Isolation and application of natural plant resistance genes

Chapter 1

Frank C. Lanfermeijer[1]
Ijaz Rasul[2]
Jacques Hille[2]

In: Principles and practice of advanced technology in plant virology, **2010**, pp. 269-295. ISBN: 978-81-308-0395-1, Editor: Aiming Wang. Trivandrum, India: Research Signpost Press.

[1]Laboratory of Plant Physiology, Centre for Ecological and Evolutionary Studies (CEES), University of Groningen, P.O. Box 14, 9750 AA, Haren, The Netherlands.

[2]Department Molecular Biology of Plants, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, P.O. Box 14, 9750 AA, Haren, The Netherlands.

ABSTRACT

Like all organisms plants are confronted with pathogens, which aim to infect and profit from them. In order to counteract upon infection plants have developed two major lines of defense. These are passive, ubiquitously present barriers to prevent entrance of pathogens and a reactive second line of defense which depends on detection and recognition of the pathogen and mounts active defensive measures against the pathogen.

The understanding of the latter type has made large progress recently. This is mainly due to the ability of phytopathologists to isolate and characterize resistance associated genes, and especially, the *Resistance* genes that encode detectors of the pathogens or their virulence factors.

In this review we describe three major methods for gene isolation: 1) product directed cloning (PDC), 2) map-based cloning (MBC) and 3) cloning through mutagenesis (CTM) and these methods are illustrated by describing the isolation protocols of some *R* genes from tomato.

Finally, a short state of the art of plant resistance is given.

1. INTRODUCTION

Before *Arabidopsis thaliana* became the major subject of molecular plant biology, research was performed on several other plant species. For instance, tomato, tobacco and maize were three of the frequently used species in plant biology. These plants were economically important crops and consequently they were chosen in many plant biology studies as the experimental subject. For the same reason diseases of those plants were important models in phytopathology. As a consequence of the interest in Solanaceous species, Tobacco Mosaic Virus (TMV) was the first plant virus, and even the first virus, isolated, characterized and crystallized ever (Creager *et al.*, 1999). Furthermore, several of the first disease resistance genes and their proteins characterized, have been isolated from Solanaceous species: the *N* gene from tobacco (Whitham *et al.*, 1994) and *Pto* from tomato (Martin *et al.*, 1993) while the first resistance gene ever isolated is the *Hm1* gene from maize (Johal & Briggs, 1992). However, due to the extended generation time of these species (several months) and the large and complex genomes (*Solanum lycopersicum*: 950 Mb, *Nicotiana tabacum*: 4400 Mb and *Zea mays*: 2500 Mb), those species lost their key role in plant biology to *Arabidopsis thaliana*, (generation time: 8 weeks and genome size: 119.2 Mb) in the age of plant molecular biology. Moreover, *S. lycopersicum*, *N. tabacum* and *Z. mays* lines that are used in agriculture are genetically complex species. *Z. mays* is considered to have arisen as an auto- or allo-tetraploid with teosinte as at least one of the progenitors, while introgression with other species cannot be excluded. The *S. lycopersicum* presently cultivated throughout the world is the result of crossing the ancestral tomato with at least 4 wild *Solanum* species and *N. tabacum* is an allotetraploid with *N. sylvestris* and *N. tomentosiformis* as ancestors. However, during the starting period of plant molecular biology, while *Arabidopsis* had not obtained its prominent role yet, molecular genetic techniques were developed and also implemented on these more genetically complex species. Therefore, some of the first *Resistance (R)* genes have been isolated from tomato, tobacco and maize and all techniques that were developed to isolate genes have been successfully applied in these species. Hence, tomato with its *R* genes is a good illustration to describe those now classical techniques.

Most of the important *R* genes have been isolated in the pre-genomics era and, therefore, several imaginative methods have been used to isolate these genes. This review describes and discusses these methods. Such a description might facilitate the isolation of *R* genes from species of which the genome has not been sequenced yet, albeit future high throughput technologies and bulk sequencing tools will probably also provide new tools. Therefore this review will also discuss and reflect on some of the possibilities and potentials of the rapid progress of genome sequencing.

In this review we discriminate three methods for gene isolation: 1) product directed cloning (PDC), 2) map-based cloning (MBC) and 3) cloning through mutagenesis (CTM). However, the boundaries between those methods are not always clear-cut. The three methods are discussed using examples of resistance genes from tomato, namely PDC with *Tm-1* and the other alleles of *Tm-2*^{2} (*Tm-2, tm-2 and lptm-2*), MBC by *Sw-5* and *Asc* and CTM by *Tm-2*^{2}.

TOMATO

Tomato is an important crop and as such diseases which threaten this species have received considerable attention from the scientific community. Some of the diseases in this crop are: Tomato Mosaic Virus infection, Alternaria Stem Canker, Late Blight, Bacterial Spot, Bacterial Speck Disease, Spotted Wilt, Powdery Mildew, Verticillium Wilt, Fusarium Crown Rot and Root Knot Disease. One of the shortcomings of agricultural crops, and especially tomato, is their limited germplasm. As a result of the continuous breeding for and cultivation of high-yield lines genetic diversity and important genes were lost. Amongst the ones lost were almost certainly also important *R* genes. Presently, the focus of breeding has broadened with interest for quality, nutritional value and resistance and tolerance against hazardous conditions. Moreover, during the last decade sustainability of agriculture has become an important issue and breeding for resistances and tolerance against less optimal conditions is an important instrument to achieve this (Witcombe *et al.*, 2008). This increased robustness is achieved by the introduction of specific traits from wild species. For instance, the cultivated tomato is crossed with wild species to increase salt-tolerance (e.g. *S. pennellii*), low temperature tolerance and

Table 1. Cloned *R* genes in *S. lycopersicum*: their origin, the relevant pathogens and their cloning methods

***R* genes**[a]	**Donor species**	**Pathogen**	**Cloning Method**[b]	**References**
TNL class				
Bs4	ncc[c]	*Xanthomonas campestris*	MBC	(Schornack *et al.*, 2004)
CNL class				
Hero	*S. pimpinellifolium*	*Globodera spec*	MBC	(Ernst *et al.*, 2002)
I-2	*S. pimpinellifolium*	*Fusarium oxysporum*	MBC	(Ori *et al.*, 1997; Simons *et al.*, 1998)
Mi-1.2	*S. peruvianum*	*Meloidogyne spec.* (nematodes) *Bemisia tabaci* (aphid)	MBC	(Milligan *et al.*, 1998; Vos *et al.*, 1998)
Prf	*S. pimpinellifolium*	*Pseudomonas syringae*	CTM	(Salmeron *et al.*, 1996)
Sw-5	*S. peruvianum*	TSWV[d]	MBC	(Brommonschenkel *et al.*, 2000; Spassova *et al.*, 2001)
Tm-2	*S. peruvianum*	ToMV[e]	PDC	(Lanfermeijer *et al.*, 2005b)
Tm-2^2	*S. peruvianum*	ToMV	CTM: TT	(Lanfermeijer *et al.*, 2003)

R genes[a]	Donor species	Pathogen	Cloning Method[b]	References
RLP class				
Cf-2	*S. peruvianum*	*Cladosporium fulvum*	MBC	(Dixon *et al.*, 1996)
Cf-4	*S. habrochaites*	*Cladosporium fulvum*	MBC	(Thomas *et al.*, 1997)
Cf-4A	*S. habrochaites*	*Cladosporium fulvum*	CTM: TT	(Takken *et al.*, 1998)
Cf-5	*S. lycopersicum var. cerasiforme*	*Cladosporium fulvum*	CTM	(Dixon *et al.*, 1998)
Cf-9	*S. pimpinellifolium*	*Cladosporium fulvum*	CTM: TT	(Jones *et al.*, 1994)
Ve1	ncc[c]	*Verticillium albo-atrum*	MBC	(Kawchuk *et al.*, 1998)
Ve2	ncc[c]	*Verticillium albo-atrum*	MBC	(Kawchuk *et al.*, 1998)
Other				
Asc	*S. lycopersicum*	*Alternaria alternata*	MBC	(Brandwagt *et al.*, 2000)
Tm-1	*S. habrochaites*	ToMV	PDC	(Ishibashi *et al.*, 2007)
Pto	S. pimpinellifolium	*Pseudomonas syringae*	MBC	(Martin *et al.*, 1993)

[a]: TNL: TIR-NBS- LRR, CNL: CC-NBS-LRR, RLP: Receptor like Protein; [b]: CTM: cloning through mutagenesis, MBC: map-based cloning, PDC: product directed cloning, TT: transposon tagging; [c]: ncc: not completely clear; [d]: Tomato Spotted Wilt Virus; [e]: Tomato Mosaic Virus.

disease resistance (e.g. *S. peruvianum*, *S. pimpinellifolium* and *S. habrochaites* (formerly known as *Lycopersicon hirsutum*)).

Disease prevention and resistance are important issues. Crops are usually grown as monocultures at high densities and genetic variance between the individuals is limited or even absent. Hence, successful invasion of such cultures by a pathogen has a detrimental effect on the harvest of the crop. Attempts to limit the impact of plant diseases on the yield of the crops can be divided in several types of measures. Sanitary measures and cross-protection are provisional methods to prevent infection by controlling handling of the plants and media, by chemical or other means of sterilization or by inoculating plants with harmless crippled pathogens. The most important, the most environment-friendly and the most effective one is the introduction of resistances into the germplasm of cultivated tomato by crossing the species with wild *Solanum* species and further breeding (Table 1).

Plants, like tomato, have two major lines of defenses against infection by pathogens. First, plants have developed passive and ubiquitously present barriers to prevent entrance of pathogens. These physical barriers consist of cell walls, cuticles and obstacles like hairs and trichomes that repel herbivores and insects and prevent the easy colonization of soft tissue (parenchyma, mesophyll). Because insects can also function as vectors for other viruses, bacteria and fungi, this line of defense has also its role in the resistance against diseases. The second line of defense is inducible, depends on detection and recognition of the pathogen, and involves active defensive measures towards the pathogen. Three types of these inducible defenses are recognized and function against most types of pathogens (bacteria, oomycetes, insects, nematodes, fungi and viruses). The three types of defense are: (I) a basal and general defense response against all microbes (or innate immunity); (II) pathogen- and host-specific resistance; and (III) post-transcriptional gene silencing.

Basal and general defense (I) depend on general microbial elicitors, which are called pathogen-associated molecular patterns (PAMPs). These elicitors usually are derivatives or degradation products of ubiquitous elements of pathogens, like flagella and transcription factors. They are usually extracellulary recognized by special plasma membrane bound receptors of the receptor-like kinase (RLK) or

receptor like proteins (RLP) types, which are named pattern recognition receptors (PRR) (Bent & Mackey, 2007; Nürnberger & Kemmerling, 2006; Schwessinger & Zipfel, 2008; Zipfel, 2008). This type of resistance has been termed non-host-specific resistance. Pathogen- and host-specific resistance (II) depends on specific gene products produced only by a specific pathogen (the avirulence protein or Avr protein; the pathogen specificity) and the recognition of these products by specific host resistance proteins (the R protein; the host specificity). In this type of resistance, the interaction between the pathogens and plants has been described by the 'gene-for-gene' hypothesis (Flor, 1971). This hypothesis is based on the observation that both the plant and the pathogen need to produce a gene product for resistance to take place. This type of resistance is usually associated with a hypersensitive response. The interaction between the Avr and R protein can take place either intra-cellularly or extra-cellularly (Jones & Dangl, 2006). Post-transcriptional gene silencing (III) is also a non-host specific type of resistance, because no specialized and pathogen-specific plant recognition proteins are necessary. Instead, the specific recognition of pathogenic RNAs is accomplished by the use of small 21–23 nucleotide RNAs, which are derived from pathogenic double-stranded RNAs (Mlotshwa *et al.*, 2008; Soosaar *et al.*, 2005; Voinnet, 2008).

2. ISOLATION OF *R* GENES

The understanding of the mechanism of pathogen- and host-specific resistance, (as cell-biological processes in general) profited from the development of the possibilities to isolate genes and, subsequently, the characterization and comprehensive analysis of their protein products. The isolation of genes from plants really took the fast lane when methods were developed that made use of the knowledge of the localization of genes on chromosomes. From the early days of genetics, plant biologists recognized that some genes inherit linked to other genes or loci while others did not. This linkage was expressed as a percentage (the fraction of recombination occurring between two genes or loci) with the smaller the percentage the stronger the linkage. It was soon realized that linkage between two genes was actually a consequence of their nearby location on the same chromosome. The

percentage by which two genes were linked was soon considered to be a reflection of the distance between the two genes on the chromosome. Using these linkage percentages genes could be mapped to chromosomes and positioned in relation to each other on a relative linkage map. These maps showed to be a valuable tool for the isolation of specific genes. Using the maps, regions could be assigned to which a gene was localized. However, until the development of molecular genetic markers these maps were of low resolution and were difficult to use as a tool in the isolation of genes. But since the development of molecular markers and molecular genetic tools the resolution of these maps has been improved drastically and in such a way that they can now be used as starting point for the isolation of genes. These maps are available for most classical model species of plant biology. Especially, linkage maps of tomato are already available for several decades.

2.1 PRODUCT-DIRECTED CLONING

2.1.1. Protein sequence based cloning

Product-directed cloning (PDC) is the oldest method to isolate a gene. Starting from a biological process the responsible protein, usually an enzyme, is isolated to high purity and its amino acid sequence is partly determined using for instance N-terminal sequencing (like Edman sequencing). Based on this amino acid sequence degenerated primers are developed and the protein encoding genes can be isolated using PCR (Fig. 1). This method is still valid, is shown by the recent isolation of the *Tm-1* gene (Ishibashi *et al.*, 2007), although the amino acid sequence data of the Tm-1 protein in this case was obtained with the contemporary method of Mass-spectrometry.

Product-Directed Cloning of *Tm-1*

Recently, the sequencing of all Tm-named genes was completed by the publication of the sequence of *Tm-1* (Ishibashi *et al.*, 2007). The *Tm-1* gene is like *Tm-2* and *Tm-2*^{2}, a gene which confers resistance against Tomato Mosaic Virus (ToMV) infection. However, the *Tm-1* is at another locus. It resides on a different chromosome, it has another Avr protein (the ToMV replicase protein) and it has

other specific functional differences (Hall, 1980; Pelham, 1966). For example *Tm-1* protects tomato protoplasts against infection with ToMV while *Tm-2* and *Tm-2*^{2} do not. The isolation of this gene is a clear example of Product-Directed Cloning (PDC).

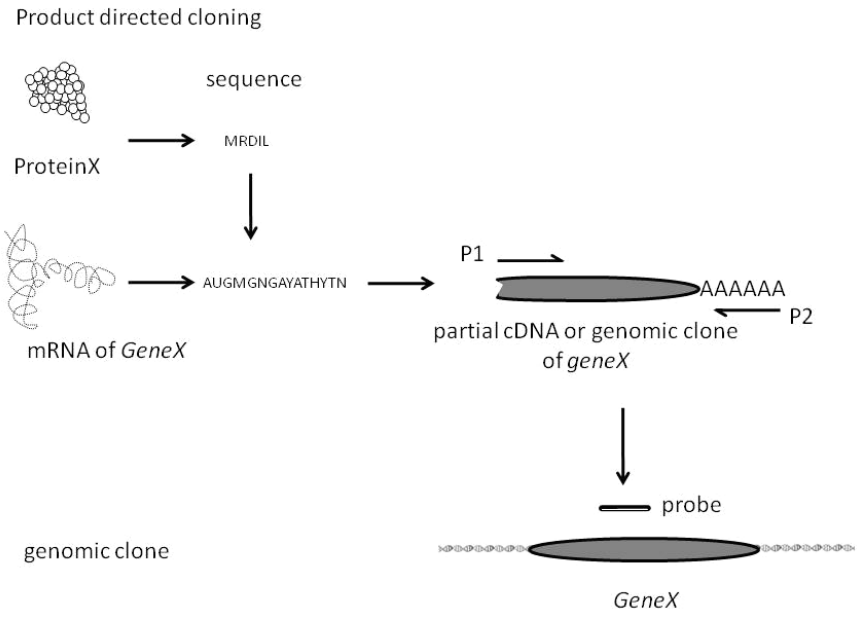


Figure 1. The principle of product directed cloning. Process associated gene products are isolated (protein or mRNA) and their amino acid or nucleotide sequence is determined. The methods to assign a specific mRNA to a process (DNA-micro-arrays, cDNA-AFLP) do not result in the complete sequence of the mRNA. Only a fragment is obtained. Hence, also in this case additional cloning is required. Subsequently, based on the sequence data PCR-primers (P1 and P2) are developed and using a cDNA pool the respective partial cDNA is isolated. Using the cDNA clone the complete gene with all its introns, exons and regulatory sequences can be isolated from a genomic DNA pool by using the cDNA as a probe. Specificity of the protocol can be increased by using additional sets of nested primers which are located inside the annealing sites of the primers P1 and P2 (not shown).

The isolation of the gene started from the observation that extracts of *Tm-1* gene containing tomato cells inhibit the *in vitro* RNA replication of WT ToMV more strongly than that of the *Tm-1*-resistance breaking mutant of ToMV. Subsequently, a protein product was isolated responsible for this inhibition and this protein was sequenced by Mass-spectrometry. The gene encoding this protein was identified and it was shown that this gene co-segregated with the *Tm-1* gene. The transgenic expression of the gene conferred resistance to ToMV in susceptible tomato plants and Virus-Induced Gene Silencing of the gene in *Tm-1* containing tomato lines resulted in ToMV susceptible plants. Hence, the newly identified gene was the *Tm-1* gene. The isolation of this gene shows a new mechanism of resistance by which the replication of WT ToMV RNA is inhibited by binding of the Tm-1 protein to the replication protein of ToMV.

2.1.2. Cloning of homologous alleles or homologous genes

Presently, more methods are available to obtain sequence information on the gene of interest due to the enormous amount of sequence data presently available. If one wants to isolate new alleles from the same species or genus one can use sequence information of already known alleles (Allele Fishing (AF) or Mining). The applicability of this technique depends on the conservation of the sequence of the protein among the varieties or species. If one wants to isolate homologous proteins from more distantly related species one has to search for conserved domains in the protein and generate a consensus sequence of these domains by aligning similar proteins from different species. The consensus sequence then is used to develop degenerated primers for PCR. The sequence differences between the homologous proteins lead to the necessity of degenerated primers which take these differences into account.

In the case of *R* genes, the situation is more complicated by the fact that some of the gene families encoding R proteins are present in a high copy number in a single species. For instance, *Arabidopsis* contains around 200 genes encoding the CC-NBS-LRR (CNL) and TIR-NBS-LRR (TNL) type of resistance proteins (Meyers *et al.*, 2003). In these families some conserved domains can be pinpointed but those

domains cannot be used to isolate specific *R* genes. These conserved domains are believed to function in signal transduction and not in the specificity of the protein towards its target-pathogen. Hence, these domains cannot be used to isolate a single *R* gene targeted against a specific pathogen. However, these conserved domains can be used to isolate a large complement of the *R* genes present in one species and can be used to study the phylogenetic and evolutionary relationship of this group of genes. The function of these proteins is detection of a single pathogen (or in some cases a few) and the specificity of the protein with which this coincides is mainly determined by the highly variable CC-, TIR- and LRR domains (Belkhadir *et al.*, 2004; Hammond-Kosack & Jones, 1997; Martin *et al.*, 2003; McHale *et al.*, 2006). These domains are highly variable as a result of the ongoing struggle of the pathogen to remain undetected and of the host to be able to detect the changing pathogens. Hence, amino acid identity between two identical R proteins of two closely related species recognizing the same pathogen will soon decline. As a result of this only in very closely related plants or varieties homologous *R* genes, or maybe even only alleles with a potentially similar specificity can be isolated.

Allele fishing for other *Tm-2²* alleles

Examples of the isolation of other alleles of the resistance locus *Tm-2²*, by sequence homology and PCR, are the isolation of the non-resistance conferring allele *tm-2* and another resistance conferring allele, *Tm-2*, from *S. lycopersicum*. The isolation of *lptm-2*, a *Tm-2²*-resembling allele from *S. peruvianum*, can be considered isolation of a similar gene from another species (Lanfermeijer *et al.*, 2005). However, one has to keep in mind that the cultivated tomato is a plant species into which large pieces of the genomes of other wild tomato species have been introduced and, thus, the exact species-boundary within this species is difficult to define. For instance, the *Tm-2²* gene originates from *S. peruvianum* (Lanfermeijer *et al.*, 2003). Another recent example of allele fishing is the exploration of *Solanum* species for *Rpi-blb* homologues (Wang *et al.*, 2008).

2.1.3. Analysis of the effect of infection on the transcriptome/proteome

Today several high through put systems exist which allow analysis of changes in almost the full complement of gene products which are present in a tissue or plant. However, these methods are mainly available for proteomically and genomically well characterized plant species. Methods like cDNA-AFLP (Gabriëls *et al.*, 2006) and DNA-arrays (Zhang *et al.*, 2006) allow you to observe changes in mRNA levels, induced by biotic stress, while Mass-spectrometry (Smith *et al.*, 2006) allows you to observe changes in the proteome. Using these methods resistance associated gene products can be identified. The sequence information obtained by these methods can then be used to isolate the encoding genes. Conformation of potential candidates as being valid *R* genes can then be obtained by silencing or transgenic expression. Moreover, expression profiling technologies, in combination with other genomic tools, will have a substantial impact on our understanding of plant-pathogen interactions and defense signaling pathways because not only the receptors but also potential signaling elements can be detected (Wan *et al.*, 2002; Wise *et al.*, 2007).

2.1.4. Analysis of synteny

Synteny is the term used when parts of chromosomes of different species display organizational similarity. In general this phenomenon is considered to be the result of rearrangements of large parts of the chromosomes during speciation with conservation of their local arrangement. Genes positioned on the syntenic regions are homologous and might display similar functions (Bortiri *et al.*, 2006). Presently, genomes are sequenced with increasing regularity and, hence, more and more genomes can be compared and their synteny can be assessed. Using synteny, genes with identical functions could be potentially tracked and characterized. It remains to be mentioned that R proteins and receptors are highly adaptive in relation to their Avr proteins or ligands and a functional homology between two syntenic genes could quickly be lost in time and phylogenetic distance (Huang *et al.*, 2005). This method is therefore more applicable to less adaptive proteins, like for instance house-keeping enzymes.

2.2. MAP BASED CLONING (MBC)

Map based cloning or positional cloning makes use of the phenomenon of linkage to the extreme. The main objective of MBC is to pinpoint the gene to a region on the chromosome so small that it can easily be cloned, sequenced and analyzed. The theoretically obtainable size of the region in which the gene is localized depends on the number of markers available on the chromosome in question (Peters *et al.*, 2003). Several types of markers are available: 1) phenotypical markers: markers which have an observable phenotype in the crossed progeny, 2) genotypical markers: markers which can be observed with standard molecular biological tools: amongst which RFLPs (Restriction fragment length polymorphism), AFLPs (Amplified fragment length polymorphism), SNPs (single nucleotide polymorphism), Indels (inserts en deletions of one or more nucleotides) and many more. While in most plant species knowledge of the existence and localization of SNPs and Indels is rare and usually obtained accidently, in *Arabidopsis* these markers can be used routinely because the existence and localization of all differences, which can be used as a marker, between the genomes of two frequently used ecotypes, Columbia and Landsberg erecta, is known. These two ecotypes are easily crossed and are frequently used in MBC-based gene cloning projects. In the meantime analysis of more ecotypes in relation to molecular markers is under way (Jander *et al.*, 2002).

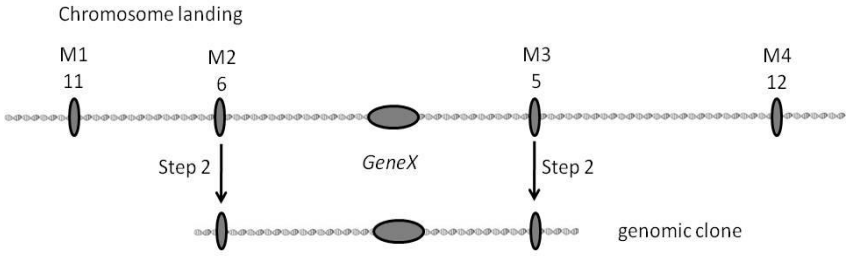
The MBC approach, however, necessitates the alleles of the two parent lines to be different and distinguishable. When there is no natural difference between the two alleles, one can generate a difference by mutagenesis of the gene by EMS, X-rays, slow neutrons or knock-out by transposons. However, when one introduces mutations one needs to confirm that one uses plant lines with single mutations. By determining the linked inheritage between the targeted gene (as tracked by its phenotype) and known markers the targeted gene is positioned relative to those markers.

After placing the locus of interest on the linkage map the next step in the process of isolation is either chromosome walking or chromosome landing (Fig. 2). Which of these two approaches can be used depends on the physical distance between either the left or the right marker with the gene of interest. If one has linked the gene of

interest with one of the markers at a physical distance less than the average size of the clones of the genomic library that one has created of the subject species, one can use chromosome landing (Fig. 2A). The markers are subsequently used to isolate the genomic clone(s) with those markers and the gene of interest. The clone is sequenced and analyzed for the presence of other genes.

Chromosome walking (Fig. 2B) is used when the average distance between the markers and the gene of interest is larger than the average size of the clones of the genomic library. First the genomic clones are isolated that contain the markers that are nearest to the gene of interest. Then the ends of the genomic clones are sequenced and both ends of the clones are used as new markers on the chromosome. The linkage between the gene of interest and the "end"-markers determine which end of the clone is closest to the gene. Subsequently, the clone-end closest to the gene of interest is used to isolate a new clone more closely to the gene of interest. These steps are repeated until the whole region between the markers is covered by genomic clones and a contig can be constructed. The contig is subsequently divided into practical sub-clones and these sub-clones which should contain the gene are used to complement the recessive genotype. The sub-clone which is able to complement the phenotype is subsequently fully sequenced and analyzed for the presence of genes.

A



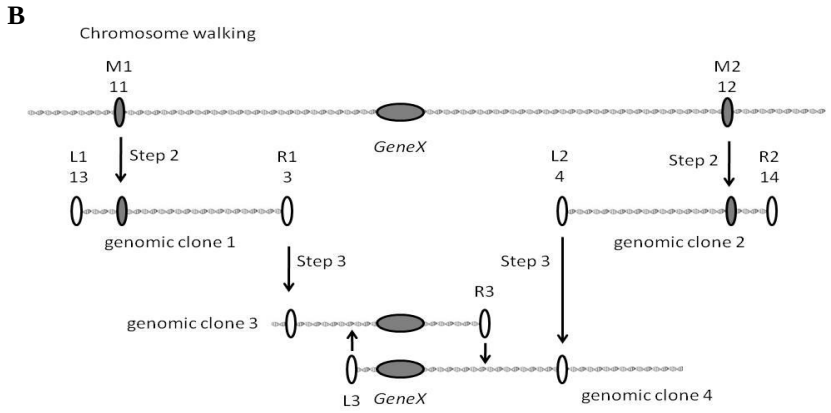


Figure 2. The principle of map based cloning. Linkage between the gene of interest (*GeneX*, large dark grey ellipse) and known markers (dark grey vertical ellipses) is determined in a large population of recombinants (Step 1, not shown). Linkage is usually expressed as the number of recombinants observed (the number below the marker).
A: Chromosome landing. Because of the small distance between *GeneX* and the markers, M2 and M3, a clone from a genomic library can be isolated which contains all three items in step 2. The genomic clone is sequenced and its genomic structure is assessed. Finally, via comparison between the sequence of the different alleles of *GeneX* and complementation of the phenotype *GeneX* is confirmed.
B: Chromosome walking. Because of the large distance between the closest known markers genomic clones 1 and 2 are isolated which contain markers M1 and M2, respectively (step 2). Subsequently, the sequences of both ends of both genomic clones (L1, R1 and L2, R2, respectively; light grey vertical ellipses) are determined and used as new markers. The linkage between those new end-markers and *GeneX* is determined in the original population of recombinants. Based on this linkage, which shows that R1 and L2 are closest to *GeneX*, genomic clones are isolated containing the two end-markers, R1 and L2 (Step 3). The new genomic clones are compared by using internal small sequences and if they overlap their genomic structure is determined. If no overlap can be established Step 3 is repeated. If overlap can be shown the clones are sequenced and its genomic structure is determined. Finally, via comparison between the sequence of the different alleles of *GeneX* and complementation of the phenotype *GeneX* is confirmed.

Because the alleles of the two parent lines differ, sequencing and comparing of the genomic clone of the library with the wild type sequence should result in identification of the gene. Although sequence homology with known genes

sometimes already gives a clue about the nature of the genes, the conclusive argument for linking a gene sequence to a trait is either complementing the mutant phenotype by the introduction of the wild type gene or by transferring the trait to another plant by introducing the gene in this plant. Moreover, especially in the case of *R* genes, multiple homologous genes can be present at the indentified locus, for instance the *Cf-4/Cf-9* and *Sw-5* loci (Brommonschenkel *et al.*, 2000; Spassova *et al.*, 2001; Thomas *et al.*, 1997), which makes it necessary to ascertain the responsible gene by one of the previous methods.

In *Arabidopsis* an additional method can be available. Because for this species plant collections exist in which there is a high chance that all genes are tagged by a T-DNAs (in other words: saturated with T-DNAs), mutant plant lines can be obtained in which the candidate gene has been disrupted. These lines can be used to analyze the phenotype or to complement the mutant phenotype with the wild type gene. In tomato successful applications of MBC are the isolations of the *Pto* gene (Martin *et al.*, 1993), the *Asc* gene (Brandwagt *et al.*, 2000) and the *Sw-5* locus (Brommonschenkel *et al.*, 2000; Spassova *et al.*, 2001).

Map based cloning of *Pto*

The *Pto* gene was introgressed from *S. pimpinellifolium*, into cultivated tomato and gives resistance to Bacterial Speck Disease. However, only when the pathogen, *Pseudomonas syringae* pv. *tomato* carries the *Avr* gene *avrPto* (the gene-for-gene concept). The gene encodes a serine-threonine protein kinase, which upon direct physical interaction with one of the two Avr proteins, AvrPto or AvrPtoB, activates several signal transduction pathways. This subsequently mounts the defense responses. *Pto* co-segregated with an RFLP marker and this marker was used to indentify a yeast artificial chromosome clone which was shown to encompass the *Pto* gene. Finally, a probe from this YAC was used to isolate the gene from a cDNA library. The fact that *Pto* belonged to a gene family which was clustered at the *Pto*-locus complicated the isolation. However, confirmation of the isolation of the *Pto* gene was obtained when susceptible tomato plants became resistant when they were transformed with the cDNA of Pto (Martin *et al.*, 1993; Pedley & Martin, 2003).

Map based cloning of the *Sw-5* gene

Resistance to Tomato Spotted Wilt Virus (TSWV) is conferred by the resistance gene, *Sw5-b*, which was introgressed from *S. peruvianum* in cultivated tomato plants. Analysis of the *Sw5* locus in tomato revealed linkage of the locus with the CT220 RFLP marker. The *Sw5* locus was identified on a genomic DNA contig of 250 kb surrounding the RFLP marker which was created from a bacterial artificial chromosome (BAC) library of a tomato cultivar containing the *Sw5* locus. The resistance locus on the contig contained two genes, named *Sw5-a* and *Sw5-b*, that both encode two CNL proteins. Conclusive identification of *Sw5-b* being the sole resistance gene was achieved by transformation of susceptible *Nicotiana tabacum*, which resulted in resistance against Tomato Spotted Wilt Virus (Brommonschenkel *et al.*, 2000; Spassova *et al.*, 2001).

Map based cloning of *Asc*

Resistance or susceptibility to *Alternaria alternata* f. sp. *lycopersici* toxin in tomato is determined by the *Asc* locus. The *Asc* resistance locus was linked to two chromosome markers which enabled the isolation of three YAC clones around the *Asc* locus. Two markers were produced, residing at one and three recombinations from *Asc*, respectively. Chromosome walking from these two markers in a Lambda phage YAC442A library produced a 70-kb contig of Lambda clones with the *Asc* locus. Finally a genomic region to be used for functional complementation was confined to 17 kb. Subsequently, functional complementation by an *Asc* containing clone in a sensitive *asc*, *asc* background confirmed the isolation of *Asc* (Brandwagt *et al.*, 2000).

Future developments

Spin offs of the present interest in genomics, transcriptomics and proteomics are the big leaps which are made in the development of new tools to study large amounts of genes and proteins. Moreover, the enormous amount of data and tools present in the scientific community (databanks, seedbanks, and mutant banks) will have large implications for the isolation of genes. For example, brute force sequencing, which

allows you to obtain large stretches of DNA sequence in a relatively short time, will reduce the tedious linkage analysis and placing your gene of interest on a small clone to only a small part of the protocol. The allowed distance between markers will increase with the development of sequencing speed and capacity, whereas the reduction of costs will remove the barriers to implement these techniques. Of course the impact of these differences on the phenotypes will have to be verified by either complementation or even "site-directed mutagenesis". Using brute force sequencing a gene was isolated involved in senescence and aging (Sturre *et al.*, 2008). After mapping the gene to a region of 150 kbp this region was completely sequenced in a 100-fold redundancy. By combining both mutant and wildtype DNA, a gene was found in which half the sequences contained an adenine and the other half a guanine. The amino acid change was confirmed in the mutant by conventional tools and, finally the involvement of the gene in the senescence phenotype was confirmed (Shirzadian-Khorramabad, 2008).

2.3. CLONING THROUGH MUTAGENESIS (CTM)

Cloning through mutagenesis depends on the introduction of a mutation which allows you to track and pinpoint the gene of interest. Several mechanisms for mutagenesis are available chemical methods like EMS mutagenesis, physical methods like fast-neutron and X-ray radiation, and methods making use of transposons and T-DNAs. All these methods introduce their characteristic type of mutations, like base-changes, small or large deletions, insertions and major DNA rearrangements. Often these mutations are accompanied by phenotypical changes which allow detecting and tracking the mutated gene and potentially allow isolation of the gene by for instance MBC. However, the phenomenon of redundancy of gene-function could obscure a phenotypical effect of the mutation. Several methods are available to isolate the mutated genes. One of those is already mentioned above. If no phenotypically different alleles are naturally present in a population, mutagenesis could induce this requisite and make MBC possible.

2.3.1. Isolating genes by either chemical or physical mutagenesis

A large number of genes have been isolated by chemical or physical mutagenesis. However, the disadvantage of chemical or physical mode of mutation is the nature of their effects on the DNA which impede easy isolation of the region in which the mutation resides. Their effects are either minor (EMS, X-ray) or lead to rearrangements and deletions of various size (fast neutron) and these changes can not be detected easily by probes or other tools, like is the case with transposon or T-DNA tagging. Usually, after the phenotype of the mutated plant has been characterized the mutagenized gene is isolated using MBC. However, with the development of DNA-arrays an additional method has become available to find the gene that has been mutagenized. If the mutations are of such a nature that stability or synthesis of the mRNA is affected and consequently drastic changes can be observed on the DNA-array the gene can be isolated using the affected mRNA. However, the latter method is until now mainly available for proteomically and genomically well characterized plant species, like Arabidopsis.

The cloning of *Prf*

In tomato resistance against Bacterial Speck Disease depends not only on the presence of *Pto* but also on the *Prf* gene. The latter gene could be isolated because of the tight linkage between this gene and the *Pto* gene. In this research tomato lines were mutagenized which resulted in two complementation groups, which were susceptible for infection. One group appeared to contain mutations in the *Pto* gene, while the other group was at a separate locus that co-segregated with *Pto*. Using probes, generated from cosmids that were available for the region in which *Pto* and thus *Prf* reside, a probe was identified which showed a major sequence rearrangement in a mutant line. The cosmid from which this probe was derived was able to complement the *prf* mutation. The same probe was used to screen a cDNA library and the *Prf* cDNA was isolated. Subsequently, the gene was isolated and characterized (Salmeron *et al.*, 1996).

2.3.2. Silencing of homologous genes

By using different methods of gene-silencing (virus-induced gene silencing (VIGS) or hairpin constructs) it is possible to determine if a related plant species contains a functional homologue of a characterized *R* gene from another species (Balaji *et al.*, 2007). Subsequently, the cDNA of the gene can be isolated using sequence homology. Additionally, by checking the broader effects of the silencing additional *R* genes with other specificities can be isolated (Balaji *et al.*, 2007; Benedito *et al.*, 2004; Peart *et al.*, 2005).

2.3.3. T-DNA-tagging

With the rise of genomics progressively more efforts are initiated to saturate the sequenced genomes with T-DNAs. The objective of these efforts is to obtain a population of individuals in which every gene is at least disrupted once by a T-DNA. This effort is close to finalization in *Arabidopsis*, ongoing for rice and other plant species will soon follow. It can be imagined that a pool of insertional mutant plants is screened for a phenotype and that subsequently the responsible gene is isolated (An *et al.*, 2005; Topping *et al.*, 1995).

2.3.4. Transposon tagging

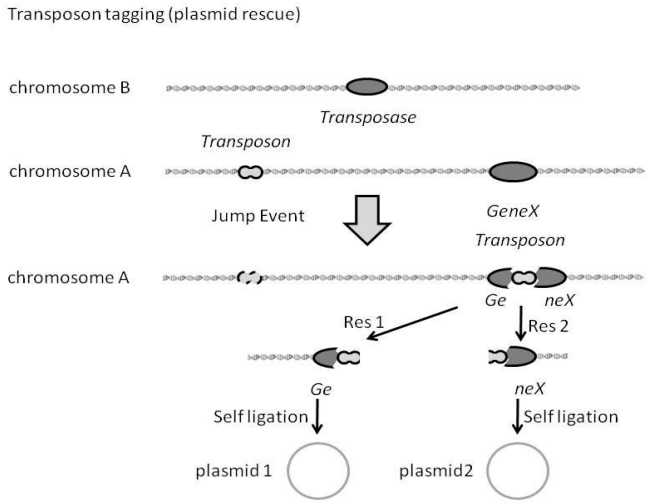
In the transposon-tagging (TT) approach one introduces an active transposon into the genome as a mutator and a tool to isolate the gene (Fig. 3). Many of these transposons originate from maize as a result of the pioneering and ground-breaking studies of Barbara McClintock (Jones, 2005). Presently, the *Ac/Ds* and *MuDR/Mu* maize transposons are the transposon systems of choice for use in transposon tagging experiments (Osborne & Baker, 1995; Walbot, 2000) not only in maize but other plant species as well because they can be functionally transferred to other plant species. Knowledge of the chromosome on which the target gene resides is beneficial. Transposons move with a higher frequency within a chromosome than between chromosomes. Hence, in order to limit the amount of work starting with a plant line in which a transposon is on the same chromosome as the target gene is beneficial. For tomato and most other model species these lines are available. Once

the transposon has inserted in the gene of interest the gene will be split and its halves will be on both sides of the transposon. The gene is disrupted and its gene product is highly likely to be non-functional. The mutant plants with the resulting changed phenotype can now be selected. When this technique was still in its infancy, a clone of a genomic library, made from the mutant plant, was isolated using a transposon based probe. This clone contains the transposon and, also, some sequence of the disrupted gene. Subsequently, a clone is isolated from a genomic (or cDNA) library that was made from the wild type plant using the clone of the mutant plant as a probe. Because the clone of the mutant plant contains, next to transposon sequence, also some DNA of the gene of interest a wild type clone with the intact gene will be found (Fig. 3A). The latter step is necessary because due to the size of the transposons or T-DNAs used, the clones of the mutant plants do not contain all sequence information of the gene of interest. The *Hm1 R* gene of maize was isolated in this way. However, in order to be cloned the insertion needs to result in a recognizable phenotype.

Presently, transposons are engineered to facilitate transposon tagging. These engineered transposons contain elements which allow selection of the plant-lines containing the transposon (antibiotic resistances), restriction enzyme sites, which allow isolation of the transposon with bordering plant sequence and, elements, which allow conversion of isolated linear transposon with the bordering plant DNA into a viable circular plasmid (origin of replication, antibiotic resistances). Usually, a few restriction enzymes have a single recognition site present in the transposon. These sites are located in such a way that they allow isolation of left or right bordering plant DNA without disturbing the essential elements, needed to propagate the rescued DNA as plasmid in *Escherichia coli*. Subsequently, those plasmids can be propagated in *E.coli* (plasmid rescue). Moreover, the *transposase* gene or *activator* gene (*Ac*), which encodes the enzyme responsible for the extraction of the transposon from and reinsertion into the DNA and which is originally present in the transposon itself, has been removed from the transposon. However, in order for the transposon to fulfill its task this *transposase* gene needs to be present at the beginning of the transposon-tagging process. Hence, transgenic lines are available in

which this gene is present integrated in the genome but isolated from the transposon. For convenience usually lines are chosen in which the activator gene is present on a chromosome different from the chromosome with the target gene. This allows separation of the activator gene and the transposon in a later generation what results in a stable insertion of the transposon in the target gene (Fig. 3B).

A



B

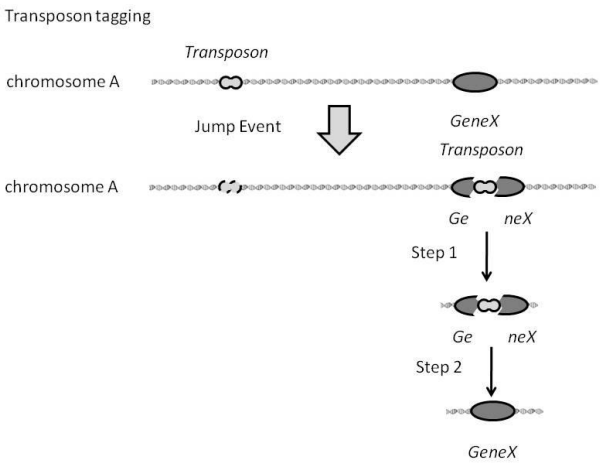


Figure 3. The principle of transposon tagging. Transposon tagging depends on the fact that disruption of the gene by a transposon results in a detectable phenotype in the plant. This approach starts with the introduction of a suitable transposon (eight-shaped symbol) in the plant and preferably in the chromosome on which the gene of interest (*GeneX*, large grey ellipse) resides. Subsequently, a descendant of this plant which shows the expected mutant phenotype is selected.
When the transposon leaves its original position a footprint is left behind (the dotted eight-shaped figure).
A: **The original protocol**. A genomic library is created from this plant and a clone isolated which contains DNA of the transposon and, hence, DNA of the plant that includes the *GeneX* (Step 1, not shown). With the use of this clone as a probe, a clone is isolated from a wildtype genomic library that contains the wildtype gene. The genomic clone is sequenced and its genomic structure is assessed. Finally, via comparison between the sequence of the different alleles of *GeneX* and complementation of the phenotype *GeneX* is confirmed (Step 2).
B: **Transposon tagging with plasmid rescue**. This protocol has some improvements. Firstly, the transposase does not reside in the transposon any more. This improves the stability of the transposon in the gene of interest because after generating progeny the transposon is separated from the transposase. Consequently, the transposon cannot jump out of the gene again. Secondly, the transposons are engineered in such a way that they contain elements which allow maintenance of plasmids that are derived from them in *E. coli*. Initially the protocol follows the same steps as the original protocol. The transposon is introduced and the descendant which shows the expected mutant phenotype is selected. The next steps differ from the original protocol. DNA is isolated from the mutant plant and this DNA is digested separately with two restriction enzymes (Res1 and Res2) which allow isolation of the transposon with the essential elements for plasmid rescue and plant DNA left and right of the transposon. The restriction products are self-ligated and transformed into *E. coli*. Only the circular DNA with the transposon and, thus, the essential elements for maintenance in *E. coli* are able to sustain growth of *E. coli* on selective plates. These plasmids contain sequence information of *GeneX* and are sequenced. The gene is reconstructed from the data obtained and finally its function is confirmed by complementation.

The next requirement of transposon tagging is a system which allows for easy selection of plant lines in which the function of the gene of interest has been disturbed or mutated. This requirement is exemplified by the successful first use of a transposon as a tool for cloning genes in the case of the bronze gene (Fedoroff *et al.*, 1984). In this case individual plants containing the disrupted gene were selected on the changed color of the maize kernel from which they germinated. Another example is the isolation of the genes responsible for the albino mutations in Arabidopsis which render the plant white and allow for selection of the plant lines

with the disrupted genes (Long *et al.*, 1993). Imagination and creativity is needed to engineer such systems of easy selection or genetic screens (Candela & Hake, 2008). As such resistance genes are good candidates because after inoculation of the transposon-tagged population with the pathogen, susceptible plant-lines can be separated from the resistant lines. The susceptible lines are the ones one needs because in these lines the resistance gene has been destroyed by the transposon and is therefore tagged and ready for isolation. Examples of the application of this strategy are the isolation of the *N*-resistance gene of tobacco (Whitham *et al.*, 1994) and the isolation of *Cf-9* of tomato (Jones *et al.*, 1994). Especially the cloning of the *N gene* is a nice example of making creative use of characteristics of the targeted gene. The *N*-resistance against Tobacco Mosaic Virus (TMV) is temperature sensitive: at temperatures above the permissive temperature (25°C) the resistance is not functional. The *N gene* was isolated by transposon tagging, infecting the plants with TMV and growing them at temperatures above the permissive temperature for a short period. After transferring the plants, which were now systemically infected, to a lower temperature massive cell death appeared but not in the plants in which the *N* gene was disrupted and, thus, tagged. The characteristic that plants in which both the *Avr* gene and the intact *R* gene show various levels of necrosis has also been employed for the isolation of the ToMV resistance gene, *Tm-2*^{2} (Lanfermeijer *et al.*, 2003). In this case a tomato line with a non-functional *Tm-2*^{2} allele (*tm*-2) and containing both the Avr protein of ToMV (the movement protein) and the *transposase* gene was crossed with a transposon-containing tomato line with *Tm-2*^{2}. The seedlings, germinated from the resulting seeds and that were able to grow contained a disrupted *Tm-2*^{2} gene. The gene could subsequently be isolated and characterized. This protocol has the advantage that due to the sole presence of the *Avr* gene, and not the whole pathogen, surviving plants are in good physical shape and therefore can easily be grown and used.

3. CHARACTERIZATION OF THE *TM-2²* GENE BY TRANSPOSON TAGGING

In order to clone the *Tm-2²* gene we employed the two-component *Ac/Ds* transposon system (Lanfermeijer *et al.*, 2003) and developed a selection method based on an observation of Weber and Pfitzner (Weber & Pfitzner, 1998). These authors observed that the cross between tomato plants containing the *Tm-2²* gene and transgenic tomato plants expressing the ToMV-MP gene results in a progeny which dies at the seedling stage. The rationale behind our selection method was that when the *Tm-2²* gene is disrupted by a transposon insertion event, seedlings containing the otherwise lethal combination of the *Tm-2²* gene and the ToMV-MP transgene will survive. In these survivors the resistance gene will be tagged by the transposon and isolation of the disrupted *R* gene will become possible.

For this protocol we could make use of a collection of 140 individual tomato lines containing a characterized copy of pJasm13, a multifunctional T-DNA/modified *Ds* transposon element construct (Gidoni *et al.*, 2003). In these lines the position of 140 individual inserts of the *Ds*$_{13\text{-}15}$ transposon was determined on the molecular genetic map of tomato. These inserts are distributed over the 12 chromosomes of tomato, allowing targeted and non-targeted transposon tagging of genes on each chromosome. From this collection we selected a tomato line in which the *Ds*$_{13\text{-}15}$ transposon was localized on chromosome 9, the chromosome on which the *Tm-2²* gene resides, and in which the distance between the *Tm-2²* gene and the transposon was the smallest. A short distance between the transposon and the targeted gene increases the chance that the targeted gene is hit by the transposon. The distance between this transposon and the resistance gene was approximated to be 2 cM by back-crossing tomato with the Ds$_{13\text{-}15}$ transposon twice with a tomato genotype which was homozygous for *tm-2*. As discussed above an activated transposon needs to be present for tagging the gene of interest. This was achieved by introducing a stabilized *transposase* gene (*sAc*) in the tomato with the *Ds*$_{13\text{-}15}$ transposon by crossing these plants with a *transposase*-containing line. Selfings from the progeny of this cross were selected for homozygosity of both *transposase* and *Tm-2²* via PCR. One of these plants was subsequently used in a cross with genotype *Ds*$_{13\text{-}15}$.

Progeny of this cross was selected via PCR for the presence of $Ds_{13\text{-}15}$. Finally, about 100 independent plants with the genotype *Ds*, –; s*Ac*, –; *Tm*-2^2 , *Tm*-2^2 were selected and used as males and females in crosses with a transgenic tomato line, which was homozygous for the ToMV *MP* gene for a large-scale tagging experiment.

From about 30000 seeds obtained from these crosses, 5 seeds germinated. Using PCR it was shown that these 5 surviving mutant plants contained the *Ds* element and the *MP* gene and only one also still contained sAc. To test whether these 5 putative mutants were really mutants in the *Tm*-2^2 gene, cuttings of these plants were inoculated with ToMV. The cuttings of all 5 putative mutants could be infected with ToMV and, subsequently further analyzed. In the surviving mutant plants, excision and reinsertion of the $DS_{13\text{-}15}$ transposon could be demonstrated by DNA blot analysis. In this analysis four of the mutant plants displayed a similar pattern and this pattern was indicative for having a transposon at its original position and another one at a new position. The remaining mutant plant had only one copy of the transposon at a position different from both the original position and the new position of the other group mutants.

Because the *Ds* element harbors an origin of replication and a chloramphenicol resistance gene, cloning of plant DNA flanking the left and the right side of the insertion was possible via plasmid rescue. In our case genomic DNA was cut with the use of two restriction enzymes, BamHI and SacI. These two restriction enzymes cut only once in the *Ds*-element, cut with at a low frequency in the tomato genome and leave the origin of replication (*ori*) and a chloramphenicol resistance gene intact. The low frequency of cutting the tomato DNA of the two restriction enzymes is advantageous for the isolation of significant stretches of plant DNA. The resulting genomic DNA fragments were circularized using a ligase and transformed into Ultra-competent *Escherichia coli* cells. Ultra-competent cells were used because the procedure results in a huge number of circularized DNA fragments of which only a few are capable of being propagated in *E. coli* due to the presence of the transposon (plasmid recue).

Four plasmid types were obtained by this plasmid rescue procedure; a BamHI-and a SacI-derived one of each mutant group. All four plasmids were sequenced and it

became evident that three of the four plasmids were mutually related. An 8 base pair repeat typical of the insertion of a *Ds*-type transposon was present in both plasmids derived from the first group of mutant plants. Insertion of a transposon-like element, like $Ds_{13\text{-}15}$, is accompanied by duplication of 8 basepairs at the insertion site. For insertion of the transposon the DNA strand has to be opened and this opening has the nature of a 8 basepair overhang for $Ds_{13\text{-}15.}$ This 8 basepair repeat is located on both sides of the transposon. Insertion of other transposons results in other characteristically sized repeats. When the transposon leaves the site of integration again, on both sides the 8 base pair duplication remains. So, a short piece of DNA is duplicated which is then referred to as a footprint.

Using the 8 basepair repeat and the sequences of the pB1, pS1 and pS2 plasmids a strand of 9.8 kb of continuous plant-derived DNA could be created. The 9.8 kb of continuous plant DNA contained an open reading frame (ORF) of 2586 basepair which corresponded with a polypeptide of 861 amino acids. This polypeptide showed the characteristics of the CNL class of R proteins (see below). This, in combination with the facts that this ORF was disrupted in both groups of mutant plants and that in both groups of mutant plants resistance was lost, strongly suggested that the *Tm-2*^{2} gene had been identified. The final confirmation that the *Tm-2*^{2} gene had been identified was obtained by the transformation of a tomato line, which was susceptible to ToMV infection, with the new gene under the control of its own promoter or with the new gene under the control of the cauliflower mosaic virus *35S* promoter. These transformations resulted in several transformants which were resistant to ToMV infection.

Analysis and sequencing of the fourth plasmid revealed that the plant DNA rescued in this plasmid had no obvious relation with the other three rescued plasmids. Also, no relation with other sequences in the databases was observed. However, because only one *Ds*-type transposon was present in the genome of the group 2 mutant plants, as determined by Southern blot analysis, it was assumed that the plant DNA in this plasmid represented one of the flanking regions of the inserted transposon in the group 2 mutant plants. However, in this mutant plant a deletion or re-arrangement had occurred that resulted in the loss of 75% of the 5'-end of the *Tm-2*^{2}

gene and an unknown amount of additional plant DNA. Such kind of rearrangements and deletions often take place during the jumping of the transposon through the host DNA and can result in complications of the transposon tagging procedure (Krishnaswamy *et al.*, 2008).

4. THE STATE OF THE ART

4.1. Disease resistance in plants

As mentioned above, pathogen- and host-specific resistance depends on two gene products to be present in the pathogen and the host. These gene products are the pathogen's Avr protein and the plant's R protein. If one of those proteins is not produced or is changed, interaction between the two does not take place, thus defense mechanisms are not mounted and the plant becomes infected. Avr proteins are very diverse in nature and function. The isolation of the *R* genes showed that most R proteins of plants are members of five families, all with the same function: the detection of the pathogen presence and, subsequently, the activation of the defense mechanism. To fulfill this role, all R proteins are made up from modules which have distinct roles in this process (Abramovitch *et al.*, 2003; Belkhadir *et al.*, 2004; Hammond-Kosack & Jones, 1997; McHale *et al.*, 2006). R proteins contain domains which function in the recognition of the pathogen and domains which function in the activation of the downstream components of the signal transduction network and are presently described as molecular switches (van Ooijen *et al.*, 2007, 2008; Rairdan & Moffett, 2006; Rairdan *et al.*, 2008; Takken *et al.*, 2006). Originally, the Avr protein and the R protein were thought to interact physically with each other. Due to the limited number of demonstrated direct interactions between host proteins and avirulence proteins, the idea has arisen that R proteins guard host proteins from hijacking and exploitation by pathogens (Van der Biezen & Jones, 1998; Ellis *et al.*, 2000; Glazebrook, 2001; Van der Hoorn *et al.*, 2002). In the 'guard' hypothesis, the effects of avirulence proteins on host proteins (the guardees) are monitored by the R proteins, which then trigger the defense reaction. Hence, direct interaction between Avr protein and R protein is not always necessary as long as the effect of the Avr protein on the targeted process or protein is detected.

Recently, it was suggested that plant R proteins can function either by directly detecting the corresponding Avr protein, the 'receptor-ligand' model, or by perceiving alterations in plant machines that are targets of Avr protein action in the promotion of pathogen virulence, the 'guard hypothesis' (Dodds *et al.*, 2006). The guard hypothesis might also explain the presence of alleles which do not confer resistance. These alleles may be either non-functional alleles or resistance proteins directed against a distinct pathogen, which is not present in the current plant culture. However, the non-random distribution of mutations in these non- resistance conferring alleles suggests that they are exposed to selective pressure. One possible explanation is that the non-resistance conferring alleles guard the same process or protein as the functional resistance alleles. However, the non-resistance conferring alleles would not detect changes induced by the pathogen, but would rather detect similar changes induced by other phenomena. In this sense, R proteins and their homologues could have a broader function in the biology of the cell (Lanfermeijer *et al.*, 2003): not only discriminating between 'self and non-self' but also between 'self and wrong-self'. The current view is that either upon direct binding of the R protein with the Avr protein or upon binding of the R protein with the host protein, which is affected by the Avr protein, the signal transduction cascade is activated.

It has become apparent that the avirulence proteins of the pathogens are necessary for virulence of the pathogen. During the last years it has even become apparent that many of the avirulence proteins are directed against elements of the innate immunity (Bent & Mackey, 2007; Jones & Dangl, 2006). Pathogens are attempting to sabotage this defense of plants by interfering, inactivating or disturbing the signal transduction. Subsequently, plants developed guards (the R proteins) which monitor the presence of these specific saboteurs. However, some of the R proteins are also directed against essential elements of the pathogen, e.g. the replicase of ToMV (The *Tm-1*-resistance gene product) (Meshi *et al.*, 1988) or the movement protein of ToMV (The *Tm-2*-resistance gene product) (Calder & Palukaitis, 1992; Weber & Pfitzner, 1998; Weber *et al.*, 1993).

Considerable progress has been made to unravel the mechanism of signal transduction. Several observations suggest that R proteins function in complexes

with homologous and heterologous proteins. First, immuno-precipitation experiments showed oligomerization of the N protein in the presence of its Avr protein (Mestre & Baulcombe, 2006). Second, various potential interactors have been identified by yeast two-hybrid experiments (De la Fuente van Bentem *et al.*, 2005). Studies with two R proteins, I-2 conferring resistance to *Fusarium oxysporum* and Mi-1 conferring resistance to root knot nematodes and potato aphids, provided the first evidence that ATP hydrolysis is involved in signal transduction via CNL proteins (Tameling *et al.*, 2002, 2006). Studies with various mutants of the protein of the *Rx* resistance of potato against Potato virus X demonstrated an important role for intramolecular interactions between the three domains of the R protein and the disruption of these interactions in activation of the resistance (Moffett *et al.*, 2002; Rairdan & Moffett, 2006; Rathjen & Moffett, 2003). Together, the studies on the *I-2*, *Mi-1* and *Rx* resistances resulted in a model for intramolecular rearrangement of subdomains as the result of R protein/ elicitor interaction (van Ooijen *et al.*, 2007, 2008; Rairdan & Moffett, 2006; Rairdan *et al.*, 2008; Takken *et al.*, 2006). Recently, it has been shown that the N protein is also able to hydrolyze ATP (Ueda *et al.*, 2006). It has been proposed that the R protein cycles through different conformations, driven by transient interaction with the elicitor and by ATP hydrolysis. The different conformations have different characteristics of interaction with the elicitor and activation of the resistance mechanisms. Subsequently, this iterative mechanism was suggested to result in signal amplification (Rairdan & Moffett, 2006). This mechanism could also explain the absence of experimental evidence for direct interactions between the R proteins and their elicitors because of the transient interaction between the two proteins during the activating cycle.

There are currently five well-defined classes of R proteins (Belkhadir *et al.*, 2004; Hammond-Kosack & Jones, 1997; Martin *et al.*, 2003; McHale *et al.*, 2006) in plants. However, new forms of resistance proteins are discovered on regular bases that do not fit in the five established classes of R proteins. Of the classic ones, the first class is represented by Pto from tomato, which confers resistance to *Pseudomonas syringae* pv. tomato (Pedley & Martin, 2003). Pto is unique in its class, a kinase, and no virus resistance genes that encode this type of R proteins have

been isolated until now. The next two classes are plasma membrane-localized proteins, which belong to the related protein families of receptor-like kinase (e.g. Xa21 protein, which is involved in rice bacterial blight disease resistance; (Song *et al.*, 1995) or receptor-like proteins (e.g. Cf proteins, which are involved in *Cladosporium fulvum* resistance in tomato; (Kruijt *et al.*, 2005), respectively. Due to their structure, these types of proteins are highly suitable for the extracellular perception of avirulence signals and the subsequent transduction of the signal intracellularly. The completion of the genome sequence of several plant species shows that plants contain a large number of genes encoding these types of proteins (Morillo & Tax, 2006). The last two large classes of resistance genes code for intracellular proteins, which have a C-terminal region of leucine-rich repeats (LRRs) and a central nucleotide binding site (NBS). At the N-terminus of these proteins are, either a coiled coil (CC) sequence or a Toll and Interleukin 1 Receptor (TIR) sequence, distinguishing the CNL or TNL classes of resistance proteins. Several *R* genes have been characterized that do not fit the above presented classification. For example, the *RTM* genes allow genome replication and cell-to-cell movement but do not allow long-distance movement of tobacco etch virus, and as such protect the plant against systemic infection. The resistance mechanism does not involve hypersensitive cell death or systemic acquired resistance (Chisholm *et al.*, 2000; Whitham *et al.*, 2000). Also, the well-known disease resistance locus *Tm-1* does not belong to the five well-defined groups of R proteins. The protein, which is encoded by this gene, interferes with the replication of ToMV RNA through binding to the replication proteins (Ishibashi *et al.*, 2007).

These pathogen- and host-specific *R* genes are good candidates of genes being lost during the process of breeding and cultivation of *S. lycopersicum*. The loss of these *R* genes has been put to a stop by the introduction of new resistances through introgression of tomato with wild *Solanum* species. Table 1 shows examples of disease resistances introduced into *S. lycopersicum* and the wild species from which they originate. The introduction of resistances through breeding into crops requires easily recognizable characteristics and markers which can be used in the breeding process. While in the pre-molecular genetic era only visible markers and traits were

used, today tools which guide the breeding process have been expanded with molecular markers

4.2. The future of the study of plant resistance

Plants and their pathogens are in a continuous arms-race and consequently breeders and growers, are in a continuous need for updates of existing resistances or new resistances. Although exploration of the wild resources could make new ones available, the pool of useful natural resistances is limited. Presently, introduction of a new resistance into a crop depends on the ability of the donor species to be crossed with the acceptor species. A second disadvantage of the classical methods is its time and labor-consuming nature. The isolation of *R* genes through the methods described in this review can accelerate the process of introduction of the highly needed resistances in plants. Firstly, isolation of *R* genes would allow the introduction of these *R* genes without the need of the repetitive and time-consuming back-crossing in order to remove the unwanted genes of the donor species from the acceptor species as is the case in classical breeding. Secondly, the isolation of resistance genes would allow the exchange of resistances between species which would normally not be exchangeable through classical breeding, as examplified by the *N* gene in tomato (Whitham *et al.*, 1996) or the *Tm-2*^{2} gene in tobacco (Lanfermeijer *et al.*, 2004). However, the barriers to exchange genes between species are not solely determined by their potential to cross and produce a viable progeny. Also, species could phylogenetically be so distant that the cell biological infrastructure, which is necessary for the R proteins to function, differs too much. Thirdly, detailed knowledge of resistance, of its mechanisms and of the structure-function relationships of R proteins could give supplementary tools for the highly needed new resistances. Either by enabling modification of resistances in such a way that resistance genes can be exchanged between genera or even families or by designer resistances with engineered specificities or improvement of existing resistances.

We expect that, just as recombination has played a major role in the evolution of genes, DNA and gene shuffling (Moose & Mumm, 2008) will play a central role in the development of applied molecular evolution technologies. This will help to

understand the relation between structure and specificity of R proteins (Van der Hoorn *et al.*, 2001; Wulff *et al.*, 2001) and will provide tools to engineer designer resistances with planned specificities or to improve existing resistances.

Recently, a study described the artificial evolution of intracellular Rx-resistance protein resulting in an expanded recognition spectrum (Farnham & Baulcombe, 2006). An *in vitro* random mutagenesis approach provided resistance to a distinct strain of *Potato virus X* and to the distantly related *Poplar mosaic virus*. This pioneering study shows how isolation and manipulation of known resistances could provide a new tool in the search for resistances against new pathogens or emerging virulent isolates of controlled pathogens.