf.lfl-1BS	1BS	28.4 (23.4–32.4)	*IWB12257* (17.73)	15.4	0.10	−0.17	0.05	−0.07	0.18
	Rf4	6BS	12.7 (1.6–18.4)	*AX-95173841* (25.60)	6.4	0.05	−0.08	−0.09	0.25	−0.09
	QRf.lfl-6BL	6BL	69.3 (65.0–76.0)	*IWB57728* (711.24)	5.0	0.04	0.05	0.10	−0.21	0.06
	Full model					0.52				

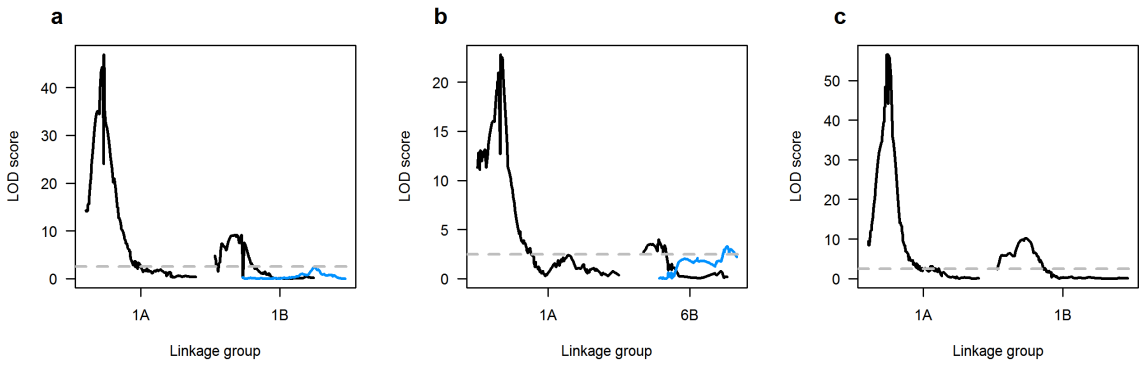

Figure 8 LOD score curves for QTL detected in PopR3 (**a**), PopR113 (**b**) and PopL19 (**c**) using iQTLm within populations. The dashed, grey lines illustrate the genome-wide thresholds for QTL detection

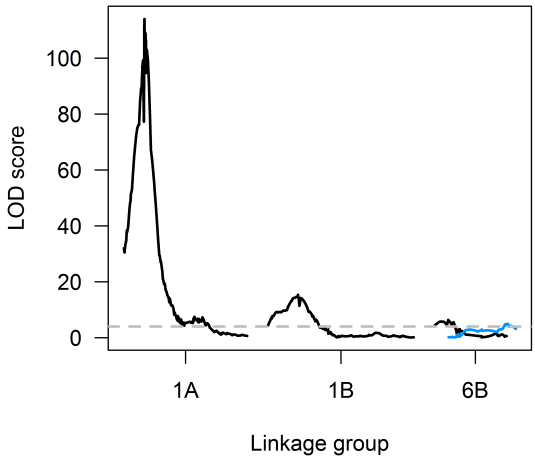

Figure 9 LOD score curves for QTL detected across the populations PopR3, PopR113 and PopL19 using iQTLm. The dashed, grey line illustrates the genome-wide threshold for QTL detection

QTL effects and effects of possible interactions were estimated using Haley-Knott regression. In PopR3, the detected loci *Rf1* and *QRf.lfl-1BS* revealed a significant interaction, which was therefore included in the multiple regression model for the estimation of QTL parameters. The regression coefficients and the percentages of explained variance are listed in Table 9. According to the estimated regression coefficients, *Rf1* exhibited an additive effect of 0.22 or 0.73 seeds per spikelet, depending on the geno-

type of *QRf.lfl-1BS*. The favourable allele at this modifier locus was inherited from R3. Classifying the PopR3 individuals on the basis of QTL genotypes showed that 14 plants were sterile despite carrying a restoring allele at the *Rf1* locus, thereby indicating incomplete penetrance of *Rf1*. Assuming fertility for these individuals, the ratio of fertile and sterile plants did not deviate from the 1:1 segregation pattern expected for a single dominant restorer locus ($P = 0.13$). The comparison of discrete QTL genotype classes furthermore confirmed that *QRf.lfl-1BS* is a modifier locus rather than a restorer locus because it did not restore fertility in any of the PopR3 descendants. In contrast to the results of the two-part model, the targeted analysis of specific QTL genotype classes using Tukey's HSD test and Fisher's exact test suggested that *QRf.lfl-1BS* not only affected the expressivity but also the penetrance of *Rf1* ($P < 0.001$, Figure 10a). The effects of *Rf1* and *QRf.lfl-1BS* in PopR3 were additionally estimated by Haley-Knott regression considering *QRf.lfl-1BL*, which was only detected using iQTLm. Including this QTL in the linear model revealed that the three-way interaction *Rf1*×*QRf.lfl-1BS*×*QRf.lfl-1BL* ($P = 0.22$) and the two-way interaction *QRf.lfl-1BS*×*QRf.lfl-1BL* ($P = 0.22$) were not significant and were therefore excluded for the estimation of QTL parameters. In contrast, the two-way interactions *Rf1*×*QRf.lfl-1BS* and *Rf1*×*QRf.lfl-1BL* were significant and were retained in the model. Considering these interactions, the effect of *Rf1* varied between 0.16 and 0.82 seeds per spikelet. Analogous to *QRf.lfl-1BS*, the favourable *QRf.lfl-1BL* allele was derived from R3. Classifying the individuals of PopR3 on the basis of QTL genotypes clearly showed that none of the 58 plants heterozygous for *QRf.lfl-1BL* restored fertility in the absence of *Rf1*. Hence, it can be proposed that this QTL is a modifier locus. A comparison of the mean seed set and the ratio of fertile and sterile individuals among the QTL genotype groups demonstrated that *QRf.lfl-1BS* affected the expressivity ($P < 0.001$) and penetrance ($P = 0.014$) of *Rf1*, whereas *QRf.lfl-1BL* only influenced its expressivity ($P = 0.042$), while the penetrance ($P = 0.36$) remained unaffected. Fertile individuals carrying either *QRf.lfl-1BS*, *QRf.lfl-1BL* or combining both modifier loci did not exhibit differences pertaining expressivity ($P \geq 0.11$) and penetrance ($P \geq 0.07$; Figure 11a).

Haley-Knott regression in PopR113 revealed a significant interaction between the detected restorer loci *Rf1* and *Rf4*, which was included in the linear model. According to the estimated regression coefficients, *Rf1* exhibited an additive effect of 0.90 or 0.60

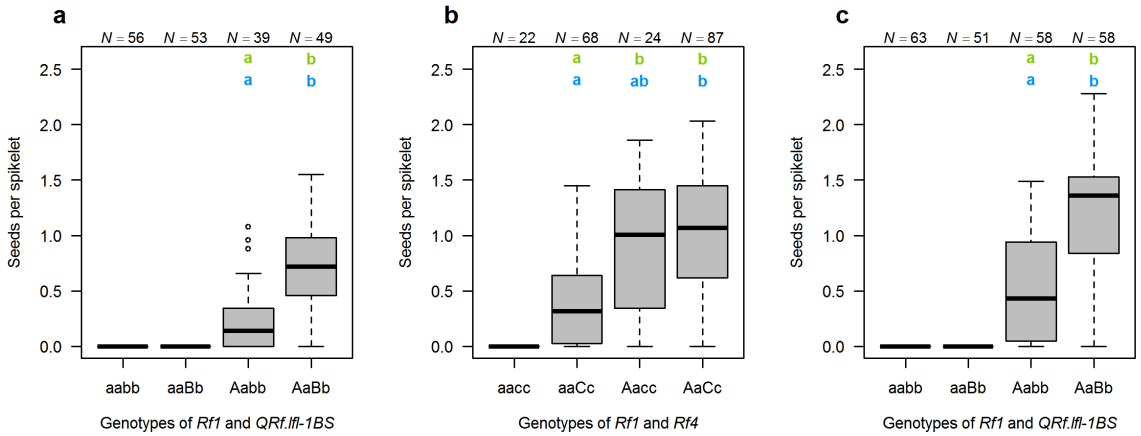

Figure 10 Seed set of PopR3 (**a**), PopR113 (**b**) and PopL19 (**c**) ordered by QTL genotype classes. The letters A/a, B/b and C/c at the bottom of the figure represent *Rf1*, *QRf.lfl-1BS* and *Rf4*, respectively. Lower-case and capital letters refer to the alleles of Sperber and the restorer parents, respectively. Different letters at the top of the figure denote significant differences for expressivity (green) and penetrance (blue)

seeds per spikelet, conditional on homozygosity or heterozygosity at the *Rf4* locus. Analogously, *Rf4* showed an additive effect of 0.43 or 0.14 seeds per spikelet (Table 9). Assigning the PopR113 descendants to QTL genotype classes revealed that two plants were sterile although they probably contained the restoring *Rf1* allele. Furthermore, 16 plants were sterile despite carrying the restoring *Rf4* allele. This discrepancy suggests incomplete penetrance of *Rf1* and *Rf4*. *Rf1* revealed a superior expressivity compared to *Rf4* ($P = 0.002$), whereas no significant difference was observed for the penetrance of the two loci ($P = 0.20$). Plants combining the favourable *Rf1* and *Rf4* alleles exhibited an increased expressivity and penetrance compared to individuals carrying only *Rf4* ($P < 0.001$). No significant differences in expressivity ($P = 0.84$) or penetrance ($P = 0.20$) were found between individuals carrying the restoring alleles for *Rf1* and *Rf4* and plants harbouring only the fertility-restoring *Rf1* allele (Figure 10b). Haley-Knott regression for this population was additionally performed considering *QRf.lfl-6BL*, which was only detected using the iQTLm algorithm. By adding this locus to the linear model, the three-way interaction *Rf1* × *Rf4* × *QRf.lfl-6BL* ($P = 0.14$) was not significant and was dropped from the model, whereas the two-way interactions *Rf1* × *Rf4*, *Rf1* × *QRf.lfl-6BL* and *Rf4* × *QRf.lfl-6BL* were considered for the estimation

of QTL parameters. Depending on *Rf4* and *QRf.lfl-6BL*, *Rf1* exhibited an additive effect ranging from 0.38 to 1.18 seeds per spikelet. The fertility-enhancing allele of *QRf.lfl-6BL* was inherited from Sperber. Grouping the individuals of PopR113 on the basis of their genotypes at the three loci suggested that *QRf.lfl-6BL* is a modifier locus because it did not restore fertility in any of the 11 individuals carrying the *QRf.lfl-6BL* allele of Sperber in the absence of *Rf1* and *Rf4*. Comparing the mean seed set and the ratio of fertile and sterile individuals among the groups showed that *Rf1* only exhibited a higher expressivity than *Rf4* if *QRf.lfl-6BL* was homozygous for the allele of Sperber ($P < 0.001$). No differences were found between the penetrance of *Rf1* and *Rf4* ($P \geq 0.19$). Irrespective of this modifier locus, the expressivity and penetrance of individuals carrying both *Rf1* and *Rf4* was higher compared to individuals harbouring only *Rf4* ($P \leq 0.039$). Under the premise that *QRf.lfl-6BL* was fixed for the comparison of plants carrying either *Rf1* or a combination of *Rf1* and *Rf4*, stacking of these two *Rf* genes did not result in a superior expressivity or penetrance ($P \geq 0.29$; Figure 11b).

In PopL19, a significant interaction between the two detected loci *Rf1* and *QRf.lfl-1BS* was observed. Depending on the *QRf.lfl-1BS* genotype, *Rf1* exhibited an additive effect of 0.50 or 1.22 seeds per spikelet (Table 9). The favourable allele at this locus was inherited from L19. Grouping the individuals of PopL19 based on the genotypes of the two loci confirmed the assumption that *QRf.lfl-1BS* is a modifier locus because it did not cause fertility restoration in any of analysed individuals carrying a non-restoring *Rf1* allele. Analogous to observations in PopR3 and PopR113, 14 plants were sterile although they were heterozygous at the *Rf1* locus, suggesting incomplete penetrance of *Rf1*. *QRf.lfl-1BS* influenced both the expressivity and penetrance of *Rf1* ($P \leq 0.008$; Figure 10c). QTL parameters were additionally estimated in the process of QTL detection using iQTLm. Percentages of explained phenotypic variance and estimated regression coefficients are summarised in Table 8.

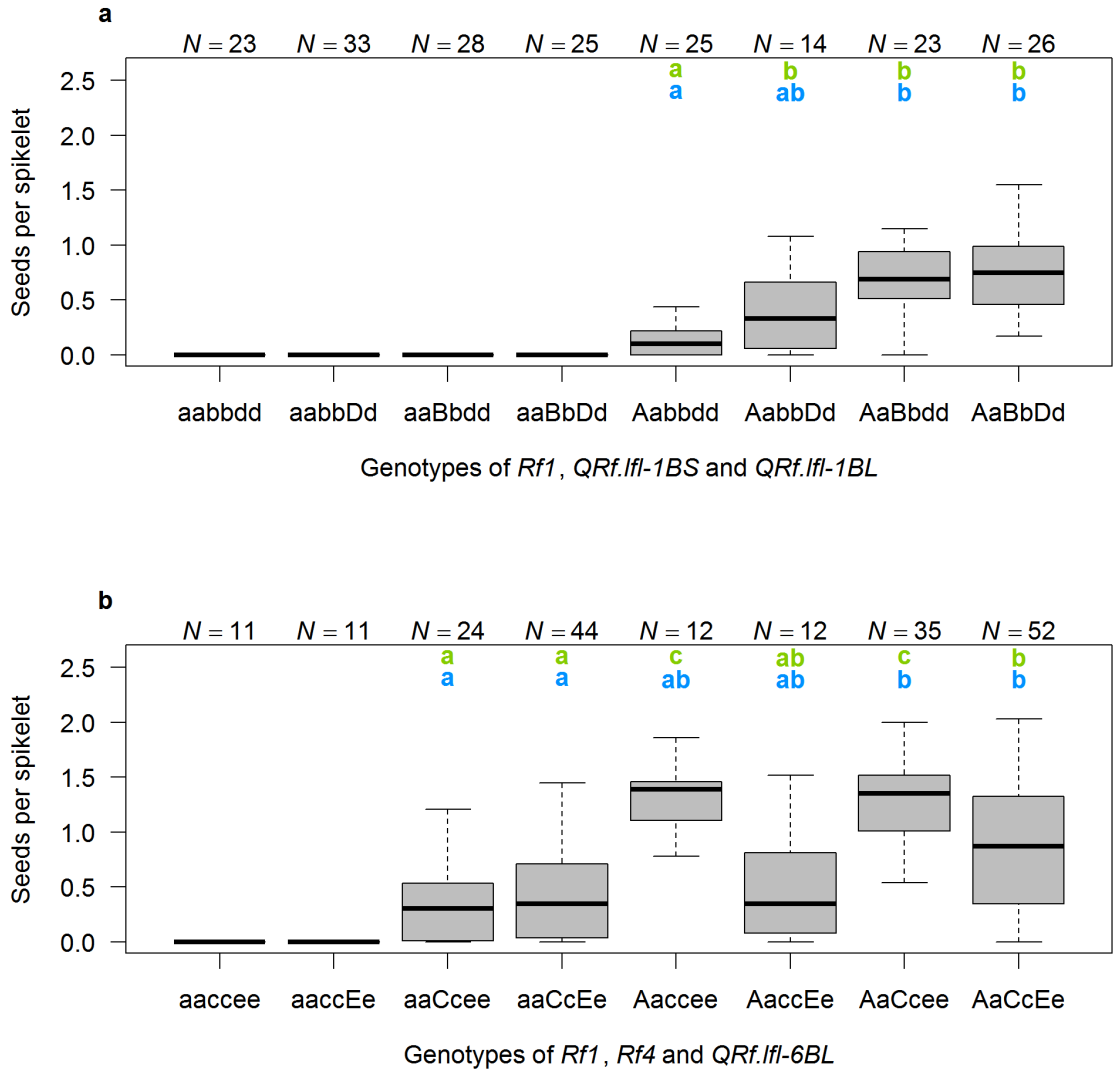

Figure 11 Seed set of PopR3 (**a**) and PopR113 (**b**) ordered by QTL genotype classes, additionally considering *QRf.lfl-1BL* and *QRf.lfl-6BL*, which were only detected by iQTLm analysis in the respective population. The letters A/a, B/b, C/c, D/d and E/e at the bottom of the subfigures represent *Rf1*, *QRf.lfl-1BS*, *Rf4*, *QRf.lfl-1BL* and *QRf.lfl-6BL*, respectively. Lower-case and capital letters refer to the alleles of Sperber and R3/R113, respectively. Different letters at the top of the subfigures denote significant differences for expressivity (green) and penetrance (blue)

Table 9 Explained phenotypic variance (R^2), estimated regression coefficients (β; seeds per spikelet) and P values for detected QTL and interactions. Alternative models were fitted additionally considering *QRf.lfl-1BL* and *QRf.lfl-6BL*, which were only detected using iQTLm

Population	Genetic factor	Chromosome	R^2	β	P
PopR3	*Rf1*	1AS	0.56	0.48	< 0.001
	QRf.lfl-1BS	1BS	0.21	0.26	< 0.001
	Rf1×*QRf.lfl-1BS*		0.11	0.51	< 0.001
	Full model		0.68		
	Rf1	1AS	0.57	0.49	< 0.001
	QRf.lfl-1BS	1BS	0.18	0.24	< 0.001
	QRf.lfl-1BL	1BL	0.03	0.09	0.022
	Rf1×*QRf.lfl-1BS*		0.10	0.48	< 0.001
	Rf1×*QRf.lfl-1BL*		0.01	0.18	0.011
	Full model		0.71		
PopR113	*Rf1*	1AS	0.33	0.75	< 0.001
	Rf4	6BS	0.05	0.28	<0.001
	Rf1×*Rf4*		0.01	-0.30	0.042
	Full model		0.37		
	Rf1	1AS	0.37	0.78	< 0.001
	Rf4	6BS	0.06	0.30	< 0.001
	QRf.lfl-6BL	6BL	0.11	-0.29	< 0.001
	Rf1×*Rf4*		0.01	-0.23	0.049
	Rf1×*QRf.lfl-6BL*		0.05	-0.57	< 0.001
	Rf4×*QRf.lfl-6BL*		0.01	0.26	0.032
	Full model		0.48		
PopL19	*Rf1*	1AS	0.56	0.86	< 0.001
	QRf.lfl-1BS	1BS	0.17	0.36	< 0.001
	Rf1×*QRf.lfl-1BS*		0.08	0.72	< 0.001
	Full model		0.67		

3.2.3 Validation of *RFL-PPR* candidate genes

Basis for the identification of candidate genes were the ten *Rf-like pentatricopeptide repeat* (*RFL-PPR*) genes on chromosome 1AS that Zhou et al. (2017) detected by capture sequencing of Sperber, R3, R113 and L19. In this preliminary study, 193 unique SNPs were detected within nine of these candidate genes. For 48 of these SNPs, competitive allele-specific marker assays were developed. The marker assays were used to genotype PopR3, PopR113, PopL19 and their parental lines. For *RFL01*, *RFL02* and *RFL06*, at least one marker was polymorphic across all three populations. For *RFL03*, *RFL04*, *RFL08* and *RFL10*, markers were only polymorphic in PopR3. The number of polymorphic assays in PopR3, PopR113 and PopL19 is summarised in Table 10. A considerable number of marker assays could not be used for linkage mapping because SNP allele calling revealed only one genotype cluster. About 61% of these non-informative marker assays revealed both alleles for all individuals of a population, indicating a large proportion of multi-locus amplification.

Table 10 Number of polymorphic marker assays for *Rf1* candidate genes

Scaffold ID	*RFL-PPR* gene	Developed assays	Polymorphic assays		
			PopR3	PopR113	PopL19
scaffold44309	*RFL01*	6	2	2	2
	RFL02	6	1	1	1
scaffold83748	*RFL03*	5	1	0	0
	RFL04	4	3	0	0
	RFL06	6	1	1	1
	RFL07	6	0	0	0
scaffold150701	*RFL08*	5	2	0	0
	RFL09	5	0	0	0
	RFL10	5	4	0	0

Genotypic information of polymorphic markers within the same gene was merged to a single genotype representing the *RFL-PPR* gene. The polymorphic *RFL-PPR* genes were integrated into the linkage maps of PopR3, PopR113 and PopL19 and the consensus map. No recombination was observed between *RFL01* and *RFL02* in any of

the three populations. In PopR3, cosegregation was also observed between *RFL04* and *RFL06* and between *RFL08* and *RFL10*, confirming the proximity of these genes on the physical map. In the same population, two recombinations were observed between *RFL03* and the two cosegregating genes *RFL04* and *RFL06*, and, as a consequence, *RFL03* was positioned distal to these two genes. The inconsistency between the *RFL-PPR* gene nomenclature and the order of these genes on the genetic map may be explained by a false scaffold orientation. In fact, the order of *RFL03–RFL06* on the genetic map of PopR3 is in agreement with the RefSeq v1.0 (IWGSC 2018). QTL analysis using the two-part model revealed that *RFL01* and *RFL02* were located inside the 1.5-LOD support intervals for *Rf1* in all three populations (Figure 12). The support intervals for *Rf1* in PopR3, PopR113 and PopL19 spanned 3.0, 3.1 and 2.0 cM, respectively. In PopR3 and PopR113, *RFL01* and *RFL02* were mapped to the locus with the maximum LOD score. Performing iQTLm analysis across the three populations located *RFL01* and *RFL02* in the 1.5-LOD support interval for *Rf1*, which spanned 0.2 cM on the consensus map. Besides these two genes, only the array-derived SNP marker *AX-94682405* was located within the *Rf1* support interval. Taken together, *RFL01* and *RFL02* were mapped to the *Rf1* locus in all populations with both methods.

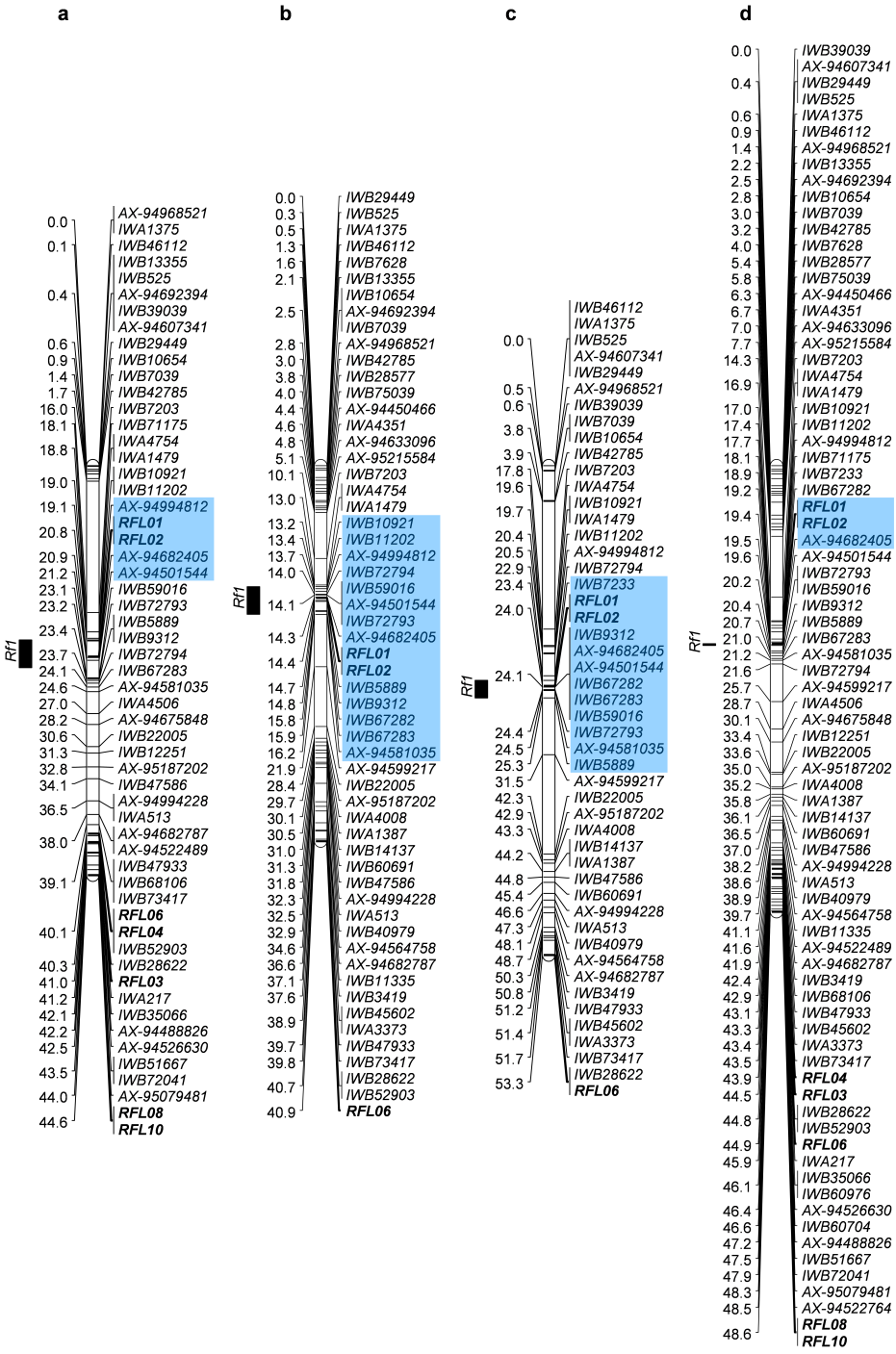

Figure 12 Positions of *restorer-of-fertility-like pentatricopeptide repeat* (*RFL-PPR*) genes on chromosome-1A linkage maps derived from PopR3 (**a**), PopR113 (**b**) and PopL19 (**c**) and on the consensus map (**d**). Black bars and blue boxes depict 1.5-LOD support intervals for *Rf1*, which were estimated using the two-part model within populations (**a**–**c**) and iQTLm across populations (**d**). Map positions are given in cM

3.3 Analysis of the distribution of *Rf1*, *QRf.lfl-1BS* and *Rf4*

To determine whether *Rf1* was exclusively introgressed from *Triticum timopheevii* or if it independently exists in lines of common wheat (*Triticum aestivum* ssp. *aestivum*), such as R113 and L19, the SNP markers in the genomic target region were analysed in all available *Rf1* donor accessions, 17 *T. timopheevii* accessions and 524 common wheat breeding lines. The distribution of *Rf1* was evaluated based on a marker haplotype comprising the 14 SNP markers that were located in any of the 1.5-LOD support intervals for *Rf1* together with three unmapped duplicate markers. The marker haplotype covered a region spanning 8.17 mega base pairs (Mbp) in the reference sequence of Chinese Spring (IWGSC 2018). For the analysis of the *Rf1* region in common wheat, 17 breeding lines were excluded from the analysis because of missing marker genotypes. Analysing the markers in the remaining 507 common wheat lines revealed minor allele frequencies (MAFs) ranging from 0.19 to 0.49. Hierarchical cluster analysis of the *Rf1* marker haplotypes of *T. timopheevii* accessions, the breeding lines and the *Rf1* donor accessions resulted in 22 groups, with each group representing a unique haplotype. The R3 haplotype was identical to those of the *Rf1* donor line R4 and 13 *T. timopheevii* accessions. The haplotypes of R113 and L19 were grouped with those of two breeding lines. For the analysed markers, R1 was identical to 216 breeding lines, and R2 was grouped together with R5. Cluster analysis further revealed that R113 and L19 were similar to *T. timopheevii* for the genomic region. For $k \leq 11$ groups, their haplotypes were assigned to the same group as 15 *T. timopheevii* accessions, R3 and 24 breeding lines. The results of the cluster analysis are summarised in Figure 13. By comparing the genetic distance between the haplotypes of R113/L19 and *T. timopheevii* ($\overline{d} = 0.31$) to the genetic distance between the haplotypes of R113/L19 and the breeding lines ($\overline{d} = 0.61$), it could be confirmed that, for the analysed marker haplotype, R113 and L19 were genetically more similar to *T. timopheevii* ($P < 0.001$).

SNP marker haplotypes of *Rf1* donor lines, *T. timopheevii* accessions and common wheat lines were also employed to evaluate whether the modifier locus *QRf.lfl-1BS* and the restorer locus *Rf4* were introgressed from *T. timopheevii*. For the analysis of the *QRf.lfl-1BS* region, a haplotype was constructed from 30 SNP markers located within the 1.5-LOD support intervals in PopR3 and PopL19, including also duplicate markers.

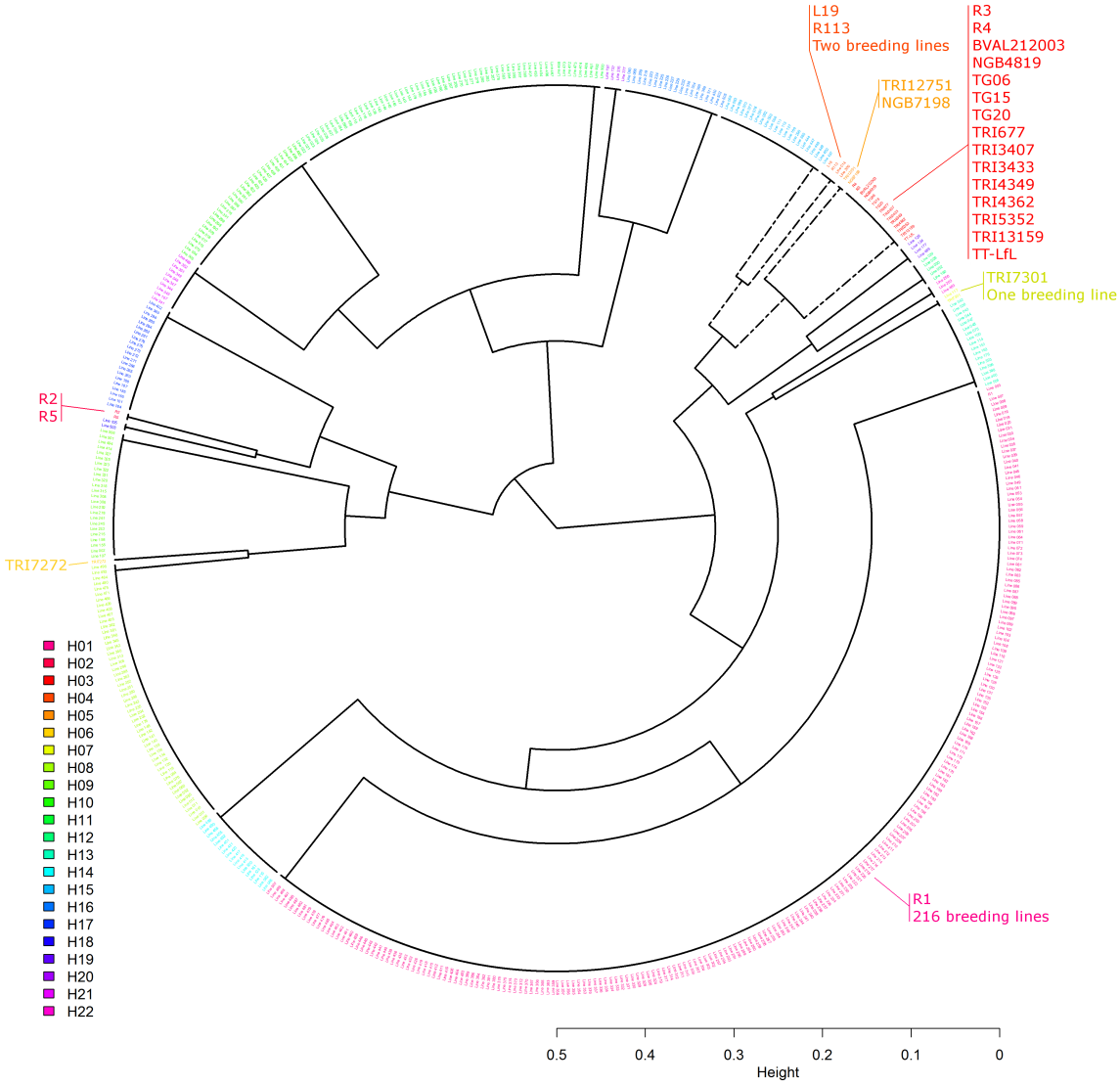

Figure 13 Phylogram for the *Rf1* marker haplotypes of all known *Rf1* donor lines, 17 *Triticum timopheevii* accessions and 507 common wheat breeding lines. Lines under the dashed branches belonged to the same group if the number of groups was restricted to $k \leq 11$

In the RefSeq v1.0, the 30 markers spanned a region of 19.9 Mbp (IWGSC 2018). Hierarchical cluster analysis resulted in 39 unique haplotypes among the two species. The haplotype of R3, which was associated with the fertility-enhancing modifier allele, was identical to the R4 haplotype. For $k \leq 10$ groups, the haplotypes of R3 and R4 formed a group with 16 *T. timopheevii* haplotypes (Figure A12). Based on the 30 SNPs, the mean genetic distance between R3 and all *T. timopheevii* accessions ($\overline{d} = 0.38$) was significantly smaller than the mean genetic distance between R3 and all common wheat lines ($\overline{d} = 0.90$; $P < 0.001$). L19, which was also identified as a carrier of a favourable *QRf.lfl-1BS* allele, exhibited an identical haplotype as R113. In contrast to R3, the haplotype of R113 and L19 was separated from the *T. timopheevii* haplotypes for $k \geq 2$ groups. Comparing the mean genetic distances corroborated that the R113/L19 haplotype was genetically more similar to the haplotypes of common wheat ($\overline{d} = 0.66$) than to those of *T. timopheevii* ($\overline{d} = 0.70$; $P < 0.001$).

To evaluate whether the restorer locus *Rf4* was introgressed from *T. timopheevii*, SNP markers in the *Rf4* support interval were used to analyse the panel comprising *Rf1* donor accessions, *T. timopheevii* and common wheat breeding lines. As a result of the low recombination rate observed on chromosome 6B, the 263 SNP markers located within the *Rf4* support interval (including duplicates) covered 690.5 Mbp of the chromosome (IWGSC 2018). Hierarchical cluster analysis revealed 335 unique haplotypes. The R113 haplotype, which was found to restore fertility in PopR113, was unique but for $k \leq 182$ groups, R113 was grouped with R1, R2, R5 and 16 *T. timopheevii* accessions (Figure A13). A comparison of mean genetic distances confirmed that, for this genomic region of chromosome 6B, R113 was genetically more similar to *T. timopheevii* ($\overline{d} = 0.20$) than to common wheat ($\overline{d} = 0.84$; $P < 0.001$).

3.4 Genetic mapping of *Rf3*

3.4.1 Marker selection and linkage mapping

The fertility-restoring locus *Rf3* was initially mapped in the BC_1 population PopPrimepi. In contrast to *Rf1*, which was mapped by interval mapping, *Rf3* was mapped using the discrete phenotypes (sterile or fertile) as a proxy for the *Rf3* genotype, thereby assuming a monogenic inheritance of fertility restoration. Five of the tested simple sequence repeat (SSR) markers on chromosome 1BS were polymorphic between Sperber and Primepi and were used for linkage mapping: *Xbarc8*, *Xbarc128*, *Xgwm264*, *Xwmc406* and *Xwmc798*. The partial linkage map for chromosome 1BS comprising *Rf3* and these five SSR markers spanned 33.7 cM (map not shown). *Rf3* was flanked by the marker loci *Xbarc128* and *Xwmc406*, located 7.2 cM distal and 14.5 cM proximal to *Rf3*, respectively. In order to enrich this map interval with molecular markers, 40 individuals of PopPrimepi that were recombinant between the flanking SSR loci were genotyped using the 15k Infinium® iSelect® SNP array. PCoA of the whole-genome SNP data revealed no population structure within the 40 individuals (Figures A2d and A3). After marker selection, 2688 polymorphic SNP markers remained for linkage mapping. Linkage group construction revealed that three SNP markers were linked to *Rf3*, namely *IWB14060*, *IWB72107* and *IWB73447*. The segregation of these markers did not deviate from the expected 1:1 ratio ($P = 1$). The SNPs *IWB14060* and *IWB72107* were used to develop cleaved amplified polymorphic sequence (CAPS) assays, designated *CAPS_IWB14060* and *CAPS_IWB72107*, respectively. The specifications for these CAPS assays is summarised in Table A2. CAPS assay development for *IWB73447* was not successful because the SNP-surrounding sequence did not match the recognition site for a restriction en-

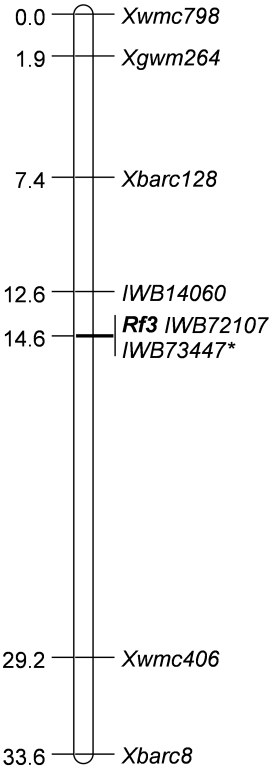

Figure 14 Position of *Rf3* on the chromosome-1B linkage map derived from PopPrimepi. Map distances are given in cM. * The position of *IWB73447* was based on 40 individuals

zyme. All individuals of PopPrimepi were genotyped with the developed CAPS markers and the original array-derived SNP genotypes of the 40 individuals were reproduced. Furthermore, the CAPS markers confirmed the assumed genotypes of the plants that were not genotyped with the SNP array. The linkage map derived from the SSR and SNP genotypes of all individuals spanned 33.6 cM. Whereas *IWB14060* was mapped 2.0 cM distal to *Rf3*, *IWB72107* cosegregated with the restorer locus (Figure 14).

3.4.2 Marker validation

The linkage between *CAPS_IWB72107* was subsequently validated in the mapping populations involving PR143, Badenkrone, Badenstern, Holstenkorn and Schwabenspelz, which were known as potential carriers of *Rf3*. In these populations, the CAPS marker was mapped 0.6, 0.4, 2.3, 1.1 and 1.2 cM apart from *Rf3*, with one recombination in each of PopPR143, PopBadenkrone, PopHolstenkorn, PopSchwabenspelz and two recombinations in PopBadenstern. All six plants that exhibited a recombination between the assumed *Rf3* genotype and *CAPS_IWB72107* were sterile. This CAPS marker was also employed to validate the results of Panayotov et al. (1975), who identified Palmaress as a carrier of restoring alleles on chromosomes 4A, 5B and 7D. Interestingly, in PopPalmaress, strong linkage (2.2 cM; two recombinations) between *CAPS_IWB72107* and a restorer gene on chromosome 1BS was observed, probably representing *Rf3*. The results of the marker validation in these six mapping populations indicated that fertility restoration was exclusively controlled by the *Rf3* locus. Furthermore, the findings confirmed the close linkage between *Rf3* and *CAPS_IWB72107*, which was previously observed in PopPrimepi. *CAPS_IWB72107* did not deviate from the expected segregation patterns in any of the six validation populations ($P \geq 0.14$). To assess the diagnostic ability of *CAPS_IWB72107*, the marker was further validated in a diversity panel comprising common wheat and European spelt (*Triticum aestivum* ssp. *spelta*) accessions with previous knowledge about their putative *Rf3* genotypes. The marker allele of Primepi was set as the reference for the prediction of a fertility-restoring *Rf3* allele. The *CAPS_IWB72107* genotype predicted the assumed genotype at the *Rf3* locus in all 29 analysed common wheat accessions (Table 11). Testcrosses between 30 spelt accessions and CMS-Sperber revealed that 20 spelt lines held the capacity to restore fertility and were therefore assumed to carry *Rf3*. The five restoring spelt accessions

Ceralio, Tauro, Titan, Zollernspelz and Züricher Oberländer Rotkorn were excluded from the diversity panel as there was no record that they belong to the *duhamelianum* variety, which is the only spelt variety known to carry a restoring *Rf3* allele (Tahir and Tsunewaki 1969). *CAPS_IWB72107* predicted the assumed *Rf3* allele in 23 of the remaining 25 spelt cultivars. Only the two spelt accessions Bauländer Spelz and Grey did not reveal restoration capacity despite carrying the marker allele associated with fertility restoration. Besides the two alleles that were previously observed in the mapping populations (Figure A11), *CAPS_IWB72107* additionally revealed a null allele for the negative controls 444-74, 539-74, 563-76, R1, R3, Samir and Sirino. Overall, these results indicate that *IWB72107* and the corresponding CAPS marker are suitable for marker-assisted selection and related applications.

Table 11 *Rf3* alleles (+ fertility-restoring allele; − non-restoring allele) and observed marker alleles for *CAPS_IWB72107* across the lines of the diversity panel. Superscript letters a and b denote fertility and sterility of hybrids from testcrosses with CMS-Sperber, respectively

Accession	Taxon	*Rf3*	*CAPS_IWB72107*
Minister	Common wheat	+	A
PR143	Common wheat	$+^{a}$	A
Primepi	Common wheat	$+^{a}$	A
Prof. Marchal	Common wheat	+	A
Alkor	European spelt (var. *duhamelianum*)	$+^{a}$	A
Altgold	European spelt (var. *duhamelianum*)	$+^{a}$	A
Badenkrone	European spelt	$+^{a}$	A
Badenstern	European spelt	$+^{a}$	A
Cosmos	European spelt (var. *duhamelianum*)	$+^{a}$	A
Divimar	European spelt (var. *duhamelianum*)	$+^{a}$	A
Ebners Rotkorn	European spelt (var. *duhamelianum*)	$+^{a}$	A
Epanis	European spelt (var. *duhamelianum*)	$+^{a}$	A
Holstenkorn	European spelt (var. *duhamelianum*)	$+^{a}$	A
Ostro	European spelt (var. *duhamelianum*)	$+^{a}$	A
Poeme	European spelt (var. *duhamelianum*)	$+^{a}$	A
Renval	European spelt (var. *duhamelianum*)	$+^{a}$	A
Rouquin	European spelt (var. *duhamelianum*)	$+^{a}$	A
Schwabenkorn	European spelt (var. *duhamelianum*)	$+^{a}$	A
Schwabenspelz	European spelt (var. *duhamelianum*)	$+^{a}$	A
10-77	Common wheat (maintainer)	−	G

Table 11 continued

Accession	Taxon	*Rf3*	*CAPS_IWB72107*
1-78	Common wheat (maintainer)	–	G
107-77	Common wheat (maintainer)	–	G
31-77	Common wheat (maintainer)	–	G
406-76	Common wheat (maintainer)	–[b]	G
435-76	Common wheat (maintainer)	–	G
441-78	Common wheat (maintainer)	–	G
444-74	Common wheat (maintainer)	–	Null
463-77	Common wheat (maintainer)	–[b]	G
50-74	Common wheat (maintainer)	–[b]	G
539-74	Common wheat (maintainer)	–	Null
563-76	Common wheat (maintainer)	–[b]	Null
609-73	Common wheat (maintainer)	–[b]	G
629-77	Common wheat (maintainer)	–	G
82-77	Common wheat (maintainer)	–	G
Bert	Common wheat (maintainer)	–	G
Granit	Common wheat (maintainer)	–	G
Mission	Common wheat (maintainer)	–[b]	G
Navojoa	Common wheat (maintainer)	–	G
PR189	Common wheat	–[a]	G
R1	Common wheat	–	Null
R3	Common wheat	–[a]	Null
Severin	Common wheat (maintainer)	–	G
Sperber	Common wheat (maintainer)	–[b]	G
Vorobey	Common wheat (maintainer)	–	G
Albin	European spelt	–[b]	G
Badengold	European spelt	–[b]	G
Bauländer Spelz	European spelt (var. *duhamelianum*)	–[b]	A
Filderweiss	European spelt	–[b]	G
Franckenkorn	European spelt (var. *duhamelianum*)	–[b]	G
Grey	European spelt	–[b]	A
Hercule	European spelt (var. *album*)	–[b]	G
Oberkulmer Rotkorn	European spelt (var. *duhamelianum*)	–[b]	G
Samir	European spelt	–[b]	Null
Sirino	European spelt	–[b]	Null

3.4.3 Validation of *RFL-PPR* candidate genes

Basis for the validation of *Rf3* candidate genes were 87 SNPs detected across 13 *RFL-PPR* genes on chromosome 1BS (Zhou et al. 2017). For 48 of these 87 SNPs, competitive allele-specific marker assays were developed and used to genotype PopPrimepi, PopBadenkrone, the corresponding parental lines and the four individuals of PopPR143, PopBadenstern and PopSchwabenspelz that exhibited a putative recombination between *Rf3* and the marker *CAPS_IWB72107*. Markers were found to be polymorphic in PopPrimepi and PopBadenkrone for *RFL12*, *RFL13*, *RFL15* and *RFL16*. The remaining marker assays were only polymorphic in PopPrimepi, with the exception of the two *RFL11* SNPs, which were monomorphic in both populations. The number of assays polymorphic in PopPrimepi and PopBadenkrone is summarised in Table 12. For about 56% of the non-informative marker assays, both alleles were observed for all individuals of a population, indicating a considerable proportion of multi-locus amplification.

Table 12 Number of polymorphic marker assays for *Rf3* candidate genes

Scaffold ID	*RFL-PPR* gene	Developed assays	Polymorphic assays	
			PopPrimepi	PopBadenkrone
scaffold35219	*RFL11*	2	0	0
	RFL12	7	4	2
	RFL13	7	5	5
	RFL14	2	1	0
	RFL15	5	2	2
scaffold5117	*RFL16*	3	1	1
scaffold108702	*RFL17*	2	1	0
	RFL18	4	2	0
	RFL20	3	3	0
	RFL21	7	3	0
	RFL22	4	4	0
	RFL23	1	1	0
scaffold77575	*RFL25*	1	1	0

Genotypic information of polymorphic markers within the same *RFL-PPR* gene was combined to construct a single genotype representing the *RFL-PPR* gene. These genotypes were integrated into the linkage maps derived from PopPrimepi and PopBadenkrone. In PopPrimepi, *Rf3* cosegregated with *RFL12–RFL16*. The remaining *RFL-PPR* genes were mapped about 17.2 cM proximal to *Rf3*. No recombination was observed between them (Figure 15). In PopBadenkrone, *RFL12*, *RFL13* and *RFL15* cosegregated with *CAPS_IWB72107* and were mapped 0.4 cM (one recombination) apart from *Rf3*. A further recombination was observed between *Rf3* and *RFL16*, which was mapped 0.7 cM apart from *Rf3*. For the three plants of PopPR143 and PopBadenstern that were sterile although *CAPS_IWB72107* indicated a fertility-restoring *Rf3* allele, the marker assays for *RFL12–RFL15* also exhibited the marker alleles of the respective *Rf3* donor. In contrast, the sterile recombinant individual of PopSchwabenspelz carried the alleles of Sperber for these candidate genes.

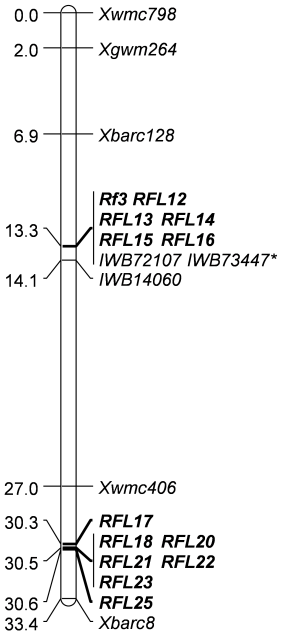

Figure 15 Positions of *RFL-PPR* genes on the chromosome-1B linkage map derived from PopPrimepi. Map positions are given in cM

3.5 Analysis of the distribution of *Rf3*

To estimate the distribution of *Rf3* alleles, 524 common wheat breeding lines, 30 spelt cultivars, 17 *T. timopheevii* accessions and the *Rf3* donor Primepi were analysed using the SNP marker *IWB72107* and the thereof derived CAPS marker, which revealed diagnostic potential in the diversity panel. The common wheat lines, the *T. timopheevii* accessions and Primepi were additionally analysed using the SNP markers *IWB14060* and *IWB73447*, which were located distal from *IWB72107* in the RefSeq v1.0 (IWGSC 2018). Four breeding lines were excluded from the analysis because of missing marker genotypes. In the common wheat panel, 8.9% of the lines carried the *IWB72107* allele that was associated with the restoring *Rf3* allele. The closely linked markers *IWB14060* and *IWB73447* exhibited MAFs of 0.40 and 0.42, respectively. The Primepi marker

haplotype was observed in the majority of the breeding lines that carried the favourable *IWB72107* allele; only four common wheat lines exhibited the minor *IWB72107* allele in another combination (Figure A14). PCoA of the whole-genome SNP data revealed that *IWB72107* alleles were not associated with any of the first four principal coordinates ($P \geq 0.11$), which jointly explained 13.5% of the genotypic variation, thereby indicating a homogeneous distribution of *IWB72107* alleles. The first two principal coordinates are illustrated in Figure 16a. Analysing the restoration capacity and *IWB72107* marker genotypes of 30 spelt cultivars indicated that 20 cultivars (66.7%) carried a fertility-restoring allele at the *Rf3* locus. The five fertility-restoring spelt accessions that were discarded from the diversity panel were included for this analysis because the marker *IWB72107* predicted the restoring *Rf3* allele for all of them. As the spelt cultivars Bauländer Spelz and Grey could not restore fertility in the testcrosses, it was assumed that these cultivars did not carry a fertility-restoring *Rf3* allele. The expected *Rf3* genotypes of the spelt cultivars were not significantly correlated to any of the first four principal coordinates ($P \geq 0.20$), which together explained 40.8% of the genotypic variation (Figure 16b). This indicates that *Rf3* alleles were homogeneously distributed within the analysed spelt population. Comparing the mean genetic distance between spelt cultivars and common wheat lines revealed no significant difference between spelt lines for which a restoring ($\overline{d} = 0.90$) or a non-restoring ($\overline{d} = 0.87$) *Rf3* genotype was assumed ($P = 0.054$; Figure A15). This formally suggests that the *Rf3* genotype of the analysed spelt cultivars may be independent from their kinship to common wheat. Among the 17 *T. timopheevii* accessions, only TRI7301 exhibited the *IWB72107* allele associated with fertility restoration. The haplotype of this accession was shared only by one common wheat line. All other *T. timopheevii* lines exhibited a haplotype that was also observed for 183 common wheat lines.

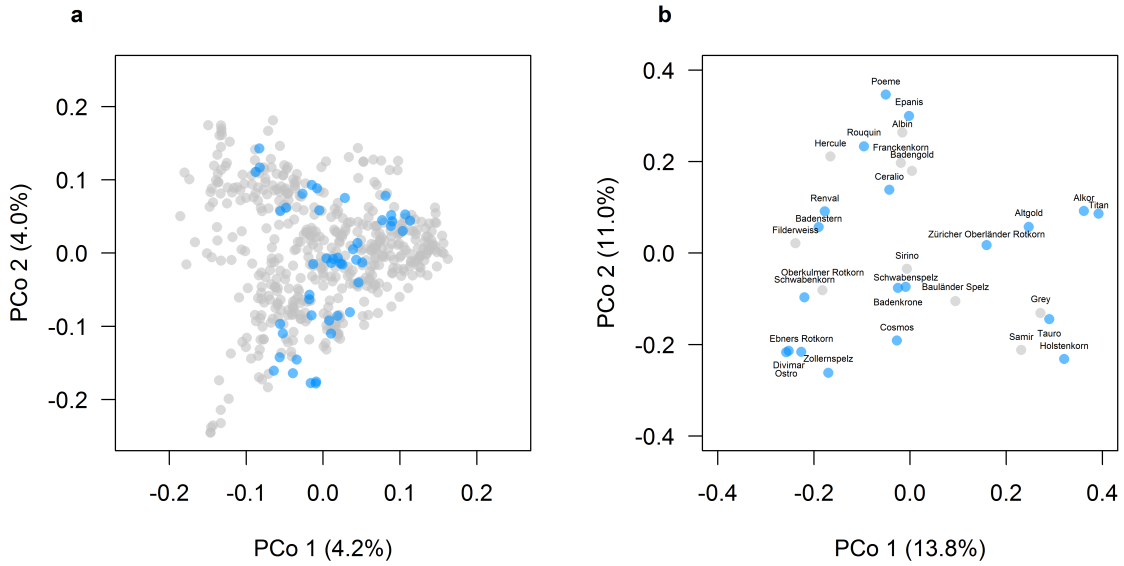

Figure 16 Principal coordinate analyses (PCoA) of a set of 520 German common wheat breeding lines (**a**) and 30 European spelt cultivars (**b**). Blue dots represent lines for which a restoring *Rf3* allele was assumed

4 Discussion

4.1 Genetic architecture of fertility restoration

4.1.1 Identity of restorer and modifier loci detected on homoeologous group 1 chromosomes

Previous studies identified chromosome 1A as an important component for fertility restoration in several restorer accessions. It was assumed that a single *restorer-of-fertility* (*Rf*) locus, namely *Rf1*, caused fertility restoration in these lines (Yen et al. 1969; Odenbach 1970; Bahl and Maan 1973; Maan et al. 1984; Maan 1985; Du et al. 1991). However, these investigations were restricted to monosomic analysis and it has not been evaluated whether or not the restoration capacity of chromosome 1A can be explained by *Rf1* in all described accessions. Analogously, monosomic studies observed restoration capacity on chromosome 1B in a variety of European spelt (*Triticum aestivum* ssp. *spelta* var. *duhamelianum*) and common wheat (*Triticum aestivum* ssp. *aestivum*) cultivars, and it was suggested that the restorer locus *Rf3* caused restoration capacity in both subspecies (Bahl and Maan 1973; Kučera 1982; Ma and Sorrells 1995).

In the present study, the putative *Rf1* locus was mapped by quantitative trait locus (QTL) analysis in populations derived from the restorer accessions R3, R113 and L19. In all three populations, QTL analysis employing the two-part model revealed that the same single-nucleotide polymorphism (SNP) marker (*AX-94682405*) exhibited the highest logarithm of the odds (LOD) score. QTL analysis across the three populations resulted in a single QTL on chromosome 1A, with a support interval comprising only the mentioned peak marker. Hence, at the given map resolution, it was confirmed that all three accessions carried the same restorer locus (*Rf1*). Linkage mapping in PopPrimepi revealed that the SNP marker *IWB72107* was closely linked to the restorer locus on chromosome 1BS, which was assumed to be *Rf3*. Strong linkage was also observed between *IWB72107* and the *Rf3* locus in multiple European spelt mapping populations. The same marker was shown to hold potential for the prediction of fertility restoration in a diverse panel of common wheat and spelt accessions. These findings corroborate the hypothesis that, in both subspecies, *Rf3* is the only hitherto described restorer

locus on chromosome 1BS against G-type cytoplasmic male sterility (CMS). The same chromosome arm was shown to harbour the locus *QRf.lfl-1BS*, which modified the effect of *Rf1* in the two populations PopR3 and PopL19. The *QRf.lfl-1BS* support intervals assessed by the two-part model included four markers common to both populations. In the RefSeq v1.0, the markers with the highest LOD scores in PopR3 and PopL19 were located 5.17 mega base pairs (Mbp) apart from each other (IWGSC 2018). This gap may be explained by the limited marker density in the region distal from the peak marker on the PopL19-derived map. The QTL analysis across populations detected a single QTL on the consensus map, supporting the assumption that the modifier loci are the same in the two populations. Interestingly, the two peak markers for *QRf.lfl-1BS* were located 3.74 Mbp and 1.44 Mbp away from *IWB72107*, which was found to be very closely linked to the *Rf3* locus. On the consensus map, the peak marker for *QRf.lfl-1BS* was located 2.86 Mbp apart from the *Rf3*-linked marker. Thus, it can be concluded that *QRf.lfl-1BS*, which modified the effect of *Rf1*, is probably tightly linked to *Rf3*. The results of the present study do not allow inferences about a possible homology between *QRf.lfl-1BS* and *Rf3*. However, in the case of homology, different alleles may be an explanation for the two distinct effects (modification and restoration) observed across several populations. To evaluate whether *Rf3* or *QRf.lfl-1BS* may be homoeologous to *Rf1* on chromosome 1AS, sequences of the 17 markers located in the *Rf1* support intervals were searched against the reference sequence of chromosome 1B (IWGSC 2018) using the Basic Local Alignment Search Tool (BLAST). For 12 of 17 hits on chromosome 1BS, the physical positions were collinear with the physical SNP positions on chromosome 1AS (data not shown). The hit on chromosome 1BS for the *Rf1* peak marker *AX-94682405* (17.79 Mbp) was located 2.79 Mbp distal from *IWB72107*, which was closely linked to *Rf3*, and 0.06–4.23 Mbp apart from the *QRf.lfl-1BS* peak markers. These findings indicate that the *Rf1* locus may be homoeologous to *QRf.lfl-1BS* or *Rf3*, but further research is required to validate this hypothesis. Moreover, Tsunewaki (2015) and Hohn and Lukaszewski (2016) reported a locus designated Rf_{multi} on chromosome 1BS, which restored male fertility in CMS lines with the plasmon of *Aegilops kotschyi*, *Aegilops mutica* and *Aegilops uniaristata* but not in G-type CMS lines. However, the relationship between *Rf1*, *Rf3*, *QRf.lfl-1BS* and Rf_{multi} has yet to be investigated.

4.1.2 Detection of restorer and modifier loci across studies

The analysed *Rf* gene resources R3, R113, L19, Primepi, Palmaress and European spelt have been the subject of interest in several genetic studies for the last six decades. In this section, the results of the preceding investigations are compared with the loci detected in the present study. For R3, the genetic architecture of restoration capacity was described by two studies employing monosomic analysis. Robertson and Curtis (1967) detected a fertility-restoring effect on chromosome 1A (*Rf1*) and modifier effects on chromosomes 1B, 2A, 3D, 6A and 6B. In the second study (Bahl and Maan 1973), R3 was found to carry restorer genes on chromosomes 1A (*Rf1*) and 7D (*Rf2*). Furthermore, monosomic condition of the chromosomes 4A, 5A and 5B inhibited fertility restoration. In contrast, the present study detected only three fertility-affecting loci in the R3-derived population: *Rf1* on chromosome 1AS and the modifier loci *QRf.lfl-1BS* and *QRf.lfl-1BL*. Chromosomes and loci influencing the restoration capacity of R113 were evaluated in two investigations. In a monosomic analysis of R113, Maan et al. (1984) concluded that chromosomes 1A and 6B harboured the fertility-restoring genes *Rf1* and *Rf4*. Several other chromosomes were found to modify fertility restoration, but most of the critical chromosomes could not be validated across environments. Analysing the restoration capacity of R113 by single marker regression revealed a restorer locus on chromosome 6BS and modifier loci on chromosomes 5AL, 5DS and 7BS (Ma and Sorrells 1995). In the present study, QTL analysis in PopR113 detected the fertility-restoring loci *Rf1* (chromosome 1AS) and *Rf4* (chromosome 6BS) and the modifier locus *QRf.lfl-6BL*. A previous monosomic analysis of L19 concluded that chromosome 1A carried *Rf1*, and that chromosomes 2A, 2B, 3A, 3B, 3D, 4B, 5B, 6A, 6B and 6D modified fertility restoration (Du et al. 1991). In contrast, the present study detected only *Rf1* and *QRf.lfl-1BS* in the population derived from L19. In PopPrimepi, fertility restoration was controlled by the single restorer locus *Rf3* on chromosome 1BS. This is in accordance with the findings of Ingold (1968), who reported a monogenic inheritance in a segregating population involving Primepi. However, Miller et al. (1974) analysed Primepi in a F_2 population and did not observe deviations from a 9:6:1 ratio of fertile, partially fertile and sterile individuals, indicating a digenic inheritance. The hypothesis of two restorer loci was corroborated by Bahl and Maan (1973), who found that genes on chromosomes 1BS and 5D of Primepi restored

fertility. The common wheat cultivar Palmaress has been reported to carry unnamed restorer loci on chromosomes 4A, 5B and 7D (Panayotov et al. 1975; Kučera 1982). In contrast, PopPalmaress revealed no deviation from a 1:1 segregation pattern expected for a monogenic inheritance of fertility restoration. Linkage mapping clearly showed that the restoration capacity of Palmaress can be explained by *Rf3*. The current study further analysed four European spelt cultivars in segregating populations. In all four populations, fertility restoration was controlled by the single restorer locus *Rf3*. These results are in line with previous studies that identified the *duhamelianum* variety of European spelt as donor for *Rf3* (Tahir and Tsunewaki 1969, 1971; Kojima et al. 1997). Besides *Rf3*, Tahir and Tsunewaki (1969) additionally observed a weak supressing effect of chromosome 7D of this spelt variety.

Discrepancies pertaining to the detection of restorer and modifier loci between the present study and previous investigations have several possible reasons. A likely explanation may be environmental factors, which were shown to significantly affect fertility restoration in several experiments (Johnson et al. 1967; Robertson and Curtis 1967; Jošt 1982; Maan et al. 1984). The observation that the same restorer line exhibited different modifier loci across experiments may also be explained by the diverse male-sterile or monosomic parents used as crossing partners. These may have modifier alleles in common with the restorer parent or they may contribute new variation that modifies fertility restoration. The detection of restorer or modifier loci by means of monosomic analysis may also be biased due to unexpected transmission rates of monosomic gametes. Furthermore, monosomic analysis does not account for possible cumulative or cancelling effects of two or more loci on the same chromosome. To some degree, this drawback can be avoided by interval mapping, especially by methods such as composite interval mapping (Zeng 1994). However, the inconsistent detection of *QRf.lfl-6BL* in the present study may be a result of the same phenomenon. In this case, simple interval mapping using the two-part model may not have detected this locus as a consequence of the linkage between *QRf.lfl-6BL* and *Rf4* and the fact that the favourable alleles of both loci were in repulsion-phase linkage. The contrasting results for the detection of this locus using the two-part model and iQTLm showed that a follow-up iQTLm analysis across all linkage groups may have the potential to discover further QTL.

4.1.3 Effects of *Rf1* and *Rf3*

The effect of *Rf1* was estimated in three populations derived from the *Rf1* donor accessions R3, R113 and L19. Classifying the individuals on the basis of QTL genotype probabilites showed that *Rf1* exhibited incomplete penetrance in all three populations. Besides *Rf1*, QTL analysis detected the modifier loci *QRf.lfl-1BS*, *QRf.lfl-1BL*, *QRf.lfl-6BL* and the restorer locus *Rf4*. All three modifier loci significantly affected the expressivity of *Rf1*, and *QRf.lfl-1BS* was additionally shown to influence its penetrance. Interestingly, the effects of *Rf1* and *Rf4* were not additive. Instead, individuals carrying a combination of both restorer loci did not exhibit a superior fertility restoration compared to plants harbouring only *Rf1*. Thus, using the example of *Rf1*, the present study emphasises the role of epistatic interactions for CMS hybrid wheat. Whereas a single modifier locus can affect fertility restoration to a large extent, stacking of restorer loci does not necessarily lead to a significant improvement.

In the current study, the effect of *Rf3* was estimated as the mean seed set of the fertile fraction. This single-locus model may oversimplify the actual genetic architecture of the restoration capacity in the *Rf3* donor accessions. Unidentified modifier loci could have affected the expressivity of *Rf3* and maybe also its penetrance. Cosegregation between molecular markers and the *Rf3* locus in PopPrimepi indicated that *Rf3* probably exhibited complete penetrance in this population. However, the other six populations segregating for *Rf3* revealed putative recombinations between *Rf3* and the marker *IWB72107*, which showed diagnostic potential for *Rf3* in the diversity panel. Interestingly, seven of eight plants that indicated a recombination were sterile. The only fertile recombinant plant was observed in PopPalmaress. As this plant was cultivated in the same pot as another putatively recombinant individual, the discrepancy between the phenotypes and marker genotypes could possibly the result of a mix-up of the two samples. The fact that the individuals that indicated a recombination between *Rf3* and *IWB72107* were mainly sterile raises the possibility that this observation could also be the result of incomplete penetrance of *Rf3*. Hence, the effect of *Rf3* could be affected by epistasis, similar as it was observed for *Rf1*. However, to investigate possible modifier loci for *Rf3*, a follow-up experiment with whole-genome marker data of a population segregating for *Rf3* would be necessary. The incomplete penetrance that

was observed for *Rf1*, *Rf4* and that may also be the case for *Rf3* was already a subject of discussion in previous studies (Maan et al. 1984; Maan 1985). This phenomenon may not be restricted to the investigated restorer loci (Ma et al. 1995) and could also occur in CMS systems in other species (Dill et al. 1997; Tang et al. 2007; Mehrajuddin et al. 2013).

Given that all individuals within a population were exposed to the same environmental condition, genetic factors must be considered to explain the incomplete penetrance of fertility restoration observed in this study. Similar as *QRf.lfl-1BS* influenced the penetrance of *Rf1*, further undetected modifier loci may have supressed the effect of this restorer locus. Nuclear-encoded sterility genes introgressed from *Triticum timopheevii* are another possible explanation for the excess of sterile individuals observed in PopR3, PopR113 and PopL19. Introgressed genes supressing the fertility of wheat were previously reported in the form of gametocidal genes discovered in introgressed genomic regions of several *Aegilops* species (Endo 1982; Tsujimoto and Tsunewaki 1984). Nuclear-encoded sterility genes originated from mutations in the common wheat genome are unlikely to have caused the high number of sterile plants because none of the well documented accessions harbouring these genes appear in the pedigree of the accessions used in the present study (Pugsley and Oram 1959; Barlow and Driscoll 1981; Bing-Hua and Jing-Yang 1986; Zhou et al. 2008; Whitford et al. 2013). However, irrespective of the origin and action of such sterility genes, the observed ratios of sterile plants among the QTL genotype classes suggest that fertility was inherited in an oligogenic or polygenic manner. Loci explaining the excess of sterile plants could have remained undetected as a consequence of their minor effects and the experimental design.

4.1.4 *RFL-PPR* genes as candidates for *Rf1* and *Rf3*

Although the restorer loci *Rf1* and *Rf3* have been the subject of multiple genetic studies, no candidate genes have been suggested so far. Positional cloning approaches in other species revealed that, in most cases, fertility-restoring genes encoded for proteins of the pentatricopeptide repeat (PPR) family (Dahan and Mireau 2013; Gaborieau et al. 2016). In the present study, *Rf-like* (*RFL*)-*PPR* genes located on the target chromosomes by Zhou et al. (2017) were mapped in PopR3, PopR113, PopL19, PopPrimepi

and PopBadenkrone. The candidate genes *RFL01* and *RFL02* were located in the *Rf1* support interval in PopR3, PopR113, PopL19 and also in the multicross-analysis (Figure 12 and Table 10). Thus, the two *RFL-PPR* genes can be proposed as candidates for *Rf1*. All other *RFL-PPR* genes on this chromosome were mapped several centimorgans (cM) proximal to *Rf1* and can be ruled out as candidates for this locus. In PopPrimepi, the five *RFL-PPR* genes *RFL12–RFL16* cosegregated with the *Rf3* locus and the array-derived SNP markers *IWB72107* and *IWB73447* (Figure 15 and Table 12). The seven genes *RFL17–RFL25* were mapped 17 cM proximal to *Rf3* and should therefore not be considered as candidates for *Rf3*. In PopBadenkrone, *RFL12*, *RFL13* and *RFL15* also cosegregated with *CAPS_IWB72107*, but a putative recombination was observed between these loci and *Rf3*. In the same population, a second recombination was observed between the mentioned cosegregating loci and *RFL16*. A likely explanation for the putative recombinations between *Rf3* and the three candidate genes *RFL12*, *RFL13* and *RFL15* in PopBadenkrone could be incomplete penetrance, which was also observed for *Rf1*. However, the additional recombination between these candidate genes and *RFL16* cannot be explained by this phenomenon. Thus, *RFL16* can be ruled out as a candidate for *Rf3*. Markers for *RFL12–RFL15* were also used to genotype the four sterile plants of PopPR143, PopBadenstern and PopSchwabenspelz that exhibited a putative recombination between *Rf3* and the marker *CAPS_IWB72107*. For the three recombinant individuals of PopPR143 and PopBadenstern, *RFL12–RFL15* showed the same segregation pattern as *CAPS_IWB72107*. This could also be explained by incomplete penetrance, as suggested for one recombination in PopBadenkrone. In contrast, the sterile recombinant individual of PopSchwabenspelz harboured the alleles of Sperber for these candidate genes. This indicates that *RFL12–RFL15* could be linked more closely to *Rf3* than the mentioned marker. Polymorphic marker assays could not be developed for *RFL11*, although Zhou et al. (2017) detected two SNPs between Sperber and Primepi within this gene. As both marker assays for *RFL11* suggested heterozygosity for all analysed individuals, it is likely that the primer sequences were not genome-specific. Although *RFL11* could not be mapped genetically, it is known to be located on the same scaffold as *RFL12–RFL15* on the physical map. Hence, it can be proposed that the *RFL-PPR* genes *RFL11–RFL15* are potential candidates for *Rf3*.

For both studied restorer loci, *Rf1* and *Rf3*, *RFL-PPR* candidate genes are organised in clusters, each spanning a genomic interval of less than 1 Mbp. Such clusters of *PPR* genes were reported for *Rf* loci in several species (Gaborieau et al. 2016). In CMS petunia (*Petunia hybrida*), the fertility-restoring gene *Rf-PPR592*, which was shown to reduce the amount of the sterility-inducing PCF protein, is located adjacent to *Rf-PPR591*, and possibly a third *PPR* gene (Bentolila et al. 2002). In Boro-Taichung (BT)-type CMS rice (*Oryza sativa*), fertility restoration can be achieved by the *Rf1* locus on chromosome 10. It was demonstrated that this locus harboured nine duplicate *PPR* genes in close proximity, of which two, *Rf1a* and *Rf1b*, were able to restore fertility by blocking the production of the toxic ATP6-ORF79 protein. Although both *PPR* genes revealed restoration capacity, it was suggested that *Rf1a* and *Rf1b* have different mechanisms for silencing *atp6-orf79* mRNA (Akagi et al. 2004; Komori et al. 2004; Wang et al. 2006). Hu et al. (2012) demonstrated that *Rf1a* is identical to *Rf5*, which can restore fertility in Hong-Lian (HL)-type CMS rice. In contrast to *Rf1* in BT-type CMS, the protein encoded by *Rf5* does not bind to sterility-conferring *atp6-orfh79* transcripts. Instead, *Rf5* interacts with the glycine-rich protein GRP162 to inhibit the translation of *atp6-orfh79* (Hu et al. 2012). Subsequent studies revealed that this multigene cluster on chromosome 10 also harbours the restorer genes *Rf4* and *Rf98*, which have shown restoration capacity in wild-abortive (WA)-type CMS rice (Tang et al. 2014; Igarashi et al. 2016). Another restorer locus that harbours a cluster of *PPR* genes is *Rfo*, which is able to restore fertility in Ogura-type CMS radish (*Raphanus sativus*). Desloire et al. (2003) detected three *PPR* genes at this locus: the two functional genes *Ppr-A*, *Ppr-B* and the pseudogene *Ppr-C*. Analysis of transgenic rapeseed (*Brassica napus*) plants expressing *Ppr-A* and *Ppr-B* revealed that PPR-B inhibited the translation of *orf138* mRNA, which is associated with CMS, whereas PPR-A had no effect on the synthesis of the sterility-inducing protein (Uyttewaal et al. 2008). A further *PPR* gene cluster was discovered at the restorer locus *Rf5* on chromosome SBI-05 of sorghum (*Sorghum bicolor*). The *Rf5* locus, which enables fertility restoration in the A1 and the A2 cytoplasm, was delimited to an interval containing seven candidate genes of the *PPR* family (Jordan et al. 2011). Another example for the clustering of *PPR* genes at fertility-restoring loci is yellow monkeyflower (*Mimulus guttatus*). Male fertility in this species was shown to be maintained by two restorer loci, *Rf1* and *Rf2*, located within an interval of 1.3 cM. A dominant fertility-restoring allele at one of these

loci is sufficient to restore fertility. Although the functional genes for *Rf1* and *Rf2* have not been detected, the genomic region contains a cluster of 17 *PPR* candidate genes with high homology to *Rf* genes of petunia and rice (Fishman and Willis 2006; Barr and Fishman 2010).

These examples show that *RFL-PPR* gene clusters, as they were observed at the *Rf1* and *Rf3* loci, commonly occur in a wide range of plant species. The distribution of multiple *RFL-PPR* genes across the genome and the local clustering of these genes can be explained by gene duplication events such as tandem and segmental duplication, which also play an important role in the evolution of plant disease resistance genes (Leister 2004; Geddy and Brown 2007). Hierarchical cluster analysis of *RFL-PPR* gene sequences in diverse species demonstrated that all *RFL-PPR* genes have an ancient common origin and that these genes have extensively evolved since speciation. In contrast, non-*RFL-PPR* genes are highly conserved across plant species (Fujii et al. 2011). Geddy and Brown (2007) suggested that duplicated *RFL-PPR* genes are subject to diversifying selection, a process which is also known for the evolution of disease resistance genes (Michelmore and Meyers 1998). In the case of a single *PPR* gene interacting with a CMS-inducing gene, duplication of the fertility-restoring *PPR* gene enables functional mutations of the duplicates while the function of one of the duplicates is conserved by a strong purifying selective pressure. For the scenario of a newly evolved CMS-conferring gene, a positive selective pressure may favour mutations in one of the duplicate genes to adapt to the novel sterility gene. Fujii et al. (2011) suggested that the amino acid residues 1, 3, and 6 within the PPR motif were subject to diversifying selection and that these residues specify the RNA-binding capacity of PPR proteins. This process of diversifying selection may facilitate the coevolution of mitochondrial-encoded sterility proteins and nuclear-encoded fertility-restoring PPR (Rf-PPR) proteins.

The hypothesis of diversifying selection also provides a basis to discuss the function of the *RFL-PPR* genes that were mapped in the present study. As the restorer loci *Rf1* and *Rf3* cause fertility restoration in the same cytoplasmic background, it can be proposed that *Rf-PPR* genes at both loci may encode for proteins binding to the same CMS-inducing RNA. One or more *RFL-PPR* genes on chromosome 1BS may also cause the modifying effect observed at *QRf.lfl-1BS* but a model explaining the

interaction between *Rf1* and *QRf.lfl-1BS* on a molecular level is lacking. As studies in rice have demonstrated, two or more paralogous *Rf-PPR* genes at the same locus could counteract the effect of a single sterility gene (Akagi et al. 2004; Komori et al. 2004; Wang et al. 2006). In this case, varying restoration capacity of wheat lines carrying the same *Rf* locus could not only be explained by epistatic effects of modifier loci but also by different *Rf-PPR* haplotypes at the respective locus. However, a *Rf-PPR* allele leading to incomplete fertility restoration is unlikely to spread because, in a hermaphrodite species, the favourable allele is assumed to be fixed (Caruso et al. 2012). All *RFL-PPR* genes that have no function in the analysed G-type cytoplasm may act as restorer genes in cryptic CMS systems that could not be studied because the restoring alleles are fixed. However, as Fujii et al. (2011) also found *RFL-PPR* genes in a dioecious species, for which a CMS system could be of no advantage, it is possible that some *RFL-PPR* genes also play a general role in suppressing the expression of non-conserved, deleterious, mitochondrial genes not associated with CMS. Although the present study suggested candidate genes for *Rf1* and *Rf3*, the causal genes at these loci have yet to be identified. Due to close proximity of the clustered candidate genes, fine-mapping of *Rf1* and *Rf3* using a larger populations would be cumbersome and eventually not sufficient to separate all candidate genes by at least one recombination. Instead, the final validation of the *RFL-PPR* candidate genes could be realised by the transfer of putative *Rf-PPR* restorer alleles into CMS lines, an approach previously applied for the validation of *Rf* genes in petunia, rice and radish (Gaborieau et al. 2016). Alternatively, a genome editing method such as CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 (CRISPR-associated 9) could be exploited to produce knock-out mutants of restorer lines and study the effect of single *RFL-PPR* genes in sterile cytoplasm (Hussain et al. 2018).

4.1.5 An improved methodology for linkage mapping of fertility-restoring loci

The restorer loci characterised in the present study are promising components for the development of an efficient CMS-Rf system in wheat. However, to achieve a stable fertility restoration, follow-up studies investigating further restorer and modifier loci may be necessary. Besides the loci analysed in the present study, several other fertility-

restoring loci were reported on chromosomes 2D, 4R, 6D, 6R, 6U, 7B and 7D (Tahir and Tsunewaki 1969; Curtis and Lukaszewski 1993; Ma et al. 1995; Kojima et al. 1997; Sinha et al. 2013). Moreover, many restorer genes in *T. timopheevii* may have not been detected yet. The results of the present study are a valuable resource to optimise future mapping strategies. The QTL analyses in populations segregating for *Rf1* demonstrated that fertility-restoring loci in wheat can be affected by epistatic interactions with modifier loci, which may not be apparent from the discrete phenotype. In contrast, in populations segregating for *Rf3*, a monogenic inheritance was assumed, and *Rf3* was mapped using only the categorical phenotype. Although this approach enabled mapping of *Rf3* and the identification of closely linked markers, a quantitative genetic approach may have identified further loci influencing the penetrance and expressivity of this restorer locus. Thus, for future linkage mapping studies, it can be suggested to investigate the genetic architecture of fertility restoration by means of QTL analysis considering quantitative phenotypic values. Furthermore, linkage mapping in this study was performed using phenotypic data from unreplicated trials in single controlled environments. However, the influence of environmental factors on fertility restoration in G-type CMS wheat was demonstrated by several studies (Johnson et al. 1967; Jošt 1982). As a consequence of the applied experimental design, the power of QTL detection and the accuracy of QTL parameter estimation were limited and it was not possible to separate the effects of QTL, environments and QTL-by-environment interactions (Lu et al. 1996; Melchinger et al. 1998; Schön et al. 2004). To circumvent these drawbacks in follow-up studies, replicated trials in multiple environments and an increased population size must be considered, especially in populations with a complex inheritance of fertility restoration. The use of immortal mapping populations, such as double haploids or recombinant inbred lines, is a common approach to enable replicated experiments. As the development of an immortal population polymorphic for *Rf* loci is not possible in a sterility-conferring cytoplasm, testcrosses between a CMS line and each individual of an euplasmic, immortal population segregating for *Rf* loci would be necessary. Such a procedure was recently used to genetically map a fertility-restoring locus in rye (*Secale cereale*; Hackauf et al. 2017). An alternative approach enabling replicated trials is the development of an immortal population in G-type cytoplasm, followed by clonal propagation of each individual. This approach was previously employed to map fertility-restoring loci in rye (Wricke et al. 1993; Miedaner et al. 2000)

and triticale (×*Triticosecale*; Würschum et al. 2017). In the present investigation, fertility restoration was assessed indirectly by determining the seed set of isolated spikes. Although this trait has shown to be a reliable proxy for the restoration of male fertility, it does not account for a possible cross-pollination between hermaphrodite and male-sterile florets within a single spike. To avoid detrimental effects of cross-pollination, fertility restoration could be defined as the ratio of viable pollen grains determined by pollen staining (Sinha et al. 2013; Würschum et al. 2017).

4.2 Distribution of *Rf1*, *Rf3*, *QRf.lfl-1BS* and *Rf4* across wheat species

Rf1 was previously detected in accessions with introgressions from *T. timopheevii* or *Triticum zhukovskyi* but also in the common wheat line R113 and its descendant L19, for which no introgression from wild wheat species is documented. To evaluate whether R113 and L19 also obtained *Rf1* from *T. timopheevii*, markers located in the target region were analysed in the *Rf1* donor lines, *T. timopheevii* accessions and a set of diverse common wheat breeding lines. The evaluation of the marker haplotypes revealed that the R3 haplotype was identical to the one of R4 and the majority of the *T. timopheevii* accessions. This confirms the assumption that *Rf1* was introgressed in these two lines. The haplotypes of R113 and L19 were grouped with those of two breeding lines. However, hierarchical cluster analysis and a comparison of mean genetic distances suggested that R113 and L19 were considerably more similar to *T. timopheevii* than to common wheat for the analysed markers. Thus, it is likely that *Rf1* was introgressed from *T. timopheevii* in both lines. As all other documented *Rf1* donor accessions are known to be derived from crosses involving *T. timopheevii* (Bahl and Maan 1973), it can be concluded that the restoring *Rf1* allele may be derived exclusively from this wild relative. Despite these putative introgressions, which were confirmed in the case of R3, markers in the *Rf1* region did not deviate from expected segregation patterns in any of the populations derived from R3, R113 and L19. This suggests that the *T. timopheevii* chromatin did not affect the transmission of gametes produced by the F_1 descendants of CMS-Sperber and the three *Rf1* donors for this genomic region. Other than these restorer accessions, the restorer lines R2 and

R5 formed a unique haplotype which was relatively distant from most *T. timopheevii* haplotypes. This could be explained by introgressions from *T. timopheevii* resources that were not represented by the sample used in this study. Contrary to expectations, the R1 haplotype was identical to those of 216 breeding lines. It is possible therefore that the analysed R1 accession obtained a non-restoring *Rf1* allele by a possible recent outcrossing event. Alternatively, this result may also suggest that the introgressed segment was too small to track its signature using the selected set of markers. The observation that 15 of 17 *T. timopheevii* accessions were monophyletic supports the hypothesis that the restoring *Rf1* allele is widely distributed and may be fixed in this species. Hence, a functional mutation in *Rf1* may drastically decrease the fitness in *T. timopheevii*. Whether the two *T. timopheevii* accessions that were not grouped in this clade carry *Rf1* alleles that also restore fertility cannot be answered yet. According to the results of the cluster analysis, several common wheat lines were genetically similar to the majority of *T. timopheevii* accessions. Whether these accession hold the capacity to restore male fertility has yet to be evaluated. Under the assumption that the restoring and non-restoring *Rf1* alleles are fixed in *T. timopheevii* and common wheat, respectively, none of the mapped SNP markers can be diagnostic for *Rf1* as the employed marker array only comprises SNP markers polymorphic in common wheat.

The same set of genotypes was evaluated for markers in the vicinity of the modifier locus *QRf.lfl-1BS*. For the selected markers, R3 and R4 were identical, as already observed for the markers in the *Rf1* region. In the phylogram, both accessions formed a clade with 16 of 17 *T. timopheevii* lines. As R3 was clearly shown to harbour the favourable *QRf.lfl-1BS* allele, the result indicates that the fertility-enhancing allele is widely distributed or fixed in *T. timopheevii*. Surprisingly, the haplotype of L19, which was shown to carry the favourable *QRf.lfl-1BS* allele, was identical to the haplotype of R113, for which no QTL was detected on this chromosome. Thus, it can be inferred that R113 carries the same *QRf.lfl-1BS* allele as its descendant L19. This locus may not have been detected in PopR113 as a result of epistasis, environmental effects or due to the experimental design, which may not have been suitable for the detection of several restorer and modifier loci within a biparental population. As the validated *QRf.lfl-1BS* donors, R3 and L19, were polyphyletic, no definitive conclusions can be drawn about the distribution of *QRf.lfl-1BS* in common wheat.

Evaluating the relationship of these lines for selected markers on chromosome 6B revealed that R113, which was found to carry *Rf4*, was assigned to the same clade as R1, R2, R5 and all *T. timopheevii* accessions. It can thus be suggested that *Rf4* in R113 was introgressed from *T. timopheevii* analogous to *Rf1*. As the G genome of *T. timopheevii* is closely related to the B genome of common wheat, bivalent formation between chromosome 6B and the homologous 6G chromosome is possible (Nath et al. 1985; Brown-Guedira et al. 1996). This hypothesis may also explain the observed segregation distortion along the whole chromosome 6B and the preferred transmission of the gamete carrying *Rf4*. It is therefore likely that R113 either carries a 6G translocation or a 6G(6B) substitution.

The present study confirmed the previous assumption that the same locus, namely *Rf3*, explains the restoration capacity observed for chromosome 1BS in several common wheat cultivars and in European spelt. To estimate the distribution of *Rf3* in these subspecies and validate whether *Rf3* may be introgressed from *T. timopheevii*, closely linked markers were used to characterise samples of common wheat breeding lines, *T. timopheevii* accessions and European spelt cultivars. In the common wheat lines, 8.9% carried the *IWB72107* allele that was shown to have the potential to predict the restoring *Rf3* allele in the diversity panel. Hence, it can be concluded that this percentage approximately reflects the frequency of the restoring *Rf3* allele in the German common wheat pool. In a diverse panel of European common wheat cultivars, Zanke et al. (2014) observed a frequency of 12% for the particular marker allele. In an initial screening for fertility restoration (data not shown), it was determined that about 14% of European common wheat cultivars had the potential to restore the fertility in crosses with G-type CMS lines. However, the genetic architecture of fertility restoration in these cultivars remained unclear. The results of the present study provide support for the hypothesis that the restoration capacity in European common wheat can be mainly explained by *Rf3*. The fact that the *IWB72107* alleles were homogeneously distributed in common wheat indicates that the *Rf3* genotype is probably independent of population structure. The relatively low minor allele frequency (MAF) of *IWB72107* further suggests that the fertility-restoring *Rf3* allele has probably no strong positive effect on the fitness and agronomic performance of common wheat. The SNP markers *IWB14060* and *IWB73447*, which were found to be closely linked to *Rf3* in PopPrimepi,

exhibited a MAF of 0.40 and 0.42 in the common wheat breeding lines, respectively. Comparing these MAFs with the estimated prevalence of restoration capacity against G-type CMS, it can be concluded that both markers are not diagnostic for *Rf3*, as they would overestimate the frequency of the fertility-restoring *Rf3* allele. Analysing the marker haplotypes of common wheat lines and *T. timopheevii* accessions revealed that the minor allele of *IWB72107*, which was shown to be associated with the restoring *Rf3* allele, was in almost complete linkage disequilibrium with the alleles of *IWB14060* and *IWB73447*. Besides TRI7301, all *T. timopheevii* accessions exhibited a marker haplotype comprising the major allele of *IWB72107*. Thus, the results of the present study provide no evidence for the hypothesis that *T. timopheevii* carries the restoring *Rf3* allele and that this allele was introgressed into common wheat. Testcrosses and genotypic analysis of European spelt suggested that 66.7% of the analysed spelt cultivars carried the restoring *Rf3* allele. Similar as observed in common wheat, the distribution of *Rf3* did not depend on population structure. Furthermore, the estimated *Rf3* genotypes of spelt cultivars were probably independent of their relationship to a common wheat reference population. This finding may support the conjecture that neither the restoring nor the non-restoring allele was introduced to spelt by recent crosses with common wheat. Due to the limited sample size of spelt cultivars, definitive conclusions about the exchange of *Rf3* alleles between the two subspecies cannot be drawn. However, the two oldest analysed spelt cultivars, Altgold and Oberkulmer Rotkorn, were found to probably carry the restoring and the non-restoring *Rf3* allele, respectively. Thus, it can be concluded that both alleles have existed in European spelt at least since the 1950s.

The fact that *Rf3* is functionally polymorphic in common wheat and European spelt implies that the gene at this locus does not have a fertility-restoring function in euplasmic accessions of these subspecies. The presence of the restoring *Rf3* allele in common wheat and European spelt may be explained by a scenario in which *Rf3* once interacted with a CMS gene orthologous to the CMS gene that is exploited in G-type cytoplasm. In this scenario, the restoring *Rf3* allele would have been fixed in both subspecies or their progenitor. A functional mutation in the CMS-conferring gene may have led to the loss of purifying selective pressure on *Rf3* and consequently to a loss of fixation. Different allele frequencies in common wheat and spelt could be explained by

a possible pleiotropic, subspecies-specific effect of *Rf3* or by genetic drift in the process of European spelt emergence. Whether the *Rf3* locus is limited to two alleles and if *Rf3* alleles differ between common wheat and spelt remains to be answered.

4.3 Prospects of hybrid wheat breeding

The main motivations for hybrid wheat breeding are grain yield heterosis and an increased yield stability. Although these advantages have been intensively studied, the long-term perspective of hybrid breeding versus line breeding in wheat depends on several factors and cannot be finally clarified yet. To compare the prospects of the two breeding methods, Longin et al. (2014) introduced the concept of the predicted future yield potential, which was defined as a function of the mean trait value of the breeding population, genetic gain and the length of a breeding cycle. They proposed that, despite an assumed mid-parent heterosis of 10–15%, line breeding may outperform hybrid breeding after few breeding cycles, provided that the same budget is spent for both methods. Based on these findings, future efforts must aim at (1) improving grain yield heterosis, (2) increasing the genetic variance of hybrid populations and (3) reducing the costs of hybrid seed production (Longin et al. 2014). The implications of these three objectives are shortly discussed in the remainder of this section.

To efficiently use heterosis in future wheat breeding programs, the establishment of heterotic groups and the identification of heterotic patterns between these groups are of substantial importance. Unlike in crops such as maize (*Zea mays*) and rye, heterotic groups have yet to be developed in wheat. Global wheat accessions share a high degree of genetic similarity, especially within the European, Asian and US-American gene pool (Boeven et al. 2016). Hence, the establishment of genetically distinct parental groups is a prerequisite to enable an optimal exploitation of heterosis. Reif (personal communication) suggested an initial pool size of 20 individuals for a long-term success of hybrid wheat. Besides the development of heterotic groups, the selection of parental lines for testcross seed production is of major importance. As it is not feasible to phenotype the hybrids of all possible combinations of parental lines, multi-stage selection is commonly applied in hybrid breeding programs. In the first stage, the potential parental lines are selected based on their *per se* performance. Subsequently, the selected

lines are evaluated for their general combining ability (GCA). In the third stage, combinations of parental lines are selected based on the GCA and the specific combining ability (SCA). The optimal allocation of test resources in such multi-stage breeding programs strongly depends on the ratio of the variance due to GCA to the variance due to SCA. In hybrid wheat, several studies observed not only significant GCA effects but also significant SCA effects for grain yield (Gyawali et al. 1968; Perenzin et al. 1998; Gowda et al. 2012; Longin et al. 2013). For the amount of the variances due to GCA versus SCA, Gowda et al. (2012) and Longin et al. (2013) estimated a ratio of 2.6 and 2.1, respectively. The relevance of SCA effects can be addressed by using many tester lines or complex testers (Hallauer and Miranda 1981; Longin et al. 2013). The development of heterotic groups could increase the ratio of variances due to GCA versus SCA, thereby improving the efficiency of GCA-based selection (Reif et al. 2007).

One of the major constraints in hybrid wheat breeding is the relatively small genetic variance of hybrid wheat populations. As the genetic variance enters the formula for selection gain twice, it has a strong impact on long-term breeding success. According to Gowda et al. (2012) and Longin et al. (2013), the genotypic variance estimates in line breeding wheat populations are about twice the genetic variance observed in hybrids. It was speculated that this discrepancy could be explained by the fact that testcross parents were selected for traits that enable an optimal cross-pollination, thereby narrowing the genetic diversity available for hybrid breeding (Longin et al. 2014). This disadvantage could be overcome by introducing favourable alleles affecting floral biology into a diverse genetic background. However, the available variance estimates must be carefully interpreted as they rely on two studies which were based on experimental populations rather than actual breeding populations.

Another disadvantage of hybrid wheat breeding is the cost of hybrid seed production. These expenses incur for the production of testcross seeds in the breeding process as well as for the large-scale production of registered hybrid varieties. Compared to line breeding, the elevated costs for testcross seed production reduce the number of tested hybrids and test environments. This negatively affects selection intensity and heritability, thereby impairing genetic gain. Efforts to optimise hybrid wheat seed production can be divided into (1) strategies to alter the floral biology for an improved cross-pollination between parental lines and (2) the development of an efficient and

stable male sterility system. Floral biology traits are important breeding goals for both the female and the male pool. Whereas the male parent must be selected for a maximum anther extrusion and the shedding of large quantities of viable pollen, the female ideotype should exhibit widely opened florets that enable an optimal pollen reception (Whitford et al. 2013). Using technologies such as chemical hybridising agents (CHA), the floral phenotype of the female parent is less crucial. This can be explained by the fact that, in contrast to fertile wheat florets, which open for up to 30 minutes at anthesis, unfertilised wheat florets additionally accept pollen for several days during a second opening (Okada et al. 2018). However, in CMS systems, the pollination capacity and the opening of florets are not only relevant for the production of hybrid seeds but also for the maintenance of the CMS line. Anther extrusion, which reflects the interplay of many important floral characteristics, is a genetically complex trait that exhibits a wide variation and high heritability, indicating that phenotypic selection for anther extrusion is a promising approach to improve cross-pollination between parental lines (Muqaddasi et al. 2017).

Besides the physical floral architecture, the second component necessary for cross-pollination of wheat is a system that prevents self-pollination of the female parent. Currently, most hybrid breeding programs make use of CHAs to control fertilisation. The application of CHAs involves a considerable effort and its success depends on many factors. As a result, the production of hybrid seeds for multi-location yield trials across two years costs about four times as much as one yield plot (Longin et al. 2014). Therefore, genetic-based systems controlling male fertility are a promising alternative for future hybrid wheat breeding programs. These can be classified into genic male sterility (GMS) systems, CMS systems and genetically engineered systems. GMS systems sensitive for temperature and photoperiod demand contrasting climatic regions for the propagation of the GMS line and hybrid seed production. Furthermore, these technologies are often associated with incomplete sterility of the GMS line and partial fertility restoration in the hybrid generation. GMS systems that are not conditional on environmental factors have been impeded by difficulties in the large-scale propagation of male-sterile lines (Whitford et al. 2013). Hybridisation technologies enabled by genetic modifications could offer an alternative. An approach that makes use of genetic modifications is seed production technology (SPT) described by Wu et al. (2016).

SPT enables the exploitation of a nuclear male fertility gene by using a transgene that contains a fertility-restoring allele and is biologically contained to the maintainer line through the action of a pollen germination inhibitor. The application of SPT in wheat was proposed by Tucker et al. (2017) and Okada et al. (2019) for the newly identified male fertlity gene *Ms1*. However, the success of such transgenic approaches also depends on the approval of the authorities. In most countries of the European Union, the production of genetically modified plants is prohibited.

CMS using the G-type cytoplasm is one of the most promising approaches to reduce the cost of hybrid seed production. However, the success of this system has been impeded by incomplete fertility restoration of the hybrids, which neutralises the beneficial effect of heterosis and can also affect the quality of hybrid wheat. As a single restorer locus is not sufficient to stably express a complete restoration across varying environments, stacking of restorer loci in elite restorer lines was suggested as a possible solution (Johnson and Patterson 1977). Although this hypothesis has not been validated, studies in other hybrid crops substantiate the necessity of gene stacking. Fertility restoration in T-type CMS maize can only be accomplished by a combination of the complementary restorer locus *Rf2* and at least one of the loci *Rf1*, *Rf8* or *Rf** (Dill et al. 1997). In CMS rye, the two restorer loci *Rfp1* and *Rfp2* are combined to achieve sufficient fertility restoration in Pampa cytoplasm (Geiger and Miedaner 2009). A recent study in rice revealed that the combination of *Rf3* and *Rf4*, which can independently restore fertility in WA-type CMS, resulted in a superior fertility restoration compared to the individual effect of *Rf3* or *Rf4* (Katara et al. 2017). Stacking of *Rf1* and *Rf3* in wheat is a plausible approach as both loci are able to restore fertility to a large extent. Furthermore, *Rf1* and *Rf3* are located on subgenomes A and B, respectively. This suggests that both loci can be easily integrated by marker-assisted backcrossing without detrimental effects of linkage drag. This could be cumbersome for restorer loci introgressed from wild relatives such as rye (*Rfc3*, *Rfc4*), *Aegilops umbellulata* (*Rf6*) or the G genome of *T. timopheevii* (*Rf4*). However, using the example of *Rf1* and *Rf4*, the results of the present study showed that the combination of two restorer loci does not necessarily exhibit a superior effect compared to a single locus. Thus, possible epistatic effects should be investigated in experimental populations before implementing restorer loci in breeding programs. Boeven et al. (2016) suggested to start the

development of heterotic pools with elite lines exhibiting a high *per se* performance for grain yield, a high level of genetic diversity and favourable floral characteristics. Lines showing a high GCA in testcrosses can then be selected to establish female and male pools. This selection step may be facilitated by a genome-based algorithm searching for heterotic patterns (Zhao et al. 2015). At this step, the sterility-inducing cytoplasm and loci affecting fertility restoration could be introgressed and fixed by backcrossing in the female and male pool, respectively. After the establishment of a CMS/maintainer pool and a restorer pool, reciprocal recurrent selection can be applied to further improve combining ability. Markers identified in the present study are a valuable resource to efficiently select for restorer and modifier loci in the process of pool development. As the array-derived markers linked to *Rf1* are not diagnostic, only a marker haplotype can be used to track this locus in diverse germplasm. Marker-assisted selection for *Rf3* can either be realised using the haplotype of closely linked markers or only *IWB72107*, which has shown diagnostic potential in the diversity panel. Markers developed for the *RFL-PPR* candidate genes also exhibited tight linkage to *Rf1* and *Rf3* and may complement the array-derived markers for marker-assisted selection. However, whether these candidates comprise the functional genes remains to be answered.

5 Summary

Hybrid breeding is a promising approach to increase the rate of genetic gain in common wheat (*Triticum aestivum* ssp. *aestivum*). Cytoplasmic male sterility (CMS) using the cytoplasm of *Triticum timopheevii* (G-type CMS) is one of the most widely discussed methods to force cross-pollination between the autogamous parental lines. However, the success of G-type CMS has been impeded by incomplete restoration of male fertility in hybrids, which counteracts the benefits of heterosis. Stacking of major *restorer-of-fertility* (*Rf*) loci in elite restorer lines is a potential solution to achieve a full fertility restoration. Two promising genetic components for this approach are the *Rf* loci *Rf1* and *Rf3* on chromosomes 1AS and 1BS, respectively. However, it has not been validated whether all fertility-restoring effects reported on the two chromosomes can be explained by these two loci. Although *Rf1* and *Rf3* have been the subject of several studies, the functional genes are unknown and molecular markers that allow an efficient selection for these loci have been unavailable. It has also been unclear how *Rf1* and *Rf3* are distributed across wheat species. In the present study, *Rf1* and *Rf3* were genetically mapped in several biparental populations. Linkage mapping also involved the validation of candidate genes of the *pentatricopeptide repeat* (*PPR*) family. Using closely linked single-nucleotide polymorphism (SNP) markers, the genomic target regions were analysed in diverse wheat accessions.

Quantitative trait locus (QTL) analyses in populations derived from the restorer accessions R3, R113 and L19 revealed that the restoration capacity of chromosome 1AS can be explained by *Rf1* in all three populations. The effect of *Rf1* was affected by modifier loci on chromosomes 1BS, 1BL and 6BL. It was also demonstrated that *Rf1* did not exhibit an inferior restoration capacity compared to the combination of *Rf1* and the restorer locus *Rf4* on chromosome 6BS. A cluster of two *PPR* genes was located within the *Rf1* support interval in all three populations, suggesting that one or both of them may cause the restoration capacity at this locus. Analysing SNP markers at the *Rf1* locus in diverse common wheat breeding lines, *T. timopheevii* accessions and all hitherto known *Rf1* donors provided evidence for the hypothesis that the restoring *Rf1* allele is widely distributed or fixed in *T. timopheevii*. Furthermore, the presence of *Rf1* in common wheat may be completely traced back to *T. timopheevii* introgres-

sions. The restorer locus *Rf3* was genetically mapped in seven populations involving the common wheat restorer lines Primepi, PR143 and Palmaress and the European spelt (*Triticum aestivum* ssp. *spelta*) cultivars Badenkrone, Badenstern, Holstenkorn and Schwabenspelz. Genetic mapping revealed that the restoration capacity was controlled by *Rf3* in all analysed accessions. In the populations derived from Primepi and Badenkrone, a cluster comprising five *PPR* candidate genes was found to be closely linked to the *Rf3* locus, indicating that one or several of these genes may explain the restoration capacity of chromosome 1BS. The characterisation of diverse common wheat breeding lines, *T. timopheevii* accessions and European spelt cultivars using an informative SNP marker and testcrosses with a CMS line indicated that approximately 8.9% of the common wheat lines and 66.7% of the spelt cultivars carried a restoring *Rf3* allele. These results provide support for the hypothesis that the restoration capacity in European common wheat can be mainly explained by *Rf3*. No evidence was found for the presence of a restoring *Rf3* allele in *T. timopheevii*. The results of the present study are a valuable resource towards a refined understanding of the restorer loci on homoeologous group 1 and their implementation in future hybrid wheat breeding programs.

6 Zusammenfassung

Hybridzüchtung ist ein vielversprechender Ansatz, um den Zuchtfortschritt bei Weichweizen (*Triticum aestivum* ssp. *aestivum*) zu beschleunigen. Cytoplasmatische männliche Sterilität basierend auf dem Cytoplasma von *Triticum timopheevii* (G-Typ-CMS) ist eine der meist diskutierten Methoden, um dabei eine Fremdbestäubung zwischen den autogamen Elterlinien zu gewährleisten. Der Erfolg von G-Typ-CMS wurde jedoch bisher durch eine unvollständige Restaurierung der männlichen Fertilität bei den Hybriden verhindert, welche den Vorteilen der Heterosis entgegenwirkt. Die Kombination von *Restorer-of-Fertility* (*Rf*)-Genen in Elite-Restorer-Linien ist eine mögliche Lösung, um eine vollständige Restaurierung der männlichen Fertilität zu erreichen. Zwei vielversprechende genetische Komponenten für diesen Ansatz sind die *Rf*-Loci *Rf1* und *Rf3* auf Chromosom 1AS beziehungsweise 1BS. Es wurde jedoch noch nicht validiert, ob alle merkmalsbezogenen Effekte auf diesen Chromosomen auf die beiden Loci zurückzuführen sind. Obwohl *Rf1* und *Rf3* Gegenstand mehrerer Studien waren, sind die funktionalen Gene unbekannt. Des Weiteren sind keine molekularen Marker verfügbar, welche eine effiziente Selektion dieser Loci ermöglichen. Es ist ebenso unbekannt, wie *Rf1* und *Rf3* in Weizenarten verteilt sind. In der vorliegenden Studie wurden *Rf1* und *Rf3* in sieben biparentalen Populationen genetisch kartiert. Im Zuge der genetischen Kartierung wurden außerdem Kandidatengene der *Pentatricopeptid-Repeat* (*PPR*)-Familie validiert. Mit Hilfe eng gekoppelter Single-Nucleotide-Polymorphism (SNP)-Marker wurden die genomischen Zielregionen in diversen Weizenakzessionen untersucht.

Quantitative-Trait-Locus (QTL)-Analysen in Nachkommenschaften der Restorer-Linien R3, R113 und L19 zeigten, dass die Restaurationskapazität von Chromosom 1AS in allen drei Populationen durch *Rf1* erklärt werden kann. Der Effekt von *Rf1* wurde von Modifier-Loci auf den Chromosomen 1BS, 1BL und 6BL beeinflusst. Es konnte außerdem gezeigt werden, dass *Rf1* keine schlechtere Restaurierung der Fertilität bewirkte als die Kombination aus *Rf1* und dem Restorer-Locus *Rf4* auf Chromosom 6BS. Ein Cluster von zwei *PPR*-Genen wurde in allen drei Populationen innerhalb des *Rf1*-Support-Intervalls lokalisiert. Dies deutet darauf hin, dass eines oder beide dieser Gene die Restaurierung der Fertilität an diesem Locus verursachen könnte. Die Untersuchung

von diversen Weichweizen-Zuchtlinien, *T. timopheevii*-Akzessionen und allen bisher bekannten *Rf1*-Donoren mit Hilfe von SNP-Markern für den *Rf1*-Locus wiesen darauf hin, dass das restaurierende *Rf1*-Allel in *T. timopheevii* weit verbreitet oder fixiert ist. Des Weiteren könnte das Vorliegen von *Rf1* in Weichweizen vollständig auf Introgressionen von *T. timopheevii* zurückzuführen sein. Die genetische Kartierung des Restorer-Locus *Rf3* erfolgte in sieben Populationen, welche mit Hilfe der Weichweizen-Restorer-Linien Primepi, PR143 und Palmaress und den Europäischen Dinkelsorten (*Triticum aestivum* ssp. *spelta*) Badenkrone, Badenstern, Holstenkorn und Schwabenspelz erstellt wurden. Die genetische Kartierung zeigte, dass die Restaurationskapazität in allen untersuchten Akzessionen von *Rf3* kontrolliert wurde. In den von Primepi und Badenkrone abgeleiteten Populationen wurde eine enge Kopplung zwischen *Rf3* und einem Cluster von fünf *PPR*-Kandidatengenen beobachtet. Dies lässt darauf schließen, dass eines oder mehrere dieser Gene die Restaurationskapazität des Chromosoms 1BS erklären könnte. Die Charakterisierung diverser Weichweizen-Zuchtstämme, *T. timopheevii*-Akzessionen und Sorten des Europäischen Dinkels mit Hilfe eines informativen SNP-Markers und Testkreuzungen mit einer CMS-Linie deuteten darauf hin, dass etwa 8.9% der Zuchtlinien und 66.7% der Dinkelsorten das restaurierende *Rf3*-Allel trugen. Diese Ergebnisse geben Grund zur Annahme, dass die Restaurationskapazität im europäischem Weichweizen hauptsächlich durch *Rf3* erklärt werden kann. In *T. timopheevii* konnte *Rf3* nicht nachgewiesen werden. Die Ergebnisse der vorliegenden Studie sind eine wertvolle Grundlage für ein verbessertes Verständnis der Restorer-Loci der homöologen Gruppe 1 und deren Nutzung in der zukünftigen Züchtung von Hybridweizen.

7 References

Abel S, Möllers C, Becker HC (2005) Development of synthetic *Brassica napus* lines for the analysis of "fixed heterosis" in allopolyploid plants. Euphytica 146:157–163

Adugna A, Nanda GS, Singh K, Bains NS (2004) A comparison of cytoplasmic and chemically-induced male sterility systems for hybrid seed production in wheat (*Triticum aestivum* L.). Euphytica 135:297–304

Ahmed TA, Tsujimoto H, Sasakuma T (2001) QTL analysis of fertility-restoration against cytoplasmic male sterility in wheat. Genes Genet Syst 76:33–38

Akagi H, Nakamura A, Yokozeki-Misono Y, Inagaki A, Takahashi H, Mori K, Fujimura T (2004) Positional cloning of the rice *Rf-1* gene, a restorer of BT-type cytoplasmic male sterility that encodes a mitochondria-targeting PPR protein. Theor Appl Genet 108:1449–1457

Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. Theor Appl Genet 119:507–517

Allen AM, Winfield MO, Burridge AJ, Downie RC, Benbow HR, Barker GLA, Wilkinson PA, Coghill J, Waterfall C, Davassi A, Scopes G, Pirani A, Webster T, Brew F, Bloor C, Griffiths S, Bentley AR, Alda M, Jack P, Phillips AL, Edwards KJ (2017) Characterization of a wheat breeders' array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). Plant Biotechnol J 15:390–401

Andolfo G, Jupe F, Witek K, Etherington GJ, Ercolano MR, Jones JDG (2014) Defining the full tomato *NB-LRR* resistance gene repertoire using genomic and cDNA RenSeq. BMC Plant Biol 14:120

Bahl PN, Maan SS (1973) Chromosomal location of male fertility restoring genes in six lines of common wheat. Crop Sci 13:317–320

Bardol N, Ventelon M, Mangin B, Jasson S, Loywick V, Couton F, Derue C, Blanchard P, Charcosset A, Moreau L (2013) Combined linkage and linkage disequilibrium QTL mapping in multiple families of maize (*Zea mays* L.) line crosses highlights complementarities between models based on parental haplotype and single locus polymorphism. Theor Appl Genet 126:2717–2736

Barlow KK, Driscoll CJ (1981) Linkage studies involving two chromosomal male-sterility mutants in hexaploid wheat. Genetics 98:791–799

Barr CM, Fishman L (2010) The nuclear component of a cytonuclear hybrid incompatibility in *Mimulus* maps to a cluster of pentatricopeptide repeat genes. Genetics 184:455–465

Batey I (2017) Measuring wheat quality. In: Langridge P (ed) Achieving sustainable cultivation of wheat. Volume 1: Breeding quality traits, pests and diseases. Burleigh Dodds Science Publishing, Cambridge, pp 213–229

Battenfield SD, Guzmán C, Gaynor RC, Singh RP, Peña RJ, Dreisigacker S, Fritz AK, Poland JA (2016) Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. Plant Genome 9

Becker HC (1987) Zur Heritabilität statistischer Maßzahlen für die Ertragssicherheit. Vortr Pflanzenzüchtg 12:134–144

Beckmann JS, Soller M (1986) Restriction fragment length polymorphisms and genetic improvement of agricultural species. Euphytica 35:111–124

Bentley AR, Mackay I (2017) Advances in wheat breeding techniques. In: Langridge P (ed) Achieving sustainable cultivation of wheat. Volume 1: Breeding, quality traits, pests and diseases. Burleigh Dodds Science Publishing, Cambridge, pp 53–72

Bentolila S, Alfonso AA, Hanson MR (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc Natl Acad Sci U S A 99:10887–10892

Bernardo R (2008) Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Sci 48:1649–1664

Bing-Hua L, Jing-Yang D (1986) A dominant gene for male sterility in wheat. Plant Breeding 97:204–209

Boeven PHG, Longin CFH, Würschum T (2016) A unified framework for hybrid breeding and the establishment of heterotic groups in wheat. Theor Appl Genet 129:1231–1245

Bonnett DG, Rebetzke GJ, Spielmeyer W (2005) Strategies for efficient implementation of molecular markers in wheat breeding. Mol Breeding 15:75–85

Borghi B, Perenzin M, Nash RJ (1988) Agronomic and qualitative characteristics of ten bread wheat hybrids produced using a chemical hybridizing agent. Euphytica 39:185–194

Brenchley R, Spannagl M, Pfeifer M, Barker GLA, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo MC, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KFX, Edwards KJ, Bevan MW, Hall N (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491:705–710

Briggle LW, Daum RJ, Stevens H (1964) Expression of heterosis in two wheat crosses. Crop Sci 4:220–223

Broman KW (2003) Mapping quantitative trait loci in the case of a spike in the phenotype distribution. Genetics 163:1169–1175

Broman KW, Sen S (2009) A guide to QTL mapping with R/qtl. Statistics for Biology and Health. Springer, New York

Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890

Brown-Guedira GL, Badaeva ED, Gill BS, Cox TS (1996) Chromosome substitutions of *Triticum timopheevii* in common wheat and some observations on the evolution of polyploid wheat species. Theor Appl Genet 93:1291–1298

Bruns R, Peterson CJ (1998) Yield and stability factors associated with hybrid wheat. Euphytica 100:1–5

Caruso CM, Case AL, Bailey MF (2012) The evolutionary ecology of cytonuclear interactions in angiosperms. Trends Plant Sci 17:638–643

CerealsDB (2016) Tools for the analysis of the wheat genome. University of Bristol. http://www.cerealsdb.uk.net. Accessed 1 Feb 2016

Chapman JA, Mascher M, Buluç A, Barry K, Georganas E, Session A, Strnadova V, Jenkins J, Sehgal S, Oliker L, Schmutz J, Yelick KA, Scholz U, Waugh R, Poland JA, Muehlbauer GJ, Stein N, Rokhsar DS (2015) A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. Genome Biol 16:26

Charcosset A, Mangin B, Moreau L, Combes L, Jourjon M, Gallais A (2001) Heterosis in maize investigated using connected RIL populations. In: Gallais A, Dillmann C, Goldringer I (eds) Quantitative genetics and breeding methods: the way ahead. Les colloques no. 96. Institut national de la recherche agronomique (INRA), Paris, pp 89–98

Charmet G, Robert N, Perretant MR, Gay G, Sourdille P, Groos C, Bernard S, Bernard M (1999) Marker-assisted recurrent selection for cumulating additive and interactive QTLs in recombinant inbred lines. Theor Appl Genet 99:1143–1148

Cheng SH, Cao LY, Zhuang JY, Chen SG, Zhan XD, Fan YY, Zhu DF, Min SK (2007) Super hybrid rice breeding in China: Achievements and prospects. J Integr Plant Biol 49:805–810

Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlub S, Le Paslier MC, Magdelenat G, Gonthier C, Couloux A, Budak H, Breen J, Pumphrey M, Liu S, Kong X, Jia J, Gut M, Brunel D, Anderson JA, Gill BS, Appels R, Keller B, Feuillet C (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell 22:1686–1701

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Clustal Omega (2016) The European Bioinformatics Institute (EMBL-EBI). https://www.ebi.ac.uk/. Accessed 1 Feb 2016

Corbellini M, Perenzin M, Accerbi M, Vaccino P, Borghi B (2002) Genetic diversity in bread wheat, as revealed by coefficient of parentage and molecular markers, and its relationship to hybrid performance. Euphytica 123:273–285

Cronn R, Knaus BJ, Liston A, Maughan PJ, Parks M, Syring JV, Udall J (2012) Targeted enrichment strategies for next-generation plant biology. Am J Bot 99:291–311

Crow JF (1998) 90 years ago: The beginning of hybrid maize. Genetics 148:923–928

Curtis CA, Lukaszewski AJ (1993) Localization of genes in rye that restore male fertility to hexaploid wheat with *timopheevi* cytoplasm. Plant Breeding 111:106–112

Curtis T, Halford NG (2014) Food security: the challenge of increasing wheat yield and the importance of not compromising food safety. Ann Appl Biol 164:354–372

Dahan J, Mireau H (2013) The Rf and Rf-like PPR in higher plants, a fast-evolving subclass of PPR genes. RNA Biol 10:1469–1476

Desloire S, Gherbi H, Laloui W, Marhadour S, Clouet V, Cattolico L, Falentin C, Giancola S, Renard M, Budar F, Small I, Caboche M, Delourme R, Bendahmane A (2003) Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family. EMBO Rep 4:588–594

Devos KM, Ma J, Pontaroli AC, Pratt LH, Bennetzen JL (2005) Analysis and mapping of randomly chosen bacterial artificial chromosome clones from hexaploid bread wheat. Proc Natl Acad Sci U S A 102:19243–19248

de Vries AP, Ie TS (1970) Electron-microscopy on anther tissue and pollen of male sterile and fertile wheat (*Triticum aestivum* L.). Euphytica 19:103–120

Dill CL, Wise RP, Schnable PS (1997) *Rf8* and *Rf** mediate unique T-*urf13*-transcript accumulation, revealing a conserved motif associated with RNA processing and restoration of pollen fertility in T-cytoplasm maize. Genetics 147:1367–1379

Du H, Maan SS, Hammond JJ (1991) Genetic analyses of male-fertility restoration in wheat: III. Effects of aneuploidy. Crop Sci 31:319–322

Endo TR (1982) Gametocidal chromosomes of three *Aegilops* species in common wheat. Can J Genet Cytol 24:201–206

Evenson RE, Gollin D (2003) Assessing the impact of the green revolution, 1960 to 2000. Science 300:758–762

Eversole K, Rogers J, Keller B, Appels R, Feuillet C (2017) Sequencing and assemly of the wheat genome. In: Langridge P (ed) Achieving sustainable cultivation of wheat. Volume 1: Breeding quality traits, pests and diseases. Burleigh Dodds Science Publishing, Cambridge, pp 25–52

FAOSTAT (2019) Food and Agriculture Organization Corporate Statistical Database. http://www.fao.org/faostat/en/#home. Accessed 4 Mar 2019

Faris JD (2014) Wheat domestication: key to agricultural revolutions past and future. In: Tuberosa R, Graner A, Frison E (eds) Genomics of Plant Genetic Resources. Springer, Dordrecht, pp 439–464

Feldman M (2001) Origin of cultivated wheat. In: Bonjean AP, Angus WJ (eds) The world wheat book. A history of wheat breeding. Lavoisier Publishing, Paris, pp 3–56

Fishman L, Willis JH (2006) A cytonuclear incompatibility causes anther sterility in *Mimulus* hybrids. Evolution 60:1372–1381

Flint-Garcia SA, Buckler ES, Tiffin P, Ersoz E, Springer NM (2009) Heterosis is prevalent for multiple traits in diverse maize germplasm. PLOS ONE 4:e7433

Foley JA, Ramankutty N, Brauman KA, Cassidy ES, Gerber JS, Johnston M, Mueller ND, O'Connell C, Ray DK, West PC, Balzer C, Bennett EM, Carpenter SR, Hill J, Monfreda C, Polasky S, Rockström J, Sheehan J, Siebert S, Tilman D, Zaks DPM (2011) Solutions for a cultivated planet. Nature 478:337–342

Fossati D, Ingold M (2001) Mountain wheat pool. In: Bonjean AP, Angus WJ (eds) The world wheat book. A history of wheat breeding. Lavoisier Publishing, Paris, pp 311–332

Freeman GF (1919) Heredity of quantitative characters in wheat. Genetics 4:1–93

Fu D, Xiao M, Hayward A, Jiang G, Zhu L, Zhou Q, Li J, Zhang M (2015) What is crop heterosis: new insights into an old topic. J Appl Genet 56:1–13

Fujii S, Bond CS, Small ID (2011) Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. Proc Natl Acad Sci U S A 108:1723–1728

Fujii S, Toriyama K (2009) Suppressed expression of retrograde-regulated male sterility restores pollen fertility in cytoplasmic male sterile rice plants. Proc Natl Acad Sci U S A 106:9513–9518

Fukasawa H (1953) Studies on restoration and substitution of nucleus of *Aegilotricum*, I. Appearance of male-sterile durum in substitution crosses. Cytologia 18:167–175

Gaborieau L, Brown GG, Mireau H (2016) The propensity of pentatricopeptide repeat genes to evolve into restorers of cytoplasmic male sterility. Front Plant Sci 7:1816

Geddy R, Brown GG (2007) Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. BMC Genomics 8:130

Geiger HH, Miedaner T (2009) Rye (*Secale cereale* L.). In: Carena MJ (ed) Cereals. Springer, New York, pp 157–181

Gillman JD, Bentolila S, Hanson MR (2007) The petunia restorer of fertility protein is part of a large mitochondrial complex that interacts with transcripts of the CMS-associated locus. Plant J 49:217–227

Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C (2010) Food security: The challenge of feeding 9 billion people. Science 327:812–818

Gowda M, Longin CFH, Lein V, Reif JC (2012) Relevance of specific versus general combining ability in winter wheat. Crop Sci 52:2494–2500

GrainGenes (2016) A Database for Triticeae and *Avena*. http://wheat.pw.usda.gov. Accessed 1 Feb 2016

Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM (2000) An introduction to genetic analysis, 7th edn. W. H. Freeman & Co Ltd, New York

Gyawali KK, Qualset CO, Yamazaki WT (1968) Estimates of heterosis and combining ability in winter wheat. Crop Sci 8:322–324

Hackauf B, Bauer E, Korzun V, Miedaner T (2017) Fine mapping of the restorer gene *Rfp3* from an Iranian primitive rye (*Secale cereale* L.). Theor Appl Genet 130:1179–1189

Haley C, Knott S (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Hallauer AR, Miranda JB (1981) Quantitative genetics in maize breeding. Iowa State University Press, Ames

Halloran GM (1975) Heterosis in wheat. Z Pflanzenzüchtg 47:18–27

Hayward CF (1975) The status and prospects for hybrid winter wheat. In: Agricultural Institute Zagreb (ed) Proceedings of the 2nd International Winter Wheat Conference. Agricultural Institute Zagreb, Zagreb, pp 84–104

Hedgcoth C, El-Shehawi AM, Wei P, Clarkson M, Tamalis D (2002) A chimeric open reading frame associated with cytoplasmic male sterility in alloplasmic wheat with *Triticum timopheevi* mitochondria is present in several *Triticum* and *Aegilops* species, barley, and rye. Curr Genet 41:357–366

Heffner EL, Jannink JL, Iwata H, Souza E, Sorrells ME (2011) Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Sci 51:2597–2606

Hohn CE, Lukaszewski AJ (2016) Engineering the 1BS chromosome arm in wheat to remove the Rf_{multi} locus restoring male fertility in cytoplasms of *Aegilops kotschyi*, *Ae. uniaristata* and *Ae. mutica*. Theor Appl Genet 129:1769–1774

Hu J, Wang K, Huang W, Liu G, Gao Y, Wang J, Huang Q, Ji Y, Qin X, Wan L, Zhu R, Li S, Yang D, Zhu Y (2012) 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

Hussain B, Lucas SJ, Budak H (2018) CRISPR/Cas9 in plants: at play in the genome and at work for crop improvement. Brief Funct Genomics ely016

Igarashi K, Kazama T, Toriyama K (2016) A gene encoding pentatricopeptide repeat protein partially restores fertility in RT98-type cytoplasmic male-sterile rice. Plant Cell Physiol 57:2187–2193

Ingold M (1968) Male sterility and restorer systems in wheat. Euphytica 17:69–74

Itabashi E, Iwata N, Fujii S, Kazama T, Toriyama K (2011) The fertility restorer gene, *Rf2*, for Lead Rice-type cytoplasmic male sterility of rice encodes a mitochondrial glycine-rich protein. Plant J 65:359–367

IWGSC (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788

IWGSC (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361:eaar7191

Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, Appels R, Pfeifer M, Tao Y, Zhang X, Jing R, Zhang C, Ma Y, Gao L, Gao C, Spannagl M, Mayer KFX, Li D, Pan S, Zheng F, Hu Q, Xia X, Li J, Liang Q, Chen J, Wicker T, Gou C, Kuang H, He G, Luo Y, Keller B, Xia Q, Lu P, Wang J, Zou H, Zhang R, Xu J, Gao J, Middleton C, Quan Z, Liu G, Wang J, International Wheat Genome Sequencing Consortium, Yang H, Liu X, He Z, Mao L, Wang J (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91–95

Jiang Y, Schmidt RH, Zhao Y, Reif JC (2017a) A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. Nat Genet 49:1741–1746

Jiang Y, Schulthess AW, Rodemann B, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, Argillier O, Stiewe G, Ganal MW, Röder MS, Reif JC (2017b) Validating the prediction accuracies of marker-assisted and genomic selection of Fusarium head blight resistance in wheat using an independent sample. Theor Appl Genet 130:471–482

Johnson JW, Patterson FL (1977) Interaction of genetic factors for fertility restoration in hybrid wheat. Crop Sci 17:695–699

Johnson VA, Schmidt JW, Mattern PJ (1967) Hybrid wheat in the United States. Plant Food Hum Nutr 14:193–211

Joppa LR, McNeal FH, Welsh JR (1966) Pollen and anther development in cytoplasmic male sterile wheat, (*Triticum aestivum* L.). Crop Sci 6:296–297

Jordan DR, Klein RR, Sakrewski KG, Henzell RG, Klein PE, Mace ES (2011) Mapping and characterization of *Rf5*: a new gene conditioning pollen fertility restoration in A1 and A2 cytoplasm in sorghum (*Sorghum bicolor* (L.) Moench). Theor Appl Genet 123:383–396

Jošt M (1982) Results of the 4th international wheat restorer germplasm screening nursery 1981. University of Zagreb, Zagreb

Jourjon MF, Jasson S, Marcel J, Ngom B, Mangin B (2005) MCQTL: multi-allelic QTL mapping in multi-cross design. Bioinformatics 21:128–130

Jupe F, Witek K, Verweij W, Śliwka J, Pritchard L, Etherington GJ, Maclean D, Cock PJ, Leggett RM, Bryan GJ, Cardle L, Hein I, Jones JD (2013) Resistance gene enrichment sequencing (RenSeq) enables reannotation of the *NB-LRR* gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. Plant J 76:530–544

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152:1203–1216

Katara JL, Verma RL, Nayak D, Ngangkham U, Ray S, Subudhi H, Behera L, Samantaray S, Rao RN, Singh ON, Mohapatra T (2017) Frequency and fertility restoration efficiency of *Rf3* and *Rf4* genes in Indian rice. Plant Breeding 136:74–82

Kazama T, Nakamura T, Watanabe M, Sugita M, Toriyama K (2008) Suppression mechanism of mitochondrial ORF79 accumulation by Rf1 protein in BT-type cytoplasmic male sterile rice. Plant J 55:619–628

Keydel F (1973) Die Restoration der Fertilität in Weizenhybriden. Bayer Landw Jb 50:424–430

Kihara H (1951) Substitution of nucleus and its effects on genome manifestations. Cytologia 16:177–193

Kihara H, Tsunewaki K (1967) Genetic principles applied to the breeding of crop plants. In: Brink RA (ed) Heritage from Mendel, University of Wisconsin Press, Wisconsin, pp 403–418

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

Koebner RMD, Summers RW (2003) 21st century wheat breeding: plot selection or plate detection? Trends Biotechnol 21:59–63

Koekemoer FP, Van Eeden E, Bonjean AP (2011) An overview of hybrid wheat production in South Africa and review of current worldwide wheat hybrid developments. In: Bonjean AP, Angus WJ, van Ginkel M (eds) The world wheat book. A history of wheat breeding. Volume 2. Lavoisier Publishing, Paris, pp 907–950

Koemel JE, Guenzi AC, Carver BF, Payton ME, Morgan GH, Smith EL (2004) Hybrid and pureline hard winter wheat yield and stability. Crop Sci 44:107–113

Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohno-Murase J, Sakai T, Kawasaki S, Imamura J (2003) Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosena radish. Plant J 34:407–415

Kojima T, Tsujimoto H, Ogihara Y (1997) High-resolution RFLP mapping of the fertility restoration *Rf3* gene against *Triticum timopheevi* cytoplasm located on chromosome 1BS of common wheat. Genes Genet Syst 72:353–359

Komori T, Ohta S, Murai N, Takakura Y, Kuraya Y, Suzuki S, Hiei Y, Imaseki H, Nitta N (2004) Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). Plant J 37:315–325

KOMUGI (2016) Wheat genetic resource database. Shared information of genetic resources (SHIGEN). http://shigen.nig.ac.jp. Accessed 1 Feb 2016

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eug 12:172–175

Krasileva KV, Buffalo V, Bailey P, Pearce S, Ayling S, Tabbita F, Soria M, Wang S, Akhunov E, Uauy C, Dubcovsky J (2013) Separating homeologs by phasing in the tetraploid wheat transcriptome. Genome Biol 14:R66

Kučera L (1982) Monosomic analysis of fertility restoration in common wheat 'Prof. Marchal'. Euphytica 31:895–900

Kumar J, Mir RR, Kumar N, Kumar A, Mohan A, Prabhu KV, Balyan HS, Gupta PK (2010) Marker-assisted selection for pre-harvest sprouting tolerance and leaf rust resistance in bread wheat. Plant Breeding 129:617–621

La Rota M, Sorrells ME (2004) Comparative DNA sequence analysis of mapped wheat ESTs reveals the complexity of genome relationships between rice and wheat. Funct Integr Genomics 4:34–46

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Langridge P (2017) Introduction. In: Langridge P (ed) Achieving sustainable cultivation of wheat. Volume 1: Breeding, quality traits, pests and diseases. Burleigh Dodds Science Publishing, Cambridge, pp XVIII–XIX

Larièpe A, Mangin B, Jasson S, Combes V, Dumas F, Jamin P, Lariagon C, Jolivot D, Madur D, Fiévet J, Gallais A, Dubreuil P, Charcosset A, Moreau L (2012) The genetic basis of heterosis: multiparental quantitative trait loci mapping reveals contrasted levels of apparent overdominance among traits of agronomical interest in maize (*Zea mays* L.). Genetics 190:795–811

Lawrie RG, Matus-Cádiz MA, Hucl P (2006) Estimating out-crossing rates in spring wheat cultivars using the contact method. Crop Sci 46:247–249

Leister D (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends in Genet 20:116–122

Lev-Yadun S, Gopher A, Abbo S (2000) The cradle of agriculture. Science 288:1602–1603

Ling HQ, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y, Gao C, Wu H, Li Y, Cui Y, Guo X, Zheng S, Wang B, Yu K, Liang Q, Yang W, Lou X, Chen J, Feng M, Jian J, Zhang X, Luo G, Jiang Y, Liu J, Wang Z, Sha Y, Zhang B, Wu H, Tang D, Shen Q, Xue P, Zou S, Wang X, Liu X, Wang F, Yang Y, An X, Dong Z, Zhang K, Zhang X, Luo MC, Dvorak J, Tong Y, Wang J, Yang H, Li Z, Wang D, Zhang A, Wang J (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496:87–90

Liu B, Asseng S, Müller C, Ewert F, Elliott J, Lobell DB, Martre P, Ruane AC, Wallach D, Jones JW, Rosenzweig C, Aggarwal PK, Alderman PD, Anothai J, Basso B, Biernath C, Cammarano D, Challinor A, Deryng D, Sanctis GD, Doltra J, Fereres E, Folberth C, Garcia-Vila M, Gayler S, Hoogenboom G, Hunt LA, Izaurralde RC, Jabloun M, Jones CD, Kersebaum KC, Kimball BA, Koehler AK, Kumar SN, Nendel

C, O'Leary GJ, Olesen JE, Ottman MJ, Palosuo T, Prasad PVV, Priesack E, Pugh TAM, Reynolds M, Rezaei EE, Rötter RP, Schmid E, Semenov MA, Shcherbak I, Stehfest E, Stöckle CO, Stratonovitch P, Streck T, Supit I, Tao F, Thorburn P, Waha K, Wall GW, Wang E, White JW, Wolf J, Zhao Z, Zhu Y (2016a) Similar estimates of temperature impacts on global wheat yield by three independent methods. Nat Clim Change 6:1130–1136

Liu F, Cui X, Horner HT, Weiner H, Schnable PS (2001) Mitochondrial aldehyde dehydrogenase activity is required for male fertility in maize. Plant Cell 13:1063–1078

Liu Z, Yang Z, Wang X, Li K, An H, Liu J, Yang G, Fu T, Yi B, Hong D (2016b) A mitochondria-targeted PPR protein restores *pol* cytoplasmic male sterility by reducing *orf224* transcript levels in oilseed rape. Mol Plant 9:1082–1084

Livers RW (1964) Fertility restoration and its inheritance in cytoplasmic male-sterile wheat. Science 144:420

Longin CFH, Gowda M, Mühleisen J, Ebmeyer E, Kazman E, Schachschneider R, Schacht J, Kirchhoff M, Zhao Y, Reif JC (2013) Hybrid wheat: quantitative genetic parameters and consequences for the design of breeding programs. Theor Appl Genet 126:2791–2801

Longin CFH, Mühleisen J, Maurer HP, Zhang H, Gowda M, Reif JC (2012) Hybrid breeding in autogamous cereals. Theor Appl Genet 125:1087–1096

Longin CFH, Reif JC, Würschum T (2014) Long-term perspective of hybrid versus line breeding in wheat based on quantitative genetic theory. Theor Appl Genet 127:1635–1641

Lu C, Shen L, Tan Z, Xu Y, He P, Chen Y, Zhu L (1996) Comparative mapping of QTLs for agronomic traits of rice across environments using a doubled haploid population. Theor Appl Genet 93:1211–1217

Luo MC, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, Huo N, Zhu T, Wang L, Wang Y, McGuire PE, Liu S, Long H, Ramasamy RK, Rodriguez JC, Van SL, Yuan L, Wang Z, Xia Z, Xiao L, Anderson OD, Ouyang S, Liang Y, Zimin AV, Pertea G, Qi P, Bennetzen JL, Dai X, Dawson MW, Müller HG, Kugler K, Rivarola-Duarte L, Spannagl M, Mayer KFX, Lu FH, Bevan MW, Leroy P, Li P, You FM, Sun Q, Liu Z, Lyons E, Wicker T, Salzberg SL, Devos KM, Dvořák J (2017) Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. Nature 551:498–502

Ma ZQ, Sorrells ME (1995) Genetic analysis of fertility restoration in wheat using restriction fragment length polymorphisms. Crop Sci 35:1137–1143

Ma ZQ, Zhao YH, Sorrells ME (1995) Inheritance and chromosomal locations of male fertility restoring gene transferred from *Aegilops umbellulata* Zhuk. to *Triticum aestivum* L. Mol Gen Genet 247:351–357

Maan SS (1985) Genetic analyses of male-fertility restoration in wheat. II. Isolation, penetrance, and expressivity of *Rf* genes. Crop Sci 25:743–748

Maan SS, Luchen KA, Bravo JM (1984) Genetic analyses of male-fertility restoration in wheat. I. Chromosomal location of *Rf* genes. Crop Sci 24:17–20

Maccaferri M, Ricci A, Salvi S, Milner SG, Noli E, Martelli PL, Casadio R, Akhunov E, Scalabrin S, Vendramin V, Ammar K, Blanco A, Desiderio F, Distelfeld A, Dubcovsky J, Fahima T, Faris J, Korol A, Massi A, Mastrangelo AM, Morgante M, Pozniak C, N'Diaye A, Xu S, Tuberosa R (2015a) A high-density, SNP-based consensus map of tetraploid wheat as a bridge to integrate durum and bread wheat genomics and breeding. Plant Biotechnol J 13:648–663

Maccaferri M, Zhang J, Bulli P, Abate Z, Chao S, Cantu D, Bossolini E, Chen X, Pumphrey M, Dubcovsky J (2015b) A genome-wide association study of resistance to stripe rust (*Puccinia striiformis* f. sp. *tritici*) in a worldwide collection of hexaploid spring wheat (*Triticum aestivum* L.). G3 5:449–465

Mangin B, Cathelin R, Delannoy D, Escalière B, Lambert S, Marcel J, Ngom B, Jourjon MF, Rahmani A, Lerous D, Jasson S (2014) MCQTL: A reference manual. Institut national de la recherche agronomique (INRA), Toulouse

Manna S (2015) An overview of pentatricopeptide repeat proteins and their applications. Biochimie 113:93–99

Mantle PG, Swan DJ (1995) Effect of male sterility on ergot disease spread in wheat. Plant Pathol 44:392–395

Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, Jakobsen KS, Wulff BBH, Steuernagel B, Mayer KFX, Olsen OA, Rogers J, Doležel J, Pozniak C, Eversole K, Feuillet C, Gill B, Friebe B, Lukaszewski AJ, Sourdille P, Endo TR, Kubaláková M, Číhalíková J, Dubská Z, Vrána J, Šperková R, Šimková H, Febrer M, Clissold L, McLay K, Singh K, Chhuneja P, Singh NK, Khurana J, Akhunov E, Choulet F, Alberti A, Barbe V, Wincker P, Kanamori H, Kobayashi F, Itoh T, Matsumoto T, Sakai H, Tanaka T, Wu J, Ogihara Y, Handa H, Maclachlan PR, Sharpe A, Klassen D, Edwards D, Batley J, Lien S, Caccamo M, Ayling S, Ramirez-Gonzalez RH, Clavijo BJ, Wright J, Martis MM, Mascher M, Chapman J, Poland JA, Scholz U, Barry K, Waugh R, Rokhsar DS, Muehlbauer GJ, Stein N, Gundlach H, Zytnicki M, Jamilloux V, Quesneville H, Wicker T, Faccioli P, Colaiacovo M, Stanca AM, Budak H, Cattivelli L, Glover N, Pingault L, Paux E, Sharma S, Appels R, Bellgard M, Chapman B, Nussbaumer T, Bader KC, Rimbert H, Wang S, Knox R, Kilian A,

Alaux M, Alfama F, Couderc L, Guilhot N, Viseux C, Loaec M, Keller B, Praud S (2014) Ancient hybridizations among the ancestral genomes of bread wheat. Science 345:1250092

Mehrajuddin, Salgotra RK, Gupta BB (2013) Interaction of restorer genes in 'WA'-type cytoplasmic male sterility system in rice (*Oryza sativa* L.). Natl Acad Sci Lett 36:259–264

Melchinger AE, Utz HF, Schön CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149:383–403

Menassa R, L'Homme Y, Brown GG (1999) Post-transcriptional and developmental regulation of a CMS-associated mitochondrial gene region by a nuclear restorer gene. Plant J 17:491–499

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Michel S, Löschenberger F, Ametz C, Pachler B, Sparry E and Bürstmayr H (2019) Simultaneous selection for grain yield and protein content in genomics-assisted wheat breeding. Theor Appl Genet. Advance online publication. https://doi.org/10.1007/s00122-019-03312-5

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8:1113–1130

Miedaner T, Glass C, Dreyer F, Wilde P, Wortmann H, Geiger HH (2000) Mapping of genes for male-fertility restoration in 'Pampa' CMS winter rye (*Secale cereale* L.). Theor Appl Genet 101:1226–1233

Miller JF, Schmidt JW, Johnson VA (1974) Inheritance of genes controlling male-fertility restoration in the wheat cultivar Primépi. Crop Sci 14:437–438

Mohr S, Schulte-Kappert E, Odenbach W, Oettler G, Kück U (1993) Mitochondrial DNA of cytoplasmic male-sterile *Triticum timopheevi*: rearrangement of upstream sequences of the *atp6* and *orf25* genes. Theor Appl Genet 86:259–268

Muqaddasi QH, Pillen K, Plieske J, Ganal MW, Röder MS (2017) Genetic and physical mapping of anther extrusion in elite European winter wheat. PLOS ONE 12:e0187744

Muranty H (1996) Power of tests for quantitative trait loci detection using full-sib families in different schemes. Heredity 76:156–165

Mühleisen J, Piepho HP, Maurer HP, Longin CFH, Reif JC (2014) Yield stability of hybrids versus lines in wheat, barley, and triticale. Theor Appl Genet 127:309–316

Nath J, Thompson JP, Gulati SC (1985) Identification of the G-genome donor to *Triticum timopheevii* by DNA:DNA hybridizations. Biochem Genet 23:125–137

NCBI (2019) National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/unigene. Accessed 4 Mar 2019

Ndolo V, Beta T (2017) The nutritional and nutraceutical value of wheat. In: Langridge P (ed) Achieving sustainable cultivation of wheat. Volume 1: Breeding quality traits, pests and diseases. Burleigh Dodds Science Publishing, Cambridge, pp 231–260

Nivison HT, Hanson MR (1989) Identification of a mitochondrial protein associated with cytoplasmic male sterility in petunia. Plant Cell 1:1121–1130

Odenbach W (1970) Versuche zur Wiederherstellung der Fertilität pollensteriler hexaploider Weizen. Z Pflanzenzüchtg 64:73–89

Oehler E, Ingold M (1966) New cases of male-sterility and new restorer source in *T. aestivum*. Wheat Inf Serv 22:1–3

Okada A, Arndell T, Borisjuk N, Sharma N, Watson-Haigh NS, Tucker EJ, Baumann U, Langridge P and Whitford R (2019) CRISPR/Cas9-mediated knock-out of *Ms1* enables the rapid generation of male-sterile hexaploid wheat lines for use in hybrid seed production. Plant Biotechnol J. Advance online publication. https://doi.org/10.1111/pbi.13106

Okada T, Jayasinghe JEARM, Nansamba M, Baes M, Warner P, Kouidri A, Correia D, Nguyen V, Whitford R, Baumann U (2018) Unfertilized ovary pushes wheat flower open for cross-pollination. J Exp Bot 69:399–412

Oury FX, Brabant P, Bérard P, Pluchard P (2000) Predicting hybrid value in bread wheat: biometric modelling based on a "top-cross" design. Theor Appl Genet 100:96–104

Panayotov I, Savov M, Gotsov K (1975) Monosomic analysis of the fertility restorer R 130 (Palmaress) for the cytoplasm of *Triticum timopheevi* Zhuk. C R Acad Bulg Sci 28:1685–1687

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290

Patterson FL, Ohm HW, Johnson JW, Wickersham DS (1996) Registration of five wheat pollen fertility restorer germplasm lines: PR143, PR189, PR267, PR270, PR302. Crop Sci 36:1424

Paux E, Roger D, Badaeva E, Gay G, Bernard M, Sourdille P, Feuillet C (2006) Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. Plant J 48:463–474

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeyer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S, Bergès H, Eversole K, Appels R, Safar J, Simkova H, Dolezel J, Bernard M, Feuillet C (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. Science 322:101–104

Perenzin M, Corbellini M, Accerbi M, Vaccino P, Borghi B (1998) Bread wheat: F1 hybrid performance and parental diversity estimates using molecular markers. Euphytica 100:273–279

Piepho HP (1998) Methods for Comparing the Yield Stability of Cropping Systems. J Agron Crop Sci 180:193–213

Plaschke J, Ganal MW, Röder MS (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet 91:1001–1007

Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M, Jannink JL (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome 5:103–113

Pugsley AT, Oram RN (1959) Genic male sterility in wheat. Austr Plant Breeding Genet Newsl 14:10–11

Randhawa HS, Mutti JS, Kidwell K, Morris CF, Chen X, Gill KS (2009) Rapid and targeted introgression of genes into popular wheat cultivars using marker-assisted background selection. PLOS ONE 4:e5752

Rathburn HB, Hedgcoth C (1991) A chimeric open reading frame in the 5' flanking region of *coxI* mitochondrial DNA from cytoplasmic male-sterile wheat. Plant Mol Biol 16:909–912

Ray DK, Mueller ND, West PC, Foley JA (2013) Yield trends are insufficient to double global crop production by 2050. PLOS ONE 8:e66428

R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Reif JC, Gumpert FM, Fischer S, Melchinger AE (2007) Impact of interpopulation divergence on additive and dominance variance in hybrid populations. Genetics 176:1931–1934

Reif JC, Melchinger AE, Frisch M (2005) Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. Crop Sci 45:1–7

Robertson LD, Curtis BC (1967) Monosomic analysis of fertility-restoration in common wheat (*Triticum aestivum* L.). Crop Sci 7:493–495

Rutkoski JE, Poland JA, Singh RP, Huerta-Espino J, Bhavani S, Barbier H, Rouse MN, Jannink JL, Sorrells ME (2014) Genomic selection for quantitative adult plant stem rust resistance in wheat. Plant Genome 7

Sage GC (1972) The inheritance of fertility restoration in male-sterile wheat carrying cytoplasm derived from *Triticum timopheevi*. Theor Appl Genet 42:233–243

Salse J, Chagué V, Bolot S, Magdelenat G, Huneau C, Pont C, Belcram H, Couloux A, Gardais S, Evrard A, Segurens B, Charles M, Ravel C, Samain S, Charmet G, Boudet N, Chalhoub B (2008) New insights into the origin of the B genome of hexaploid wheat: evolutionary relationships at the SPA genomic region with the S genome of the diploid relative *Aegilops speltoides*. BMC Genomics 9:555

Schmidt JW, Johnson VA, Maan SS (1962) Hybrid-wheat. Nebr Exp Stat Quart 9:9

Schön CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE (2004) Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. Genetics 167:485–498

Shewry PR (2009) Wheat. J Exp Bot 60:1537–1553

Shull GH (1952) Beginnings of the heterosis concept. In: Gown JW (ed) Heterosis. Iowa State College Press, Ames, pp 14–48

Sinha P, Tomar SMS, Vinod, Singh VK, Balyan HS (2013) Genetic analysis and molecular mapping of a new fertility restorer gene *Rf8* for *Triticum timopheevi* cytoplasm in wheat (*Triticum aestivum* L.) using SSR markers. Genetica 141:431–441

Snape J, Pánková K (2007) *Triticum aestivum* (Wheat). eLS

Somers DJ, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor Appl Genet 109:1105–1114

Song J, Carver BF, Powers C, Yan L, Klápště J, El-Kassaby YA, Chen C (2017) Practical application of genomic selection in a doubled-haploid winter wheat breeding program. Mol Breeding 37:117

Song J, Hedgcoth C (1994) A chimeric gene (*orf256*) is expressed as protein only in cytoplasmic male-sterile lines of wheat. Plant Mol Biol 26:535–539

Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin, Mahmoud A, Ma X, Gustafson PJ, Qi LL, Echalier B, Gill BS, Matthews DE, Lazo GR, Chao S, Anderson OD, Edwards H, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorak J, Zhang D, Nguyen HT, Peng J, Lapitan NLV, Gonzalez-Hernandez JL, Anderson JA, Hossain K, Kalavacharla V, Kianian SF, Choi DW, Close TJ, Dilbirligi M, Gill KS, Steber C, Walker-Simmons MK, McGuire PE,

Qualset CO (2003) Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 13:1818–1827

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: Join Map. Plant J 3:739–744

Tahir CM, Tsunewaki K (1969) Monosomic analysis of *Triticum spelta* var. *duhamelianum*, a fertility-restorer for *T. timopheevi* cytoplasm. Jpn J Genet 44:1–9

Tahir CM, Tsunewaki K (1971) Distribution of fertility-restoring genes in hexaploid wheats for *Ae. ovata* and *T. timopheevi* cytoplasms. Jpn J Breed 21:49–51

Tang H, Luo D, Zhou D, Zhang Q, Tian D, Zheng X, Chen L, Liu YG (2014) The rice restorer *Rf4* for wild-abortive cytoplasmic male sterility encodes a mitochondrial-localized PPR protein that functions in reduction of *WA352* transcripts. Mol Plant 7:1497–1500

Tang HV, Pedersen JF, Chase CD, Pring DR (2007) Fertility restoration of the sorghum A3 male-sterile cytoplasm through a sporophytic mechanism derived from sudangrass. Crop Sci 47:943–950

The Triticeae Toolbox (2016) Wheat Coordinated Agricultural Project. https://triticeaetoolbox.org. Accessed 1 Feb 2016

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

Tilman D, Balzer C, Hill J, Befort BL (2011) Global food demand and the sustainable intensification of agriculture. Proc Natl Acad Sci U S A 108:20260–20264

Tomar RSS, Deshmukh RK, Naik K B, Tomar SMS, Vinod (2014) Development of chloroplast-specific microsatellite markers for molecular characterization of alloplasmic lines and phylogenetic analysis in wheat. Plant Breeding 133:12–18

Tsujimoto H, Tsunewaki K (1984) Gametocidal genes in wheat and its relatives. I. Genetic analyses in common wheat of a gametocidal gene derived from *Aegilops speltoides*. Can J Genet Cytol 26:78–84

Tsunewaki K (2015) Fine mapping of the first multi-fertility-restoring gene, Rf_{multi}, of wheat for three *Aegilops* plasmons, using 1BS-1RS recombinant lines. Theor Appl Genet 128:723–732

Tucker EJ, Baumann U, Kouidri A, Suchecki R, Baes M, Garcia M, Okada T, Dong C, Wu Y, Sandhu A, Singh M, Langridge P, Wolters P, Albertsen MC, Cigan AM, Whitford R (2017) Molecular identification of the wheat male fertility gene *Ms1* and its prospects for hybrid breeding. Nat Commun 8:869

UN DESA (2019) United Nations Department of Economic and Social Affairs. Population Division. https://esa.un.org/unpd/wpp. Accessed 4 Mar 2019

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer 3. New capabilities and interfaces. Nucleic Acids Res 40:e115

URGI (2016) Unité de Recherche Génomique Info. Institut National de la Recherche Agronomique (INRA). https://urgi.versailles.inra.fr. Accessed 1 Feb 2016

Uyttewaal M, Arnal N, Quadrado M, Martin-Canadell A, Vrielynck N, Hiard S, Gherbi H, Bendahmane A, Budar F, Mireau H (2008) Characterization of *Raphanus sativus* pentatricopeptide repeat proteins encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. Plant Cell 20:3331–3345

van Ooijen JW (2006) JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen

Virmani SS, Edwards IB (1983) Current status and future prospects for breeding hybrid rice and wheat. Adv Agron 36:145–214

Vishwakarma MK, Arun B, Mishra VK, Yadav PS, Kumar H, Joshi AK (2016) Marker-assisted improvement of grain protein content and grain weight in Indian bread wheat. Euphytica 208:313–321

Voorrips RE (2002) Software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78

Wang J, Chapman SC, Bonnett DG, Rebetzke GJ, Crouch J (2007) Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. Crop Sci 47:582–588

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, International Wheat Genome Sequencing Consortium, Lillemo M, Mather D, Appels R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo MC, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E (2014) Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. Plant Biotechnol J 12:787–796

Wang Z, Zou Y, Li X, Zhang Q, Chen L, Wu H, Su D, Chen Y, Guo J, Luo D, Long Y, Zhong Y, Liu YG (2006) Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell 18:676–687

Whitford R, Fleury D, Reif JC, Garcia M, Okada T, Korzun V, Langridge P (2013) Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. J Exp Bot 64:5411–5428

Wilson JA (1962) Material prepared through DeKalb Agricultural Assoc. Inc. Wheat Newsl 9:28–29

Wilson JA, Ross WM (1962) Male sterility interaction of the *Triticum aestivum* nucleus and *Triticum timopheevi* cytoplasm. Wheat Inf Serv 14:29–30

Winfield MO, Allen AM, Burridge AJ, Barker GLA, Benbow HR, Wilkinson PA, Coghill J, Waterfall C, Davassi A, Scopes G, Pirani A, Webster T, Brew F, Bloor C, King J, West C, Griffiths S, King I, Bentley AR, Edwards KJ (2016) High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. Plant Biotechnol J 14:1195–1206

Wricke G, Wilde P, Wehling P, Gieselmann C (1993) An isozyme marker for pollen fertility restoration in the Pampa cms system of rye (*Secale cereale* L.). Plant Breeding 111:290–294

Wright S (1978) Evolution and the genetics of populations, volume 4. The University of Chicago Press, Chicago

Wu Y, Fox TW, Trimnell MR, Wang L, Xu RJ, Cigan AM, Huffman GA, Garnaat CW, Hershey H and Albertsen MC (2016) Development of a novel recessive genetic male sterility system for hybrid seed production in maize and other cross-pollinating crops. Plant Biotechnol J 14:1046–1054

Würschum T, Leiser WL, Weissmann S, Maurer HP (2017) 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 130:1253–1266

Xie C, Gessler DDG, Xu S (1998) Combining different line crosses for mapping quantitative trait loci using the identical by descent-based variance component method. Genetics 149:1139–1146

Xu P, Yang Y, Zhang Z, Chen W, Zhang C, Zhang L, Zou S, Ma Z (2008) Expression of the nuclear gene TaF_Ad is under mitochondrial retrograde regulation in anthers of male sterile wheat plants with *timopheevii* cytoplasm. J Exp Bot 59:1375–1381

Xu S (1998) Mapping quantitative trait loci using multiple families of line crosses. Genetics 148:517–524

Yadav PS, Mishra VK, Arun B, Chand R, Vishwakarma MK, Vasistha NK, Mishra AN, Kalappanavar IK, Joshi AK (2015) Enhanced resistance in wheat against stem rust achieved by marker assisted backcrossing involving three independent *Sr* genes. Curr Plant Biol 2:25–33

Yadawad A, Gadpale A, Hanchinal RR, Nadaf HL, Desai SA, Biradar S, Naik VR (2017) Pyramiding of leaf rust resistance genes in bread wheat variety DWR 162 through marker assisted backcrossing. Indian J Genet Pl Br 77:251–257

Yen FS, Evans LE, Larter EN (1969) Monosomic analysis of fertility restoration in three restorer lines of wheat. Can J Genet Cytol 11:531–546

Zanke CD, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, Argillier O, Stiewe G, Hinze M, Neumann K, Ganal MW, Röder MS (2014) Whole genome association mapping of plant height in winter wheat (*Triticum aestivum* L.). PLOS ONE 9:e113287

Zeileis A (2004) Econometric computing with HC and HAC covariance matrix estimators. J Stat Softw 11:1–17

Zeileis A, Hothorn T (2002) Diagnostic checking in regression relationships. R News 2:7–10

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

Zeven AC (1967) Transfer and inactivation of male sterility and sources of restorer genes in wheat. Euphytica 16:183–189

Zhao Y, Li Z, Liu G, Jiang Y, Maurer HP, Würschum T, Mock HP, Matros A, Ebmeyer E, Schachschneider R, Kazman E, Schacht J, Gowda M, Longin CFH, Reif JC (2015) Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. Proc Natl Acad Sci U S A 112:15624–15629

Zhao Y, Mette MF, Gowda M, Longin CFH, Reif JC (2014) Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. Heredity 112:638–645

Zhao Y, Zeng J, Fernando R, Reif JC (2013) Genomic prediction of hybrid wheat performance. Crop Sci 53:802–810

Zhou K, Wang S, Feng Y, Ji W, Wang G (2008) A new male sterile mutant LZ in wheat (*Triticum aestivum* L.). Euphytica 159:403–410

Zhou R, Himmelbach A, Hartl L, Stein N (2017) Target enrichment sequencing of wheat pentatricopeptide repeat (PPR) gene family - Searching wheat fertility restorer gene candidates. In: Buerstmayr H, Lang-Mladek C, Steiner B, Michel S, Buerstmayr M, Lemmens M, Vollmann J, Grausgruber H (eds) Proceedings of the 13th international wheat genetics symposium. University of Natural Resources and Life Sciences, Vienna, p 204

8 Appendix

Table A1 Number of array-derived single-nucleotide polymorphism (SNP) markers discarded at six selection steps. For PopPrimepi, only 40 individuals were genotyped using the SNP array. Duplicate markers were not discarded for this population

	PopR3	PopR113	PopL19	PopPrimepi
Initial marker number	17267	17267	17267	13006
Missing for parents	1197	1231	1109	119
Monomorphic for parents	9769	9361	9585	8695
Heterozygous for parents	39	58	30	13
Missing genotypes > 10%	1240	1388	1430	484
Minor allele frequency < 0.1	196	215	179	1007
Duplicated	2475	2530	2389	0
Map construction	359	418	445	2685
Number of mapped markers	1992	2066	2100	3

Table A2 Specifications of cleaved amplified polymorphic sequence (CAPS) assays developed for *IWB14060* and *IWB72107*. The restoring *Rf3* allele was associated with the SNP alleles G and A, respectively. The annealing temperature is denoted by T_a

CAPS name	Primer sequences	T_a	Enzyme	Fragment size
CAPS_IWB14060	FW: GAGGAGGATCTCGTGGGG REV: GTGGGAGCTCGCCGAATG	57 °C	*BsrDI*	A: 115 bp G: 76 and 39 bp
CAPS_IWB72107	FW: ATGATCTGCTGGACGTGGTC REV: GTCCTTCCTTCCACGTGAGA	58 °C	*HhaI*	A: 167 bp G: 92 and 75 bp

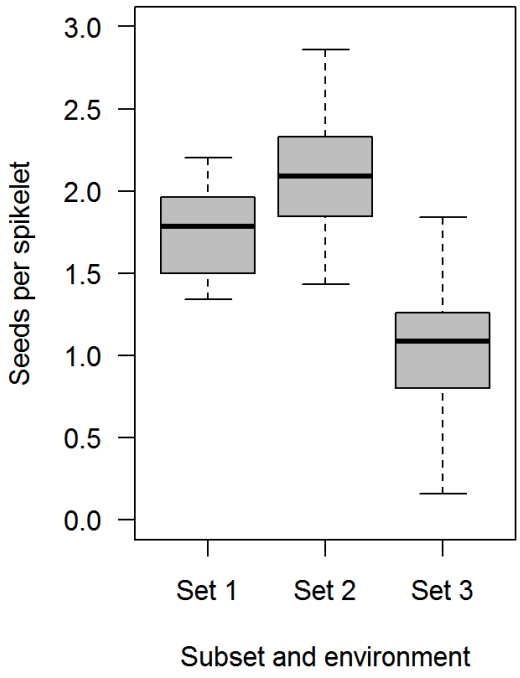

Figure A1 Seed set of the fertile fractions of three subsets of PopPrimepi

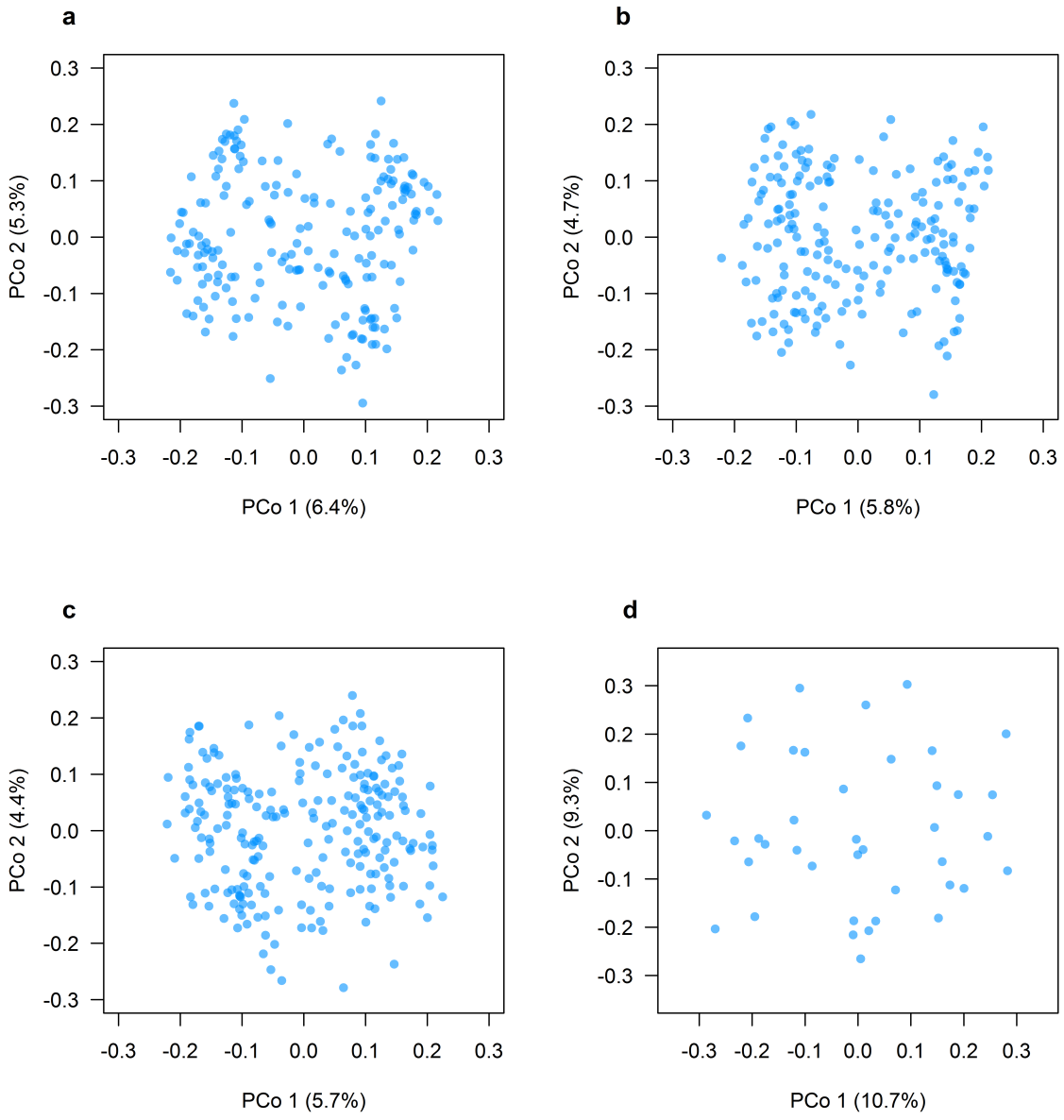

Figure A2 Principal coordinate analysis (PCoA) of the populations PopR3 (**a**), PopR113 (**b**), PopL19 (**c**) and 40 selected individuals of PopPrimepi (**d**)

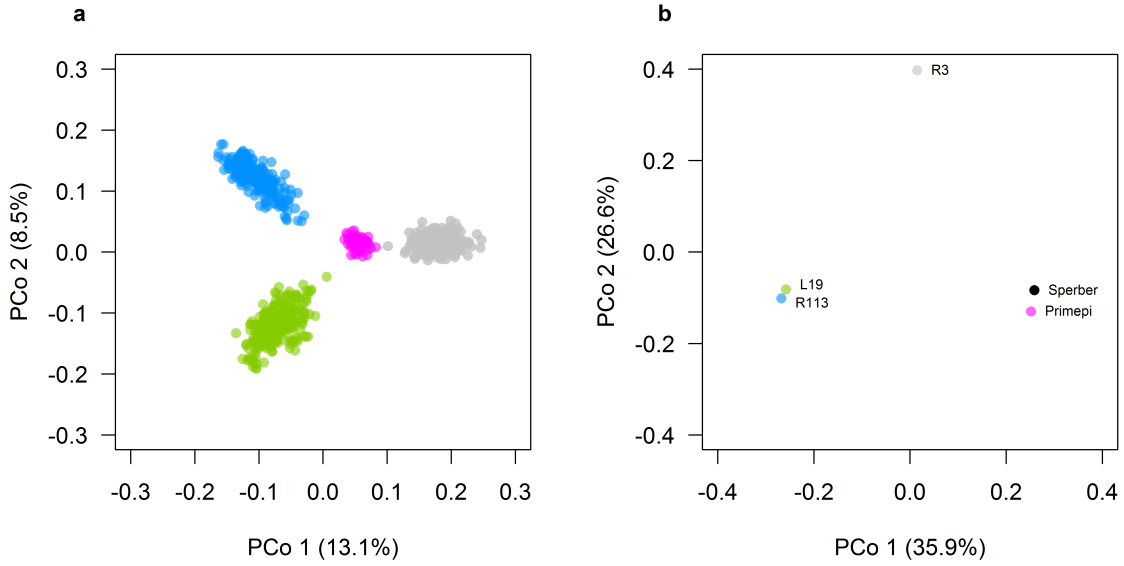

Figure A3 Subfigure **a** illustrates the PCoA of the populations PopR3 (grey), PopR113 (blue), PopL19 (green) and 40 selected individuals of PopPrimepi (magenta). The PCoA of the parental lines is depicted in subfigure **b**

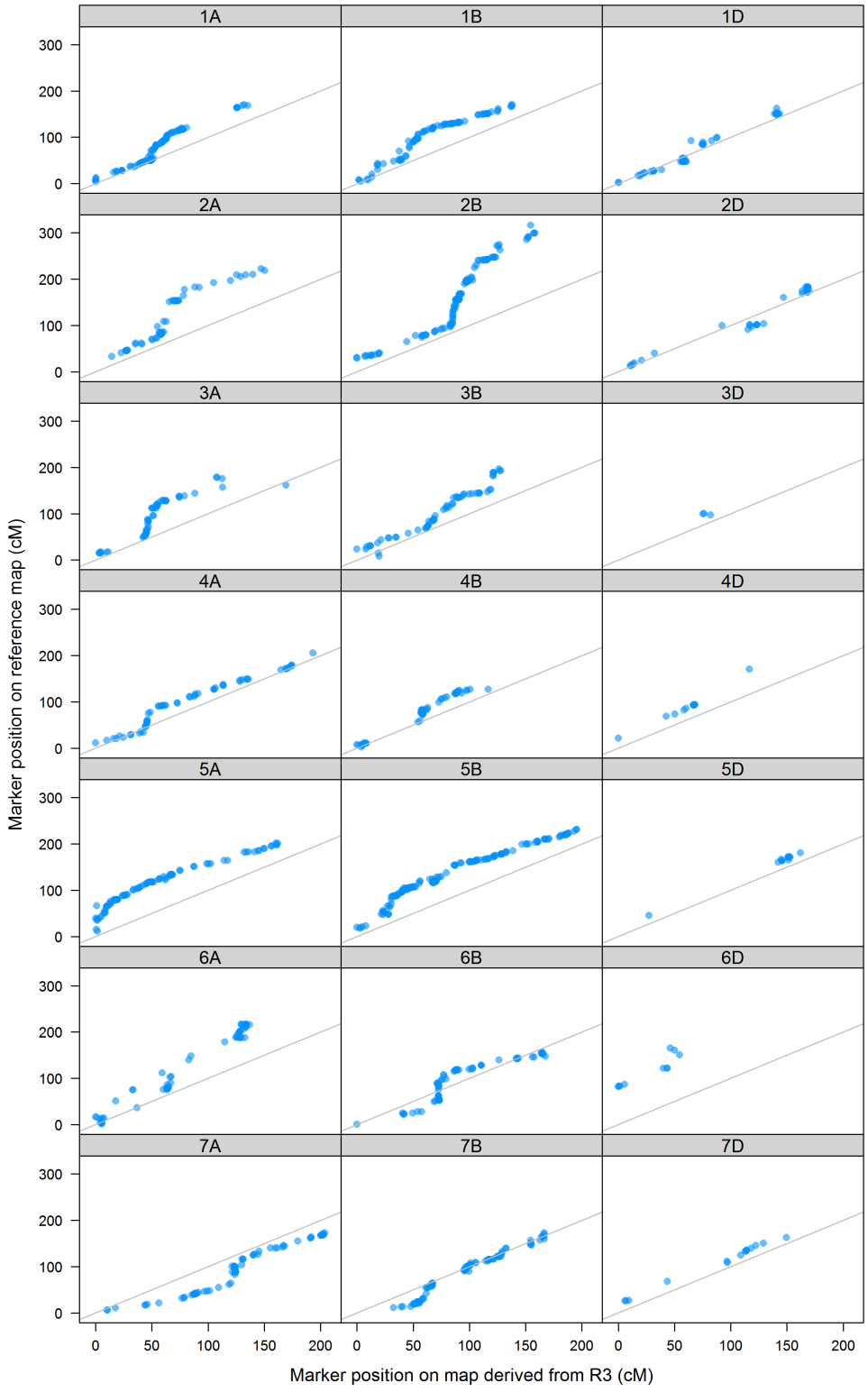

Figure A4 Marker positions of one of the employed reference maps (Maccaferri et al. 2015b) plotted against marker positions of the map derived from PopR3. Map distances between linkage groups of the same chromosome were adopted from the reference map

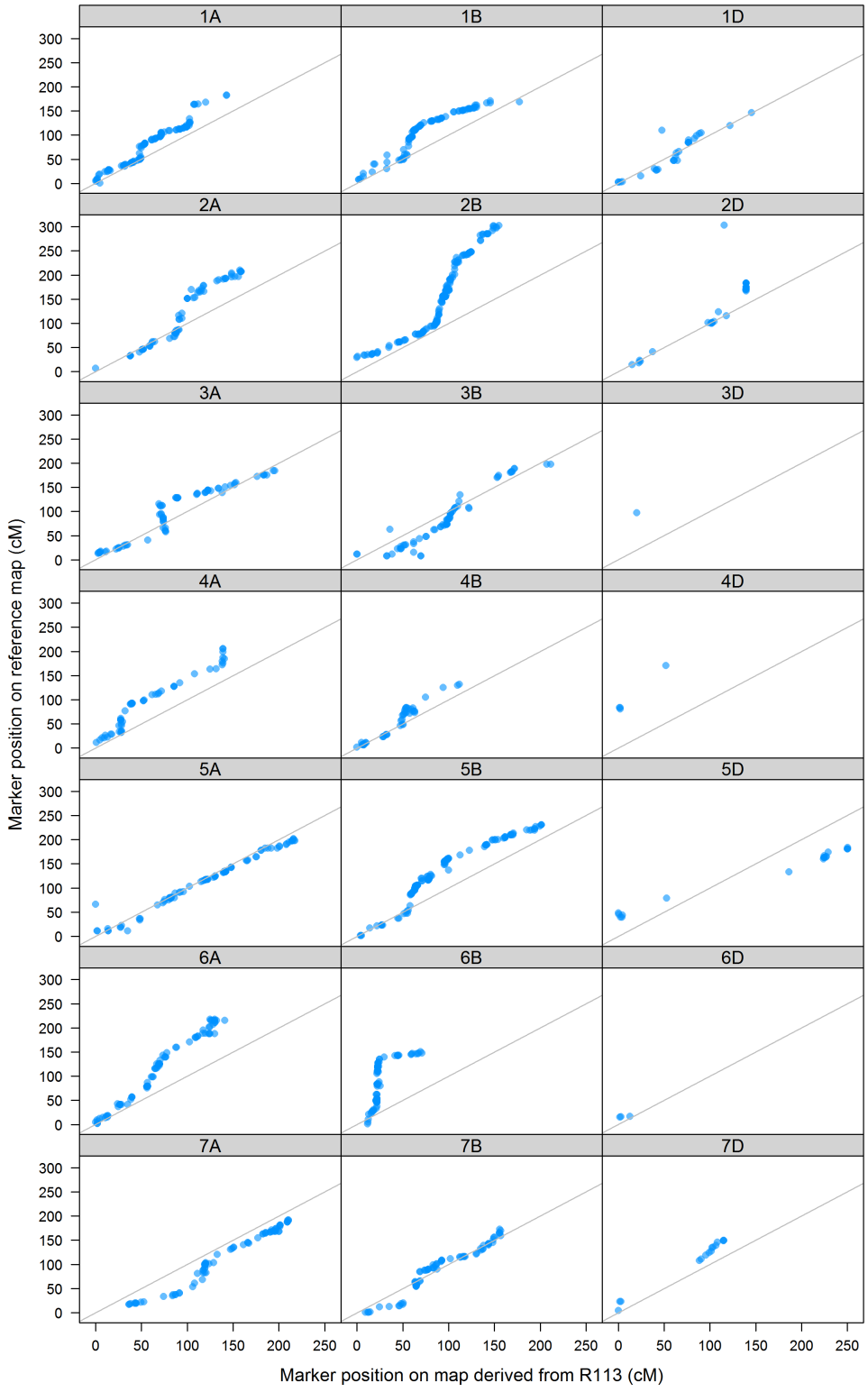

Figure A5 Marker positions of one of the employed reference maps (Maccaferri et al. 2015b) plotted against marker positions of the map derived from PopR113. Map distances between linkage groups of the same chromosome were adopted from the reference map

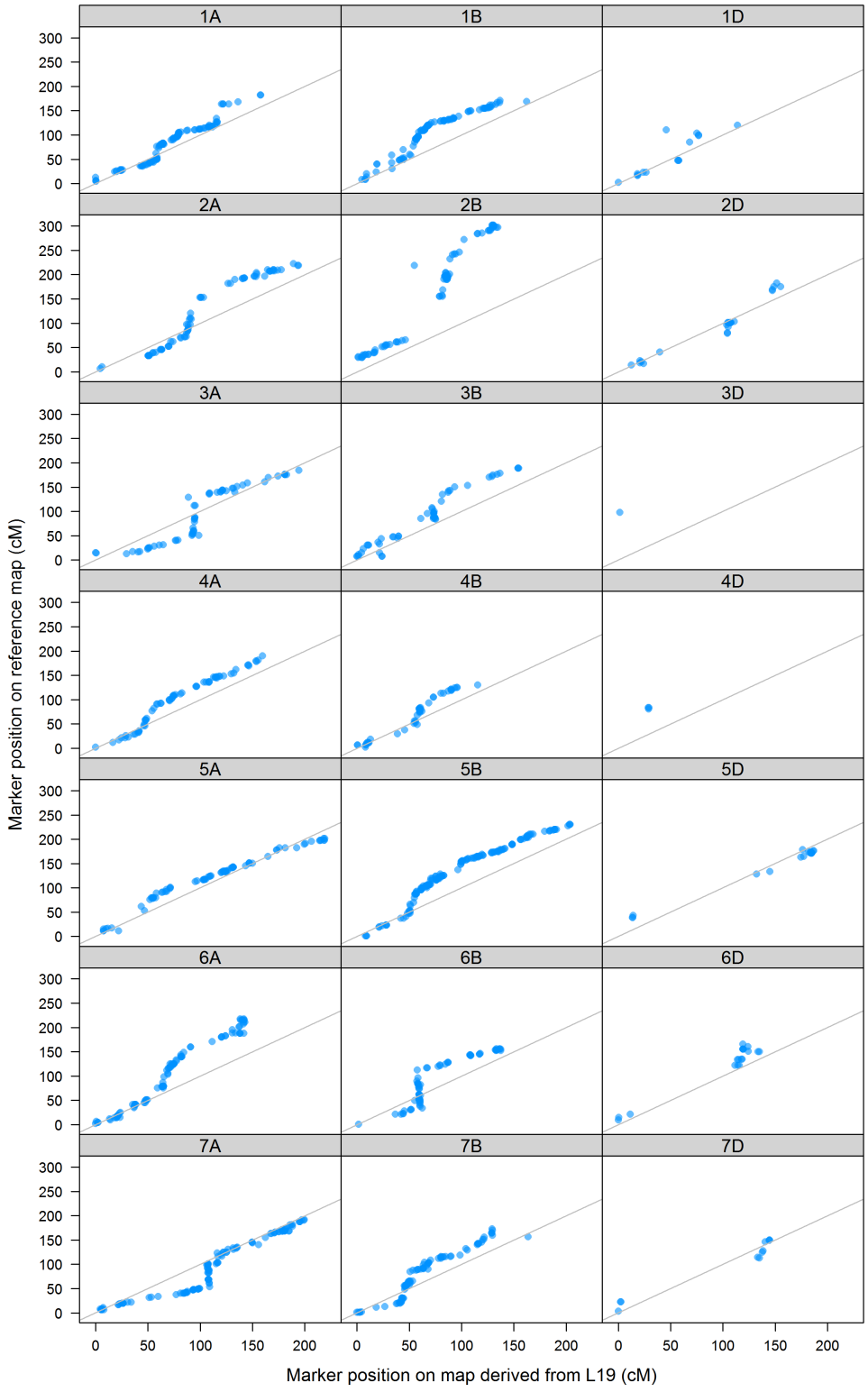

Figure A6 Marker positions of one of the employed reference maps (Maccaferri et al. 2015b) plotted against marker positions of the map derived from PopL19. Map distances between linkage groups of the same chromosome were adopted from the reference map

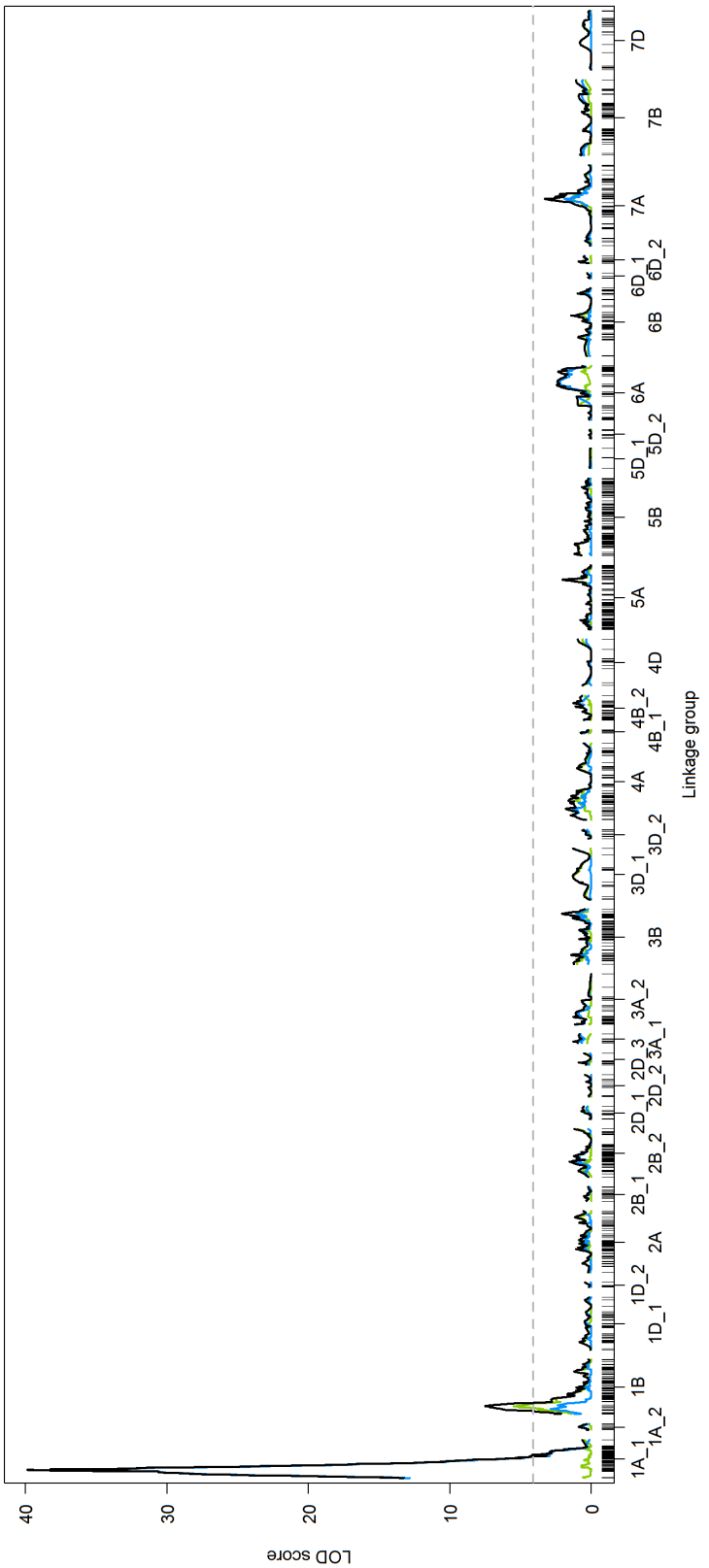

Figure A7 Genome-wide logarithm of the odds (LOD) score curve estimated for PopR3 using the two-part model. LOD scores for the two-part model, the model for the binary phenotype and the model for the quantitative phenotype of the fertile fraction are illustrated by black, blue and green lines, respectively. The dashed, grey line represents the threshold for quantitative trait locus (QTL) detection

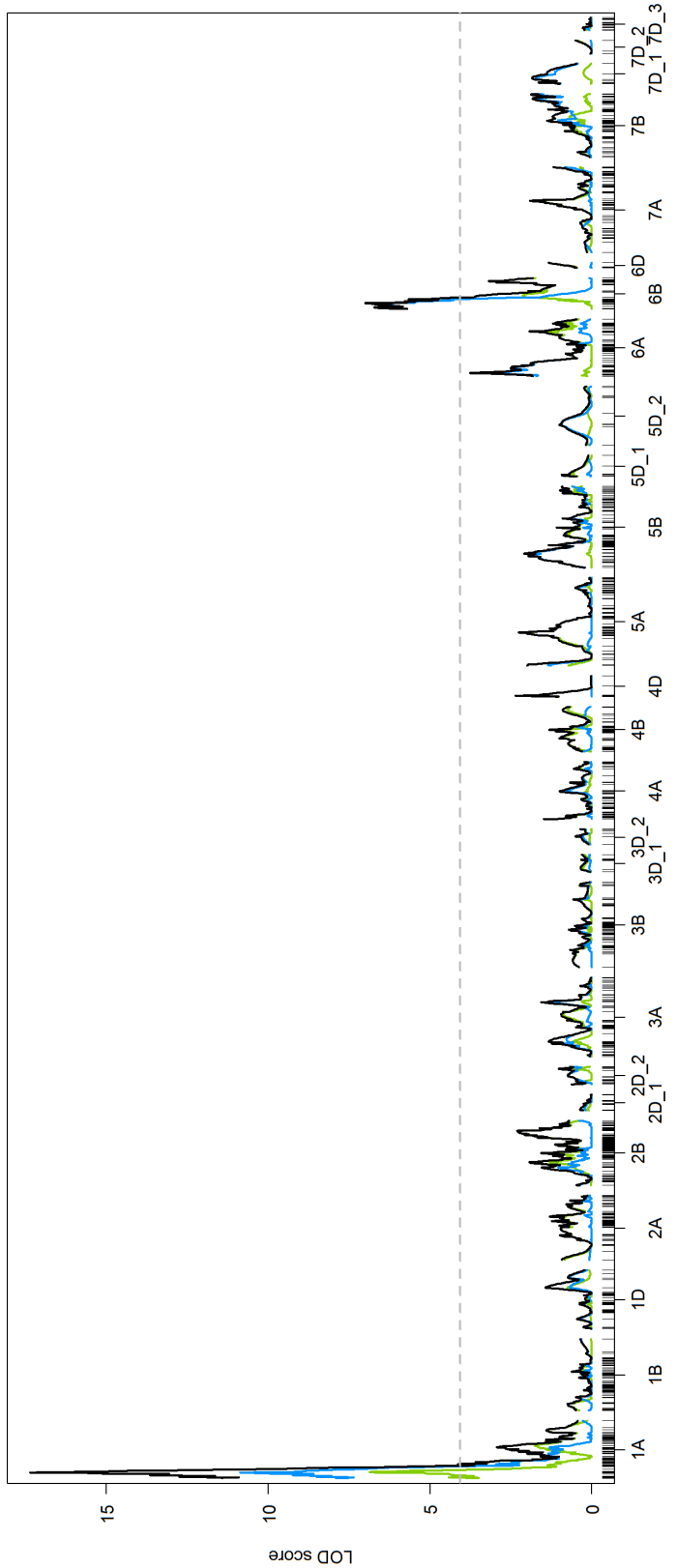

Figure A8 Genome-wide LOD score curve estimated for PopR113 using the two-part model. LOD scores for the two-part model, the model for the binary phenotype and the model for the quantitative phenotype of the fertile fraction are illustrated by black, blue and green lines, respectively. The dashed, grey line represents the threshold for QTL detection

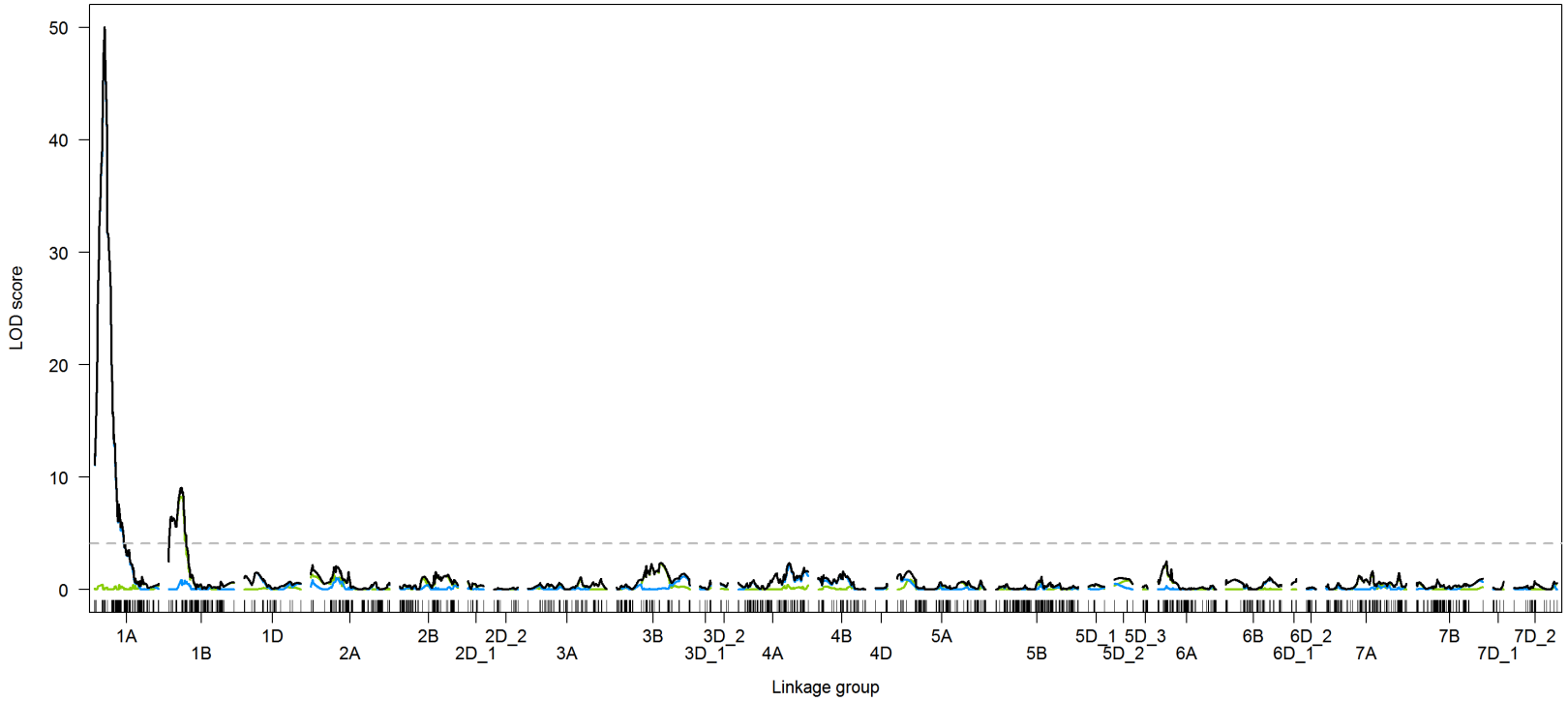

Figure A9 Genome-wide LOD score curve estimated for PopL19 using the two-part model. LOD scores for the two-part model, the model for the binary phenotype and the model for the quantitative phenotype of the fertile fraction are illustrated by black, blue and green lines, respectively. The dashed, grey line represents the threshold for QTL detection

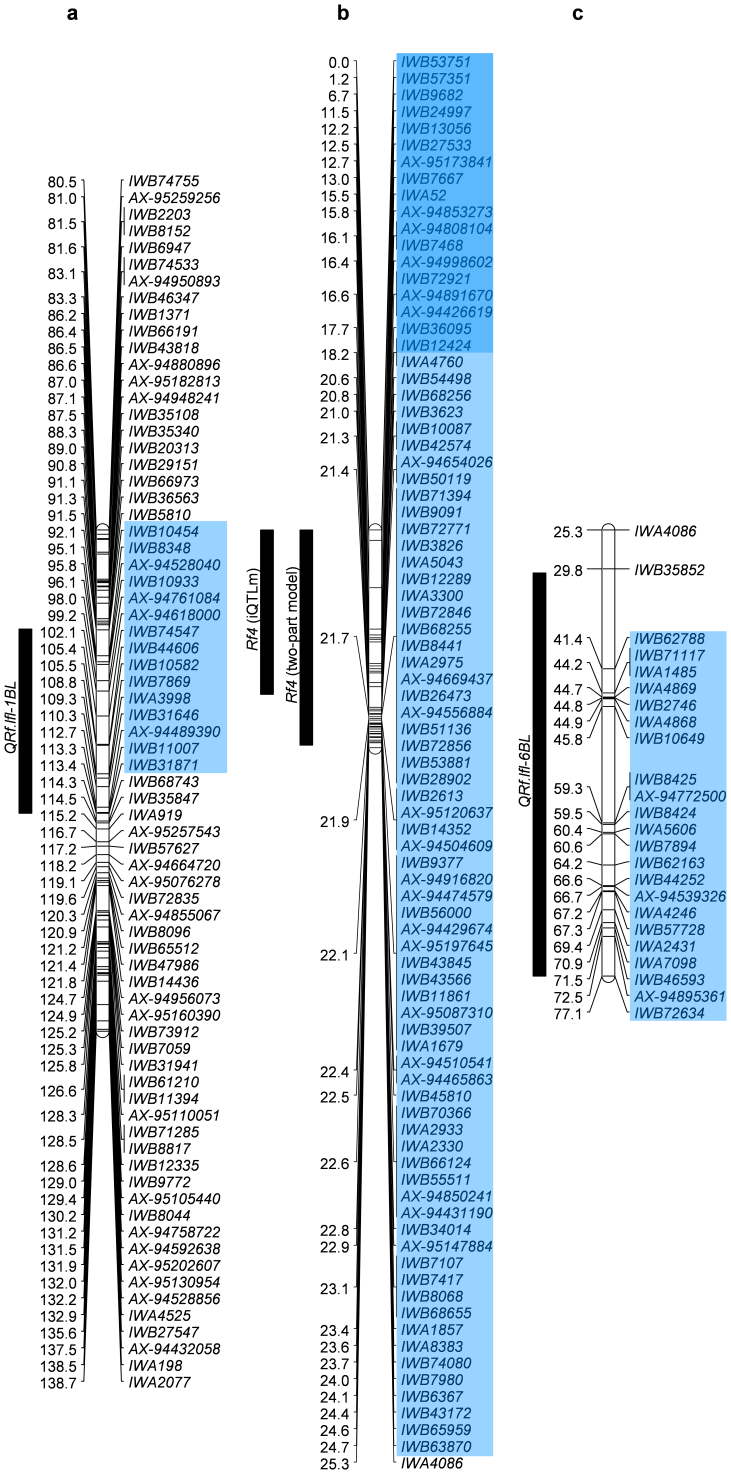

Figure A10 Positions of *QRf.lfl-1BL* (**a**), *Rf4* (**b**) and *QRf.lfl-6BL* (**c**), which were detected on chromosome 1BL in PopR3 (*QRf.lfl-1BL*) and on chromosome 6B in PopR113 (*Rf4*, *QRf.lfl-6BL*). Black bars and blue boxes illustrate 1.5-LOD support intervals, which were estimated using iQTLm (all three loci) and the two-part model (*Rf4*). The dark-blue box depicts overlapping support intervals for *Rf4* using the two QTL detection methods. Map positions are given in centimorgan (cM)

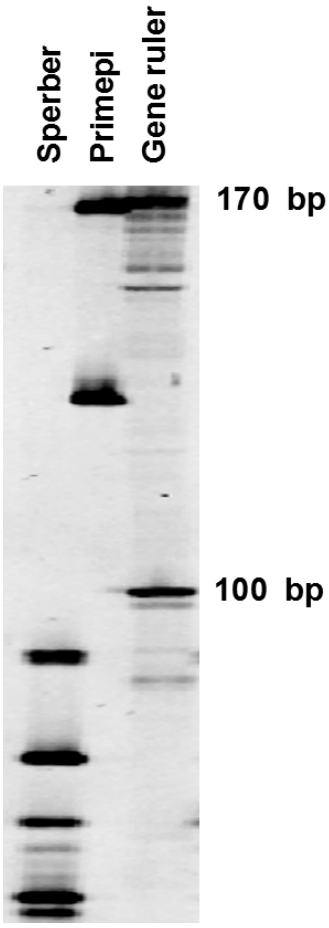

Figure A11 Products of *CAPS_IWB72107* resolved on polyacrylamide gel

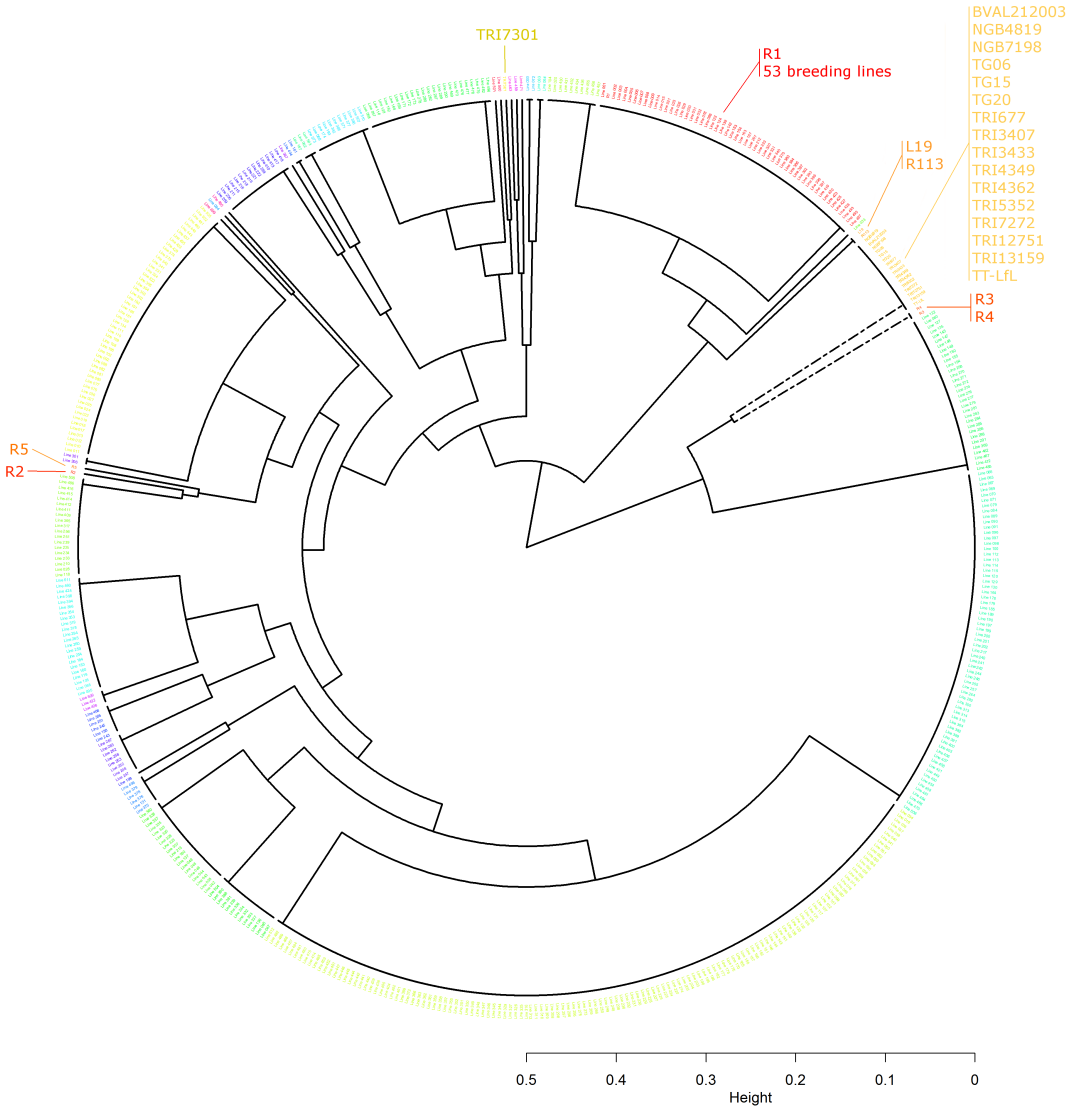

Figure A12 Phylogram for the *QRf.lfl-1BS* marker haplotypes of the *Rf1* donor lines, 17 *Triticum timopheevii* accessions and 507 common wheat breeding lines. Lines under the dashed branches belonged to the same group if the number of groups was restricted to $k \leq 10$

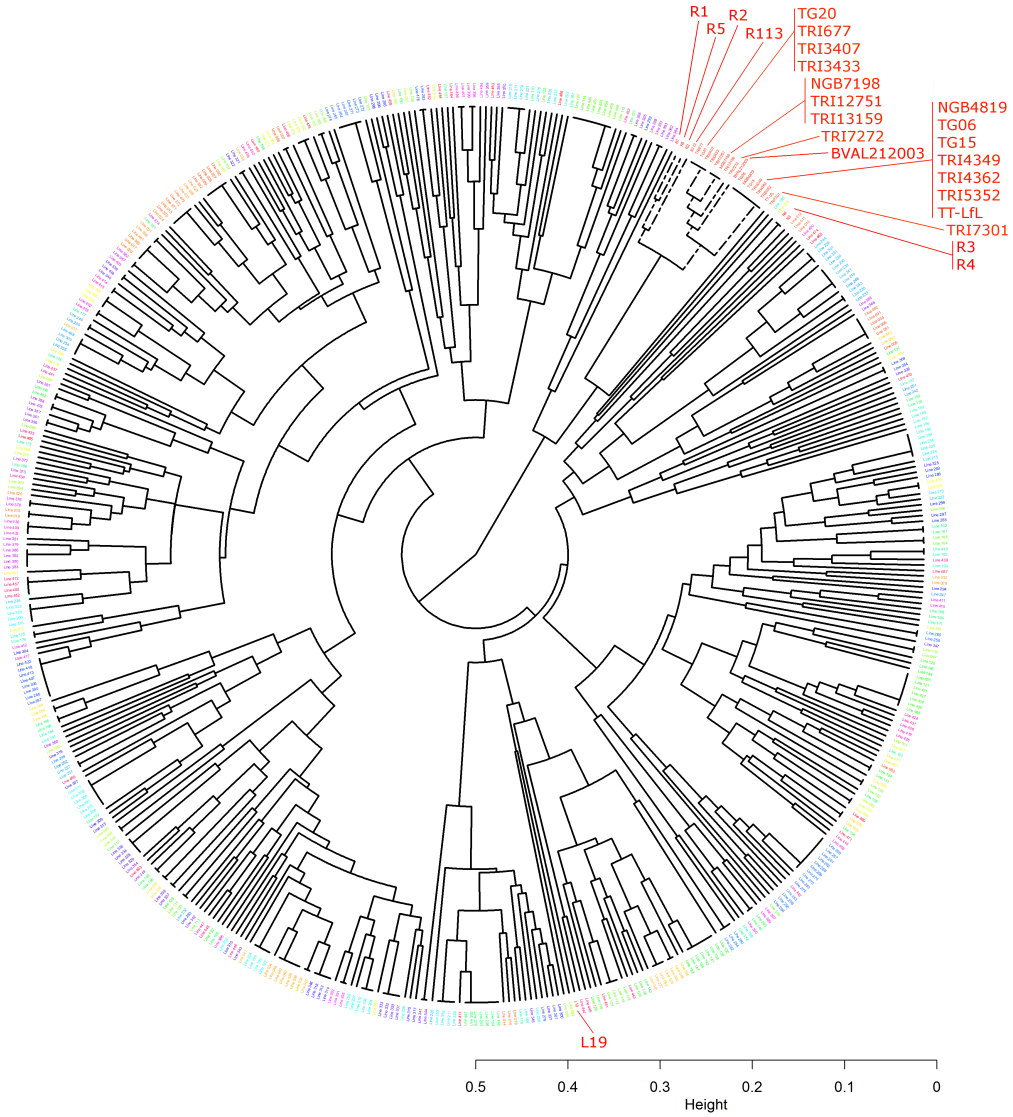

Figure A13 Phylogram for the *Rf4* marker haplotypes of the *Rf1* donor lines, 17 *T. timopheevii* accessions and 507 common wheat breeding lines. Lines under the dashed branches belonged to the same group if the number of groups was restricted to $k \leq 182$

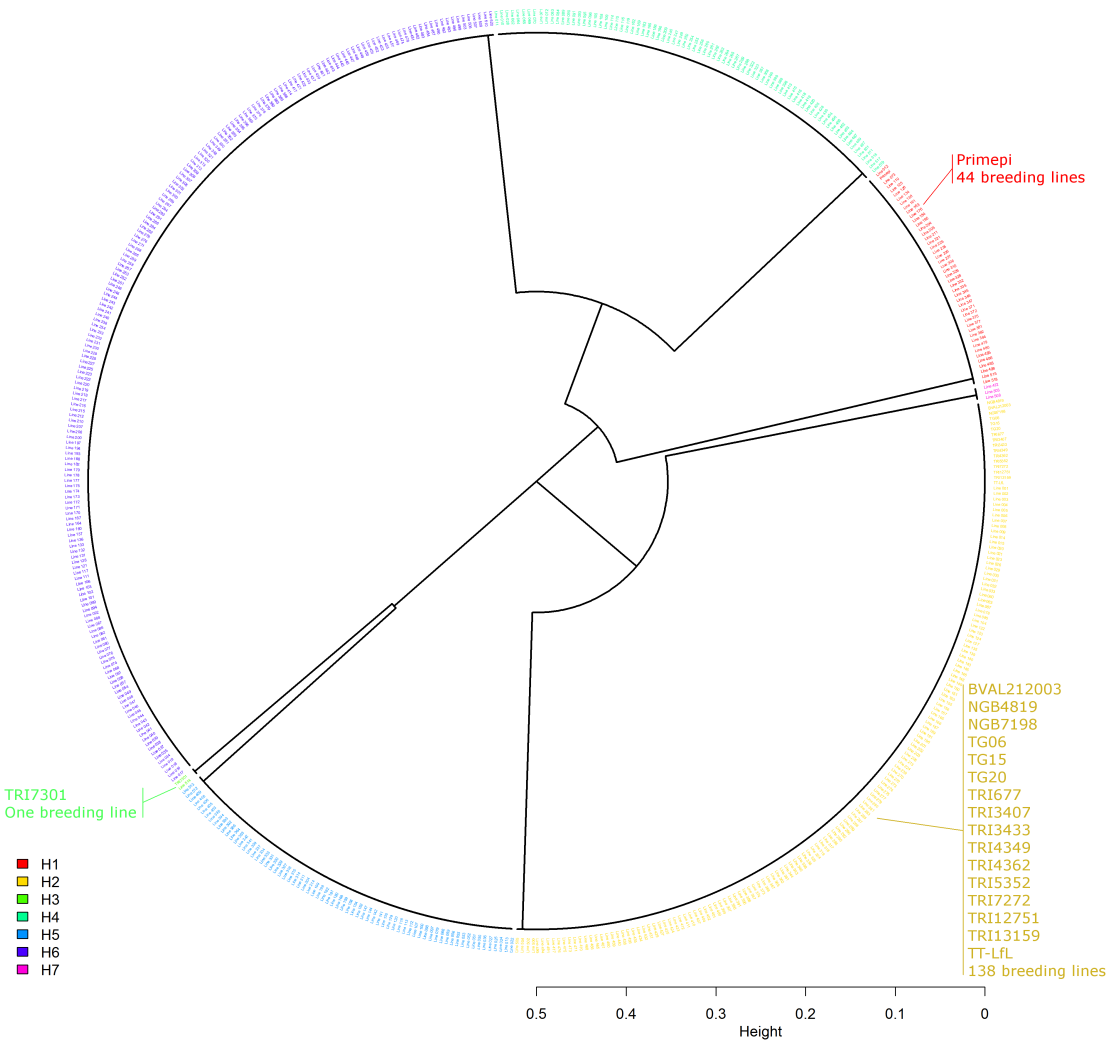

Figure A14 Phylogram for the *Rf3* marker haplotypes of the 520 common wheat breeding lines, 17 *T. timopheevii* accessions and Primepi

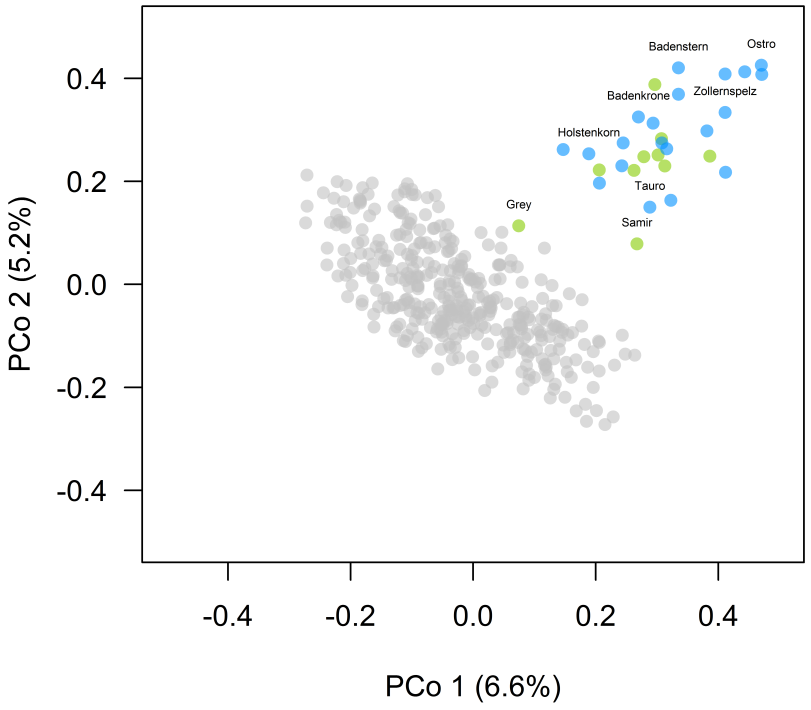

Figure A15 PCoA of 30 European spelt cultivars and 368 common wheat lines. Spelt accessions were classified as carriers of a restoring (blue) or a non-restoring (green) allele at the *Rf3* locus. Common wheat lines are depicted in grey

9 Publications

Findings presented in this thesis were published in the articles and conference proceedings listed below. The candidate contributed to the collection of phenotypic and genotypic data, performed data analysis, created graphs and tables, interpreted the results and wrote and revised the manuscripts.

Articles

Geyer M, Bund A, Albrecht T, Hartl L, Mohler V (2016) Distribution of the fertility-restoring gene *Rf3* in common and spelt wheat determined by an informative SNP marker. Mol Breeding 36:167. https://doi.org/10.1007/s11032-016-0592-6

Geyer M, Albrecht T, Hartl L, Mohler V (2018) Exploring the genetics of fertility restoration controlled by *Rf1* in common wheat (*Triticum aestivum* L.) using high-density linkage maps. Mol Genet Genomics 293:451–462. https://doi.org/10.1007/s00438-017-1396-z

Conference proceedings

Geyer M, Mohler V, Albrecht T, Bund A, Hartl L (2016) Linkage mapping of the restorer gene *Rf3* in winter wheat and spelt wheat (Talk). In: Vereinigung der Pflanzenzüchter und Saatgutkaufläute Österreichs (ed) 66. Jahrestagung 2015 Raumberg-Gumpenstein. University of Natural Resources and Life Sciences, Vienna, p 59

Geyer M, Bund A, Albrecht T, Hartl L, Mohler V (2017) Linkage mapping of fertility-restoring genes in common wheat and spelt (Poster). In: Buerstmayr H, Lang-Mladek C, Steiner B, Michel S, Buerstmayr M, Lemmens M, Vollmann J, Grausgruber H (eds) Proceedings of the 13th international wheat genetics symposium. University of Natural Resources and Life Sciences, Vienna, p 422

Geyer M, Albrecht T, Hartl L, Mohler V (2018) Epistatic interactions and distribution of the fertility-restoring locus *Rf1* in common wheat (Talk). In: Vereinigung der Pflanzenzüchter und Saatgutkaufläute Österreichs (ed) 68. Jahrestagung 2017 Raumberg-Gumpenstein. University of Natural Resources and Life Sciences, Vienna, pp 29–30

10 Acknowledgements

First, I would like to express my deep gratitude to Prof. Dr. Volker Mohler for his dedicated supervision of this work. His valuable advice, strong support and steady encouragement has been greatly appreciated.

My gratitude also goes to Prof. Dr. Chris-Carolin Schön for giving me the opportunity to pursue my thesis at the Chair of Plant Breeding. Her inspiring ideas and her encouragement to continuous improvement have been a valuable contribution to this work.

Sincere thanks to Dr. Lorenz Hartl, who gave me the opportunity to work at the Institute for Crop Science and Plant Breeding, conceived the initial design of the study and administered the project.

Many thanks go to Prof. Dr. Caroline Gutjahr for serving on the graduate committee.

I have sincerely appreciated the excellent technical assistance of Ruth Torrijos Polo, Sabine Schmidt and Andreas Klankermeier in the cumbersome collecton of phenotypic data. I am also very thankful to Petra Greim for her valuable contribution to the genotyping work.

Grateful thanks go to my colleague Dr. Theresa Albrecht for giving me scientific advice and sharing her experience with R. I also want to thank Adalbert Bund, Dr. Annette Block, Dr. Günther Schweizer and Dr. Bianca Büttner for the fruitful discussions.

Many thanks go to Dr. Finn Borum for providing seeds of the analysed *Rf1* donors. Many thanks also go to Prof. Dr. Nils Stein and Dr. Ruonan Zhou for their cooperation in the validation of candidate genes.

Finally, I want to thank my family, especially Silke and Leo, for their endless encouragement, support and understanding throughout these years.

This thesis was part of the project “CMS-Hybridweizen (AZ-1066-13)”, which was funded by the Bavarian Research Foundation.