

Chapter 2

Mining the Genus *Solanum* for Increasing Disease Resistance

Jack H. Vossen, Kwang-Ryong Jo and Ben Vosman

Abstract Plant Breeding is the art of selecting and discarding genetic material to achieve crop improvement. Favourable alleles resulting in quality improvement or disease resistance must be added, while unfavourable alleles must be removed. The source for novel alleles can be other varieties, landraces or crop wild relatives. The identification of allelic variation is referred to as allele mining. Before allelic variation can be used for breeding purposes several steps need to be taken. First of all an inventory is needed of the available genetic resources. Phenotypic screens are needed to uncover potential expected and even unanticipated alleles. Next, using genetic and molecular tools, the alleles responsible for the identified traits must be traced and distinguished in order to be introgressed into new varieties.

In this review we focus on the identification of novel disease resistance traits in the agronomically important genus *Solanum*. The fact that *R* genes are present in multigene clusters within the genome, which often include many paralogs necessitates thorough discussion on the distinction between alleles and paralogs. Often such a distinction cannot easily be made. An overview is given of how natural resources can be tapped, e.g. how germplasm can be most efficiently screened. Techniques are presented by which alleles and paralogs can be distinguished in functional and/or genetic screens, including also a specific tagging of alleles and paralogs. Several examples are given in which allele and paralog mining was successfully applied. Also examples are presented as to how allele mining supported our understanding about the evolution of *R* gene clusters. Finally an outlook is provided how the research field of allele mining might develop in the near future.

B. Vosman (✉) · J. H. Vossen · K.-R. Jo
Wageningen UR Plant Breeding, Wageningen University & Research Centre,
P.O. Box 16, 6700 AA Wageningen, The Netherlands
e-mail: ben.vosman@wur.nl

K.-R. Jo
Research Institute of Agrobiology,
Academy of Agricultural Sciences, Pyongyang,
Democratic People's Republic of Korea

2.1 Introduction

2.1.1 General Introduction

Alleles are different forms of a gene and affect a particular process in different ways. Different combinations of alleles may result in different phenotypes. Plant breeders try to improve varieties by introducing new alleles, resulting in higher yields and better quality or resistance characteristics. Identifying new, promising alleles is not an easy task. In the post-genomics era, mining of a crop's (wild) gene pool for novel and superior alleles for agronomically important traits is becoming more and more feasible. Genebanks all over the world contain huge untapped resources of distinct alleles that may have potential application in crop breeding programs. This hidden diversity, which can consist of naturally occurring sequence variation in coding or regulatory regions of genes, can be explored by allele mining (Ramkumar et al. 2010; Varshney et al. 2005, 2009). The variation includes single nucleotide polymorphisms (SNPs) as well as insertions and deletions (InDels), which have the possibility to change the resulting phenotype. This may be by altering the amount of protein or its structure and/or function (Ramkumar et al. 2010). The recent rapid advancements in the field of genomics leads to the accumulation of enormous amounts of sequence information and fast evolving bioinformatic tools which pave the road for identifying, characterizing, isolating, and deploying previously unknown or under-utilized sources of genetic variation.

In this chapter we consider allele mining as the research field that aims at unlocking the genetic diversity existing in genetic resource collections (genebanks) and artificially created mutant populations by identifying allelic variants of genes and loci. Since resistance genes occur in clusters, where allelic relationships are often not clear (Sanchez et al. 2006; Millett et al. 2007) and because paralogs in a cluster can have different functions, the scope of this chapter is broader than allele mining alone. To deal with this we introduce the concept of paralog mining. Paralog mining is the identification of a gene within a cluster of highly homologous genes with different, often unknown, functions. Paralog mining can be used as a tool to generate molecular markers and in combination with functional screens it can be used to identify new genes conferring resistance to a particular pathogen. In this review we discuss how allele and paralog mining can help to improve disease resistance in *Solanum* crops.

2.1.2 Solanaceae Resources

The family of *Solanaceae* is of high economic importance and is composed of more than 3,000 species which include important crop and model plants such as potato (*Solanum tuberosum*), tomato (*Solanum lycopersicum*) and eggplant (*Solanum melongena*) (Knapp 2002), but also wild species occurring in very different habitats

(Spooner and Hijmans 2001; Spooner et al. 2004). About 15,000 wild potato accessions are being maintained in large collections worldwide and the establishment of core and mini collections that enable an effective use of the existing variation in gene banks while maintaining the variability, as has been proposed before (Frankel and Brown 1984; Hoekstra 2009). Allele mining requires the assembly of a reasonably sized core germplasm collection usually comprising ~1,000 accessions representative of genetic diversity existing in the global population (Hofinger et al. 2009). Such collections can effectively be constructed using the Focussed Identification of Germplasm Strategy (FIGS) approach (Mackay et al. 2004; Bhullar et al. 2009). About 15,000 wild potato accessions are being maintained in large collections worldwide and the establishment of core and mini collections that enable an effective use of the existing variation in gene banks while maintaining the variability, as has been proposed before (Frankel and Brown 1984; Hoekstra 2009).

The genome sequence of potato (Potato Genome Sequencing Consortium et al. 2011) and tomato (The Tomato Genome Consortium et al. 2012) will facilitate mining for novel alleles or paralogs of resistance® genes. These may be found in the largely untapped resources of crossable species within the genus *Solanum* allowing their exploitation in breeding programs. Also, insight into sequence diversity at the *R* gene loci in wild *Solanum* species with different resistance response against economically important diseases will result in a better understanding of the mechanism of *R* gene functionality and evolution but can also help to identify new alleles or paralogs with different race specificities, and develop allele-specific diagnostic markers for marker assisted breeding.

2.1.3 Resistance Genes

If a gene is responsible for the resistance of a particular plant to a particular pathogen, this gene is called a resistance® gene. To date, more than 100 *R* genes which confer resistance to a diversity of pathogens including bacteria, fungi, oomycetes, viruses, insects and nematodes have been identified and/or cloned from various plants, by a wide variety of methods including map-based cloning, transposon tagging, and similarity based DNA library screening (Sanchez et al. 2006; Ingvarsdén et al. 2008; Vleeshouwers et al. 2011a). An overview of mapped and cloned *R* genes from Solanaceae is given in Fig. 2.1.

R genes often encode receptors for pathogen derived ligands and they are classified based on the combination of different domains (e.g. CC = coiled coil, TIR = toll interleukin receptor, Protein Kinase, NBS = nucleotide binding site, Lec (lectin), and LRRs = leucine rich repeats). Five classes can be identified, transmembrane proteins with extracellular LRRs (receptor like proteins, RLPs), transmembrane proteins with extracellular LRRs and intracellular protein kinase (receptor like kinases, RLKs), transmembrane proteins with extracellular “lectin like” domain and intracellular protein kinase (lectin receptor kinases, LecRKs), and intracellular NBS-LRR proteins which can be divided in CC-NBS-LRR and TIR-NBS-LRR (Dubery et al.

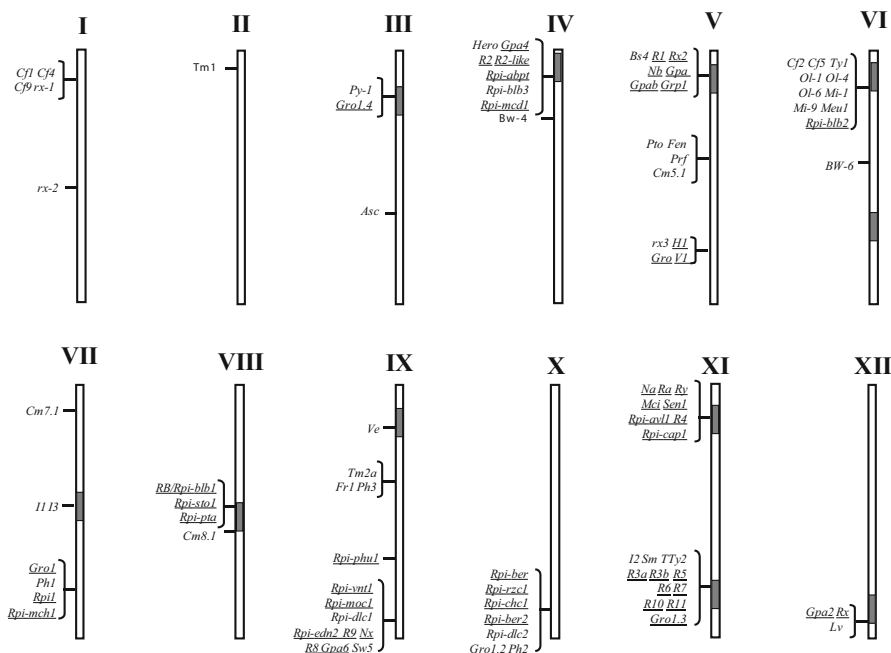


Fig. 2.1 Genetic locations of disease resistance traits in Solanaceae. Twelve linkage groups are shown and the position of *R* genes is indicated. The *R* genes for potato are underlined and those for other species, mainly tomato, are not underlined. Map segments having QTL for resistance to *Phytophthora infestans* in potato are shown in black color

2012). The NBS-LRR class is the most abundant and has been extensively studied (Hulbert et al. 2001). Although NBS-LRR genes are assumed to cause pathogen race specific (or vertical) resistance, it has also been suggested that members of the NBS-LRR gene family are candidates for quantitative trait loci (QTL) that are responsible for horizontal resistance (Rietman et al. 2012; Sanz et al. 2012; Gebhardt and Valkonen 2001). Most characterized plant NBS-LRR genes are physically clustered in the plant genome. The homologous sequences in such a cluster are referred to as paralogs (Gebhardt and Valkonen 2001) and paralogs can confer resistance to different isolates of the same pathogen (Dodds et al. 2001; Li et al. 2011; Lokossou 2010) or to different pathogens (Dodds et al. 2001; van der Vossen et al. 2000). Some paralogs may also be considered as molecular fossils of evolution, whose activity is unclear or even absent, e.g. many pseudogenes have been found. In most *R* gene clusters the number of paralogs is very high and often an allelic relationship is hard to determine (Kuang et al. 2004). However, as the genome structure between species in the *Solanaceae* family is highly conserved, positional conservation of *R* gene clusters (synteny) is observed across *Solanaceous* species (Grube et al. 2000; Park et al. 2009, Fig. 2.1). Therefore, even when relatively unknown genetic sources are used, it is likely that the genes conferring resistance are linked to syntenic clusters of *R* genes known from well-studied species like potato and tomato.

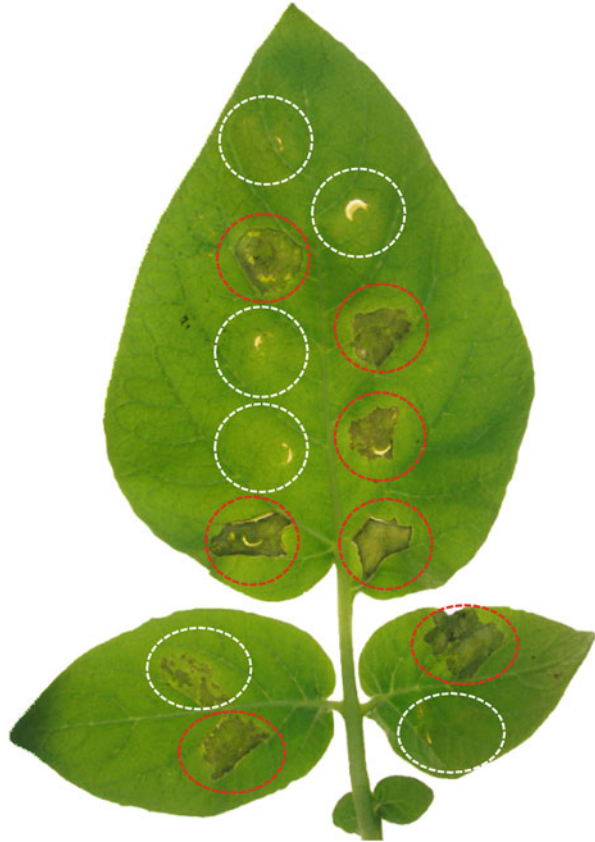
Not just the 2006; Millett et al. identification of new alleles is important, also the functional characterisation of the identified alleles is extremely important to assess the added value of the new allele over alleles that are already present in crop plants. Many approaches have already been used and especially the currently booming research field of effector genomics, through which the identification of *Avr* genes is greatly accelerated, offers fast functional assays to distinguish the activity of newly identified *R* gene alleles and paralogs. So, allele mining approaches coupled with effector profiling enable the discovery of novel *R* genes at an unprecedented rate (Vleeshouwers et al. 2008, 2011a).

2.2 Functional Resistance Screens

2.2.1 *Screening for Disease Resistant Accessions in Gene Bank Material*

Several methods are available to carry out phenotypic screens for disease resistance in gene bank collections. Here we use the evaluation of potato germplasm for late blight resistance as an example. Inoculation of entire *in vitro* plantlets or inoculation of detached leaves can be used as high throughput screening methods (Vleeshouwers 2011b). In case of race specific resistance, the selection of the pathogen isolates is an important issue in the identification of major *R* genes. If the selected isolate happens to be compatible with the *R* gene (s) in a particular accession, these *R* genes may be overlooked. Multiple isolates can be used to distinguish the different *R* genes in a particular resistant accession (Huang et al. 2005; Verzaux 2010). Complementary to working with the entire pathogen, effector responsiveness can be used to identify and classify *R* gene alleles in a germplasm core collection (Rietman et al. 2010). In such an effectoromics approach, effectors or potential *Avr* genes from the pathogen, are expressed in the plant using agro-infiltration or through inoculation with recombinant Potato Virus X, referred to as agro-infection. Upon recognition of the effector by the *R* gene expressed by the plant, a defence reaction, referred to as hypersensitive response (HR), is initiated which is visible as a necrotic lesions in the infiltrated leaf (Fig. 2.2). The agro-infiltration test appears to be applicable and reliable for many genotypes and the variation in the HR in different genetic backgrounds is limited. The use of specific pathogen isolates and the use of specific effectors can also be employed to identify functional groups of *R* genes in germplasm of crop wild relatives. Within groups of functionally similar *R* genes, a true allele mining approach can be pursued in order to identify (sequence) variation. The functional grouping of *R* genes can also be employed to reduce the redundancy that is inevitably present in germplasm collections. Another virtue of the effectoromics approach was shown recently. Potato plants which have shown durable resistance to late blight contained stacks of different *R* genes (Verzaux 2010; Kim et al. 2012). The polygenic nature of the resistances could easily be characterised using the segregation patterns of the

Fig. 2.2 Effector induced hypersensitive response (HR) in *Solanum tuberosum* genotype MaR8. Available effectors from *Phytophthora infestans* were applied using agroinfiltration in leaves of the resistant plant MaR8. The dotted circles surround the infiltrated leaf area. The red dotted circles surround the effectors that elicited an HR. These effectors are selected for validation of *R-Avr* interactions by additional genetic analysis



different effector responses. Effectors which displayed HR response in germplasm screens are potential *Avr* gene(s) recognized by the cognate *R* gene. These potential *R-Avr* interactions should be validated by additional genetic studies. Ideally, by cosegregation of responses to the effector with resistance to *P. infestans* isolates in segregating populations.

2.2.2 QTL Mapping/LD Mapping

Plant pathogen resistance, at the phenotypic level, often does not behave as a single *R* gene but as a quantitative trait that is controlled by multiple genetic and environmental factors (Trognitz et al. 2002; Bai 2003). Understanding the molecular basis for quantitative traits will facilitate diagnosis and facilitate the combination of superior alleles in crop improvement programs. The possible approaches to mapping genes that underlie quantitative traits fall broadly into two categories: candidate gene studies, which use either association or resequencing approaches, and linkage studies,

which include both QTL mapping and genome-wide association studies (GWAS). In this review, we do not discuss GWAS further because of the extensive review by Hirschhorn and Daly (2005). Linkage disequilibrium (LD) mapping, or association analysis based on candidate genes is also considered as an allele mining approach (Malosetti et al. 2007).

Some cases of close linkage between an *R* gene and quantitative trait loci (QTL) for pathogen resistance supports the hypothesis that qualitative and quantitative resistance have a similar molecular basis (Leonards-Schippers et al. 1994), thereby suggesting that genes showing sequence similarity to *R* genes are candidates for being factors underlying quantitative resistance (Rickert et al. 2003; Rietman et al. 2010). Candidate genes participating in the control of the quantitative resistance to pathogens are those involved in the disease response network; (i) *R* genes which recognize the pathogen and trigger the resistance response, (ii) genes which are involved in signal transduction pathways and (iii) the large group of pathogenesis related (PR) genes which are expressed in response to pathogen attack and are involved in the execution phase of the defence response (reviewed by Gebhardt and Valkonen 2001).

The genetic dissection of complex plant traits in QTLs first became possible with the advent of DNA-based markers (Osborn et al. 1987). The first genes and their allelic variants underlying plant QTLs have been identified by positional cloning (reviewed in Salvi and Tuberosa 2005). Positional QTL cloning is a labor- and time-consuming process which requires the generation and analysis of large experimental mapping populations. An alternative to positional cloning of QTLs may be the allele mining approach, which is based on the knowledge of a gene's function in controlling a characteristic of interest on the one hand, and genetic co-localization of a functional candidate gene with QTL of interest on the other (Pflieger et al. 2001; Faino et al. 2011). However, in this approach substantial *a priori* knowledge is required. DNA variation for genes fulfilling these criteria has been examined in natural populations of accessions related by descent for associations with positive or negative characteristic values (Li et al. 2005; Gonzalez-Martinez et al. 2007). Finding such associations indicates that DNA variation either at the candidate locus itself or at a physically linked locus is causal for the phenotypic variation, but defined prove for the involvement of the gene is still circumstantial.

2.3 Techniques for Allele Mining

Dependent on the research question but also dependent on genetic, genomic and financial resources available, several techniques can be used for allele mining, ranging from a rapid and inexpensive polymerase chain reaction (PCR) to next gen sequencing and everything in between. For some applications (partial) sequence information or only molecular polymorphism of the alleles is sufficient. For other applications actual cloning of the entire allele is required. Generally, all DNA based tools require the careful selection of target genes. The target gene model might require verification, and successively, a careful design of selective primers will allow the

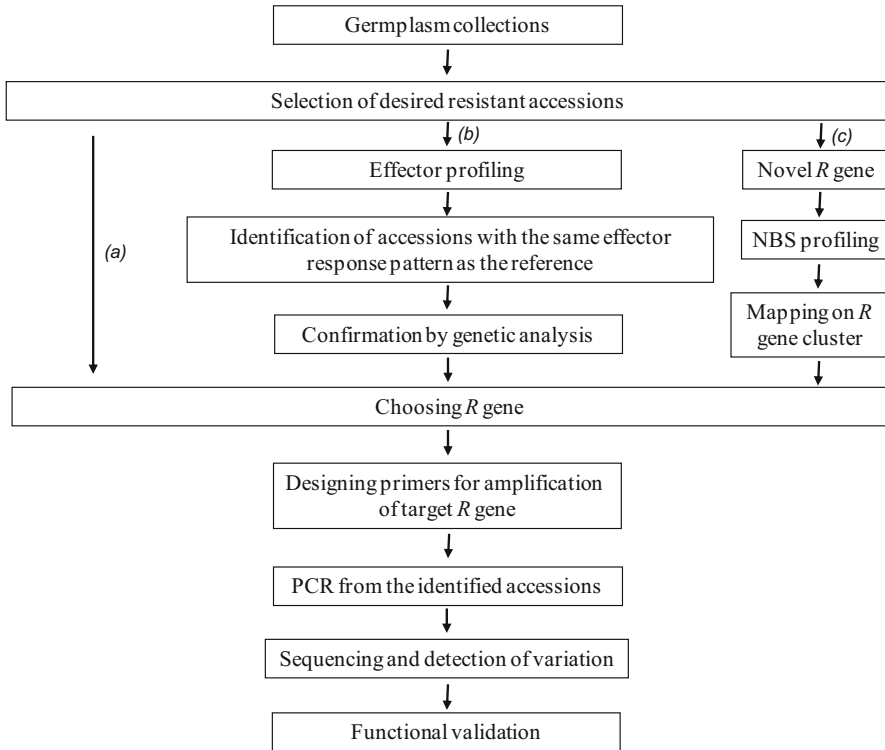


Fig. 2.3 Pipeline of allele mining of *R* genes. **a** Indicates allele mining in case that effector tools are not available for pathogen under study, **b** joint effector profiling-allele mining approach, and **c** novel *R* gene discovery in the combination of NBS profiling and allele mining

identification of novel alleles at candidate loci in the entire or core germplasm collection. In Fig. 2.3 a pipeline for novel allele discovery from germplasm collections is presented, including a combination of different approaches.

2.3.1 Molecular Tools for Allele Tagging

All molecular marker techniques include a PCR amplification of one or multiple alleles or paralogs. In order to identify polymorphisms between amplified alleles, single-strand specific nucleases could be applied. Using this technique, that is often used in TILLING approaches, nicking of heteroduplexes of PCR products can be easily detected. A recent development is the use of high resolution melting point analysis in order to screen for mismatches between amplified alleles in a high throughput fashion. Especially suitable for the highly polymorphic and duplicated *R* genes, is the NBS profiling technique (van der Linden et al. 2004). It is a powerful tool to

identify specific fragments of candidate *R* genes or *R* gene homologs throughout the genome by using degenerated primers that anneal to conserved sequences in the NBS domain of the NBS-LRR class of *R* genes. A high throughput application of this technique is to study fragment length polymorphisms as molecular markers. Also, PCR amplification of specific *R* genes is possible if primers are located in unique regions in order to target the specific paralog. The results may be visible as a DNA fragment of a specific size on an agarose gel. However, when gene specific markers are used in different germplasm material, often a-specific annealing of the primers can occur and therefore it will always be necessary to sequence the resulting PCR fragments to confirm their identity and homogeneity.

2.3.1.1 NBS Profiling

Many plant *R* genes are a member of a multigene cluster composed of multiple copies with high sequence similarity (Song et al. 2003). The NBS region of (NBS-LRR) *R* genes and their analogs (RGAs) contain highly conserved common motifs like the P-loop, the kinase -2 motif and the GLPL motif (Meyers et al. 2003; Monosi et al. 2004). These conserved motifs within the NBS-LRR genes have been used successfully to sequence (parts of) NBS regions from various plant species (Collins et al. 1998; Pflieger et al. 1999; Zhang and Gassmann 2007). NBS profiling uses the conserved motifs for efficient tagging of NBS-LRR type of *R* genes and their analogs (Van der Linden et al. 2004, 2005). The technique involves three different steps. (1) Digestion of genomic DNA with a restriction enzyme and ligation of adaptors to compatible restriction ends. (2) PCR amplification of NBS containing fragments using an NBS primer and an adaptor primer. (3) Separation of amplified fragments by polyacrylamide gel electrophoresis. The technique produces a multilocus profile of the genome.

NBS profiling can easily be adapted to target other conserved gene families, which is referred to as motif-directed profiling (Van der Linden et al. 2004, 2005). Also NBS profiling can be adapted to target particular *R* gene clusters. *R* genes from the same cluster usually have similarities in their sequences not shared with other *R* genes (McDowell and Simon 2006; Meyers et al. 2005), allowing the design of specific primers for a particular *R* gene cluster. NBS profiling could therefore also be adapted to reach high fragment saturation in an *R* gene cluster of interest (Verzaux et al. 2011, 2012; Jo et al. 2011). This technique is referred to as cluster directed profiling.

2.3.1.2 (Eco)-tilling

Eco tilling is a molecular method to screen germplasm core and mini collections. This technique is distinct from the TILLING approach since TILLING screens identify novel alleles that are induced by mutagenesis (Till et al. 2003; Barkley and Wang 2008) whereas eco-tilling identifies naturally occurring alleles in germplasm

(Barone et al. 2009). Both approaches employ a similar screening method to identify variation in alleles. Polymorphisms in PCR amplified DNA fragments are detected in hetroduplexes of the amplicons using single strand specific nucleases, high resolution melting point analysis or deep sequencing in next generation sequencing.

2.3.1.3 Amplification of Specific Allelic Variants

Family members with very similar sequences may have dispersed around the genome into non syntenous loci or may have remained within a genetic locus but has multiplied resulting in tandem or inverted repeats. In general, sequences in coding regions will be more conserved than primers in flanking sequences. Dependent of the downstream application (sequence comparison, in plant expression), primers are chosen in- or outside coding sequence to amplify the entire gene, or part of the gene or only the open reading frame. Because even single nucleotide polymorphisms can be relevant differences between alleles, preferably the DNA polymerase will contain proofreading activity. Also, because often long stretches of the target gene are amplified, a long range polymerase chain reaction (LR-PCR) polymerase is preferred. Examples of enzymes that harbour both characteristics are *Pfu*-Turbo from Invitrogen or Phusion from Fermentas. One approach is to amplify the entire coding sequence of the *R* gene of interest using primers annealing to start and stop codon regions. Subsequently, the amplicon is sequenced and for expression studies it can be cloned in a vector that harbours heterologous regulatory sequences. For some accessions a possible lack of amplification can be expected due to absence of a coding gene or to low sequence homology at the primer annealing sites. A drawback of this approach is that the promotor and terminator regions of the novel alleles are missing, so variation in these regulatory regions are neglected. For 'true' allele mining, the use of primers matching the promotor and terminator regions is feasible when sequence conservation is sufficient. Accessions may also first be screened for the presence of the known *R* gene with a diagnostic molecular marker obtained from haplotype studies at the *R* gene locus and next for the presence of new alleles of the known *R* gene (Bhullar et al. 2009) to identify stronger alleles. Song et al. (2003) showed that allele mining could be used to clone the functional *RB* allele from a cluster with two highly similar paralogs. Also Wang et al. (2008) and Lokossou et al. (2010) could specifically amplify the target allele rather than paralogous genes in a *Rpi-blb1* allele mining study. Latha et al. (2004) exploited allele mining to identify stress tolerance genes in *Oryza* species and related germplasm. A common feature of the three genes investigated was that they were members of multigene families. Primers based on the 5' and 3' untranslated region of genes were found to be sufficiently conserved over the entire range of germplasm in rice to which the concept of allelism is applicable, while the primers based on the start and stop codon amplified sequences from additional loci (Latha et al. 2004).

is the cloning of the *Rpi-vnt1.1* gene (Pel et al. 2009). NBS profiling revealed a fragment that was co-segregating with resistance in a F1 population. The sequence of this NBS profiling band was similar to a known *R* gene (*Tm-2²*). The mined allele

Genomics of Plant Genetic Resources
Volume 2. Crop productivity, food security and
nutritional quality

Tuberosa, R.; Graner, A.; Frison, E. (Eds.)
2014, XIX, 515 p. 60 illus. in color., Hardcover
ISBN: 978-94-007-7574-9