**Title:**

Rapid gene cloning in cereals

**Authors:**

Jan Bettgenhaeuser and Simon G. Krattinger

**Affiliation:**

Biological and Environmental Science and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia

**Corresponding author:**

simon.krattinger@kaust.edu.sa; +966 (0) 12 808 2668

**Keywords:**

Genomics, complexity reduction, whole-genome sequencing, gene cloning, induced mutation

**Abstract**

The past two decades were characterized by a genomics revolution that entailed profound changes to crop research, plant breeding, and agriculture. Today, high-quality reference sequences are available for all major cereal crop species. Large re-sequencing and pan-genome projects start to reveal a more comprehensive picture of the genetic makeup and the diversity among domesticated cereals and their wild relatives. These technological advancements will have a dramatic effect on dissecting genotype-phenotype associations and on gene cloning. In this review, we will highlight the status of the genomic resources available for various cereal crops and we will discuss their implications for gene cloning. A particular focus will be given to the cereal species barley and wheat, which are characterized by very large and complex genomes that have been inaccessible to rapid gene cloning until recently. With the advancements in genomics and the development of several rapid gene cloning methods, it has now become feasible to tackle the cloning of most agriculturally important genes, even in wheat and barley.

**Key message**

The large and complex genomes of many cereals hindered cloning efforts in the past. Advances in genomics now allow the rapid cloning of genes from humanity's most valuable crops.

## Introduction

The blueprints of life are encoded in the DNA of the estimated 8.7 million species on earth (Mora et al. 2011). Each of these genomes is composed of thousands to billions of nucleotides arranged in a unique sequence. In 1977, the 5,375 bp genome of the bacteriophage ΦX174 was the first full genome to be sequenced (Sanger et al. 1977). Since then, advancements in genomics and DNA sequencing technologies have made it possible to decrypt the nucleotide sequences of thousands of organisms. Today, the genomes of more than 6,000 eukaryotes and 150,000 prokaryotes have been sequenced to various qualities (NCBI genome list https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/). While the motivation for the early attempts of genome sequencing was mostly of a basic scientific nature, genomics has nowadays found its way into many practical applications, including plant breeding (Hickey et al. 2017). Today, the identification and classification of the functional elements within a genome and the association of phenotypic variation with specific nucleotide polymorphisms or genes present the main challenges in genomics (Bevan et al. 2017).

Genomic and biotechnological applications, including marker-assisted selection, genomic selection, and genome editing, have become indispensable tools in many private and public plant breeding programs (Desta and Ortiz 2014; Poland and Rutkoski 2016). However, a deeper understanding of the functional nucleotide polymorphisms that shape agriculturally important traits in crops will be pivotal in order to unleash the full potential of these technologies. A large proportion of the global food and feed production stems from a single plant family named Poaceae (grasses). This family contains the cereal crops wheat, rice, maize, barley, sorghum, and millets, which provide more than 50% of the globally consumed calories (Nelson et al. 2010). Around 45 - 60 million years ago, modern cereals shared a common ancestor (The International Brachypodium Initiative 2010). In other words, a large share of today's global food supply, and hence food security, depends on a handful of closely related grass species. Unsurprisingly, research on deciphering the genetic and molecular bases of key agricultural traits in cereals has become increasingly important and a better understanding of functional cereal genomics will be essential in order to breed the cereal cultivars capable of feeding 9-10 billion people by 2050 (Bevan et al. 2017; Godfray et al. 2010).

A surprising feature of cereal genomes is that they substantially vary in size between different cereal species, despite their close relatedness. For example, the roughly 380 million nucleotides (Mb) that make up an entire rice genome (International Rice Genome Sequencing Project 2005; Stein et al. 2018) account for only 2.4% of the bread wheat genome (The International Wheat Genome Sequencing Consortium et al. 2018). The lack of a correlation between the complexity of an organism and its DNA content is known as C-value paradox (Thomas Jr. 1971), which can mainly be explained by differences in the fraction of highly repetitive elements and in ploidy. Today, high-quality reference genomes have been completed for all major cereal species. The first crop genome to be sequenced at high quality was the *japonica* rice cultivar Nipponbare in 2005 (International Rice Genome Sequencing Project 2005). The rice genome sequence was soon followed by high-quality assemblies of sorghum (730 Mb) (Paterson et al. 2009), maize (2.3 Gb)

(Schnable et al. 2009), barley (4.79 Gb) (Mascher et al. 2017), pearl millet (1.79 Gb) (Varshney et al. 2017) and wheat (15.8 Gb) (The International Wheat Genome Sequencing Consortium et al. 2018). Completion of these high-quality reference genomes marked milestones for cereal genetics and genomics and they will play a pivotal role in increasing our understanding of the genetic and molecular control of agronomically important traits.

But how has the availability of these reference genomes influenced the identification of causal nucleotide polymorphisms and genes? A simple literature search revealed that the number of publications referring to 'gene cloning' in rice has surged since the completion of the rice genome in 2005 (Figure 1) (Ni et al. 2009). A similar trend can be observed for the model plant species *Arabidopsis thaliana*. Surprisingly, the completion of the 2.3 Gb maize genome in 2009 had a less dramatic effect on the number of cloned genes that followed in the years after the maize reference sequence was published. This might indicate that the large genomes (>1 Gb) of maize, barley, and wheat still pose a major hurdle for gene cloning despite the availability of high-quality reference sequences. A reference sequence is only one stepping stone for the functional characterization of genomes and additional innovations are needed to clone genes from large and complex genomes. In this review, we will focus on recent genomic advancements that made gene cloning more rapid and cost-efficient. In particular, numerous rapid gene cloning approaches have been published during the past two years that will facilitate gene cloning from cereal species with large and repeat-rich genomes.

## Historic perspective on gene cloning

Before venturing into the future, let's have a look at the past. Linking phenotypic variation within crop plants with causal nucleotide polymorphisms has become an integral part of plant research and breeding. With the advent of molecular marker technologies and the use of sophisticated statistical tools, it has become feasible to dissect the genetic components of complex traits. Two of the most widely used approaches to study genetic patterns in cereals are quantitative trait locus (QTL) analyses and genome-wide association studies (GWAS). The progress in DNA sequencing has facilitated the transition from single molecular markers to high-throughput genotyping. Back in the 1990s and early 2000s, the development of Restriction Fragment Length Polymorphism (RFLP), Single Sequence Repeat (SSR) or Single Nucleotide Polymorphism (SNP) markers was a major effort. Today, high-throughput marker platforms and genotyping protocols have been developed for many cereal crops, which allow for a rapid and inexpensive genotyping of mapping populations and association panels with thousands of markers (Hussain et al. 2017; Rasheed et al. 2017). QTL analyses and GWAS in cereals generally result in the identification of molecular markers in linkage with a trait of interest. Although these methods seldom enable the direct identification of causal genes, mapping of major QTLs can be very accurate and causal genes are often found within less than 1 cM of peak markers (Price 2006). Positional cloning, also referred to as map-based cloning, has been the method of choice to identify the exact stretch of nucleotides (e.g. the gene) that controls a trait of interest

(Krattinger et al. 2009). This approach is exemplified by the cloning of *Xa21*, a broad-spectrum resistance gene against *Xanthomonas oryzae* pv. *oryzae* (*Xoo*), causal agent of the bacterial blight disease in rice. *Xa21* was originally introgressed into cultivated rice from a wild African rice species (*Oryza longistaminata*) by recurrent backcrossing (Khush et al. 1991). The gene was mapped to a 1.2 cM genetic interval on chromosome 11 using RFLP and RAPD (Random Amplified Polymorphic DNA) markers (Ronald et al. 1992). A linked marker was used to probe bacterial artificial chromosomes (BAC) and cosmid libraries and subclones of the identified BACs and cosmid clones were used for complementation in a susceptible rice line (Song et al. 1995; Wang et al. 1995). *Xa21* was identified on a 9.6 kb subclone containing a receptor-like kinase (Song et al. 1995). While this example demonstrates the power of positional cloning for isolating genes without *ab initio* knowledge of the gene product, it also highlights the laborious and time-consuming nature of positional cloning projects. The *Xa21* resistance gene was first identified in 1985, when *O. longistaminata* resistance to Indian *Xoo* isolates was evaluated (Khush et al. 1989). Therefore, ten years passed between the first description of *Xa21*-mediated resistance and the identification of the causal receptor-like kinase in 1995. Notably, *Xa21* was cloned before a high-quality reference genome of rice became available.

**Positional cloning 2.0 – how to avoid spending a decade on a gene cloning project?**

As illustrated by the cloning of *Xa21*, positional cloning has been the method of choice for gene cloning in the past. Positional cloning builds on genetic mapping and molecular approaches that result in the stepwise localization of a gene. A critical step is the generation of segregating high-resolution mapping populations that are genotyped with molecular markers and phenotyped for the trait of interest. Careful selection of parental lines and the creation of suitable mapping populations are important criteria that often decide on the success of failure of a gene cloning project. Generating mapping populations takes time, particularly for cereal species with rather long generation times. However, several advancements can help to significantly speed up this process. These include the development of methodologies for double haploid (DH) production of cereals (e.g. wheat; Niu et al. 2014), which significantly accelerates the process of obtaining homozygous progeny. Speed breeding is another recent protocol with enormous potential (Watson et al. 2018). By artificially prolonging the photoperiod and by optimizing the light spectrum supplied to plants, the generation time of many crop plants could be halved. The combination of speed breeding and DH production offers the possibility to rapidly develop mapping populations. Rapid genotyping of mapping populations with hundreds to thousands of molecular markers has become routine with the SNP arrays and genotyping-by-sequencing (GBS) protocols available today (Hussain et al. 2017; Rasheed et al. 2017).

Once a gene has been mapped to a small genetic interval, the corresponding physical interval is pieced together by chromosome walking, a process that involved multiple rounds of BAC library screening. This transition from genetic to physical maps (from cM to bp) has been one of the rate-limiting steps in many map-based cloning projects, particularly in cereals with large genomes. In wheat for example, hundreds of

disease resistance genes have been described and also mapped to a certain extent (McIntosh et al. 2008), but only a few dozens of them have been cloned (Keller et al. 2018). BAC clones can store a piece of DNA 100 -200 kb in size. Around 500,000 BAC clones are thus required to represent an entire hexaploid wheat genome. Also, the presence of three highly similar sub-genomes in bread wheat can confound BAC library screenings. While the availability of reference genomes partially overcomes this problem, it has also been observed that gene content and gene order can differ considerably between different cereal cultivars (Chia et al. 2012; Jordan et al. 2015; Mago et al. 2014). For positional cloning projects, it is thus important to generate sequence information from a donor line that carries the gene of interest. The major limitation of BAC clones is the size of the piece of foreign DNA that can be taken up by the vector and that can consequently be stored in a single *E. coli* culture. Most importantly, this biological limitation cannot be overcome with novel sequencing technologies. As long as no alternative vector systems are available that would allow for the storage of significantly longer DNA pieces, chromosome walking can only progress in steps of 100-200 kb at the time. It would thus be more accurate to refer to 'chromosome creeping' in wheat and barley.

Thind et al. (2017) set out to overcome this problem by developing an approach that enabled the rapid establishment of physical intervals from any donor line. 'Targeted chromosome-based cloning via long-range assembly' (TACCA) used a combination of short-read Illumina sequencing and chromosome contact maps of *in vitro* reconstituted chromosomes (Dovetail Genomics Chicago approach; Putnam et al. 2016) to generate large scaffolds from isolated chromosomes (Figure 2). As a proof of concept, the leaf rust resistance gene *Lr22a* of wheat was cloned. *Lr22a* was first mapped to a 0.09 cM interval on wheat chromosome 2D using a traditional positional cloning approach. However, the corresponding physical interval was not established by isolating relatively small BAC clones. Instead, chromosome 2D of the *Lr22a*-continaing wheat line 'CH Campala *Lr22a*' was isolated through flow cytometry (Vrána et al. 2000). Non-amplified high molecular weight DNA was then used for sequencing and Chicago long-range scaffolding. The resulting chromosome assembly had a scaffold N50 of 9.79 Mb, meaning that half of the assembly was covered by scaffolds of 9.79 Mb or longer. This is 50-100 times longer than a BAC clone. Hence, instead of spending several years trying to painstakingly piece together a physical interval with BAC clones, TACCA allows the generation of megabase-sized scaffolds spanning the target region within a few months. All the *Lr22a* flanking markers identified through genetic mapping were localized on a single 6.39 Mb scaffold and because the assembly was generated from an *Lr22a* donor line, all the candidate genes were immediately identified. This allowed Thind et al. (2017) to show that the *Lr22a* leaf rust resistance is caused by a nucleotide binding, leucine-rich repeat (NLR)-encoding gene. Compared to BAC clones, TACCA is not limited in terms of insert size other than the size of the sorted chromosome. It is thus possible that future sequencing or assembly protocols will result in even more impressive assembly statistics. Additionally, this approach showed that the sequencing of an entire wheat chromosome of 800 Mb has become an attractive methodology, even if one is only interested in a particular region of the assembly that might correspond to as little as 1% of the chromosome.

In summary, double haploid production, speed breeding, high-throughput molecular marker technologies, and TACCA have significantly reduced the time required for positional cloning projects.

## Gene cloning without fine-mapping - exploiting induced mutations for direct gene cloning

A popular approach to study the effect of single genes and to induce variation in breeding germplasm is through the induction of mutations (Schneeberger 2014; Uauy et al. 2017). DNA-altering agents such as radiation or chemicals, which can induce phenotypic variation in crop plants, have been known since the early 20$^{th}$ century. Pioneering work by Lewis John Stadler in the 1920s demonstrated that the exposure of barley seedlings to X-ray and radium caused visible alterations such as white seedlings (Stadler 1928). Later, it was discovered that certain chemicals like ethyl-methanesulfonate (EMS) and sodium azide had similar effects, and groundbreaking work by Barbara McClintock in the 1940s led to the discovery of 'jumping genes' in maize (McClintock 1950). Radiation and chemical mutagens induce random DNA changes across the genome that can range from SNPs (EMS) to large chromosomal rearrangements or deletions (radiation). On the other hand, the insertion of transposons (jumping genes) can disrupt gene function, a fact that has been systematically exploited by generating hundreds of thousands of T-DNA insertion and transposon lines, particularly in rice and maize (An et al. 2005; Vollbrecht et al. 2010). The first gene to be cloned by transposon tagging in plants was the *bronze* (*bz*) locus of maize, which controls anthocyanin pigmentation of maize kernels (Fedoroff et al. 1984; Rhoades 1952). An insertion of the *Activator* (*Ac*) transposable element produces the mutant phenotype. This allowed the cloning of *Ac* containing DNA fragments from maize plants displaying the mutant *bz* phenotype. *Ac* flanking sequences were used to probe homozygous *bz* and *Bz* and heterozygous *Bz*/*bz* maize lines. Hybridization of homozygous *bz* and *Bz* lines yielded single 10.8 kb and 6.3 kb fragments, respectively, while hybridization of the heterozygous line produced both fragments. This provided the stepping stone to isolate the non-mutant *Bz* allele that confers purple maize kernels (Fedoroff et al. 1984). This approach is not unique to *Ac* elements and can be used with other transposable elements, as demonstrated by the cloning of the *Knotted* gene in maize using a *Dissociation* (*Ds*) transposable element (Hake et al. 1989).

There are several advantages of using induced mutagenesis over natural variation to elucidate gene function: (*i*) Natural mutations occur infrequently, whereas mutagens can rapidly induce an almost infinite number of polymorphisms at virtually any position within a genome. (*ii*) Phenotypic alterations observed in mutagenized plants are often the result of recessive mutations in single genes, which results in a simple Mendelian inheritance of the trait (Schneeberger 2014). (*iii*) Mutagenized plants are phenotypically identical to their respective wild-type parents with the exception of traits that have been altered as a direct consequence of the mutagenesis. This makes phenotypic screens relatively simple and less dependent on genetic backgrounds and environmental conditions. (*iv*) The vast majority of the genotypic differences between the mutagenized plants and untreated parent are the direct result of the mutagenesis. The polymorphism density in single plants after EMS treatments is an order of magnitude lower than natural

polymorphism rates between two cereal cultivars. The number of induced mutations after EMS treatment in the diploid cereal species rice and sorghum lies in the range of 5-11 SNPs per Mb on average in a single plant (Henry et al. 2014; Jiao et al. 2016), whereas a comparison between the closely related *indica* rice cultivars Zhenshan 97 and Minghui 63 revealed one SNP every 3.65 kb (= 274 SNPs / Mb) (Zhang et al. 2016). Polyploid genomes tolerate a higher mutation density. For example, EMS-induced mutation density in tetraploid and hexaploid wheat ranged from 20-42 SNPs per Mb (Henry et al. 2014; Krasileva et al. 2017; Slade et al. 2005; Uauy et al. 2009). Despite the higher polymorphism density it can be more challenging to visualize mutant phenotypes in polyploid species because of genetic redundancy. While the effects of X-rays, EMS and transposons on plant phenotypes were already described in the first half of the 20$^{th}$ century, it took several decades from Stadler's and McClintock's groundbreaking experiments to develop sophisticated forward and reverse genetic approaches that allowed the identification of causal DNA polymorphisms underlying mutant phenotypes.

In the year 2000, McCallum et al. (2000) described Targeting Induced Local Lesions IN Genomes (TILLING) in Arabidopsis, a reverse genetic approach that allowed the detection of EMS-induced polymorphisms in specific gene sequences by using a polymerase chain reaction (PCR)-based screen. In this initial experiment, McCallum et al. (2000) screened an EMS-treated Arabidopsis population with primers designed on DNA methyltransferase genes and they used denaturing high-performance liquid chromatography (DHPLC) to identify single point mutations in pools containing DNA of several independent $M_2$ plants. TILLING proved to be a powerful reverse genetic approach that enabled the targeting of any gene sequence in a genome, allowing the study of phenotypic effects of induced mutations. After Arabidopsis, multiple TILLING populations were subsequently generated for many cereal crops, including barley (Gottwald et al. 2009; Talamè et al. 2008), sorghum (Xin et al. 2008) rice (Till et al. 2007), maize (Till et al. 2004), and wheat (Rawat et al. 2012; Uauy et al. 2009). For example, the TILLING population of sorghum has aided in understanding the pathways controlling lignification. Lignocelluloses in plant cell walls represent the biggest proportion of carbon fixed by plants and they are therefore of particular interest in biofuel production (Wang et al. 2015). Reduced lignification increases the digestibility when the crops are used as feed for livestock (Sattler et al. 2010). Reduced lignification in maize and sorghum is accompanied by a brown midrib (*bmr* in sorghum) phenotype of the leaves (Sattler et al. 2010). Several *bmr* mutants have been identified both in maize and sorghum. For example, the sorghum *bmr2* gene encodes a 4-coumarate-CoA ligase and its identification was aided by the sorghum TILLING population (Saballos et al. 2012). The TILLING population provided an additional *bmr2-2* mutant to the previously known *bmr2-ref* mutant. A map-based cloning approach with *bmr2-ref* delineated a 262 kb interval containing 22 annotated genes on sorghum chromosome 4. An obvious candidate gene in the region contained missense mutations in both mutant lines. As no transformation protocol existed for sorghum, the identification of a second mutant line from the TILLING population with a mutation in the same gene greatly facilitated the gene identification in this case (Saballos et al. 2012). Six additional *bmr* mutants have since then been identified in the sorghum TILLING population, which are not allelic to previously described *bmr* mutants (Sattler et al.

2014). This demonstrates the power of TILLING populations to provide novel mutations and expand on natural phenotypic variation.

Mutant identification in TILLING populations however was initially restricted by the size of the PCR products (usually a few hundred base pairs) and by the number of PCR primers that could be tested on a population. Pooling, automation, and the modification of PCR protocols allowing for longer PCR products (Uauy et al. 2009) resulted in a certain increase of the throughput, but a genome-wide assessment of induced mutations within a mutant line or even an entire population was not feasible until very recently. Abe et al. (2012) developed the first whole-genome approach in cereals, named MutMap, which allows rapid identification of causal, EMS-induced polymorphisms (Figure 2). For MutMap, a plant with an interesting phenotype that has been selected from a large mutant population is crossed to its respective wild-type parent. In the $F_2$, the recessive mutant phenotype will segregate in a 3:1 wild-type/mutant ratio. The DNA of 20 $F_2$ plants showing the mutant phenotype is bulked, the DNA bulk sequenced to ~12x coverage using short-read sequencing, and reads are aligned to a reference sequence. The majority of the non-causal SNPs will be recovered in a 1:1 wild-type/mutant ratio, which translates into a SNP index of 0.5. The only exception is the region linked to the phenotype of interest, where the SNP index will increase and ultimately reach a value of 1 in and next to the causal polymorphism. By plotting the SNP index across all the chromosomes, causal mutations can thus be easily identified. Abe et al. (2012) subjected 12,000 plants of the Japanese *japonica* rice cultivar Hitomebore to EMS treatment. Re-sequencing of five randomly selected mutants revealed that each mutant plant carried around 1,500 SNPs that were introduced by the EMS treatment (= 4 SNPs / Mb). A proof of concept was established by applying MutMap to two mutants showing a pale green phenotype. In both cases, the authors identified seven and five linked SNPs with a SNP index of 1. In one of the two cases, a mutation in a gene encoding a chlorophyll *a* oxygenase (*OsCAO1*) was shown to cause the pale green phenotype. In a follow-up publication, Fekih et al. (2013) modified the MutMap protocol to allow identification of causal mutations without the need of crossings. MutMap+ is thus also feasible for mutations that cause sterility, developmental lethality, or effects on flowering time that can impede crossings.

Whole genome re-sequencing is a feasible strategy for small cereal genomes. Advances in DNA sequencing technologies have even enabled the cataloging of induced polymorphisms in entire mutant populations. For example, Li et al. (2017a) comprehensively assessed polymorphism in 1,504 fast-neutron-induced mutant lines of the rice cultivar Kitaake. In total, they found 91,513 mutations affecting 32,307 genes, which results in an average of 61 mutations per line. Li et al. (2017a) focused on one particular mutant line that showed dwarfism and shorter panicles. This particular line contained 76 mutations and by using a small segregating population, the authors could show that the mutant phenotype was caused by an inversion on chromosome 5 that affected the *Dwarf 1/RGA1* gene. Similarly, whole-genome sequencing of 256 EMS-induced mutant lines of the sorghum inbred line BTx623 identified more than 1.8 million SNPs (Jiao et al. 2016). For the cereal species maize, barley, and wheat with their large genomes, whole-genome

re-sequencing of hundreds or even thousands of plants remains challenging. For these species, a genome complexity reduction of some sort is usually required. For example, RNA-sequencing is a widely used form of complexity reduction, as only a small proportion of a cereal genome harbors protein-coding genes. By comparing the transcriptomes of mutant and wild-type bulks (bulk segregant RNA-seq = BSR-seq), (Liu et al. 2012) mapped and cloned the *glossy3* gene in maize. Mutants of *glossy3* display altered accumulation of epicuticular waxes. Krasileva et al. (2017) developed an exome capture assay based on 82,511 transcripts of wheat. The capture array was applied to sequence the coding regions of 1,535 EMS mutants derived from the tetraploid wheat cultivar Kronos and 1,200 mutant lines in the background of the hexaploid wheat cultivar Cadenza. This led to the identification of 48,172 and 73,895 gene models with at least one mutation in Kronos and Cadenza, respectively. By using this approach, the 'size' of the hexaploid Cadenza genome was reduced to about 162 Mb, which corresponds to 1% of a hexaploid wheat genome. An example that highlights the usefulness of this TILLING resource is the cloning of the *Stb6* gene, which provides resistance against the fungal *Septoria tritici* blotch disease (Saintenac et al. 2018). Cadenza carries *Stb6*, which allowed the authors to rapidly identify missense and nonsense mutations for two of the candidate genes identified through genetic mapping. Mutations in a gene encoding a wall-associated receptor kinase (WAK) showed increased susceptibility compared to the wild-type control, proving that this WAK gene is *Stb6*. In contrast to whole-genome re-sequencing, the success of this approach largely depends on the quality of the capture array and hence, on the transcriptomics data or gene annotations used to design the array. As the example of *Stb6* shows, the re-sequencing of TILLING populations, be it based on whole-genome sequencing or reduced representation libraries, has opened up entirely new possibilities for gene cloning. Traditionally, TILLING populations were used for reverse genetic approaches and the availability of whole-genome sequencing data will greatly facilitate the identification of mutants in candidate genes, as exemplified by *Stb6*. In addition, the availability of sequence information from entire mutant population will also allow to use TILLING populations for forward genetic approaches in the near future, e.g. to start from an interesting phenotype detected in certain mutant lines and to clone the underlying genes through association mapping, as it was exemplified in rice by MutMap and MutMap+.

Another strategy to reduce the complexity of large cereal genomes is through a candidate gene approach. For example, the majority of disease resistance genes cloned in plants so far code for intracellular immune receptors of the NLR family (Kourelis and van der Hoorn 2018). NLRs perceive pathogen-derived virulence effectors that are secreted into the plant cell during the infection process. This recognition often triggers a strong defense reaction referred to as hypersensitive response (Krattinger and Keller 2016). The knowledge on the importance of NLR-triggered immunity for plant defense was exploited in a protocol coined Resistance gene enrichment Sequencing (RenSeq) (Jupe et al. 2013). RenSeq works by the design of a capture array that specifically allows for the enrichment of NLR sequences. Steuernagel et al. (2016) combined RenSeq with forward genetics using EMS mutagenesis to develop MutRenSeq. For MutRenSeq, an NLR capture array was designed based on sequence information from domesticated wheat and its wild relatives. The capture array was used to enrich for the NLR complement of a resistant wild-type parent and

several independent susceptible loss-of-function mutants. Alignment of the mutant reads to the wild-type parent allowed the identification of contigs that carried EMS-induced polymorphisms in all independent mutants. Steuernagel et al. (2016) demonstrated the usefulness of MutRenSeq by cloning the two genes, *Sr22* and *Sr45,* that both provide resistance against the fungal stem rust disease of wheat. For *Sr22*, 1,300 EMS-treated $M_2$ families derived from the stem rust resistant, *Sr22* carrying wheat cultivar Schomburgk were assessed for susceptibility. The NLR complement of Schomburgk and six independent, susceptible loss-of-function mutants were sequenced using Illumina short-read sequencing. Twenty-three contigs showed a SNP in two of the six mutants when compared to the Schomburgk wild-type sequence and only a single contig showed mutations in five of the six mutants. This latter contig could be physically linked to a neighboring contig with a polymorphism in the remaining sixth mutant line. Hence, by using six independent mutants, Steuernagel et al. (2016) could unambiguously identify a single gene that was shown to correspond to *Sr22*. In comparison to exome capture, RenSeq allows researchers to even further reduce the genome fraction, which in turn reduces the costs for sequencing. The NLR complement sequenced by Steuernagel et al. (2016) consisted of 8,235 NLR-containing contigs spanning 14.5 Mb, which corresponds to as little as 0.09% of a hexaploid wheat genome. Hence, this approach is highly suitable to cost-effectively and rapidly identify NLR-encoding disease resistance genes. However, as the examples of *Xa21* and *Stb6* show, not all disease resistance genes in cereals encode NLRs.

Besides using capture arrays, the genomes of many cereal species can be reduced by making use of the natural DNA partitioning, e.g. the chromosomes. Through flow cytometry, single chromosomes from many cereal species can be isolated at high purity (Vrána et al. 2000). For hexaploid wheat, chromosome sorting reduces the size of the 'genome' by a factor of 21, also eliminating a lot of the problems caused by the high sequence relatedness of the homoeologous chromosomes. Sánchez-Martín et al. (2016) turned chromosome sorting into a neat approach to isolate causal mutations. Their approach, named MutChromSeq, starts by identifying loss-of-function mutants in an EMS-treated population. Chromosomes harboring the respective gene are then isolated from several independent mutants and the respective wild-type parent and subjected to short-read sequencing. Comparable to the MutRenSeq pipeline, the MutChromSeq pipeline will call the sequence variants between the wild-type contigs and mutant reads. Contigs that contain independent sequence polymorphisms in all the mutants are candidates for further validation. The MutChromSeq pipeline was used to clone the barley *Eceriferum-q* gene that is involved in the accumulation of epicuticular wax and the *Pm2* powdery mildew resistance gene of hexaploid wheat. In both cases, the short-read sequencing produced tens of thousands of unordered contigs. Comparable to MutRenSeq, five to six independent loss-of-function mutants were sufficient to identify a single or two candidate contigs that had a SNP in all the mutants. This approach demonstrated that the *Eceriferum-q* gene codes for a lipase/carboxyl transferase and *Pm2* for a NLR. In contrast to MutRenSeq, MutChromSeq makes no prior assumptions about the gene product and beside the chromosome location, no additional genetic mapping is required.

## Phenotyping – the new bottleneck of gene cloning

The examples highlighted above demonstrate the enormous progress that has been made over the past years to develop genomic approaches that allow rapid and cost-effective gene identification based on induced mutagenesis. The challenge of induced mutagenesis is that rare single plants with a mutant phenotype need to be identified from thousands of mutagenized plants, of which most will retain a wild-type phenotype for the trait of interest. In addition, functional redundancy might prevent the identification of mutants in polyploid species. Induced mutagenesis thus works well for clear phenotypes that are caused by single genes. The examples of the pale green rice mutants, the *Eceriferum-q* gene of barley or the wheat stem rust and powdery mildew NLRs mentioned in the previous paragraph are perfect examples for this. These traits are caused by single genes with strong effects and mutations were thus easy to identify. Many agriculturally important traits however are quantitatively inherited, which means that several additively-acting genes, each with a minor effect, contribute to the overall phenotype (Bazakos et al. 2017). Prime examples are grain yield, drought tolerance, or quantitative disease resistance. Identifying minor variations affecting quantitative traits in mutant populations can be very challenging, even more so if there is genotype-environment interaction. *Lr34*, *Lr46*, and *Lr68* are adult plant resistance genes against leaf rust, a fungal pathogen of wheat. While *Lr34* and *Lr46* have a greater effect than *Lr68* on leaf rust resistance in most parts of the world (Herrera-Foessel et al. 2012), this relationship is reversed in Uruguay and Argentina, where *Lr68* displayed the strongest phenotypic effect of the three genes (Silva et al. 2015). It is conceivable that environmental effects could be exploited for the cloning of genes like *Lr68*. Choosing the right environmental conditions will maximize the effect of a QTL and provide more robust phenotypic data.

As described in this review, advances in whole-genome sequencing, genome complexity reduction, marker systems, and genomic analyses have accelerated the cloning of cereal genes over the past decade. In parallel, phenotyping systems are being developed for small-scale greenhouse (e.g. Hartmann et al. 2011, Clark et al. 2013, and Czedik-Eysenberg et al. 2018) to large-scale field applications (e.g. Crain et al. 2016, Haghighattalab et al. 2016, and Virlet et al. 2017). These phenotyping systems aid the detection of small phenotypic differences between plants, which may have been undetectable by the human eye. For example, smut fungi affect many cereal species, including barley and maize. These fungal pathogens produce obvious spore-filled sori in floral organs, but do not cause obvious symptoms during the plant's vegetative phase. Czedik-Eysenberg et al. (2018) developed an automated phenotyping system that is able to detect subtle growth differences of infected plants prior to flowering and reliably predict infection outcome. Obtaining more robust phenotypic data like these will ultimately improve the detection and cloning of genes, especially QTLs.

## What does the near future hold for gene cloning?

At last, what does a look into the crystal ball tell us about gene cloning in the future? With the recent addition of high-quality barley and wheat reference genomes, cereal genomics has ultimately entered the post-genomics era with unprecedented consequences for gene cloning. The genome structure and gene content of single reference accessions of rice, sorghum, maize, barley, and wheat are known. This will for example greatly facilitate the development of molecular markers for positional cloning projects because it has become possible to orient oneself within a genome. The next leap forward will consist in the comprehensive assessment of sequence diversity between different accessions of a cereal species, including single nucleotide variants, copy number variation, and structural rearrangements. As an example, high-quality genome sequences of various wild and domesticated rice species were recently published (Stein et al. 2018; Zhao et al. 2018) and re-sequencing data from more than 3,000 *O. sativa* accessions are available (Wang et al. 2018), allowing for the first time to assess the structural and sequence diversity in rice among thousands of accessions. Similar projects are currently underway for other cereal species, including wheat and barley. A deeper understanding of the genomic variation across entire species will for example allow to develop improved capture arrays for the sequencing of TILLING populations or for MutRenSeq. In addition, extensive re-sequencing data on thousands of cereal accessions might greatly enhance the power of GWAS, which could ultimately allow the direct identification of candidate genes from association studies. An interesting approach that goes in this direction has been developed by Arora et al. (2018). They used association genetics in combination with complexity reduction (RenSeq) to clone several stem rust resistance genes from a panel comprising 174 accessions of the diploid wild wheat progenitor *Aegilops tauschii*. In contrast to other gene cloning approaches, this approach coined Association genetics with R gene enrichment Sequencing (AgRenSeq) worked without generating mapping populations or mutant populations. Also, once the panel has been established and sequenced, the same accessions can be used over and over again to identify NLR-encoding resistance genes against different diseases.

It can also be anticipated that sequencing technologies will further improve in the near future with regard to read-length and sequence accuracy. At the same time, the cost of sequencing is likely to drop even further. For example, a high-quality genome of *Arabidopsis thaliana* has recently been generated from a single MinION Nanopore flow cell. The 111 Mb genome was assembled into 62 contigs with a contig N50 of 12.3 Mb. This study highlights the fact that is has become possible to rapidly derive high-quality genomes at relatively low costs (Michael et al. 2018).The continuing progress in sequencing technology will greatly affect gene cloning methods that rely on *de novo* or re-sequencing. MutMap, TACCA and MutChromSeq have already demonstrated that sequencing of entire chromosomes and genomes has become a feasible, cost-effective and attractive option, even if one is ‘just’ interested in a small region of a particular chromosome. Introducing long-read sequencing to MutMap for example, in combination with the multiple high-quality rice genome available (Stein et al. 2018; Zhao et al. 2018), is likely to greatly simplify the identification of causative SNPs. A protocol to integrate single-molecule real time (SMRT) sequencing into RenSeq has already been established (Witek et al. 2016) and long-read sequencing technologies might also be integrated into MutChromSeq. This step could greatly reduce the number of contigs obtained for

MutRenSeq and MutChromSeq and thereby simplify the discovery of causal mutations. The use of short-read sequencing technologies for MutRenSeq and MutChromSeq bears the risk that a gene gets assembled in two separate contigs, which often happens if a gene contains large introns. This can greatly impede gene identification. In the future, it might even become feasible to cost-effectively re-sequence entire maize, barley, and wheat genomes for gene cloning projects.

Most of the advances discussed in this review relate to the identification and cloning of genes underlying a trait with a Mendelian inheritance (i.e. controlled by a single gene) or QTLs of major effect. A remaining frontier is the cloning of minor effect QTLs. While these have only a minor effect individually, additive minor effect QTLs can play a major role in many agriculturally important traits. For example, in maize both grain yield and flowering time are complex traits (Buckler et al. 2009; Schön et al. 2004). Currently, the cloning of quantitative genes is still best achieved by positional cloning, following an initial QTL or GWAS study. A combination of GWAS and positional cloning was for example used to clone the race non-specific rice blast resistance gene *bsr-d1* from the durably resistant rice cultivar Digu (Li et al. 2017b). To clone *bsr-d1*, the genomic sequence of Digu was compared to 66 rice accessions lacking broad-spectrum resistance, which led to the identification of 2,576 SNPs that were unique to Digu. Association analysis of these SNPs in a RIL population derived from a cross of Digu with the susceptible rice line LTH revealed that a single SNP located in the promoter region of a $C_2H_2$-type transcription factor gene was associated with the rice blast resistance. As minor effect QTLs lack an easily identifiable phenotypic effect, approaches such as MutMap or MutChromSeq which rely on the phenotypic screening of large mutant populations, are challenging in this regard. The cloning of minor QTLs will thus much depend on progress on precision phenotyping beside genomic advancements.

## Conclusion

In summary, progress in genetics has revolutionized gene cloning in the past five years. Gene cloning projects are no longer matters of ten years or longer. Hence, our excuses not to clone genes in cereals are dwindling. It is therefore the right time to consider the cloning of most agriculturally important genes in an international effort. A comprehensive catalogue of cloned genes, along with the knowledge of the functional polymorphism, could greatly increase the accuracy of genomics-assisted breeding.

## Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

## Figures

**Figure 1.** Cumulative number of publications that included the words ‘map-based’, ‘gene’, and ‘cloning’ in the title or abstract. The arrows indicate the year of the release of the first high-quality reference sequence for each species. To generate the data, a search with these keywords and the respective species was performed on the NCBI PubMed advanced search builder (https://www.ncbi.nlm.nih.gov/pubmed/advanced).

**Figure 2.** An overview of the steps involved in deciding on an approach to rapidly clone a gene of interest in cereals. These approaches are not necessarily mutually exclusive and can often be combined. For example, long-range assembly could also be used for MutChromSeq. Some approaches are only suitable for specific gene families or genes with a strong phenotype. Depending on the plant studied, genome complexity reduction may need to be performed. Many aspects, e.g. whether WGS can be performed, are subjective and determined by the resources available. Some approaches include the creation of time- or cost-intensive resources (e.g. long-range chromosome assembly for TACCA or sequenced diversity panel for AgRenSeq), which can be re-used very cheaply. Should a chosen approach not lead to the identification of the causal gene (e.g. MutRenSeq does not yield a causal NLR-encoding gene), other approaches may offer more success (e.g. MutChromSeq on the mutants identified for MutRenSeq). Abbreviations: NLR = nucleotide binding, leucine-rich repeat encoding gene; WGS = whole-genome sequencing; TACCA = Targeted chromosome-based cloning via long-range assembly.

## References


Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, Matsumura H, Yoshida K, Mitsuoka C, Tamiru M, Innan H, Cano L, Kamoun S, Terauchi R (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol 30:174-178

An G, Lee S, Kim SH, Kim SR (2005) Molecular genetics using T-DNA in rice. Plant Cell Physiol 46:14-22

Arora S, Steuernagel B, Chandramohan S, Long Y, Matny O, Johnson R, Enk J, Periyannan S, Hatta MAM, Athiyannan N, Cheema J, Yu G, Kangara N, Ghosh S, Szabo LJ, Poland J, Bariana H, Jones JDG, Bentley AR, Ayliffe M, Olson E, Xu SS, Steffenson BJ, Lagudah E, Wulff BBH (2018) Resistance gene discovery and cloning by sequence capture and association genetics. bioRxiv

Bazakos C, Hanemian M, Trontin C, Jiménez-Gómez JM, Loudet O (2017) New strategies and tools in quantitative genetics: How to go from the phenotype to the genotype. Annu Rev Plant Biol 68:435-455

Bevan MW, Uauy C, Wulff BB, Zhou J, Krasileva K, Clark MD (2017) Genomic innovation for crop improvement. Nature 543:346-354

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Villeda HS, Sofia da Silva H, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of maize flowering time. Science 325:714-718

Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, Gore M, Guill KE, Holland J, Hufford MB, Lai J, Li M, Liu X, Lu Y, McCombie R, Nelson R, Poland J, Prasanna BM, Pyhäjärvi T, Rong T, Sekhon RS, Sun Q, Tenaillon MI, Tian F, Wang J, Xu X, Zhang Z, Kaeppler SM, Ross-Ibarra J, McMullen MD, Buckler ES, Zhang G, Xu Y, Ware D (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 44:803-807

Clark RT, Famoso AN, Zhao K, Shaff JE, Craft EJ, Bustamante CD, McCouch SR, Aneshansley DJ, Kochian LV (2013) High-throughput two-dimensional root system phenotyping platform facilitates genetic analysis of root growth and development. Plant, Cell Environ 36:454-466

Crain JL, Wei Y, Barker J, Thompson SM, Alderman PD, Reynolds M, Zhang N, Poland J (2016) Development and deployment of a portable field phenotyping platform. Crop Sci 56

Czedik-Eysenberg A, Seitner S, Güldener U, Koemeda S, Jez J, Colombini M, Djamei A (2018) The 'PhenoBox', a flexible, automated, open-source plant phenotyping solution. New Phytologist 219:808-823

Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci 19:592-601

Fedoroff NV, Furtek DB, Nelson OE (1984) Cloning of the *bronze* locus in maize by a simple and generalizable procedure using the transposable controlling element *Activator* (*Ac*). Proc Natl Acad Sci USA 81:3825-3829

Fekih R, Takagi H, Tamiru M, Abe A, Natsume S, Yaegashi H, Sharma S, Sharma S, Kanzaki H, Matsumura H, Saitoh H, Mitsuoka C, Utsushi H, Uemura A, Kanzaki E, Kosugi S, Yoshida K, Cano L, Kamoun S, Terauchi R (2013) MutMap+: genetic mapping and mutant identification without crossing in rice. PLoS One 8:e68529

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

Gottwald S, Bauer P, Komatsuda T, Lundqvist U, Stein N (2009) TILLING in the two-rowed barley cultivar 'Barke' reveals preferred sites of functional diversity in the gene *HvHox1*. BMC Res Notes 2:258

Haghighattalab A, González Pérez L, Mondal S, Singh D, Schinstock D, Rutkoski J, Ortiz-Monasterio I, Singh RP, Goodin D, Poland J (2016) Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. Plant Methods 12:35

Hake S, Vollbrecht E, Freeling M (1989) Cloning *Knotted*, the dominant morphological mutant in maize using *Ds2* as a transposon tag. EMBO J 8:15-22

Hartmann A, Czauderna T, Hoffmann R, Stein N, Schreiber F (2011) HTPheno: An image analysis pipeline for high-throughput plant phenotyping. BMC Bioinformatics 12:148

Henry IM, Nagalakshmi U, Lieberman MC, Ngo KJ, Krasileva KV, Vasquez-Gross H, Akhunova A, Akhunov E, Dubcovsky J, Tai TH, Comai L (2014) Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. Plant Cell 26:1382-1397

Herrera-Foessel SA, Singh RP, Huerta-Espino J, Rosewarne GM, Periyannan SK, Viccars L, Calvo-Salazar V, Lan C, Lagudah ES (2012) *Lr68*: a new gene conferring slow rusting resistance to leaf rust in wheat. Theor Appl Genet 124:1475-1486

Hickey JM, Chiurugwi T, Mackay I, Powell W, Implementing Genomic Selection in CGIAR Breeding Programs Workshop Participants (2017) Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. Nat Genet 49:1297-1303

Hussain W, Baenziger PS, Belamkar V, Guttieri MJ, Venegas JP, Easterly A, Sallam A, Poland J (2017) Genotyping-by-sequencing derived high-density linkage map and its application to QTL mapping of flag leaf traits in bread wheat. Scientific Reports 7

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793-800

Jiao Y, Burke JJ, Chopra R, Burow G, Chen J, Wang B, Hayes C, Emendack Y, Ware D, Xin Z (2016) A sorghum mutant resource as an efficient platform for gene discovery in grasses. Plant Cell

Jordan KW, Wang S, Lun Y, Gardiner L-J, MacLachlan R, Hucl P, Wiebe K, Wong D, Forrest KL, Consortium I, Sharpe AG, Sidebottom CHD, Hall N, Toomajian C, Close T, Dubcovsky J, Akhunova A, Talbert L, Bansal UK, Bariana HS, Hayden MJ, Pozniak C, Jeddeloh JA, Hall A, Akhunov E (2015) A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. Genome Biol 16

Jupe F, Witek K, Verweij W, Sliwka 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

Keller B, Wicker T, Krattinger SG (2018) Advances in wheat and pathogen genomics: Implications for disease control. Annual Review of Phytopathology 56:67-87

Khush GS, Bacalangco E, Ogawa T (1991) A new gene for resistance to bacterial blight from *O. longistaminata*. Rice Genet Newslett 7:121-122

Khush GS, Mackill DJ, Sidhu GS (1989) Breeding rice for resistance to bacterial blight. Proceedings of the International Workshop on Bacterial Blight of Rice. International Rice Research Institute, Manila, Philippines

Kourelis J, van der Hoorn RAL (2018) Defended to the Nines: 25 Years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function. Plant Cell 30:285-299

Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, Clissold L, Simmonds J, Ramirez-Gonzalez RH, Wang X, Borrill P, Fosker C, Ayling S, Phillips AL, Uauy C, Dubcovsky J (2017) Uncovering hidden variation in polyploid wheat. Proc Natl Acad Sci USA 114:E913-E921

Krattinger S, Wicker T, Keller B (2009) Map-based cloning of genes in Triticeae (wheat and barley). In: Muehlbauer G, Feuillet C (eds) Genetics and genomics of the Triticeae. Springer, New York, NY, pp 337-357

Krattinger SG, Keller B (2016) Molecular genetics and evolution of disease resistance in cereals. New Phytologist 212:320-332

Li G, Jain R, Chern M, Pham NT, Martin JA, Wei T, Schackwitz WS, Lipzen AM, Duong PQ, Jones KC, Jiang L, Ruan D, Bauer D, Peng Y, Barry KW, Schmutz J, Ronald PC (2017a) The sequences of 1504 mutants in the model rice variety Kitaake facilitate rapid functional genomic studies. Plant Cell 29:1218-1231

Li W, Zhu Z, Chern M, Yin J, Yang C, Ran L, Cheng M, He M, Wang K, Wang J, Zhou X, Zhu X, Chen Z, Wang J, Zhao W, Ma B, Qin P, Chen W, Wang Y, Liu J, Wang W, Wu X, Li P, Wang J, Zhu L, Li S, Chen X (2017b) A natural allele of a transcription factor in rice confers broad-spectrum blast resistance. Cell 170:114-126 e115

Liu S, Yeh CT, Tang HM, Nettleton D, Schnable PS (2012) Gene mapping via bulked segregant RNA-Seq (BSR-Seq). PLoS One 7:e36406

Mago R, Tabe L, Vautrin S, Šimková H, Kubaláková M, Upadhyaya N, Berges H, Kong X, Breen J, Doležel J, Appels R, Ellis JG, Spielmeyer W (2014) Major haplotype divergence including multiple germin-like protein genes, at the wheat *Sr2* adult plant stem rust resistance locus. BMC Plant Biol 14:379

Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J, Bayer M, Ramsay L, Liu H, Haberer G, Zhang XQ, Zhang Q, Barrero RA, Li L, Taudien S, Groth M, Felder M, Hastie A, Simkova H, Stankova H, Vrana J, Chan S, Munoz-Amatriain M, Ounit R, Wanamaker S, Bolser D, Colmsee C, Schmutzer T, Aliyeva-Schnorr L, Grasso S, Tanskanen J, Chailyan A, Sampath D, Heavens D, Clissold L, Cao S, Chapman B, Dai F, Han Y, Li H, Li X, Lin C, McCooke JK,

590 Tan C, Wang P, Wang S, Yin S, Zhou G, Poland JA, Bellgard MI, Borisjuk L, Houben A, Dolezel J, Ayling
591 S, Lonardi S, Kersey P, Langridge P, Muehlbauer GJ, Clark MD, Caccamo M, Schulman AH, Mayer KFX,
592 Platzer M, Close TJ, Scholz U, Hansson M, Zhang G, Braumann I, Spannagl M, Li C, Waugh R, Stein N
593 (2017) A chromosome conformation capture ordered sequence of the barley genome. Nature 544:427-433
594 McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting Induced LocalLesions IN Genomes
595 (TILLING) for plant functional genomics. Plant Physiol 123:439-442
596 McClintock B (1950) The origin and behavior of mutable loci in maize. Proc Natl Acad Sci USA 36:344-355
597 McIntosh R, Yamazaki Y, Dubcovsky J, Rogers J, Morris C, Somers D, Appels R, Devos K (2008)
598 Catalogue of gene symbols for wheat. In: Appels R, Eastwood R, Lagudah E, Langridge P, Mackay M,
599 McIntyre L, Sharp P (eds) Proceedings of the 11th International Wheat Genetics Symposium. Sydney
600 University Press, Sydney
601 Michael TP, Jupe F, Bemm F, Motley ST, Sandoval JP, Lanz C, Loudet O, Weigel D, Ecker JR (2018) High
602 contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. Nat Commun 9
603 Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How Many Species Are There on Earth and in
604 the Ocean? PLoS Biol 9
605 Nelson GC, Rosegrant MW, Palazzo A, Gray I, Ingersoll C, Robertson R, Tokgoz S, Zhu T, Sulser TB,
606 Ringler C, Msangi S, You L (2010) Food security, farming, and climate change to 2050: scenarios, results,
607 policy options. International Food Policy Research Institute, Washington, D.C., USA
608 Ni J, Pujar A, Youens-Clark K, Yap I, Jaiswal P, Tecle I, Tung CW, Ren L, Spooner W, Wei X, Avraham S,
609 Ware D, Stein L, McCouch S (2009) Gramene QTL database: development, content and applications.
610 Database 2009:bap005
611 Niu Z, Jiang A, Abu Hammad W, Oladzadabbasabadi A, Xu SS, Mergoum M, Elias EM (2014) Review of
612 doubled haploid production in durum and common wheat through wheat × maize hybridization. Plant
613 Breeding 133:313-320
614 Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U,
615 Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus
616 FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov
617 AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann
618 MC, Ming R, Peterson DG, Mehboob ur R, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009)
619 The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551-556
620 Poland J, Rutkoski J (2016) Advances and Challenges in Genomic Selection for Disease Resistance.
621 Annual Review of Phytopathology 54:79-98
622 Price AH (2006) Believe it or not, QTLs are accurate! Trends Plant Sci 11:213-216
623 Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet
624 CW, Haussler D, Rokhsar DS, Green RE (2016) Chromosome-scale shotgun assembly using an in vitro
625 method for long-range linkage. Genome Res 26:342-350
626 Rasheed A, Hao Y, Xia X, Khan A, Xu Y, Varshney RK, He Z (2017) Crop breeding chips and genotyping
627 platforms: Progress, challenges, and oerspectives. Mol Plant 10:1047-1064
628 Rawat N, Sehgal SK, Joshi A, Rothe N, Wilson DL, McGraw N, Vadlani PV, Li W, Gill BS (2012) A diploid
629 wheat TILLING resource for wheat functional genomics. BMC Plant Biol 12:205
630 Rhoades MM (1952) The effect of the Bronze locus on anthocyanin formation in maize. Am Nat 86:105-
631 108
632 Ronald PC, Albano B, Tabien R, Abenes L, Wu KS, McCouch S, Tanksley SD (1992) Genetic and physical
633 analysis of the rice bacterial blight disease resistance locus, *Xa21*. Mol Gen Genet 236:113-120
634 Saballos A, Sattler SE, Sanchez E, Foster TP, Xin Z, Kang C, Pedersen JF, Vermerris W (2012) *Brown*
635 *midrib2* (*Bmr2*) encodes the major 4-coumarate:coenzyme A ligase involved in lignin biosynthesis in
636 sorghum (*Sorghum bicolor* (L.) Moench). Plant J 70:818-830
637 Saintenac C, Lee WS, Cambon F, Rudd JJ, King RC, Marande W, Powers SJ, Bergès H, Phillips AL, Uauy
638 C, Hammond-Kosack KE, Langin T, Kanyuka K (2018) Wheat receptor-kinase-like protein Stb6 controls
639 gene-for-gene resistance to fungal pathogen *Zymoseptoria tritici*. Nat Genet 50:368-374
640 Sánchez-Martín J, Steuernagel B, Ghosh S, Herren G, Hurni S, Adamski N, Vrána J, Kubaláková M,
641 Krattinger SG, Wicker T, Doležel J, Keller B, Wulff BB (2016) Rapid gene isolation in barley and wheat by
642 mutant chromosome sequencing. Genome Biol 17:221
643 Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes JC, Hutchison CA, Slocombe PM, Smith M
644 (1977) Nucleotide sequence of bacteriophage φX174 DNA. Nature 265:687-695

Sattler SE, Funnell-Harris DL, Pedersen JF (2010) Brown midrib mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues. Plant Sci 178:229-238
Sattler SE, Saballos A, Xin Z, Funnell-Harris DL, Vermerris W, Pedersen JF (2014) Characterization of novel sorghum *brown midrib* mutants from an EMS-mutagenized population. G3 (Bethesda) 4:2115-2124
Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112-1115
Schneeberger K (2014) Using next-generation sequencing to isolate mutant genes from forward genetic screens. Nat Rev Genet 15:662-676
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
Silva P, Calvo-Salazar V, Condón F, Quincke M, Pritsch C, Gutiérrez L, Castro A, Herrera-Foessel S, von Zitzewitz J, Germán S (2015) Effects and interactions of genes *Lr34*, *Lr68* and *Sr2* on wheat leaf rust adult plant resistance in Uruguay. Euphytica 204:599-608
Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270:1804-1806
Stadler LJ (1928) Mutations in barley induced by X-rays and radium. Science 68:186-187
Stein JC, Yu Y, Copetti D, Zwickl DJ, Zhang L, Zhang C, Chougule K, Gao D, Iwata A, Goicoechea JL, Wei S, Wang J, Liao Y, Wang M, Jacquemin J, Becker C, Kudrna D, Zhang J, Londono CEM, Song X, Lee S, Sanchez P, Zuccolo A, Ammiraju JSS, Talag J, Danowitz A, Rivera LF, Gschwend AR, Noutsos C, Wu CC, Kao SM, Zeng JW, Wei FJ, Zhao Q, Feng Q, El Baidouri M, Carpentier MC, Lasserre E, Cooke R, Rosa Farias DD, da Maia LC, Dos Santos RS, Nyberg KG, McNally KL, Mauleon R, Alexandrov N, Schmutz J, Flowers D, Fan C, Weigel D, Jena KK, Wicker T, Chen M, Han B, Henry R, Hsing YC, Kurata N, de Oliveira AC, Panaud O, Jackson SA, Machado CA, Sanderson MJ, Long M, Ware D, Wing RA (2018) Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza. Nat Genet 50:285-296
Steuernagel B, Periyannan SK, Hernández-Pinzón I, Witek K, Rouse MN, Yu G, Hatta A, Ayliffe M, Bariana H, Jones JD, Lagudah ES, Wulff BB (2016) Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. Nat Biotechnol 34:652-655
Talamè V, Bovina R, Sanguineti MC, Tuberosa R, Lundqvist U, Salvi S (2008) TILLMore, a resource for the discovery of chemically induced mutants in barley. Plant Biotechnol J 6:477-485
The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763-768
The International Wheat Genome Sequencing Consortium, IWGSC RefSeq principal investigators, Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, investigators Iw-gap, Pozniak CJ, Stein N, Choulet F, Distelfeld A, Eversole K, Poland J, Rogers J, Ronen G, Sharpe AG, Whole-genome s, assembly, Pozniak C, Ronen G, Stein N, Barad O, Baruch K, Choulet F, Keeble-Gagnere G, Mascher M, Sharpe AG, Ben-Zvi G, Josselin AA, Hi Cd-bs, Stein N, Mascher M, Himmelbach A, Whole-genome assembly quality c, analyses, Choulet F, Keeble-Gagnere G, Mascher M, Rogers J, Balfourier F, Gutierrez-Gonzalez J, Hayden

M, Josselin AA, Koh C, Muehlbauer G, Pasam RK, Paux E, Pozniak CJ, Rigault P, Sharpe AG, Tibbits J, Tiwari V, Pseudomolecule a, Choulet F, Keeble-Gagnere G, Mascher M, Josselin AA, Rogers J, RefSeq genome s, gene a, Spannagl M, Choulet F, Lang D, Gundlach H, Haberer G, Keeble-Gagnere G, Mayer KFX, Ormanbekova D, Paux E, Prade V, Simkova H, Wicker T, Automated a, Choulet F, Spannagl M, Swarbreck D, Rimbert H, Felder M, Guilhot N, Gundlach H, Haberer G, Kaithakottil G, Keilwagen J, Lang D, Leroy P, Lux T, Mayer KFX, Twardziok S, Venturini L, Manual gene c, Appels R, Rimbert H, Choulet F, Juhasz A, Keeble-Gagnere G, Subgenome comparative a, Choulet F, Spannagl M, Lang D, Abrouk M, Haberer G, Keeble-Gagnere G, Mayer KFX, Wicker T, Transposable e, Choulet F, Wicker T, Gundlach H, Lang D, Spannagl M, Phylogenomic a, Lang D, Spannagl M, Appels R, Fischer I, Transcriptome a, data RN-s, Uauy C, Borrill P, Ramirez-Gonzalez RH, Appels R, Arnaud D, Chalabi S, Chalhoub B, Choulet F, Cory A, Datla R, Davey MW, Hayden M, Jacobs J, Lang D, Robinson SJ, Spannagl M, Steuernagel B, Tibbits J, Tiwari V, van Ex F, Wulff BBH, Whole-genome m, Pozniak CJ, Robinson SJ, Sharpe AG, Cory A, Histone mark a, Benhamed M, Paux E, Bendahmane A, Concia L, Latrasse D, tags BACcMI-BW-GP, Rogers J, Jacobs J, Alaux M, Appels R, Bartos J, Bellec A, Berges H, Dolezel J, Feuillet C, Frenkel Z, Gill B, Korol A, Letellier T, Olsen OA, Simkova H, Singh K, Valarik M, van der Vossen E, Vautrin S, Weining S, Chromosome LTCm, physical mapping quality c, Korol A, Frenkel Z, Fahima T, Glikson V, Raats D, Rogers J, mapping RH, Tiwari V, Gill B, Paux E, Poland J, Optical m, Dolezel J, Cihalikova J, Simkova H, Toegelova H, Vrana J, Recombination a, Sourdille P, Darrier B, Gene family a, Appels R, Spannagl M, Lang D, Fischer I, Ormanbekova D, Prade V, family CBFg, Barabaschi D, Cattivelli L, Dehydrin gene f, Hernandez P, Galvez S, Budak H, family NLRg, Steuernagel B, Jones JDG, Witek K, Wulff BBH, Yu G, family PPRg, Small I, Melonek J, Zhou R, Prolamin gene f, Juhasz A, Belova T, Appels R, Olsen OA, family WAKg, Kanyuka K, King R, Stem solidness QTLt, Nilsen K, Walkowiak S, Pozniak CJ, Cuthbert R, Datla R, Knox R, Wiebe K, Xiang D, Flowering locus Cgt, Rohde A, Golds T, Genome size a, Dolezel J, Cizkova J, Tibbits J, MicroRna, t RNAa, Budak H, Akpinar BA, Biyiklioglu S, Genetic m, mapping, Muehlbauer G, Poland J, Gao L, Gutierrez-Gonzalez J, N'Daiye A, libraries BAC, chromosome s, Dolezel J, Simkova H, Cihalikova J, Kubalakova M, Safar J, Vrana J, Bac pooling BAClr, access, Berges H, Bellec A, Vautrin S, sequence I, data r, access, Alaux M, Alfama F, Adam-Blondon AF, Flores R, Guerche C, Letellier T, Loaec M, Quesneville H, Physical m, sequences BA-b, sequencing AB, assembly, Pozniak CJ, Sharpe AG, Walkowiak S, Budak H, Condie J, Ens J, Koh C, Maclachlan R, Tan Y, Wicker T, sequencing BB, assembly, Choulet F, Paux E, Alberti A, Aury JM, Balfourier F, Barbe V, Couloux A, Cruaud C, Labadie K, Mangenot S, Wincker P, D D, mapping Dp, Gill B, Kaur G, Luo M, Sehgal S, mapping ALp, Singh K, Chhuneja P, Gupta OP, Jindal S, Kaur P, Malik P, Sharma P, Yadav B, mapping ASp, Singh NK, Khurana J, Chaudhary C, Khurana P, Kumar V, Mahato A, Mathur S, Sevanthi A, Sharma N, Tomar RS, B DBB, maps DI-BW-GPp, Rogers J, Jacobs J, Alaux M, Bellec A, Berges H, Dolezel J, Feuillet C, Frenkel Z, Gill B, Korol A, van der Vossen E, Vautrin S, mapping ALp, Gill B, Kaur G, Luo M, Sehgal S, mapping DSp, sequencing BAC, assembly, Bartos J, Holusova K, Plihal O, sequencing DB, assembly, Clark MD, Heavens D, Kettleborough G, Wright J, A physical mapping BACsa, annotation, Valarik M, Abrouk M, Balcarkova B, Holusova K, Hu Y, Luo M, sequencing BB, assembly, Salina E, Ravin N, Skryabin K, Beletsky A, Kadnikov V, Mardanov A, Nesterov M, Rakitin A, Sergeeva E, sequencing BB, assembly, Handa H, Kanamori H, Katagiri S, Kobayashi F, Nasuda S, Tanaka T, Wu J, mapping Ap, sequencing BAC, Appels R, Hayden M, Keeble-Gagnere G, Rigault P, Tibbits J, B physical mapping BACs, assembly, Olsen OA, Belova T, Cattonaro F, Jiumeng M, Kugler K, Mayer KFX, Pfeifer M, Sandve S, Xun X, Zhan B, sequencing DB, assembly, Simkova H, Abrouk M, Batley J, Bayer PE, Edwards D, Hayashi S, Toegelova H, Tulpova Z, Visendi P, mapping DLp, sequencing BAC, Weining S, Cui L, Du X, Feng K, Nie X, Tong W, Wang L, Figures, Borrill P, Gundlach H, Galvez S, Kaithakottil G, Lang D, Lux T, Mascher M, Ormanbekova D, Prade V, Ramirez-Gonzalez RH, Spannagl M, Stein N, Uauy C, Venturini L, Manuscript writing t, Stein N, Appels R, Eversole K, Rogers J, Borrill P, Cattivelli L, Choulet F, Hernandez P, Kanyuka K, Lang D, Mascher M, Nilsen K, Paux E, Pozniak CJ, Ramirez-Gonzalez RH, Simkova H, Small I, Spannagl M, Swarbreck D, Uauy C (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361

Thind AK, Wicker T, Šimková H, Fossati D, Moullet O, Brabant C, Vrána J, Doležel J, Krattinger SG (2017) Rapid cloning of genes in hexaploid wheat using cultivar-specific long-range chromosome assembly. Nat Biotechnol 35:793-796

Thomas Jr. CA (1971) The genetic organization of chromosomes. Annu Rev Genet 5:237-256

Till BJ, Cooper J, Tai TH, Colowit P, Greene EA, Henikoff S, Comai L (2007) Discovery of chemically induced mutations in rice by TILLING. BMC Plant Biol 7:19

Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, Young K, Bowers E, Codomo CA, Enns LC, Odden AR, Greene EA, Comai L, Henikoff S (2004) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4

Uauy C, Paraiso F, Colasuonno P, Tran RK, Tsai H, Berardi S, Comai L, Dubcovsky J (2009) A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. BMC Plant Biol 9:115

Uauy C, Wulff BBH, Dubcovsky J (2017) Combining traditional mutagenesis with new high-throughput sequencing and genome editing to reveal hidden variation in polyploid wheat. Annu Rev Genet 51:435-454

Varshney RK, Shi C, Thudi M, Mariac C, Wallace J, Qi P, Zhang H, Zhao Y, Wang X, Rathore A, Srivastava RK, Chitikineni A, Fan G, Bajaj P, Punnuri S, Gupta SK, Wang H, Jiang Y, Couderc M, Katta M, Paudel DR, Mungra KD, Chen W, Harris-Shultz KR, Garg V, Desai N, Doddamani D, Kane NA, Conner JA, Ghatak A, Chaturvedi P, Subramaniam S, Yadav OP, Berthouly-Salazar C, Hamidou F, Wang J, Liang X, Clotault J, Upadhyaya HD, Cubry P, Rhoné B, Gueye MC, Sunkar R, Dupuy C, Sparvoli F, Cheng S, Mahala RS, Singh B, Yadav RS, Lyons E, Datta SK, Hash CT, Devos KM, Buckler E, Bennetzen JL, Paterson AH, Ozias-Akins P, Grando S, Wang J, Mohapatra T, Weckwerth W, Reif JC, Liu X, Vigouroux Y, Xu X (2017) Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. Nat Biotechnol 35:969-976

Virlet N, Sabermanesh K, Sadeghi-Tehran P, Hawkesford MJ (2017) Field Scanalyzer: An automated robotic field phenotyping platform for detailed crop monitoring. Funct Plant Biol 44

Vollbrecht E, Duvick J, Schares JP, Ahern KR, Deewatthanawong P, Xu L, Conrad LJ, Kikuchi K, Kubinec TA, Hall BD, Weeks R, Unger-Wallace E, Muszynski M, Brendel VP, Brutnell TP (2010) Genome-wide distribution of transposed *Dissociation* elements in maize. Plant Cell 22:1667-1685

Vrána J, Kubaláková M, Simková H, Číhalíkovái J, Lysák MA, Dolezel J (2000) Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). Genetics 156:2033-2041

Wang GL, Holsten TE, Song WY, Wang HP, Ronald PC (1995) Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa-21* disease resistance locus. Plant J 7:525-533

Wang P, Dudareva N, Morgan JA, Chapple C (2015) Genetic manipulation of lignocellulosic biomass for bioenergy. Curr Opin Chem Biol 29:32-39

Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F, Mansueto L, Copetti D, Sanciangco M, Palis KC, Xu J, Sun C, Fu B, Zhang H, Gao Y, Zhao X, Shen F, Cui X, Yu H, Li Z, Chen M, Detras J, Zhou Y, Zhang X, Zhao Y, Kudrna D, Wang C, Li R, Jia B, Lu J, He X, Dong Z, Xu J, Li Y, Wang M, Shi J, Li J, Zhang D, Lee S, Hu W, Poliakov A, Dubchak I, Ulat VJ, Borja FN, Mendoza JR, Ali J, Li J, Gao Q, Niu Y, Yue Z, Naredo MEB, Talag J, Wang X, Li J, Fang X, Yin Y, Glaszmann JC, Zhang J, Li J, Hamilton RS, Wing RA, Ruan J, Zhang G, Wei C, Alexandrov N, McNally KL, Li Z, Leung H (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature 557:43-49

Watson A, Ghosh S, Williams MJ, Cuddy WS, Simmonds J, Rey MD, Asyraf Md Hatta M, Hinchliffe A, Steed A, Reynolds D, Adamski NM, Breakspear A, Korolev A, Rayner T, Dixon LE, Riaz A, Martin W, Ryan M, Edwards D, Batley J, Raman H, Carter J, Rogers C, Domoney C, Moore G, Harwood W, Nicholson P, Dieters MJ, DeLacy IH, Zhou J, Uauy C, Boden SA, Park RF, Wulff BBH, Hickey LT (2018) Speed breeding is a powerful tool to accelerate crop research and breeding. Nat Plants 4:23-29

Witek K, Jupe F, Witek AI, Baker D, Clark MD, Jones JDG (2016) Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. Nat Biotechnol 34:656-660

Xin Z, Wang ML, Barkley NA, Burow G, Franks C, Pederson G, Burke J (2008) Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. BMC Plant Biol 8:103

Zhang J, Chen LL, Xing F, Kudrna DA, Yao W, Copetti D, Mu T, Li W, Song JM, Xie W, Lee S, Talag J, Shao L, An Y, Zhang CL, Ouyang Y, Sun S, Jiao WB, Lv F, Du B, Luo M, Maldonado CE, Goicoechea JL, Xiong L, Wu C, Xing Y, Zhou DX, Yu S, Zhao Y, Wang G, Yu Y, Luo Y, Zhou ZW, Hurtado BE, Danowitz A, Wing RA, Zhang Q (2016) Extensive sequence divergence between the reference genomes of two elite *indica* rice varieties Zhenshan 97 and Minghui 63. Proc Natl Acad Sci USA 113:E5163-5171

Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, Zhan Q, Lu Y, Zhang L, Huang T, Wang Y, Fan D, Zhao Y, Wang Z, Zhou C, Chen J, Zhu C, Li W, Weng Q, Xu Q, Wang ZX, Wei X, Han B, Huang X (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat Genet 50:278-284

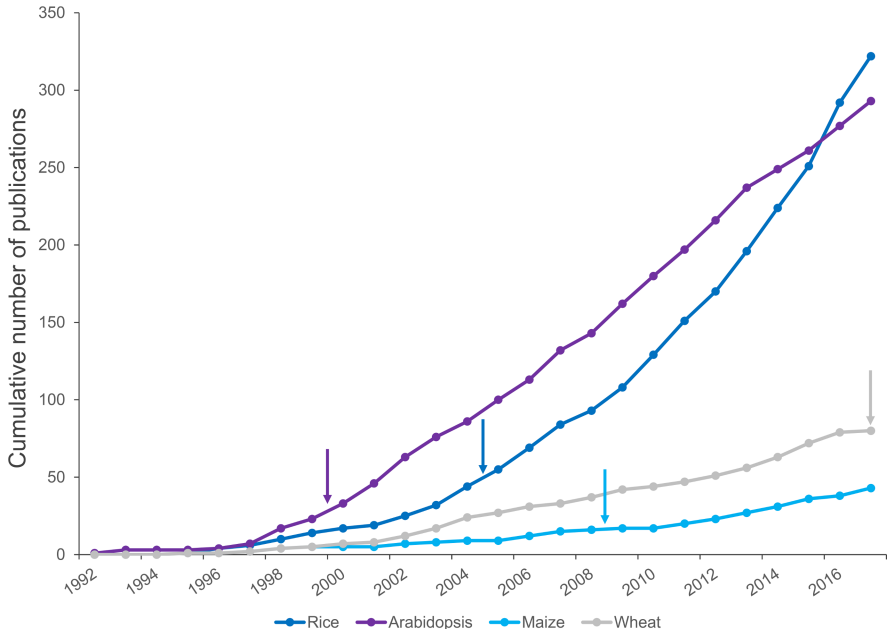


**Figure 1.** Cumulative number of publications that included the words ‘map-based’, ‘gene’, and ‘cloning’ in the title or abstract. The arrows indicate the year of the release of the first high-quality reference sequence for each species. To generate the data, a search with these keywords and the respective species was performed on the NCBI PubMed advanced search builder (https://www.ncbi.nlm.nih.gov/pubmed/advanced).

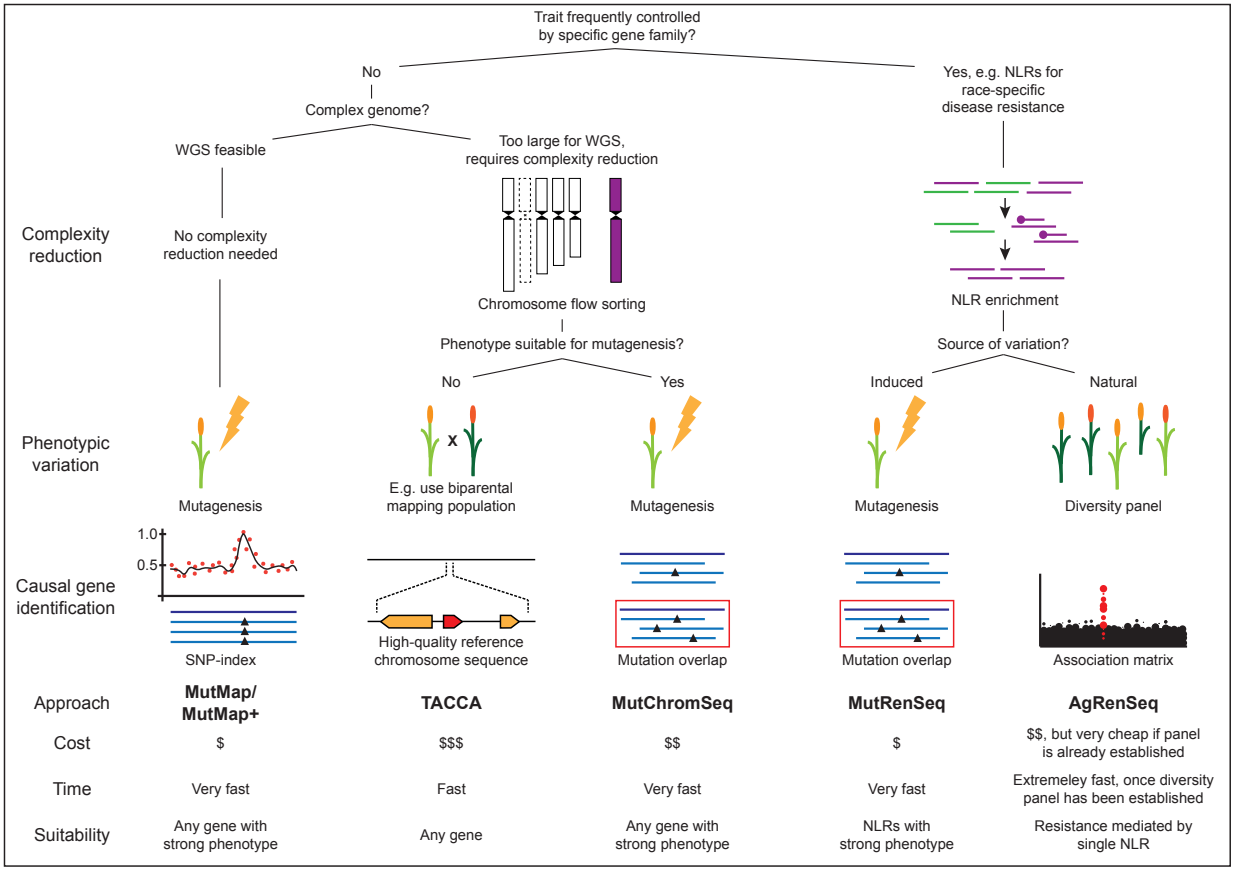


**Figure 2.** An overview of the steps involved in deciding on an approach to rapidly clone a gene of interest in cereals. These approaches are not necessarily mutually exclusive and can often be combined. For example, long-range assembly could also be used for MutChromSeq. Some approaches are only suitable for specific gene families or genes with a strong phenotype. Depending on the plant studied, genome complexity reduction may need to be performed. Many aspects, e.g. whether WGS can be performed, are subjective and determined by the resources available. Some approaches include the creation of time- or cost-intensive resources (e.g. long-range chromosome assembly for TACCA or sequenced diversity panel for AgRenSeq), which can be re-used very cheaply. Should a chosen approach not lead to the identification of the causal gene (e.g. MutRenSeq does not yield a causal NLR-encoding gene), other approaches may offer more success (e.g. MutChromSeq on the mutants identified for MutRenSeq). Abbreviations: NLR = nucleotide binding, leucine-rich repeat encoding gene; WGS = whole-genome sequencing; TACCA = Targeted chromosome-based cloning via long-range assembly.