



# **Bioinformatic-Based Approaches for Disease-Resistance Gene Discovery in Plants**

Andrea Fernandez-Gutierrez 🗅 and Juan J. Gutierrez-Gonzalez \*🗅

Departamento de Biología Molecular, Universidad de León, 24071 León, Spain; aferng33@estudiantes.unileon.es \* Correspondence: jgutg@unileon.es; Tel.: +34-987-293195

Abstract: Pathogens are among the most limiting factors for crop success and expansion. Thus, finding the underlying genetic cause of pathogen resistance is the main goal for plant geneticists. The activation of a plant's immune system is mediated by the presence of specific receptors known as disease-resistance genes (R genes). Typical R genes encode functional immune receptors with nucleotide-binding sites (NBS) and leucine-rich repeat (LRR) domains, making the NBS-LRRs the largest family of plant resistance genes. Establishing host resistance is crucial for plant growth and crop yield but also for reducing pesticide use. In this regard, pyramiding R genes is thought to be the most ecologically friendly way to enhance the durability of resistance. To accomplish this, researchers must first identify the related genes, or linked markers, within the genomes. However, the duplicated nature, with the presence of frequent paralogues, and clustered characteristic of NLRs make them difficult to predict with the classic automatic gene annotation pipelines. In the last several years, efforts have been made to develop new methods leading to a proliferation of reports on cloned genes. Herein, we review the bioinformatic tools to assist the discovery of R genes in plants, focusing on well-established pipelines with an important computer-based component.

Keywords: NBS-LRR genes; pathogen resistance; R genes; NLRs; genome complexity reduction

## 1. Introduction

Crops have long been experiencing an increase in the frequency and range of pests and diseases they are exposed to [1]. The reasons are three-fold. First, the world has become more global than ever before. Plant parts produced at a specific location travel long distances to reach the areas of consumption, which makes exhaustive control of pathogens difficult and makes local plants more prone to having contact with a wide diversity of pathogens. Second, many crops tend to have a very narrow allele pool, including R alleles [2]. This is a consequence of the bottleneck associated with their domestication and the monopoly that certain elite varieties have in a particular region. In an effort to increase the genetic pool of resistance, novel alleles are often sourced from wild related species or local landraces, which is a time-consuming endeavor. Third, the ongoing changes in weather are causing the expansion of climatic niches to novel areas. This brings new diseases to parts of the world not necessarily familiarized with them. The associated problems are either a shortage of genetic resistance within local varieties or a lack of experience among local farmers in controlling these new threats. Contrary to the application of harmful chemical pesticides, the use of R genes in breeding programs represents an environmentally friendly solution to plant pest control.

Plants have different defense mechanisms to counteract pest attacks, reviewed elsewhere (ex. [3]). There are two main types of defenses: mechanical and non-mechanical. The mechanical defenses are based on external impenetrable barriers such as bark and waxy cuticles. Among the non-mechanical defenses, plants have developed race-specific and non-race specific resistances. Non-race specific resistance is based on the recognition of pathogen-associated molecular patterns (PAMPs) which are conserved and widely distributed within a determined class of microbes. On the other hand, race-specific resistance



Citation: Fernandez-Gutierrez, A.; Gutierrez-Gonzalez, J.J. Bioinformatic-Based Approaches for Disease-Resistance Gene Discovery in Plants. *Agronomy* **2021**, *11*, 2259. https://doi.org/10.3390/ agronomy11112259

Academic Editor: Fengjie Sun

Received: 7 October 2021 Accepted: 5 November 2021 Published: 9 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). relies on molecules with antimicrobial properties, such as secondary metabolites, and molecules that trigger a hypersensitive response, leading to rapid cell death in response to infection with an avirulent pathogen [4]. This R-gene mediated response prevents the spread of the infection. The R genes encode plant receptors able to put forth race-specific resistance against pathogens. They trigger the main gene-mediated resistance. Consequently, the identification of R genes in plant genomes has become crucial, as R genes are economically and environmentally valuable traits to include in breeding programs. Although single R genes may confer durable resistance, gene pyramiding of resistance is currently the most sustainable and effective action to prevent and control the spreading of diseases in crops.

A large number of R genes that have been identified to date encode intracellular immune receptor proteins with a nucleotide binding site (NBS) and leucine-rich repeats (LRR). Collectively they are also known as NLR proteins or NLRs [3]. An N-terminal coiled-coil (CC) domain is also present in many members of this class. Other common domains are a Toll-interleukin region (TIR), N-terminal RPW8 (RNL) and a receptor-like kinase domain (RLK) [5]. There are other non-NLR resistance genes, such as patternrecognition receptors (PRR), receptor-like kinases (RLK) and receptor-like proteins RLP. A comprehensive collection of experimentally validated plant NLRs has been recently gathered and contains 442 NLRs from 31 different genera [6]. Within the genomes, they typically appear in clusters, which contain several copies of high-homologous duplicated genes. This redundancy is thought to facilitate rapid R gene evolution and adaptation to new strains. Thus, on one hand, the NLR gene sequences tend to be highly conserved among plant species. This mainly applies to the NBS domains, and not so much to the LRR domains involved in pathogen recognition. On the other hand, because they are under high evolutionary pressure to survive driven by the plant-pathogen interaction force, R genes present great diversity and variability. In fact, R genes may be more structurally and functionally diverse than previously anticipated [6], which includes tandem duplications, transposon-mediated insertion/deletions, extensive sequence diversity and copy number variation between different haplotypes [7].

Global population growth sustainability requires a similar crop production increase to meet the emergent demand [8]. However, crop productivity is greatly threatened by pests and diseases [9,10]. Because traditional breeding and farming practices alone may not be enough to keep up with the needs, researchers are looking at complementary synergistic alternatives [11]. In relatively few years, sequencing technologies have experienced an escalation in their capacities coupled with a tremendous reduction in cost. Bioinformatics tools have progressed in a parallel manner [12]. NLRs are among the most economically valuable genes, and therefore they are a frequent target in breeding programs. Herein, we briefly outline traditional map-based approaches to clone resistance genes, followed by a more extensive review on novel methods with an important high-throughput data analysis component.

#### 2. Traditional Map-Based Cloning

Map-based cloning, or positional cloning, aims to identify the genetic basis of a phenotype by studying the association of genes to markers whose physical location in the genome are known. In this approach, a candidate region must be progressively narrowed down until the causing gene is found, which usually involves the development of high-resolution genetic and physical maps. Although positional cloning does not require prior knowledge of the sequence of the gene of interest, precise genetic map construction encompasses time-consuming development of structured mapping populations such as near isogenic lines, usually with thousands of individuals. It also relies on high-density genetic maps, which are only affordable for chromosome regions with high recombination rates. Nevertheless, recombination events occur at higher frequencies in the telomeres, while they are almost absent in the centromeric regions, which makes traditional cloning of genes in centromeres extremely difficult [13,14]. In turn, modern positional cloning can directly

3 of 16

extract information from sequenced whole genomes, refraining from the need to develop physical maps de novo to scrutinize all genes present in the candidate region. Apart from sequenced genomes, other high-throughput omics data, such as transcriptome assembly or expression of transcriptomes, can assist in the identification of target genes [15,16], for instance, by searching for expression patterns consistent with the onset and development of the disease. Some of the classical successful efforts to clone R genes are outlined in the next paragraphs.

Stem rust caused by the obligate biotroph fungus *Puccinia graminis* is one of the most important foliar diseases of barley and wheat. Positional cloning has effectively isolated key R genes, such as the barley stem rust resistance *rpg4/Rpg5* locus [17]. Both genes are tightly linked in the genome. Using high-resolution mapping populations, authors were able to separate them, and unambiguously identified *Rpg5* as a gene encoding an NLR with an integrated kinase domain. The identity of the *rpg4* was confirmed some years later, which turned to be an actin-depolymerizing factor-like protein [18].

Wild relatives are often selected as sources of novel resistance genes to be introgressed into elite cultivars. For instance, in another classical work Periyannan et al. [19] cloned the Sr33 gene, previously introgressed from the wild relative Aegilops tauschii into bread wheat. The Sr33 confers resistance to diverse stem rust races. They used a single-chromosome substitution line, which has the wheat chromosome 1D replaced by the corresponding homologous chromosome that harbors the Sr33 gene from Ae. tauschii. The introgressed line was then used to generate a recombinant inbred line (RIL) family segregating for Sr33. For fine-mapping the region, two mapping populations of 85 recombinant inbred lines and  $1150 \text{ F}_2$  lines, respectively, from the cross between the introgressed line and the cultivar Chinese Spring were screened, finding 30 individuals with recombination events between the target flanking markers. A physical map that covered the candidate locus was created with the help of a BAC library. The map was determined to contain several genes, including several resistance gene analogs (RGAs). To determine the RGA behind the Sr33 gene, resistant wheat was mutagenized with EMS, identifying nine mutants that had lost Sr33 resistance. The Sr33 gene was found to be orthologous to the barley Mla powdery mildew resistance genes, which provides resistance to Blumeria graminis f. sp. hordei.

The screening of BAC libraries can be cumbersome when the region to fine map is large. Targeted chromosome-based cloning via long-rage assembly has been proposed to simplify the process [20]. Here, lossless genome-complexity reduction is carried out by chromosome flow-sorting and selecting the chromosome where the gene of interest has been previously mapped. Using this approach, Thind et al. [21] cloned the *Lr22a* leaf-rust resistance gene. Leaf rust is caused by the fungus *P. triticina*, which is another devastating disease of wheat with the potential to reduce yields by more than 50% [22]. *Lr22a*, which was previously introgressed from *Ae. tauschii* [23], and mapped to the short arm of chromosome 2D [24], confers resistance to a wide range of the pathogen isolates. Authors first isolated the 2D chromosome by flow cytometry and then de novo assembled it. A high-resolution mapping population allowed narrowing down the genetic interval to 0.09 cM (438 kb), which contained nine genes. The *Lr22a* gene was finally accredited to an NLR that had mutations between the wild-type-resistant and five independent susceptible EMS mutants.

Despite the achievements, traditional introgression breeding of R genes into elite cultivars is a time-consuming process and is usually coupled with undesirable side effects [11]. First, due to gene dragging it can also incorporate other non-beneficial or even deleterious linked genes. This is aggravated by the fact that undomesticated wild relatives are often used as the source of R genes. Second, the transfer of resistance genes from wild relatives by hybridization is also challenging and time consuming due to the lack of pairing between homoeologous chromosomes, which restrict chromosome recombination [25]. Lastly, because pathogens have high mutation rates, they can rapidly evolve to bypass the action of single R genes.

## 3. Bioinformatic-Based Approaches and Pipelines

Positional cloning involving high-resolution mapping populations and chromosome walking is resource-demanding, both money and time wise, even when a reference genome sequence is available. There is no doubt that accessibility of sequencing has ushered in a new era to gene discovering. Nowadays, there is a good chance that a crop of interest has a high-quality genome assembly. However, even if that is the case, it may not be informative because the targeted gene is not present in the sequenced cultivar. Thus, the finding of R genes may be limited by the absence of cultivar-specific genome assemblies. Another major disadvantage of the map-based cloning strategy is that it is based on recombination. Thus, genes located in areas of reduced recombination are not accessible by this technique.

In recent years, some novel approaches to gene cloning with extensive bioinformatics loads have come to light. They usually entail (i) a genome-complexity reduction, using just the subset of the genome that is of interest, (ii) sequencing and assembly of that genome subset and (iii) a bioinformatics pipeline to highlight a group of genes, in silico detection based on domain recognition, or a combination of both. In addition, these new approaches are usually reference-free, independent of fine-mapping and do not require the generation of a physical map spanning the map interval. In this review, we will divide the bioinformatics approaches into two groups: NLR annotation tools and discovering pipelines.

## 3.1. NLR Annotation Tools

Earlier, the decision to call a particular motif-containing protein an NLR was merely manual, making the process very slow. NLRs belong to large multi-gene families, which apart from the NBS and LRR domains may also include other non-canonical domains [5]. This, together with the observation that NLR clusters often contain NLR pseudogenes [20] makes in silico genomic prediction of NLR genes challenging. Annotation tools aim at highlighting specific genes in some sort of assembled contiguous sequence. In the following paragraphs we summarize the most broadly employed NLR automated annotation tools. The main advantages and disadvantages found among all these bioinformatic tools are compiled in Table 1.

Tool	Dependencies	Advantages	Disadvantages	Input	Reference
NLR-parser	motif alignment and MAST	Discrimination of pseudogenes	Predefined gene models needed	Amino acids	[26]
NLR- annotator	meme-suite, NLR-parser	Independent of gene expression, highest domain annotation accuracy, high sensitivity, high specificity	Partial or pseudogenized genes represented, duplication NLRs with multiple NB-ARC domains	Transcript/genomic	[27]
DRAGO2	HMMER, COILS, TMHMM	High sensitivity, web-based interface	Medium domain annotation accuracy	Transcript/amino acids	[28]
NLGenome- Sweeper	BLAST+, MUSCLE, SAMtools, bedtools, HMMER, InterProScan, TransDecoder	High specificity, previous gene predictions not required, good performance for RNL genes	Duplication NLRs with multiple NB-ARC domains, low domain annotation accuracy, very high computational cost	Transcript/genomic	[29]
RRGPredictor	InterProScan	High specificity, alignment or sequence homology not needed	High computational cost	Transcript/amino acids	[30]
NLRtracker	InterProScan	Output of extracted NB-ARC domain, classification NLRs into subgroups, high specificity	Not enough information available	Transcript/amino acids	[6]

#### Table 1. Principal features of NLR-annotation tools.

#### 3.1.1. NLR-Parser

NLR-parser is an automatic tool implemented in java to detect and support the annotation of NLR-encoding genes [26]. It uses motif alignment and a search tool (MAST [31]) to search for a set of 20 conserved motifs found in NLRs. Some of these motifs occur in other protein sequences. To properly classify proteins as NLRs, the NLR-parser uses a set of rules to find combinations of those motifs occurring only in NLRs. Because MAST requires a protein as input, the nucleotide sequence of each fragment to test is first translated into all six reading frames to search for potential NLR motifs. NLR-parser is also able to discriminate pseudogenes as it looks for the complete set of motifs that define an NLR protein. The NLR MAST-parser is implemented as a java program and it has been included in other tools as a part of their pipelines, such as AgRenSeq and NLR-annotator, which will be commented below.

NLR-parser has been successfully used in several annotation schemes. For instance, Wang and collaborators [32] accomplished large-scale identification and functional analysis of NLR genes in a particular rice cultivar. The cultivar was selected for its durable broad-spectrum resistance to blast, a devastating disease caused by the fungus *Magnaporthe oryzae*. They de novo sequenced the cultivar's genome and were able to annotate 455 NLRs. The NLR genes were predicted using hmmscan [33] and NLR-parser. They cloned and tested 219 of those NLRs in susceptible cultivars, and 90 of them showed strong resistance to more than one strain. However, none of the tested NLRs showed resistance to all pathogen strains assayed, suggesting that several NLRs are required for broad resistance. This aspect has been broadly documented [34]. Interestingly, authors established that cultivar's broad resistance was due, rather than to the number of stacked NLRs, to their acting synergistically as interacting pairs. Within a pair, one NLR gene (the *helper*) is thought to activate plant defense signaling after detecting the pathogen, while the other (*sensor*) would recognize pathogen effectors to prevent autoimmunity when the pathogen is not present [35].

#### 3.1.2. NLR-Annotator

Low expression coupled with sequence homology may obstruct the precise annotation of NLR genes. Steuernagel et al. [27] have developed an extension of the NLR-parser [26] termed NLR-annotator, a bioinformatics tool for de novo identification and genome annotation of NLRs independent of gene expression support. The pipeline is implemented in java and has three steps. In the first, the input sequences are split into overlapping fragments. Sequences that can be used as input are genomic contigs and scaffolds, transcriptome assemblies and raw long-read sequencing data. In the second, the NLR-parser script creates a *xml*-interface. Lastly, the third step takes this *xml* file as input to annotate the NLR loci, generating the coordinates and orientations on the input sequences.

Authors tested the tool in nine high-quality and well-annotated reference plant genome assemblies, among them Arabidopsis, soybean, tomato and Brachypodium. Despite the fact that NLR-annotator uses stringent parameters to prevent false positives, they were able to confirm a great number of them. The authors also tried the tool in the intricate hexaploid wheat genome [13], finding 3400 full-length NLR loci. Importantly, a great majority of those NLRs (88%) had low basal expression. The authors also pointed at the potential practical advantage of using NLR-annotator in conjunction with R genes that have been previously mapped on physical positions on chromosomes but that have not yet been cloned. Using this approach, they could find putative candidate genes for many of those, including stem rust, leaf rust, powdery mildew and yellow rust resistance genes [27].

#### 3.1.3. DRAGO2

The Plant Resistance Genes Database (PRGdb) is a comprehensive open online platform for analysis and prediction of plant disease resistance genes through a user-friendly interface [28]. The database hosts both bulk data files and curated gene annotations. In its current version, PRGdb has 177,072 annotated candidate Pathogen Receptor Genes (PRGs), as well as 153 reference R genes. There are a total of 99 species with annotated PRGs represented in the database.

In addition to a BLAST search tool that makes users able to browse their own sequences, a new bioinformatics tool, termed DRAGO2, was implemented and included as part of the PRGdb tool set. This tool automatically predicts and annotates Pathogen Receptor Genes (PRGs) from DNA and amino acid sequences. The core of the DRAGO2 pipeline is a perl script that predicts putative PRGs from transcriptome or proteome sequenced *fasta* files. It has been trained to detect LRR, K, NBS, CC and TIR domains. Authors validated the pipeline on the well-curated Arabidopsis proteome. The tool was able to predict more than 1700 putative PRGs. In an independent comparison, Kourelis et al. [6] found DRAGO2 to have the highest sensitivity among five other similar tools, as detailed further in the text.

#### 3.1.4. NLGenomeSweeper

NLGenomeSweeper is another pipeline to annotate functional NLR disease resistance genes in genome assemblies. It performs a BLAST-aided identification of complete NB-ARC domains, the most conserved domain in NLR genes [29]. The pipeline allows automatic identification of candidates using a two-pass strategy. The first pass aims at a coarse identification of putative NBS-LRRs. This pass uses the alignment tool tBLASTn [36] to search the assembly with the Pfam profile NB-ARC domain and other consensus sequences. Output sequences obtained in this step are then used to build an analysis-specific profile. In the second pass, the NBS-LRR candidates are polished by these new specific profiles and other class-specific consensus sequences.

The pipeline was tested on the Arabidopsis and sunflower (*Helianthus annuus*) genomes. NLGenomeSweeper could identify 152 putative NBS-LRR proteins; 140 of them matched the manually annotated NLR set from Arabidopsis, which contains 146 genes (96% sensitivity). Thus, there were 12 additional candidates. Six of them correspond to true complete CNL or TNL genes, which have been added to the updated annotation. The other six are partial gene fragments or pseudogenes and were regarded as false positives. On the same set, NLR-annotator Steuernagel et al. [27] identified the same except for two of the RNL genes.

In contrast, the sunflower NLR set is less studied, and thus it is prone to novel inclusions. Its reference genome annotation includes 352 genes [37]. Using the sunflower genome, NLGenomeSweeper identified 503 NLRs, while NLR-annotator found 603. The differences may be attributed to truncated domains, large introns, and fragments originated from structural variations or misassemblies of certain regions [29]. As for the RNL genes, while NLR-annotator could only identify two out of the ten RNL genes, NLGenomeSweeper found eight.

#### 3.1.5. RRGPredictor

The RRGPredictor pipeline [30] is a tool for identification of plant pattern recognition receptors (PRRs), without the need for an alignment tool or sequence homology methods. It relies on the presence and architecture of the main domains within proteins. The pipeline makes use of two perl scripts. The first, *RRG\_DomainDetect.pl*, starts with a *tsv* file generated by InterProScan [38] and filters out the domains of interest, which are selected by the user, to different output files. The second script, *ClassRRG.pl*, employs two processes. Initially, all lists generated after running the first script are compared among them, selecting sequence IDs if they intersect in the lists. Then, these sequences are compared and classified, eliminating duplications. Finally, separate files for each of the user-selected domains are generated with non-duplicated sequences.

The protocol was tested on 24 plant and algae reference genomes, including Arabidopsis. The later was chosen for a comparison with other similar tools, including DRAGO2. For many of the classes selected for comparison, DRAGO2 (three classes) and RRGPredictor (five) detected a higher number of sequences. Additionally, the sensitivity, or the capacity to detect true sequences, and the specificity of RRGPredictor was higher than for the other tools.

## 3.1.6. NLRtracker

Kourelis et al. [6] have published a comprehensive curated collection of experimentally validated NLRs, which in the current version includes a total of 442 NLRs, representing 31 genera of flowering plants including Arabidopsis, Glycine, Medicago, Malus, Prunus, Solanum, Oryza, Triticum and Hordeum. Based on the core features found in the collection, they developed NLRtracker, a pipeline that uses InterProScan [38] and predefined NLR motifs [39] to search and annotate NLR genes.

To benchmark the protocol, the developers compared NLRtracker with other existing NLR-annotation pipelines. Benchmarking was performed by determining their sensitivity and accuracy in finding NLR domains. They initially tested five of the most popular NLR annotation tools: DRAGO2, NLGenomeSweeper, NLR-annotator, RGAugury and RRGPredictor. They found DRAGO2 and NLR-annotator to have the highest sensitivity, retrieving 99.3 and 97.4% of the genomic sequences. With regards to annotation specificity, NLR-annotator had the highest with 86.9%, followed by RRGPredictor with 62.2% [6]. The developers also compared NLRtracker to the other NLR-annotation tools in terms of sensitivity and specificity on the Arabidopsis, tomato and rice reference genomes. Sensitivity, or total percentage of NLRs retrieved out of the total NLR data collection, was higher in NLRtracker, followed by DRAGO2. Notably, three of the pipelines: NLR-annotator, NLGenomeSweeper and NLRtracker, reached 100% specificity, defined as the total number of sequences annotated as NLRs that are in fact true NLRs.

## 3.2. NLR Discovering Pipelines

The annotation tools are intended to detect unambiguous motifs on a longer sequence. In contrast, discovering pipelines are more elaborated, and usually combine a phase of data sampling to generate a reduced amount of sequence with a detection phase, in which the desired genes are highlighted. As with the NLR annotation tools, here we will review the most frequently adopted pipelines. Figure 1 compares and summarizes the methods.

## 3.2.1. RenSeq

Resistance gene enrichment sequencing (RenSeq) combines gene enrichment with sequencing to highlight NLR genes [40]. Complexity reduction is accomplished by means of family-specific exome capture library construction enriched for R genes. To demonstrate the approach, authors used the Agilent SureSelect Target Enrichment System to preferentially select NB-LRR sequences with the help of biotinylated oligonucleotide baits. These customized baits were designed to selectively capture DNA fragments that contained NLR motifs. The protocol ends with high-throughput sequencing of the captured DNA fragments. The main advantage of RenSeq over other methods is the selective attention given to NLR sequences, the largest resistance gene family, greatly reducing data amount and complexity and simplifying downstream analysis. This, however, comes with an associated tradeoff. Only genes targeted by the baits can be pulled out and studied, leaving non-NLR resistance genes out of the picture.

The author's proof-of-concept customized design included about 50 k oligos designed based on 523 NB-LRR-like potato and tomato sequences, two of the most important *Solanum* crops. The recovered genomic fragments were paired-end sequenced with Illumina technology, de novo assembled in contigs, and then searched in for specific sequence motifs putatively characteristic of NB-LRR proteins [41], using a Motif Alignment and Search Tool (MAST) sequence homology search algorithm [39]. A total number of 755 potato NB-LRRs were identified, increasing the number of previously described NLRs by 72%. For tomato, an in silico version of the approach was implemented and used to search for sequence fragments within the assembled tomato chromosomes [42] with matches to the bait-library with at least 80% identity. Among the sequences found (394), putatively encoding NB-LRR

loci from the tomato genome, 67 had not been previously characterized. Nevertheless, because of the use of short Illumina PE 76 bp sequencing, paralogue discrimination was challenging. In an improved version, using the longer MiSeq PE 250 bp reads and the two tomato species that at the moment have been sequenced, Andolfo et al. [43] were able not only to correct about 25% of the erroneously described NLRs, but also to identify 105 novel NLR genes.



**Figure 1.** Overview of protocols for NLR discovering pipelines. Equal and differential steps are lined up to highlight the similarities/differences. WT: wild type. NGS: next generation sequencing. SNV: single nucleotide variant. A figure legend is on the upper right corner.

An extra piece of information comes from the circumstance that the hybridizing fragments are sequenced, and thus, they can be used to identify molecular markers linked to resistance. In fact, Jupe et al. [40] used RenSeq and segregating populations to develop a SNP-calling pipeline to highlight SNPs within the NB-LRR gene sequences that co-segregated with resistance to late blight pathogen *Phytophthora infestans*. These markers can be used for numerous applications, including marker assisted selection (MAS). Recently, Barbey et al. [44] have applied RenSeq to the genomes of commercial octoploid strawberry and two other diploid relatives. Results were used to better characterize the R-gene complement in the genomes of this important berry. In another example, RenSeq markers obtained in a similar manner were used to fine-map the *Rpi-rzc1*, a gene from another potato wild relative that confers broad spectrum resistance to potato late blight [45]. Researchers could narrow down the genomic sector containing the gene to a 1 cM distance.

Variations of the original method have been proposed over time. First, a comparable approach, termed MapRenSeq, was used to genetically map a new wheat leaf rust and stripe rust R locus (*LrAp*), previously introgressed from *Ae. Peregrina* [46]. In this scheme,

a bulked segregant analysis is combined with short read NLR enrichment by RenSeq to narrow down candidate regions in the genome. De novo assembly of the short reads generated, and the subsequent search for polymorphisms between resistant and susceptible pools, resulted in the development of five trait-associated SNP markers that mapped to the long arm of wheat chromosome 6B. These markers will aid in the ongoing efforts to clone the *LrAp* gene, as well as in marker-assisted gene pyramiding.

A second variation of the method came with the circumstance that sequencing of the NLR-exome capture library is typically undertaken with short-read high-throughput sequencing technology. Nevertheless, NLR paralogs tend to appear in high copy number and have highly similar coding sequences, which may hamper the assemblage de novo if short reads are used. Witek et al. [47] have proposed using PacBio SMRT sequencing instead. In an initial step, using a mapping population and short-read RenSeq combined with bulked segregant analysis, they mapped a gene for resistance to potato late blight disease to chromosome 4, between 3.5–8.5 Mb. In a second step, authors used a *Solanum* NLR bait library to capture NLRs from two DNA libraries and sequenced them using SMRT technology. They termed this approach SMRT RenSeq. An additional advantage of using this technology is derived from the average read length (more than 10 kb), compared to the size of the average NLR (3.2 kb), which allows most RenSeq molecules to have multiple sequence passes. These multiple passes are later used to correct errors, which are frequent in this technology. They also demonstrated that SMRT RenSeq captures longer (>1 kb) flanking promoter and terminator sequences.

Recently, long read sequencing in combination with RenSeq has been applied to construct the pan-NLRome of Arabidopsis [48]. The species-wide repertoire of NLR genes was generated with a diversity panel of 64 highly curated accessions, with half of the NLRs being present in most accessions and a range of 167–251 NLRs per accession.

NLRs are also implicated in nematode resistance. For instance, the *H*2 gene, which originates from a wild-type relative, has been linked to resistance against the potato cyst nematode *Globodera pallida*. Strachan and collaborators [49] used a third variation of the original RenSeq method in an attempt to identify sequence polymorphisms associated with this resistance. A drawback of RenSeq is that it can only detect linkage within the proximity of known R-gene loci. To overcome this limitation, authors conducted generic-mapping enrichment sequencing (GenSeq), which can complement and confirm RenSeq results. GenSeq [50] performs enrichment sequencing of any target gene, not just NLRs, anchored to the genome of interest. Both approaches, RenSeq and GenSeq, independently identified SNPs linked to the H2 resistance. Lately, developed allele-specific KASP markers could map the *H*2 locus down to a 4.7 Mb interval on the distal short arm of potato chromosome 5, the first step towards cloning the gene.

#### 3.2.2. MutRenSeq

Although RenSeq has been routinely used to identify NB-LRR gene families in plant genomes, the identification of the particular NLR that is responsible for the resistance is not always straightforward or even feasible. Steuernagel and collaborators [51] designed a clever and cost-effective method that combines RenSeq with chemical mutagenesis and screening for loss-of-function mutants. When applied to finding R genes, a resistant wild-type plant is mutagenized, typically with ethyl methane sulfonate (EMS), and the M<sub>2</sub> mutants screened for individuals with loss-of-resistance phenotype. Because R genes are dominant and suppressor screens tend to recover mutations that occur in R genes instead of in another secondary site [51], candidate genes can be easily isolated if the same gene is mutated in all or most of the loss-of-function individuals.

Authors demonstrated the method with rapid cloning of two wheat stem rust (*P. graminis* sp *tritici*) resistance genes, *Sr*22 and *Sr*45, which had been previously introgressed into hexaploidy wheat from their respective diploid A- and D-genome relatives. Complexity reduction was carried out by target enrichment of genomic DNA with customized *Triticeae* NLR-specific baits. Libraries were constructed for all mutant individuals plus the wild type,

and were high-throughput sequenced. The library from the disease-resistant wild type is usually sequenced at a higher depth and/or with longer reads because it has to be de novo assembled into contigs, while NLR-enriched mutant libraries are typically sequenced with much more reduced coverage and reads mapped to the newly constructed reference wild-type assembly. These mapped reads from the mutant individuals are then used to highlight polymorphisms between them and the wild type. The polymorphisms induced by EMS are single nucleotide variants (SNVs), typically G/C to A/T nucleotide transitions.

A bioinformatics pipeline was designed to facilitate the task of highlighting those EMS-induced SNVs between wild type and mutants [52]. Initially, raw reads of each mutant and wild type are aligned to the wild-type assembly using a short-read aligner. Second, SAMtools [53] are used to filter for reads mapped as a proper pair, that is, in the right orientation and distance. SAMtools are also used to convert the alignment data to *mpileup* format for downstream processing. The java program *Pileup2XML* is then used to prefilter *mpileup* files for potential variations and report those as XML format. Third, NLR-parser [26] is used to filter the wild-type de novo assembly for contigs with NLR signatures. This step helps to filter for off-target sequences always present in target enrichment data. Finally, the *MutantHunter* java program integrates all information and reports wild-type contigs with independent variations to several EMS-mutant lines. The contigs where most mutants have a variation to the wild type are the most likely candidates for independent testing.

For the first wheat stem rust resistance gene, *Sr*22, six independent susceptible EMSmutant plants were obtained, with a number of single-point mutations ranging from 44 to 84. After running the bioinformatics pipeline, a single contig was found that contained independent non-synonymous point mutations in five of the six loss-of-function plants. This contig turned out to be a fragment of the gene. A search was conducted to find the remainder of the gene in other contigs. The fragment was found in one contig that also happened to have a nucleotide variation precisely in the sixth EMS-induced mutant line. Both contigs were then merged and the full sequence of the gene was completed by chromosome walking. The locus encoded a putative CC-NB-LRR gene with four exons. Its function was later confirmed by transformation of an independent stem-rust susceptible cultivar with the *Sr*22 clone. All developed transgenic lines were resistant to the disease.

For the *Sr45* gene, six other different susceptible mutant lines were identified after screening of the EMS-mutagenized resistant wild type. Data processing revealed a single 5266-bp contig with independent single nucleotide variations in all mutants, four nonsense and two missense changes. Further inspection determined that the *Sr45* candidate contig encodes another CC-NLR protein. The gene sequence, including the 5' and 3' UTRs, was completed with chromosome walking, and revealed to contain two introns and three exons.

Following the MutRenSeq protocol, Marchal and collaborators [54] isolated and characterized three major yellow rust resistance genes from wheat: *Yr5*, *Yr7* and *YrSP*. The disease, caused by the fungus *Puccinia striiformis* sp. *tritici*, is a major rust disease in regions with cool and moist climate over the growing season. Using nine, ten and four independent EMS-mutagenized susceptible plants, respectively, authors identified a single candidate contig for each of the three loci. They could establish that the underlying genes were part of a cluster located on chromosome 2B. The three genes encode highly homologous NLR proteins with a non-canonical zinc-finger BED domain. Using the sequence information from these new genes, markers were developed to assist gene stacking in breeding programs.

#### 3.2.3. MutChromSeq

Complexity reduction methods that use a biotinylated bait library as a part of the process to capture R genes sequences (RenSeq, MutRenSeq) are powerful at data cutback; however, they are biased in the sense that only genes that are captured by the bait can be studied. Sometimes R genes are not NLRs, but they fit into other various kinds of proteins. This makes designing proper baits challenging if the aim is to target multiple types of R genes. Mutant Chromosome Sequencing (MutChromSeq) employs a different approach to

genome-complexity reduction, based on flow cytometric chromosome sorting [55]. Because the separation is based on chromosome molecules, it does not exclude any sequence from being targeted. Among its advantages are being lossless and sequence-unbiased and being able to potentially capture all R genes. This is especially relevant in species with large genomes, such as wheat and oats, where whole genome sequencing would be less practical. Some of the drawbacks of this genome-complexity reduction approach are: (i) it relies on the fact that only a few mutants are produced and that those mutated allelic variants produce a similar phenotype for easy identification by screening. In addition, only genes not essential for the survival of the plant can be targeted; (ii) it is limited to species from which chromosomes can be flow-sorted and that are amenable to mutagenesis, that is, if a protocol can be set up that induces a good enough density of mutations without killing the organism. The protocol is very laborious and does not always work. Additionally, isolation of individual chromosomes is a complex technique, and it may not be available or fine-tuned for the species of interest, and (iii) the separation of chromosomes usually comes with contaminants from other chromosomes, which can hamper downstream analysis. For instance, de novo assembly of contigs is much more challenging in polyploids if sequences from different homeologs are present.

The concept of MutChromSeq is essentially the same as MutRenSeq. Like MutRenSeq, the protocol starts with a disease-resistance wild-type individual plant and several loss-of-function EMS-induced mutants. Different from this, mitotic chromosomes of M<sub>3</sub> roots from wild-type and mutants are flow-sorted to separate the chromosome of interest, in which the R gene has been previously mapped. From this point on, the steps are similar for both methods; that is, sequencing of the wild type and de novo assembly, followed by sequencing of mutants at a lower depth and alignment of the reads to the reference wild type for variant calling. A set of java programs is available to assist the implementation of the method [56], which includes preprocessing of the SAM tools pileup format (*Pileup2XML*) and the core program, *MutChromSeq*, to call candidate contigs. Candidate contigs and putative SNPs are visually inspected with the help of a genome viewer, such as Integrative Genome Viewer (*IGV*) [57] or similar, followed by a confirmation, typically through Sanger sequencing. Because the focus is on only a particular chromosome, sequencing and analysis costs are greatly reduced.

Developers initially tested MutChromSeq on barley and wheat. For wheat, they selected six EMS-derived susceptible mutants of a dominant powdery mildew resistance gene (*Pm2*), originally mapped to chromosome 5D. The disease is caused by *Blumeria graminis* sp. *tritici*, an obligate, host-specific fungus that infects wheat leaves. After running the MutChromSeq pipeline, a unique true contig, that is, a contig that is not an artifact, was found with several SNVs in six mutant lines. All mutants were found to have either nonsense or missense usual G/C to A/T transitions. The contig was further dissected and found to contain an NLR-class gene, with CC, NBS and LRR domains.

The leaf rust caused by *P. hordei* is the most widespread and damaging foliar disease in barley [58]. The *Rph1* is a CC-NLR that has been mapped to chromosome 2H and confers resistance in several barley cultivars [59]. Authors successfully cloned the gene using sodium azide as the mutant agent and applying the MutChromSeq pipeline. A single candidate gene was identified and further confirmed harboring mutations in five individuals.

However, not all R genes enclose the canonical disease-resistance domains. A putative chimeric protein with serine/threonine kinase and several C2 domains has been recently cloned through MutChromSeq [60]. The underlying gene, *Pm4*, has a unique domain architecture and confers resistance to wheat powdery mildew. Another unusual characteristic of *Pm4* comes from the observation that it undergoes constitutive alternative splicing leading to two different interacting isoforms, both essential for resistance. Neither the *Pm4* nor a close homologue is present in the Chinese Spring wheat reference genome, demonstrating that MutChromSeq is a sequence-unbiased non-reference approach to finding R genes. Similarly, Kolodziej and collaborators [61] proved the involvement of an ankyrin (ANK)-transmembrane, another non-canonical domain, in race-specific leaf rust resistance

in wheat. The ANK proteins are typically involved in protein–protein interactions and plant immunity [62]. To clone the R gene behind this resistance, they subjected seven EMS-derived mutant seedlings to the MutChromSeq pipeline, which highlighted a single gene (Lr14a) with non-synonymous mutations in all lines.

As long as the requirements stated above are met, MutChromSeq can be applied to find any kind of mutated gene or genomic sequence capable of causing an identifiable phenotype. Thus, the pipeline is not restricted to just R genes. For instance, Sánchez-Martín et al. [55] aimed at identifying a previously cloned gene in barley which is required for wax accumulation on leaves [63]. The gene is termed *Eceriferum-q* and is known to map to chromosome 2H. They analyzed six EMS-derived mutants of a waxy wild type. A candidate contig was found that had 11 nonsense or missense point transitions, typical EMS DNA modifications. It contained one exon with 100% identity to the cloned *Eceriferum-q*. Sanger sequencing later confirmed both the identity of the candidate gene and the point mutations.

#### 3.2.4. AgRenSeq

Mutant generation and screening can be tedious and is not suitable for all genes. For instance, traits regulated for more than one gene would typically require other approaches. Arora et al. [64] developed a method that combines genome-wide association studies (GWAS) with R enrichment and sequencing. Genome-wide association studies (GWAS) use high-throughput genomic technologies to scan entire genomes for genetic variants associated with a disease or any other trait. Several of the advantages of AgRenSeq are derived from the association genetics step, in particular the accumulation of long-time historical recombination events within natural populations, acknowledged to increase the precision of gene-marker associations.

The traditional GWAS methodology is leveraged on the presence of a reference genome. This can be a problem for the study of genes that have diverged from the reference. An additional complication in the study of R genes sometimes comes from the development of resistant lines, a required preceding step. These are often derived from introgressions from distant wild-type genotypes, whose development is time consuming. To circumvent the requirement of a reference genome the use of *kmers* to genotype the diversity panel has been proposed [65], and combined with R-gene enrichment and sequencing to render AgRenSeq [64]. Another advantage to using *kmers* is that they can be generated directly from raw sequence reads. If there are *kmers* in the panel that are significantly associated with the trait of interest, those *kmers* can be used to assemble the reads from which they were derived, and thus reconstruct the sequence of the candidate gene.

Authors demonstrated the approach by cloning four stem rust resistance genes (Sr33, Sr45, Sr46 and SrTA1662) from Aegilops tauschii, the wild progenitor of bread wheat D genome. They designed a RenSeq bait library optimized to capture Ae. tauschii and developed a panel of 195 Ae. tauschii accessions that were phenotyped with races of the wheat stem rust pathogen. The capture library was sequenced with Illumina short-read technology, de novo assembled and scanned with NLR-parser [26]. A kmer-based association genetics analysis was conducted on the panel to identify correlations between *kmer* presence/absence and resistance to the disease. To reconstruct full NLRs, the *kmers* were projected onto the NLR contigs assembled from the Illumina reads. An association matrix was then formed according to *kmer* sequence identities to the NLRs from a given accession, and their correlation with the disease phenotype. Associations above a statistic threshold highlight an R-gene candidate contig. The Sr33 and Sr45 had been previously cloned [19,51] and served as positive confirmations. The newly identified SrTA1662 encodes a CC-NLR protein with 83% amino acid identity to Sr33. Additional support came when authors, using a recombinant inbred line population, found that the gene mapped to the expected genomic interval. The fourth gene, Sr46, which turned out to be another CC-NLR, was validated by fine-mapping and sequencing of candidate genes in the region in three EMS mutants that had lost resistance. Further confirmation was obtained when it was expressed as a transgene and conferred rust resistance in an otherwise susceptible background.

More recently, whole genome shotgun sequencing of a panel of 242 *Ae. tauschii* accessions was used for isolation of a novel *Puccinia graminis* resistance gene as well as for mapping of genes for several other traits [66]. Authors used *kmer*-based association mapping to identify discrete genomic regions with candidate genes for disease and pest resistance.

# 4. Remarks and Perspectives

To secure global food supply in the upcoming years we must develop crops that are resistant to a broader range of pests and diseases. In the last decade, gene-based plant pathology has seen remarkable innovation, parallel to the advancement of genomic and bioinformatics tools. This review represents a current trend by which novel and long-time known disease resistances are being revealed at the gene level. We have summarized the main tools and approaches currently being used, with an emphasis on successful cases. The reduction in costs experienced by NGS has been a game changer for the many analyses that are boosted by genome-wide screenings. In conjunction with bioinformatics, both have made possible numerous plant biology advances in the last decade, including the cloning of R-genes. In this regard, genome complexity reduction methods such as target sequence capture and chromosome sequencing have been transformative because they allowed researchers to clone genes with relatively little funding. However, we have now reached the threshold where sequencing entire genomes is going to displace complexity reduction approaches. Generating a high-quality reference sequence will very soon be a standard procedure in any lab. Sequencing diversity panels or mining the gene banks will probably follow.

**Author Contributions:** Conceptualization, J.J.G.-G.; investigation, J.J.G.-G. and A.F.-G.; writing original draft preparation, J.J.G.-G. and A.F.-G.; writing—review and editing, J.J.G.-G. and A.F.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: Authors would like to express their gratitude to Burkard Steuernagel for his detailed comments and suggestions during the development of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

#### References

- 1. Mcdonald, B.A.; Stukenbrock, E.H. Rapid emergence of pathogens in agro-ecosystems: Global threats to agricultural sustainability and food security. *Philos. Trans. R. Soc. B Biol. Sci.* 2016, 371, 20160026. [CrossRef]
- Viruel, J.; Kantar, M.B.; Gargiulo, R.; Hesketh-Prichard, P.; Leong, N.; Cockel, C.; Forest, F.; Gravendeel, B.; Pérez-Barrales, R.; Leitch, I.J.; et al. Crop wild phylorelatives (CWPs): Phylogenetic distance, cytogenetic compatibility and breeding system data enable estimation of crop wild relative gene pool classification. *Bot. J. Linn. Soc.* 2021, 195, 1–33. [CrossRef]
- 3. Jones, J.; Dangl, J. The plant immune system. Nature 2006, 444, 323–329. [CrossRef]
- 4. Kumar, J.; Ramlal, A.; Kumar, K.; Rani, A.; Mishra, V. Signaling Pathways and Downstream Effectors of Host Innate Immunity in Plants. *Int. J. Mol. Sci.* 2021, 22, 9022. [CrossRef]
- Cesari, S.; Bernoux, M.; Moncuquet, P.; Kroj, T.; Dodds, P.N. A novel conserved mechanism for plant NLR protein pairs: The "integrated decoy" hypothesis. *Front. Plant Sci.* 2014, 5, 606. [CrossRef] [PubMed]
- Kourelis, J.; Sakai, T.; Adachi, H.; Kamoun, S. RefPlantNLR: A comprehensive collection of experimentally validated plant NLRs. *BioRxiv* 2020. [CrossRef]
- Smith, S.M.; Pryor, A.J.; Hulbert, S.H. Allelic and Haplotypic Diversity at the Rp1 Rust Resistance Locus of Maize. *Genetics* 2004, 167, 1939–1947. [CrossRef]
- Calicioglu, O.; Flammini, A.; Bracco, S.; Bellù, L.; Sims, R. The Future Challenges of Food and Agriculture: An Integrated Analysis of Trends and Solutions. *Sustainability* 2019, 11, 222. [CrossRef]
- 9. Myers, S.S.; Smith, M.R.; Guth, S.; Golden, C.D.; Vaitla, B.; Mueller, N.D.; Dangour, A.D.; Huybers, P. Climate Change and Global Food Systems: Potential Impacts on Food Security and Undernutrition. *Annu. Rev. Public Health* **2017**, *382*, 59–77. [CrossRef]

- 10. Kamatham, S.; Munagapati, S.; Manikanta, K.N.; Vulchi, R.; Chadipiralla, K.; Indla, S.H.; Allam, U.S. Recent advances in engineering crop plants for resistance to insect pests. *Egypt. J. Biol. Pest Control* **2021**, *31*, 120. [CrossRef]
- 11. van Wersch, S.; Tian, L.; Hoy, R.; Li, X. Plant NLRs: The Whistleblowers of Plant Immunity. *Plant Commun.* 2020, *1*, 100016. [CrossRef]
- Gutierrez-Gonzalez, J.J.; Garvin, D.F. De Novo Transcriptome Assembly in Polyploid Species; Eds Gasparis, Sebastian. Oat Methods Protoc. 2017, 1536, 209–221. [CrossRef]
- 13. The International Wheat Genome Sequencing Consortium (IWGSC). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*. **2018**, *361*, eaar7191. [CrossRef] [PubMed]
- Gutierrez-Gonzalez, J.J.; Mascher, M.; Poland, J.; Muehlbauer, G.J. Dense genotyping-by-sequencing linkage maps of two Synthetic W7984×Opata reference populations provide insights into wheat structural diversity. *Sci. Rep.* 2019, *9*, 1793. [CrossRef] [PubMed]
- 15. Walkowiak, S.; Gao, L.; Monat, C.; Haberer, G.; Kassa, M.T.; Brinton, J.; Ramirez-Gonzalez, R.H.; Kolodziej, M.C.; Delorean, E.; Thambugala, D.; et al. Multiple wheat genomes reveal global variation in modern breeding. *Nature* **2020**, *588*, 277–283. [CrossRef]
- 16. Gutierrez-Gonzalez, J.J.; Garvin, D.F. Subgenome-specific assembly of vitamin E biosynthesis genes and expression patterns during seed development provide insight into the evolution of oat genome. *Plant Biotechnol. J.* **2016**, *14*, 2147–2157. [CrossRef]
- Brueggeman, R.; Druka, A.; Nirmala, J.; Cavileer, T.; Drader, T.; Rostoks, N.; Mirlohi, A.; Bennypaul, H.; Gill, U.; Kudrna, D.; et al. The stem rust resistance gene Rpg5 encodes a protein with nucleotide-binding-site, leucine-rich, and protein kinase domains. *Proc. Natl. Acad. Sci. USA* 2008, 105, 14970–14975. [CrossRef] [PubMed]
- Wang, X.; Richards, J.; Gross, T.; Druka, A.; Kleinhofs, A.; Steffenson, B.; Acevedo, M.; Brueggeman, R. The rpg4-mediated resistance to wheat stem rust (*Puccinia graminis*) in barley (*Hordeum vulgare*) requires Rpg5, a second NBS-LRR gene, and an actin depolymerization factor. *Mol. Plant Microbe Interact.* 2013, 26, 407–418. [CrossRef]
- Periyannan, S.; Moore, J.; Ayliffe, M.; Bansal, U.; Wang, X.; Huang, L.; Deal, K.; Luo, M.; Kong, X.; Bariana, H.; et al. The Gene *Sr33*, an Ortholog of Barley Mla Genes, Encodes Resistance to Wheat Stem Rust Race Ug99. *Science* 2013, 341, 786–788. [CrossRef] [PubMed]
- 20. Thind, A.K.; Wicker, T.; Krattinger, S.G. Rapid Identification of Rust Resistance Genes Through Cultivar-Specific De Novo Chromosome Assemblies. *Methods Mol. Biol.* 2017, 1659, 245–255. [CrossRef]
- Thind, A.; Wicker, T.; Šimková, H.; Fossati, D.; Moullet, O.; Brabant, C.; Vrána, J.; Doležel, J.; Krattinger, S.G. Rapid cloning of genes in hexaploid wheat using cultivar-specific long-range chromosome assembly. *Nat. Biotechnol.* 2017, 35, 793–796. [CrossRef] [PubMed]
- 22. Huerta-Espino, J.; Singh, R.P.; Germán, S.; McCallum, B.D.; Park, R.F.; Chen, W.Q.; Bhardwaj, S.C.; Goyeau, H. Global status of wheat leaf rust caused by *Puccinia triticina*. *Euphytica* **2011**, 179, 143–160. [CrossRef]
- Dyck, P.L.; Kerber, E.R. Inheritance in hexaploid wheat of adult-plant leaf rust resistance derived from Aegilops squarrosa. *Can. J. Genet. Cytol.* 1970, 12, 175–180. [CrossRef]
- 24. Hiebert, C.W.; Thomas, J.B.; Somers, D.J.; McCallum, B.D.; Fox, S.L. Microsatellite mapping of adult-plant leaf rust resistance gene Lr22a in wheat. *Theor. Appl. Genet.* 2007, 115, 877–884. [CrossRef] [PubMed]
- 25. Gutierrez-Gonzalez, J.J.; Garvin, D.F. Reference Genome-Directed Resolution of Homologous and Homeologous Relationships within and between Different Oat Linkage Maps. *Plant Genome* **2011**, *4*, 178–190. [CrossRef]
- 26. Steuernagel, B.; Jupe, F.; Witek, K.; Jones, J.D.; Wulff, B.B. NLR-parser: Rapid annotation of plant NLR complements. *Bioinformatics* **2015**, *31*, 1665–1667. [CrossRef]
- Steuernagel, B.; Witek, K.; Krattinger, S.G.; Ramirez-Gonzalez, R.H.; Schoonbeek, H.-J.; Yu, G.; Baggs, E.; Witek, A.I.; Yadav, I.; Krasileva, K.V.; et al. The NLR-Annotator Tool Enables Annotation of the Intracellular Immune Receptor Repertoire. *Plant Physiol.* 2020, 183, 468–482. [CrossRef]
- Osuna-Cruz, C.M.; Paytuvi-Gallart, A.; Di Donato, A.; Sundesha, V.; Andolfo, G.; Cigliano, R.A.; Sanseverino, W.; Ercolano, M.R. PRGdb 3.0: A comprehensive platform for prediction and analysis of plant disease resistance genes. *Nucleic Acids Res.* 2018, 46, D1197–D1201. [CrossRef]
- 29. Toda, N.; Rustenholz, C.; Baud, A.; Le Paslier, M.-C.; Amselem, J.; Merdinoglu, D.; Faivre-Rampant, P. NLGenomeSweeper: A Tool for Genome-Wide NBS-LRR Resistance Gene Identification. *Genes* **2020**, *11*, 333. [CrossRef]
- 30. Silva, R.J.S.; Micheli, F. RRGPredictor, a set-theory-based tool for predicting pathogen-associated molecular pattern receptors (PRRs) and resistance (R) proteins from plants. *Genomics* **2020**, *112*, 2666–2676. [CrossRef]
- 31. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. Nucleic Acids Res. 2015, 43, W39–W49. [CrossRef]
- Wang, L.; Zhao, L.; Zhang, X.; Zhang, Q.; Jia, Y.; Wang, G.; Li, S.; Tian, D.; Li, W.H.; Yang, S. Large-scale identification and functional analysis of NLR genes in blast resistance in the Tetep rice genome sequence. *Proc. Natl. Acad. Sci. USA* 2019, 116, 18479–18487. [CrossRef]
- 33. Finn, R.D.; Mistry, J.; Tate, J.; Coggill, P.; Heger, A.; Pollington, J.E.; Gavin, O.L.; Gunasekaran, P.; Ceric, G.; Forslund, K.; et al. The Pfam protein families database. *Nucleic Acids Res.* **2010**, *38*, D211–D222. [CrossRef]
- 34. Barragan, A.C.; Weigel, D. Plant NLR diversity: The known unknowns of pan-NLRomes. Plant Cell 2021, 33, 814-831. [CrossRef]
- 35. Wu, C.-H.; Abd-El-Haliem, A.; Bozkurt, T.O.; Belhaj, K.; Terauchi, R.; Vossen, J.H.; Kamoun, S. NLR network mediates immunity to diverse plant pathogens. *Proc. Natl. Acad. Sci. USA* 2017, *114*, 8113–8118. [CrossRef]

- 36. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [CrossRef]
- 37. Neupane, S.; Andersen, E.J.; Neupane, A.; Nepal, M.P. Genome-Wide Identification of NBS-Encoding Resistance Genes in Sunflower (*Helianthus annuus* L.). *Genes* 2018, *9*, 384. [CrossRef] [PubMed]
- 38. Finn, R.D.; Attwood, T.K.; Babbitt, P.C.; Bateman, A.; Bork, P.; Bridge, A.J.; Chang, H.-Y.; Dosztányi, Z.; El-Gebali, S.; Fraser, M.; et al. InterPro in 2017—Beyond protein family and domain annotations. *Nucleic Acids Res.* 2017, 45, D190–D199. [CrossRef] [PubMed]
- Jupe, F.; Pritchard, L.; Etherington, G.J.; MacKenzie, K.; Cock, P.J.A.; Wright, F.; Sharma, S.K.; Bolser, D.; Bryan, G.J.; Jones, J.D.G.; et al. Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genom.* 2012, 13, 75. [CrossRef] [PubMed]
- Jupe, F.; Witek, K.; Verweij, W.; Śliwka, J.; Pritchard, L.; Etherington, G.J.; Maclean, D.; Cock, P.J.; Leggett, R.M.; Bryan, G.J.; et al. 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.* 2013, *76*, 530–544. [CrossRef] [PubMed]
- 41. Bailey, T.L.; Gribskov, M. Methods and statistics for combining motif match scores. J. Comput. Biol. 1998, 5, 211–221. [CrossRef] [PubMed]
- 42. Tomato Genome Consortium (TGC). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **2012**, *485*, 635–641. [CrossRef] [PubMed]
- 43. Andolfo, G.; Jupe, F.; Witek, K.; Etherington, G.J.; Ercolano, M.R.; Jones, J.D. Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. *BMC Plant Biol.* **2014**, *14*, 120. [CrossRef] [PubMed]
- 44. Barbey, C.R.; Lee, S.; Verma, S.; Bird, K.A.; Yocca, A.E.; Edger, P.P.; Knapp, S.J.; Whitaker, V.M.; Folta, K.M. Disease Resistance Genetics and Genomics in Octoploid Strawberry. *G3 Genes Genomes Genet.* **2019**, *9*, 3315–3332. [CrossRef]
- 45. Brylińska, M.; Tomczyńska, I.; Jakuczun, H.; Wasilewicz-Flis, I.; Witek, K.; Jones, J.D.G.; Śliwka, J. Fine mapping of the Rpi-rzc1 gene conferring broad-spectrum resistance to potato late blight. *Eur. J. Plant Pathol.* **2015**, *143*, 193–198. [CrossRef]
- 46. Narang, D.; Kaur, S.; Steuernagel, B.; Ghosh, S.; Bansal, U.; Li, J.; Zhang, P.; Bhardwaj, S.; Uauy, C.; Wulff, B.B.H.; et al. Discovery and characterisation of a new leaf rust resistance gene introgressed in wheat from wild wheat Aegilops peregrina. *Sci. Rep.* **2020**, *10*, 7573. [CrossRef]
- 47. Witek, K.; Jupe, F.; Witek, A.I.; Baker, D.; Clark, M.D.; Jones, J.D.G. Accelerated cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing. *Nat. Biotechnol.* **2016**, *34*, 656–660. [CrossRef]
- 48. Van de Weyer, A.L.; Monteiro, F.; Furzer, O.J.; Nishimura, M.T.; Cevik, V.; Witek, K.; Jones, J.D.G.; Dangl, J.L.; Weigel, D.; Bemm, F. A Species-Wide Inventory of NLR Genes and Alleles in Arabidopsis thaliana. *Cell* **2019**, *178*, 1260–1272. [CrossRef]
- Strachan, S.M.; Armstrong, M.R.; Kaur, A.; Wright, K.M.; Lim, T.Y.; Baker, K.; Jones, J.; Bryan, G.; Blok, V.; Hein, I. Mapping the *H2* resistance effective against *Globodera pallida* pathotype Pa1 in tetraploid potato. *Theor. Appl. Genet.* 2019, *132*, 1283–1294. [CrossRef]
- 50. Chen, X.; Lewandowska, D.; Armstrong, M.R.; Baker, K.; Lim, T.-Y.; Bayer, M.; Harrower, B.; McLean, K.; Jupe, F.; Witek, K.; et al. Identification and rapid mapping of a gene conferring broad-spectrum late blight resistance in the diploid potato species Solanum verrucosum through DNA capture technologies. *Theor. Appl. Genet.* **2018**, *131*, 1287–1297. [CrossRef]
- Steuernagel, B.; Peiyannan, S.K.; Hernández-Pinzón, I.; Witek, K.; Rouse, M.N.; Yu, G.; Hatta, A.; Ayliffe, M.; Bariana, H.; Jones, J.D.G.; et al. Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat. Biotechnol.* 2016, 34, 652–655. [CrossRef] [PubMed]
- 52. Steuernagel, B.; Witek, K.; Jones, J.D.G.; Wulff, B.B.H. MutRenSeq: A method for rapid cloning of plant disease resistance genes. *Methods Mol. Biol.* **2017**, *1659*, 215–229. [PubMed]
- 53. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009, 25, 2078–2079. [CrossRef]
- Marchal, C.; Zhang, J.; Zhang, P.; Fenwick, P.; Steuernagel, B.; Adamski, N.M.; Boyd, L.; McIntosh, R.; Wulff, B.B.H.; Berry, S.; et al. BED-domain-containing immune receptors confer diverse resistance spectra to yellow rust. *Nat. Plants* 2018, 4, 662–668. [CrossRef]
- 55. Sánchez-Martín, J.; Steuernagel, B.; Ghosh, S.; Herren, G.; Hurni, S.; Adamski, N.; Vrána, J.; Kubaláková, M.; Krattinger, S.G.; Wicker, T.; et al. Rapid gene isolation in barley and wheat by mutant chromosome sequencing. *Genome Biol.* 2016, 17, 221. [CrossRef]
- Steuernagel, B.; Vrána, J.; Karafiátová, M.; Wulff, B.B.H.; Doležel, J. Rapid Gene Isolation Using MutChromSeq. *Methods Mol. Biol.* 2017, 1659, 231–243. [CrossRef]
- 57. Robinson, J.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [CrossRef]
- Park, R.F.; Golegaonkar, P.G.; Derevnina, L.; Sandhu, K.S.; Karaoglu, H.; Elmansour, H.M.; Dracatos, P.M.; Singh, D. Leaf rust of cultivated barley: Pathology and control. *Annu. Rev. Phytopathol.* 2015, *53*, 565–589. [CrossRef] [PubMed]
- Dracatos, P.M.; Bartoj, J.; Elmansour, H.; Singh, D.; Karafiátová, M.; Zhang, P.; Steuernagel, B.; Svačina, R.; Cobbin, J.C.A.; Clark, B.; et al. The Coiled-Coil NLR *Rph1*, Confers Leaf Rust Resistance in Barley Cultivar Sudan. *Plant Physiol.* 2019, 179, 1362–1372. [CrossRef]

- Sánchez-Martín, J.; Widrig, V.; Herren, G.; Wicker, T.; Zbinden, H.; Gronnier, J.; Spörri, L.; Praz, C.R.; Heuberger, M.; Kolodziej, M.C.; et al. Wheat Pm4 resistance to powdery mildew is controlled by alternative splice variants encoding chimeric proteins. *Nat. Plants* 2021, 7, 327–341. [CrossRef]
- 61. Kolodziej, M.C.; Singla, J.; Sánchez-Martín, J.; Zbinden, H.; Šimková, H.; Karafiátová, M.; Doležel, J.; Gronnier, J.; Poretti, M.; Glauser, G.; et al. A membrane-bound ankyrin repeat protein confers race-specific leaf rust disease resistance in wheat. *Nat. Commun.* **2021**, *12*, 956. [CrossRef] [PubMed]
- 62. Vo, K.T.X.; Kim, C.Y.; Chandran, A.K.N.; Jung, K.-H.; An, G.; Jeon, J.-S. Molecular insights into the function of ankyrin proteins in plants. *J. Plant Biol.* 2015, *58*, 271–284. [CrossRef]
- 63. Schneider, L.M.; Adamski, N.M.; Christensen, C.E.; Stuart, D.B.; Vautrin, S.; Hansson, M.; Uauy, C.; von Wettstein-Knowles, P. The Cer-cqu gene cluster determines three key players in a beta-diketone synthase polyketide pathway synthesizing aliphatics in epicuticular waxes. *J. Exp. Bot.* **2016**, *67*, 2715–2730. [CrossRef] [PubMed]
- Arora, S.; Steuernagel, B.; Gaurav, K.; Chandramohan, S.; Long, Y.; Mathy, O.; Johnson, R.; Enk, J.; Periyannan, S.; Singh, N.; et al. Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nat. Biotechnol.* 2019, 37, 139–143. [CrossRef] [PubMed]
- Lees, J.; Vehkala, M.; Välimäki, N.; Harris, S.R.; Chewapreecha, C.; Croucher, N.J.; Marttinen, P.; Davies, M.R.; Steer, A.C.; Tong, S.Y.C.; et al. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nat. Commun.* 2016, 7, 12797. [CrossRef]
- 66. Gaurav, K.; Arora, S.; Silva, P.; Sánchez-Martín, J.; Horsnell, R.; Gao, L.; Brar, G.S.; Widrig, V.; Raupp, J.; Singh, N.; et al. Evolution of the bread wheat D-Subgenome and enriching it with diversity from *Aegilops tauschii*. *Biorxiv* 2021. [CrossRef]