Molecular evolution of the disease resistance gene *Rx* in *Solanum*

Patrick
Butterbach

Promotor:

prof. dr. ir. Jaap Bakker
Hoogleraar in de Nematologie
In het bijzonder de fysiologie en moleculaire ecologie van nematoden
Wageningen Universiteit

Co-promotoren:

dr. ir. Aska Goverse
Universitair docent bij het Laboratorium voor Nematologie, Wageningen Universiteit
dr. ir. Erin Bakker
Postdoctorale onderzoeker bij het Laboratorium voor Nematologie,
Wageningen Universiteit

Promotiecommissie:

prof. dr. W. J. Stiekema, Wageningen Universiteit
prof. dr. V. M. Williamson, University of California, Davis, Verenigde Staten
dr. C. Gebhardt, Max-Plank-Institut für Züchtungsforschung, Keulen, Duitsland
dr. R. G. van den Berg, Wageningen Universiteit

Dit onderzoek is uitgevoerd binnen de onderzoekschool ‘Production Ecology and Resource Conservation’

Molecular evolution of the disease resistance gene *Rx* in *Solanum*

Patrick
Butterbach

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit
prof. Dr. M. J. Kropff
in het openbaar te verdedigen
op vrijdag 21 september 2007
des middags om 13.30u in de Aula

Butterbach, P. (2007)

Molecular evolution of the disease resistance gene *Rx* in *Solanum*

PhD thesis Wageningen University, Wageningen, The Netherlands, with references, with summaries in English, Dutch and German

ISBN: 90-8504-724-7

Contents

Chapter 1

General Introduction

Potato

With its richness in minerals, vitamins and some essential amino acids the starchy potato (*Solanum tuberosum* L.) exhibits premium dietary value at a high agricultural yield (www.cipotato.org). Today it is the fourth most important food crop with an annual yield of about 300 million tons over the world. While initially treated with suspicion - not at least because of its dangerous toxic leaves, unusual fruit bodies and for European standards unusual subterranean crop type - the world wide distribution of potato as crop plant nowadays is the result of a long process of successive reputation increases since Spanish conquistadors brought the first potatoes to Europe in the 16th century from the Andean region of South America. There, the indigenous inhabitants cultivated a wide variety of different landraces indicating a longer history of domestication that started at least 7,000 years ago (Hawkes, 1990). The first evidence for the use of potato tubers by humans comes from fossilized remains of potato tubers found on the floor of a cave at the Chilca Canyon in Peru (Ugent, 1982). Tubers found in the oldest layers date back to 13,000 years ago. Analysis of their starch granules shows that they resemble already more those of modern cultivated potatoes than known wild *solanaceous* relatives.

Disease resistance in potato

With the import of potatoes, pathogens causing various diseases were introduced in Europe. A very well documented historical example is the epidemics of potato late blight that caused ´the Great Famine´ in Ireland in the 1840s (Salaman, 1985). Infestation with the causal agent, the oomycete *Phytophthora infestans*, resulted in the loss of almost the complete potato harvests in several forthcoming years. Subsequently, food shortage led to starvation and one million deaths throughout the country and another 1.5 million people have emigrated mainly to the United States. To address this problem, breeding programs on improving resistance to late blight in potato started shortly afterwards in England and Germany by introgression of traits from potato germplasm imported from South America (Ross, 1986).

Several other microbial diseases are known to cause serious crop losses in potato. The fungus *Rhizoctonia solani* causing the black scurf becomes an increasing threat especially in irrigated potato plantations. A bacterial pathogen infecting potato is *Erwinia carotovora* causing the soft rot. Potato cyst nematodes like *Globodera rostochiensis* and *G.*

pallida are estimated to cause 10% of the annual yield losses worldwide. Due to the formation of cysts, they are able to survive in the soil for decades. Finally, several viruses are known to cause damage in potato plants. The most abundant are the potato leafroll virus (PLRV) and the potato viruses Y, M, S, A and X. Potato virus X (PVX) is a member of the potexvirus group and has a worldwide distribution nowadays. It is known to have a wide host range infecting species from eleven plant families (Fribourg, 1980).

In potato, disease resistance loci against various major pathogens have been identified and mapped throughout the potato genome (reviewed by Gebhardt and Valkonen, 2001). This has facilitated the isolation and characterization of an increasing number of single dominant resistance genes, including the virus resistance genes *Rx1* (chrXII) from *S. tuberosum* ssp. *andigena* and *Rx2* (chrV) from *S. acaule* (Bendahmane et al., 1999; Bendahmane et al., 2000), the nematode resistance genes *Gpa2* (chrXII) from *S. tuberosum* ssp. *andigena* (Van der Vossen et al., 2000) and *Gro1.4* (chrVII) (Paal et al., 2004), and the late blight resistance genes *R1* (chrV) and *R3a* (chrXI) from *S. demissum* (Ballvora et al., 2002; Huang et al., 2005), and *Rpi-blb1* (chrVIII) and *Rpi-blb2* (chrVI) from *S. bulbocastanum* (Van der Vossen et al., 2003; Van der Vossen et al., 2005).

_R_ genes

Plants like potato have to combat a large array of pathogenic invaders. To counteract these attacks, they protect themselves behind several overlapping layers of defence mechanisms ranging from basal countermeasures (*e.g.* cell wall reinforcement) over general defence systems reacting on pathogen associated molecular patterns (PAMPs like flagellin) to highly pathotype-specific systems such as resistance genes (*R* genes) (Jones and Dangl, 2006). They encode proteins that upon recognition of the corresponding avirulence gene (*avr* gene) product from the pathogen activate a defence response in the plant to prevent further spreading of the invader. This is often accomplished by the induction of a signal cascade leading to a rapid local cell death response or a so called hypersensitive response (HR) at the site of infection (Hammond-Kosack and Jones, 1996; Nimchuk et al., 2003; Takken et al., 2006).

Also motivated by the potential application to improve crop plants, many *R* genes have been isolated from different plant species in the recent past (Hammond-Kosack and Jones, 1997; Hulbert et al., 2001; Nimchuk et al., 2003; Parker, 2003). Most of these *R*

genes can be classified into distinct groups based on their structural domains (Cannon et al., 2002; Martin et al., 2003). The major structural class in plants consist of *R* genes encoding a leucine rich repeat domain (LRR) and a nucleotide binding site (NBS) (Martin et al., 2003; Belkhadir et al., 2004). The NBS domain plays a role in downstream signalling and activation of the defence response (Van der Biezen and Jones, 1998; Takken et al., 2006). For several R proteins (Ellis et al., 1999; Chin et al., 2001; Dodds et al., 2001; Moffett et al., 2002) it has been shown that the LRR domain is involved in direct or indirect recognition of the pathogen elicitor. Direct interaction between R proteins and Avr proteins has been detected only in a few cases (Jia et al., 2000; Deslandes et al., 2003; Dodds et al., 2006). This led to the 'guard hypothesis' suggesting an indirect activation of a defence response with the R protein embedded in a bigger protein complex. The R protein does not recognize the pathogen effector, but instead it monitors the status of associated ('guarded') plant proteins targeted by the pathogen effectors.

Genomic organization of *R* genes

Despite the large number and biodiversity of pathogens and pests, a relatively limited number of about ~ 200 NBS-LRR genes were detected in the *Arabidopsis* genome (Meyers et al., 2003). While single *R* gene loci do exist (*Rpm1*, *flax L*), the majority of *R* genes belong to large gene families located in complex loci harbouring several tandemly repeated *R* gene homologues (Michelmore and Meyers, 1998; reviewed by Hulbert et al., 2001). These *R* gene clusters or so called ´Hotspots´ for disease resistance are thought to be the result of gene duplication events (Baumgarten et al., 2003). A combination of different mechanisms that create variation (*e.g.* point mutations), brings the *R* genes under diversifying selection resulting in the continuous generation of novel recognition specificities (Parniske et al., 1997).

Various mechanisms determining the life span of *R* genes are combined in a so-called birth-and-death process (Michelmore and Meyers, 1998) in which unequal crossing overs, gene conversions and selection play a important role. Thereby, *R* genes undergo heterogenous rates of evolution (Caicedo et al., 2004; Ding et al., 2007) which was found for a set of NBS-LRR genes in lettuce containing the rapidly evolving Type I and the slowly removing Type II genes (Kuang et al., 2004). Even within the same cluster some *R* gene alleles have been retained for relatively long time, whereas others have been lost and

replaced by more recently emerged lineages (Kuang et al., 2004).

Co-evolution between pathogens and plants

Long before the discovery of the molecular mechanisms involved in resistance and virulence, certain host-pathogen interactions have been identified to be highly specific. In experiments with different races of the flax rust fungus *Melampsora lini* and flax Flor (1947) showed that resistance and virulence is dependent on single dominant factors present in the pathogen and the host plant, a phenomenon known as the gene-for-gene concept (Flor, 1971). Since then, the gene-for-gene concept has been widely accepted as a fitting model for disease resistance in plants and many combinations of resistance genes and their cognate avirulence genes have been identified and cloned since then (Staskawicz, 2001).

The arms race hypothesis

The classic model that describes the evolution of gene-for-gene interactions is the evolutionary 'arms race' between a parasite and a host plant (Dawkins and Krebs, 1979). The 'arms-race'-theory assumes that the parasite is always capable to overcome resistance by developing a new way to circumvent recognition by the host. To counter this, the host population constantly has to evolve new recognition specificities in resistance genes to match the virulent pathotypes. Because of their selective advantage for the plant, these new *R* alleles should rapidly increase their frequency within a population resulting in a so-called ´selective sweep´. In contrast, driven by the emergence of new virulence, defeated *R* gene alleles immediately decline in natural host plant populations. As a consequence, under arms race dynamics single *R* gene alleles are not very persistent over time but constantly replaced (Fig. 1). That implies the necessity of a dynamic source continuously generating novel *R* gene alleles, which is seen in the birth and death process of *R* genes.

The trench warfare hypothesis

The arms-race model assuming frequent selective sweeps should result in high rates of turnover for *R* gene alleles. However, stable polymorphisms for resistance are often observed. At the simple locus of *Rpm1* in *Arabidopsis thaliana* (Stahl et al., 1999), an NBS-LRR class gene which confers resistance to *Pseudomonas syringae*, a homologue was

found in a resistant genotype of the related species *A. lyrata*. Sequence comparison with homologous flanking regions from resistant and susceptible *A. thaliana* accessions suggests a long-term maintenance of resistant and susceptible haplotypes of *Rpm1* that predate speciation of *A. thaliana* and *A. lyrata*. The authors therefore proposed an alternative model predicting allele cycling caused by frequency dependent balanced selection and introduced the term ´trench warfare´ for the co-evolution of resistance and virulence alleles. Here, the dynamics of *R* genes and *Avr* genes follow repeated advances and retreats of resistance and virulence alleles (Frank, 1992; Stahl et al., 1999). Together with the recognition specificity the resistance alleles are maintained as balanced polymorphisms in natural populations over long periods of time (Fig. 1).

The discovery of functional *R* gene alleles with a long life-span that eventually predates host speciation events is a strong evidence for the trench warfare hypothesis. Examples for such ancient *R* gene specificities are *Pto* from *Lycopersicon pimpinellifolium* and *LhirPto* from *L. hirsutum* (Riely and Martin, 2001). It has been shown that these highly similar orthologues encode for a protein kinase that in both cases elicits a hypersensitive response when co-expressed with the avirulence factor from *P. syringae*. The *Cladosporium fulvum* resistance genes *Cf-4* and *Cf-9* from *Lycopersicon pimpinellifolium* have several functional orthologues that were found in five other wild species of tomato (Kruijt et al., 2004). It is believed that the nearly identical orthologues are derived from a common ancestor within the genus *Lycopersicon*.

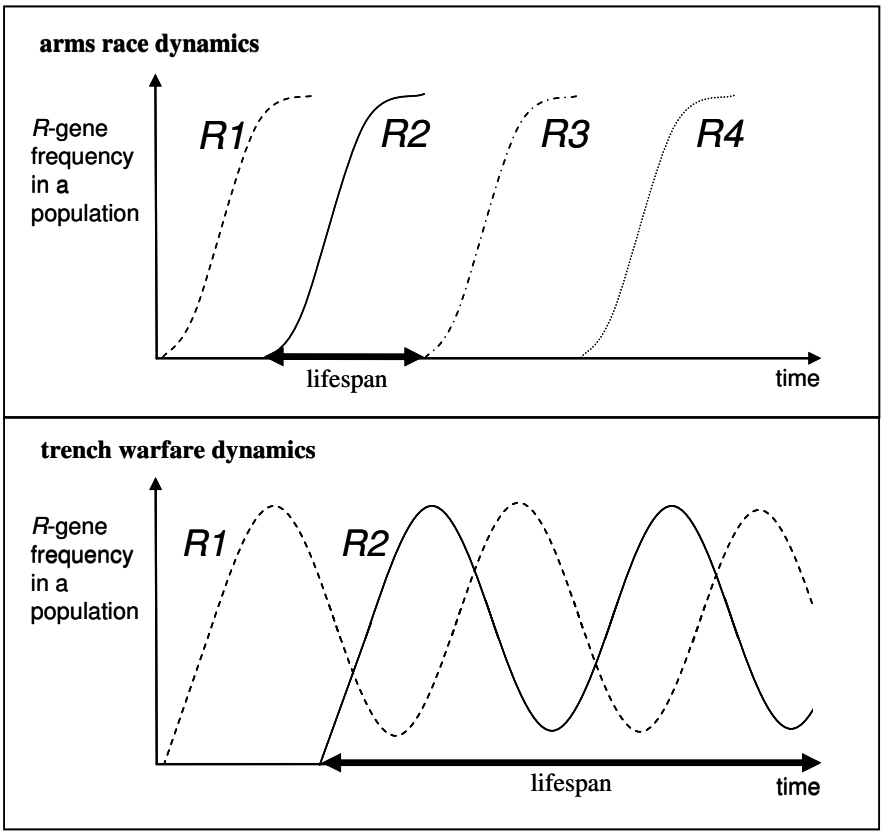


Figure 1: Schematic representation of the evolutionary dynamics of *R* genes according to the arms race (upper panel) and trench warfare theory (lower panel). Frequencies of *R*-gene alleles within a population rise and fall over time as a consequence of selection pressure imposed by the corresponding pathogen alleles.

Outline of this thesis

The aim of the research described in this thesis was to study the molecular mechanisms underlying the evolution of *R* gene recognition specificities using the virus resistance gene *Rx1* from potato. *Rx1* and it close homologue *Gpa2* are both located in a single *R* gene cluster of about 110 kb on chromosome XII in potato and share an overall homology in the amino acid sequence of about 87%. Interestingly, these two *R* genes confer resistance to taxonomically unrelated pathogens, a virus and a nematode. However, *Rx1* results in extreme resistance to potato virus X (PVX), whereas *Gpa2* gives a much slower and milder resistance response to the potato cyst nematode *Globodera pallida* (Van der Vossen et al., 2000). Therefore, the *Rx1/Gpa2* locus in potato is an excellent model system to study the evolutionary dynamics of *R* gene clusters in plants.

In **Chapter 2**, the genetic and physical mapping of additional *Rx1* and *Gpa2* homologues is described. Analysis of the three *S. tuberosum* ssp. *tuberosum* haplotypes, homeologous to the *S. tuberosum* ssp. *andigena* haplotype that harbours the *Gpa2/Rx1* resistance gene cluster, revealed nine additional *Gpa2/Rx1* homologues in the two diploid potato clones SH and RH. The resistance gene homologues (RGH) were identified with a specific primer pair based on conserved motifs of the LRR domain from *Gpa2* and *Rx1*. Sequence analysis of the RGHs revealed that they are highly similar to *Gpa2* and *Rx1* with sequence identities ranging from 93% to 100%. A modified AFLP method was used to facilitate the genetic mapping of the RGHs. They are all located in the *Gpa2/Rx1* cluster on chromosome XII.

The cluster containing *Rx1* and *Gpa2* has been introgressed into cultivated potato from *S. tuberosum* ssp. *andigena* (Bendahmane et al., 1999). An additional functional orthologue called *Rx2* was found (Bendahmane et al., 2000) in *S. acaule,* a wild relative of cultivated potato (Ritter et al., 1991). A functional screening of 51 accessions from wild potato species resulted in the identification and characterization of the two functional *Rx* orthologues *Rx3* and *Rx4* from *S. lepthophyes/sparsipilum* and *S. albicans*, respectively. Sequence analysis of the functional *Rx1* genes and their homologues showed extreme sequence conservation of the LRR domain in *Rx1 – 4* allowing the design of an *Rx*-specific

primer set, which was used for an allelic mining strategy in *Solanum*. The detection of functional *Rx* orthologues in distantly related species points out the ancient origin of the *Rx1* gene (**Chapter 3**).

From 10 *Solanum* species, sequence analysis has been carried out on the open reading frames of 75 highly similar *Rx1/Gpa2* homologues (**Chapter 4**). This analysis includes the four functional orthologous virus resistance genes *Rx1-4*, the truncated Rx_{edn}, that harbours *Rx* resistance specificity and the nematode resistance gene *Gpa2*. We identified a local structural variability in the sequence region encoding LRR 10, dividing the homologues in 3 structural subclasses influencing the predicted horseshoe structure. The occurrence of similar sequence exchange tracks in homologues isolated from different *Solanum* species indicates that interallelic exchanges took place before speciation. Analysis for positive selection revealed very significant positive selection on codon sites within the LRR domain, but also within the CC- and NBS- domain.

Chapter 5 describes a comparative sequence analysis between the *Rx1/Gpa2* clusters derived from five different haplotypes from *S. tuberosum* ssp *andigena*, *S. tuberosum* ssp *tuberosum*, S. *lepthophyes* and *S. sparsipilum*, which allowed us to determine the genomic organisation of this region in two haplotypes derived from *S. tuberosum* ssp *tuberosum.* Furthermore, three additional close homologues of *Rx* were identified in *S. lepthophyes/sparsipilum*, but it was not possible to determine the position and orientation of the *Rx* homologues due to low sequence homology in this genomic region. However, patterns of extreme sequence conservation were detected in the 5' and 3' UTR regions of the *Rx/Gpa2* homologues for each haplotype. In addition, it is shown that the *Rx/Gpa2* cluster contains a gene encoding for a sugar transporter, various retrotransposon elements and a set of genes homologous to the zink-finger protein RMA1 from *Arabidopsis*.

Finally, an outlook is presented on the functional constraints of the molecular mechanisms underlying the evolutionary dynamics of the *Rx/Gpa2* cluster as described in this thesis (**Discussion).**

Literature

Ballvora, A., Ercolano, M.R., Weiß, J., Meksem, K., Bormann, C.A., Oberhagemann, P., Salamini, F., and Gebhardt, C. (2002). The *R1* gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. Plant Journal **30,** 361.

Baumgarten, A., Cannon, S., Spangler, R., and May, G. (2003). Genome-level evolution of resistance genes in *Arabidopsis thaliana*. Genetics **165,** 309.

Belkhadir, Y., Subramaniam, R., and Dangl, J.L. (2004). Plant disease resistance protein signaling: NBS-LRR proteins and their partners. Current Opinion in Plant Biology **7,** 391.

Bendahmane, A., Kanyuka, K., and Baulcombe, D.C. (1999). The *Rx* gene from potato controls separate virus resistance and cell death responses. Plant Cell **11,** 781.

Bendahmane, A., Kanyuka, K., Baulcombe, D.C., Querci, M., and Bendahmane, A. (2000). *Agrobacterium* transient expression system as a tool for the isolation of disease resistance genes: Application to the *Rx2* locus in potato. Plant Journal **21,** 73.

Caicedo, A.L., Olsen, K.M., Purugganan, M.D., Stinchcombe, J.R., and Schmitt, J. (2004). Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. Proceedings of the National Academy of Sciences of the United States of America **101,** 15670.

Cannon, S.B., Zhu, H., Baumgarten, A.M., Spangler, R., May, G., Cook, D.R., and Young, N.D. (2002). Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. Journal of Molecular Evolution **54,** 548.

Chin, D.B., Arroyo-Garcia, R., Ochoa, O.E., Kesseli, R.V., Lavelle, D.O., and Michelmore, R.W. (2001). Recombination and spontaneous mutation at the major cluster of resistance genes in lettuce (*Lactuca sativa*). Genetics **157,** 831.

Dawkins, R., and Krebs, J.R. (1979). Arms races between and within species. Proceedings of the Royal Society of London - Biological Sciences **205,** 489.

Deslandes, L., Olivier, J., Peeters, N., Feng, D.X., Khounlotham, M., Boucher, C., Somssich, I., Genin, S., and Marco, Y. (2003). Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. Proceedings of the National Academy of Sciences of the United States of America **100,** 8024.

Ding, J., Cheng, H., Jin, X., Araki, H., Yang, Y., and Tian, D. (2007). Contrasting patterns of evolution between allelic groups at a single locus in *Arabidopsis*. Genetica **129,** 235.

Dodds, P.N., Lawrence, G.J., and Ellis, J.G. (2001). Contrasting modes of evolution acting on the complex *N* locus for rust resistance in flax. Plant Journal **27,** 439.

Dodds, P.N., Lawrence, G.J., Catanzariti, A.M., Teh, T., Wang, C.I.A., Ayliffe, M.A., Kobe, B., and Ellis, J.G. (2006). Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. Proceedings of the National Academy of Sciences of the United States of America **103,** 8888.

Ellis, J.G., Lawrence, G.J., Luck, J.E., and Dodds, P.N. (1999). Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. Plant Cell **11,** 495.

Flor, H.H. (1971). Current status of the gene-for-gene concept. [Plants, breeding, disease resistance]. Annu Rev Phytopathol **9,** 275.

Frank, S.A. (1992). Models of plant-pathogen coevolution. Trends in Genetics **8,** 213.

Fribourg, C.E. (1980). History and distribution of potato viruses in Latin America. FITOPATOLOGIA **15,** 13.

Gebhardt, C., and Valkonen, J.P.T. (2001). Organization of genes controlling disease resistance in the potato genome. Annual Review of Phytopathology **39,** 79.

Hammond-Kosack, K.E., and Jones, J.D.G. (1996). Resistance gene-dependent plant defense responses. Plant Cell **8,** 1773.

Hammond-Kosack, K.E., and Jones, J.D.G. (1997). Plant disease resistance genes. Annual Review of Plant Biology **48,** 575.

Hawkes, J. (1990). The potato. Evolution, biodiversity and genetic resources. (Belhaven, London).

Huang, S., Van der Vossen, E.A.G., Kuang, H., Vleeshouwers, V.G.A.A., Zhang, N., Borm, T.J.A., Van Eck, H.J., Baker, B., Jacobsen, E., and Visser, R.G.F. (2005). Comparative genomics enabled the isolation of the *R3a* late blight resistance gene in potato. Plant Journal **42,** 251.

Hulbert, S.H., Webb, C.A., Smith, S.M., and Sun, Q. (2001). Resistance gene complexes: Evolution and utilization. Annual Review of Phytopathology **39,** 285.

Jia, Y., McAdams, S.A., Bryan, G.T., Hershey, H.P., and Valent, B. (2000). Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. EMBO Journal **19,** 4004.

Jones, J.D.G., and Dangl, J.L. (2006). The plant immune system. Nature **444,** 323-329.

Kruijt, M., Brandwagt, B.F., and De Wit, P.J.G.M. (2004). Rearrangements in the *Cf-9* disease resistance gene cluster of wild tomato have resulted in three genes that mediate Avr9 responsiveness. Genetics **168,** 1655.

Kuang, H., Woo, S.S., Meyers, B.C., Nevo, E., and Michelmore, R.W. (2004). Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. Plant Cell **16,** 2870.

Martin, G.B., Bogdanove, A.J., and Sessa, G. (2003). Understanding the Functions of Plant Disease Resistance Proteins. Annual Review of Plant Biology **54,** 23.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. (2003). Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell **15,** 809.

Michelmore, R.W., and Meyers, B.C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Research **8,** 1113.

Moffett, P., Farnham, G., Peart, J., and Baulcombe, D.C. (2002). Interaction between domains of a plant NBS-LRR protein in disease resistance-related cell death. EMBO Journal **21,** 4511.

Nimchuk, Z., Eulgem, T., Holt Iii, B.F., and Dangl, J.L. (2003). Recognition and Response in the Plant Immune System. Annual Review of Genetics **37,** 579.

Paal, J., Henselewski, H., Muth, J., Meksem, K., Menéndez, C.M., Salamini, F.,

Ballvora, A., and Gebhardt, C. (2004). Molecular cloning of the potato *Gro1-4* gene conferring resistance to pathotype Ro1 of the root cyst nematode *Globodera rostochiensis*, based on a candidate gene approach. Plant Journal **38,** 285.

Parker, J.E. (2003). Plant recognition of microbial patterns. Trends in Plant Science **8,** 245.

Parniske, M., Golstein, C., Thomas, C.M., Wulff, B.B.H., Hammond-Kosack, K.E., Jones, D.A., Harrison, K., and Jones, J.D.G. (1997). Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. Cell **91,** 821.

Riely, B.K., and Martin, G.B. (2001). Ancient origin of pathogen recognition specificity conferred by the tomato disease resistance gene *Pto*. Proceedings of the National Academy of Sciences of the United States of America **98,** 2059.

Ritter, E., Debener, T., Barone, A., Salamini, F., and Gebhardt, C. (1991). RFLP mapping on potato chromosomes of two genes controlling extreme resistance to potato virus X (PVX). Molecular and General Genetics **227,** 81.

Ross, H. (1986). Potato breeding. Problems and perspectives. Adv. Plant. Breed. **Suppl., 13,** pp. 82.

Salaman, R.N. (1985). The potato famine: It´s causes and consequences. In : The History and Social Influence of the Potato, revised impression (Hawkes, J.G., ed.). (Cambridge: Cambridge University Press, pp 289-316).

Stahl, E.A., Kreitman, M., Bergelson, J., Dwyer, G., and Mauricio, R. (1999). Dynamics of disease resistance polymorphism at the *Rpm1* locus of Arabidopsis. Nature **400,** 667.

Staskawicz, B.J. (2001). Genetics of plant-pathogen interactions specifying plant disease resistance. Plant Physiology **125,** 73.

Takken, F.L., Albrecht, M., and Tameling, W.I. (2006). Resistance proteins: molecular switches of plant defence. Current Opinion in Plant Biology **9,** 383.

Ugent, D. (1982). The potato.

Van der Biezen, E.A., and Jones, J.D. (1998). The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. Current biology : CB **8**.

Van der Vossen, E., Te Lintel Hekkert, B., Wouters, D., Pereira, A., Stiekema, W., Sikkema, A., Gros, J., Stevens, P., Muskens, M., and Allefs, S. (2003). An ancient *R* gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. Plant Journal **36,** 867.

Van der Vossen, E.A.G., Gros, J., Sikkema, A., Muskens, M., Wouters, D., Wolters, P., Pereira, A., and Allefs, S. (2005). The *Rpi-blb2* gene from *Solanum bulbocastanum* is an *Mi-1* gene homolog conferring broad-spectrum late blight resistance in potato. Plant Journal **44,** 208.

Van der Vossen, E.A.G., Rouppe van der Voort, J.N.A.M., Kanyuka, K., Bendahmane, A., Sandbrink, H., Baulcombe, D.C., Bakker, J., Stiekema, W.J., and Klein-Lankhorst, R.M. (2000). Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: A virus and a nematode. Plant Journal **23,** 567-576.

Chapter 2

Genetic and physical mapping of homologues of the virus resistance gene *Rx1* and the cyst nematode resistance gene *Gpa2* in potato

Erin Bakker, Patrick Butterbach, Jeroen Rouppe van der Voort, Edwin van der Vossen, Joke van Vliet, Jaap Bakker and Aska Goverse

has been published as Bakker et al. 2003 TAG 106: 1524-31

Abstract

Nine resistance gene homologues (RGHs) were identified in two diploid potato clones (SH and RH) with a specific primer pair based on conserved motifs in the LRR domain of the potato cyst nematode resistance gene *Gpa2* and the potato virus X resistance gene *Rx1*. A modified AFLP method was used to facilitate the genetic mapping of the RGHs in the four haplotypes under investigation. All nine RGHs appeared to be located in the *Gpa2/Rx1* cluster on chromosome XII. Construction of a physical map using bacterial artificial chromosome (BAC) clones for both the *Solanum tuberosum* ssp. *tuberosum* and the *S. tuberosum* ssp. *andigena* haplotype of SH showed that the RGHs are located within a stretch of less than 200 kb. Sequence analysis of the RGHs revealed that they are highly similar (93 to 95%) to *Gpa2* and *Rx1*. The sequence identities among all RGHs range from 85 to 100%. Two pairs of RGHs are identical or nearly so (100 and 99,9%), with each member located in a different genotype. Southern blot analysis on genomic DNA revealed no evidence for additional homologues outside the *Gpa2/Rx1* cluster on chromosome XII.

Introduction

Plants are constantly under attack by a wide range of pathogens and pests. To defend themselves, plants have evolved an innate surveillance system consisting of a large set of resistance genes. Most resistance genes are single dominant and confer resistance in a gene-for-gene specific manner (Flor, 1971). More than 30 resistance genes (*R* genes) have been cloned from different plant species and they can be divided into four classes based on common structural motives (Takken and Joosten, 2000). The majority of the *R* genes are characterized by a leucine rich repeat (LRR) and a nucleotide binding site (NBS) domain. *R* genes that belong to this superfamily confer resistance to completely unrelated taxonomic groups like bacteria, fungi, viruses and nematodes (Mindrinos et al., 1994; Whitham et al., 1994; Lawrence et al., 1995; Vos et al., 1998; Milligan et al., 1998; Bendahmane et al., 1999; Van der Vossen et al., 2000).

In potato, nineteen *R* genes have been mapped to eleven chromosomal regions (Gebhardt and Valkonen, 2001). Four of them, *Rx1*, *Gpa2*, *Rx2* and *R1,* have been isolated and they all belong to the NBS-LRR class. *Rx1* and *Gpa2* originate from *Solanum tuberosum* ssp.

andigena and have been identified by map based cloning (Bendahmane et al., 1999; Van der Vossen et al., 2000). *Rx2* has been isolated from *S. aucaule* using a PCR based approach (Bendahmane et al., 2000) and *R1* has been isolated from *S. tuberosum* by using a combination of map based cloning and the candidate gene approach (Ballvora et al., 2002). *Rx1* and *Gpa2* are highly homologous, yet they confer resistance to two completely unrelated pathogens *viz.* potato virus X and the potato cyst nematode *Globodera pallida*. The genes are tightly linked on chromosome XII of potato (Bendahmane et al., 1997; Rouppe van der Voort et al., 1999). Sequencing a 187 kb region revealed that *Gpa2* and *Rx1* are part of a complex locus containing at least two other closely related resistance gene homologues (RGHs): *RGC1* and the pseudogene *RGC3* (Bendahmane et al., 1997; Rouppe van der Voort et al., 1999; Bendahmane et al., 1999; Van der Vossen et al., 2000). In the LRR domain, more variation is observed between the RGHs than in the NBS domain. The mean K_a/K_s ratio for the LRR region is larger than one, whereas in the NBS regions the ratio is smaller than one, indicating that the LRR domain is subject to diversifying selection and that specificity is determined by this domain (Van der Vossen et al., 2000).

In this paper we describe the use of LRR-specific primer combinations for both the identification and mapping of nine *Gpa2*/*Rx1* homologues in two diploid potato clones (SH and RH). The value of using LRR-specific primers for the dissection of *R* gene clusters in different haplotypes will be discussed.

Results

Identification of *Gpa2*/*Rx1* homologues in SH and RH

To identify homologues of *Gpa2* and *Rx1*, a specific primer set was designed based on conserved regions in the LRR domain. This primer set was used to screen a BAC library of the diploid potato clone SH, which harbours the *R* genes *Gpa2* and *Rx1* on a *S. tuberosum* ssp. *andigena* introgression segment. For each positive BAC clone, amplification products were cloned and 12 transformants were used for reamplification. After *Taq*I digestion, resistance gene homologues (RGHs) were selected based on their unique digestion pattern (Fig. 2a). This resulted in the identification of three RGHs: *SH-RGH5* (BAC clone 7E16), *SH-RGH6* and *SH-RGH 7* (both located on BAC clones 25G18 and 36G3). Also *Gpa2*

(BAC 85N1), *Rx1* (BAC 43) and *SH-RGH1* (BAC 60K18) were detected based on their expected *Taq*I digestion pattern (Bendahmane et al., 1999; Van der Vossen et al., 2000). No additional RGHs were found after screening genomic DNA of SH.

The diploid potato clone RH was studied using genomic DNA. PCR analysis resulted in three bands at 800 bp, 1300 bp and 1600 bp. The complete PCR product was cloned and restriction analysis of the *Taq*I patterns of the reamplified PCR products of 25 clones resulted in the identification of six *Gpa2*/*Rx1* homologues: *RH-RGH1* to *RH-RGH6* (Fig. 2b). Interestingly, *RH-RGH4* and *RH-RGH5* resembled the digestion patterns of *SH-RGH5* and *SH-RGH6*, respectively. In total, twelve RGHs were identified in potato including *Gpa2, Rx1* and *SH-RGH1* using a *Gpa2/Rx1* cluster-specific primer set (Table 2). Another *Gpa2/Rx1* homologue present in the cluster, the pseudogene *SH-RGH3*, was not amplified with the cluster primer set. The annealing site of the reverse primer is not present in *SH-RGH3*, because the 3' end of its sequence is truncated.

The *Gpa2*/*Rx1* homologues of SH and RH were sequenced to determine their sequence resemblance with *Rx1* and *Gpa2*. Comparison of the nucleotide sequence of the LRR domains of the RGHs showed that they are highly similar to each other with sequence identities ranging from 85.5% to 100% (Table 3). The sequences of *RH-RGH4* and *RH-RGH5* were indeed virtually identical to those of *SH-RGH5* and *SH-RGH6,* respectively (99.9% and 100% sequence identity).

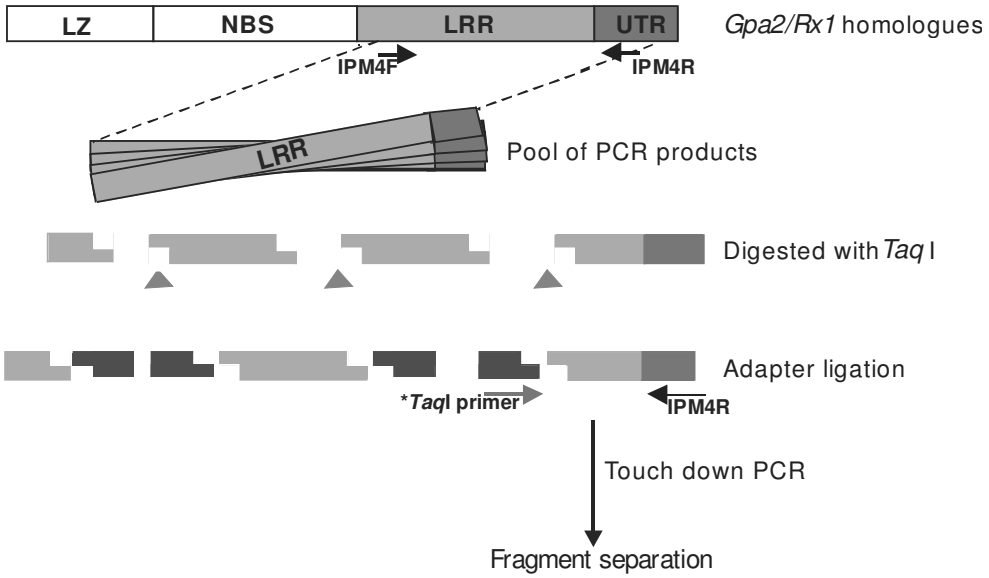


Figure 1 Schematic depiction of the method used to genetically map the *Gpa2/Rx1* homologues. The LRR-specific primers iPM4F and iPM4R are used to amplify a pool of RGH sequences from either genomic DNA or BAC DNA. Subsequently, the PCR products are used as template in a modified AFLP reaction. After digestion with the restriction enzyme *Taq*I the products are ligated to a *Taq*I adapter. A second touch down PCR reaction is performed using the *Taq*I primer and the original iPM4R primer.

Table 1 Primer sequences based on BAC end sequences to detect overlap between BAC clones.

BAC end	primers
7E16sp6	5'- cgg ggt gta atg tga tga gc-3'
	5'- ggc ctg caa gtc tgt gca c-3'
7E16t7	5'- gtt cgt atg agc gag tat gg-3'
	5'- tcc acg atg gtc tcc tcg-3'
25G18sp6	5'-cca att tca agc ttc ttc ata g-3'
	5'-cag tca agg tgc ttt gga gg-3'
25G18t7	5'- gtt acc tgc tat gtg agc tc-3'
	5'- cat cag ctg cct tgc agt tg-3'
36G3sp6	5'- gcc caa cat gat agg tcg c-3'
	5'- ctt ggt atc aga gca cag ag -3'
36G3t7	5'-tgt atg aat tgg gtc att ccg-3'
	5'- gcc caa tat tcc tcc atc tg-3'

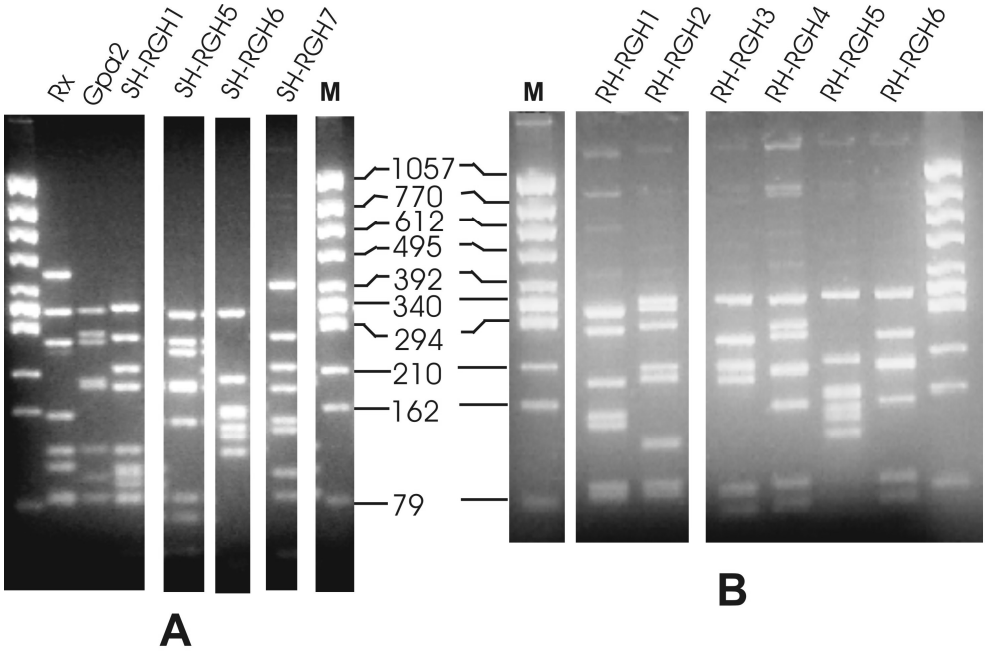


Figure 2 Identification of Gpa2/Rx1 resistance gene homologues (RGHs) in the diploid potato clone SH (2A) and RH (2B). RGHs were selected based on differences in their *Taq*I digestion pattern after PCR amplification with specific LRR primers. The *Taq*I digestion patterns are related to a DNA base pair ladder (M). Bands larger then 612 bp are the result of partially digested DNA.

Table 2 Resistance gene homologues (RGHs) identified in BAC and genomic DNA of the diploid potato clones SH and RH using a primer pair based on the LRR domain of *Gpa2* and *Rx1*.

RGH	PCR product (kb)	BAC clones	gDNA
SH-RGH1	1.6	85N1	+
Gpa2	1.6	85N1, 60K18	-[a)]
Rx1	1.6	BAC 43	-
SH-RGH5	1.5	7E16	+
SH-RGH6	1.3	25G18, 36G3	+
SH-RGH7	1.6	25G18, 36G3	-
RH-RGH1	1.6	-	+
RH-RGH2	1.6	-	+
RH-RGH3	1.6	-	+
RH-RGH4	1.5	-	+
RH-RGH5	1.3	-	+
RH-RGH6	0.8	-	+

a) not detected in genomic DNA

Table 3 Sequence identities (%) of the LRR domains of the RGHs in SH and RH

RGH	Gpa2	Rx1	SH-RGH1	SH-RGH3*	SH-RGH5	SH-RGH6	SH-RGH7	RH-RGH1	RH-RGH2	RH-RGH3	RH-RGH4	RH-RGH5
Rx1	93.5											
SH-RGH1	93.6	93.2										
SH-RGH3*	86.5	86.8	87.3									
SH-RGH5	92.9	92.8	93.5	87.4								
SH-RGH6	94.3	93.2	93.1	92.6	92.4							
SH-RGH7	92.8	93.3	91.9	86.5	91.9	93.1						
RH-RGH1	92.9	93.8	91.8	87.0	91.6	92.8	98.5					
RH-RGH2	93.7	94.8	93.1	88.3	93.5	94.8	93.4	92.2				
RH-RGH3	94.1	93.7	93.9	87.7	93.1	94.5	91.2	93.3	93.9			
RH-RGH4	93.1	92.9	93.5	87.5	99.9	92.6	92.1	91.7	93.4	93.1		
RH-RGH5	94.3	93.2	93.1	92.6	92.4	100.0	93.0	92.8	94.7	94.5	92.5	
RH-RGH6	93.6	94.0	85.5	85.5	93.3	90.8	92.3	92.2	94.1	94.3	93.3	90.8

Genetic mapping of the *Gpa2*/*Rx1* homologues in SH × RH

Gpa2/*Rx1*-specific primers were used to perform PCR on the parents SH and RH and a progeny of 100 individuals to determine the genetic position of the *Gpa2*/*Rx1* homologues in the potato genome. The selected BAC clones harbouring *Gpa2*/*Rx1* homologues were included as a control. On the resulting pools of PCR products, a modified AFLP analysis was performed that makes use of sequence polymorphisms in the last *Taq*I restriction sites of the PCR products. Using the sequence information of the RGHs we calculated the length of the expected fragments from the *Taq*I restriction site closest to the 3' end of the PCR products. In this way we were able to determine the positions in the gels of the RGHs. Figure 3 shows the autoradiogram with segregating bands for *Gpa2, Rx1, SH-RGH5, SH-RGH6* and *RH-RGH1-RH-RGH6* at the expected positions in the gel (summarised in Table 4). *SH-RGH1* and *SH-RGH7* cannot be linked to a segregating band, because they comigrate with a thick band in the gel. However, *SH-RGH1* is present on the same BAC as *Gpa2,* and *SH-RGH7* on the same BAC as *SH-RGH6*. All the segregating RGHs were linked to the *Gpa2Rx1* resistance gene cluster with a logarithm of odds (LOD) score between 8.2 and 12.6. Other DNA fragments visible on the autoradiogram either cosegregate with RGHs, or do not segregate at all. They may be artefacts caused by the double PCR and the limited complexity of the template used.

SH-RGH5 and *SH-RGH6* are in repulsion with *Gpa2*. This implies that *SH-RGH5*, *SH-RGH6* and *SH-RGH7* are all derived from the *S. tuberosum* ssp. *tuberosum* haplotype and that no additional RGHs have been found on the *Gpa2*/*Rx1* cluster of the *S. tuberosum* ssp. *andigena* introgression segment. Furthermore, *RH-RGH1, 2, 3* and *6* are in coupling with each other and in repulsion with *RH-RGH4* and *5* (Table 4). From these data we can conclude that all the identified homologues are genetically linked to the *Gpa2*/*Rx1* cluster on chromosome XII in SH and in RH. Based on these and other data (see also below), we postulated the position of the RGHs as shown in Figure 4.

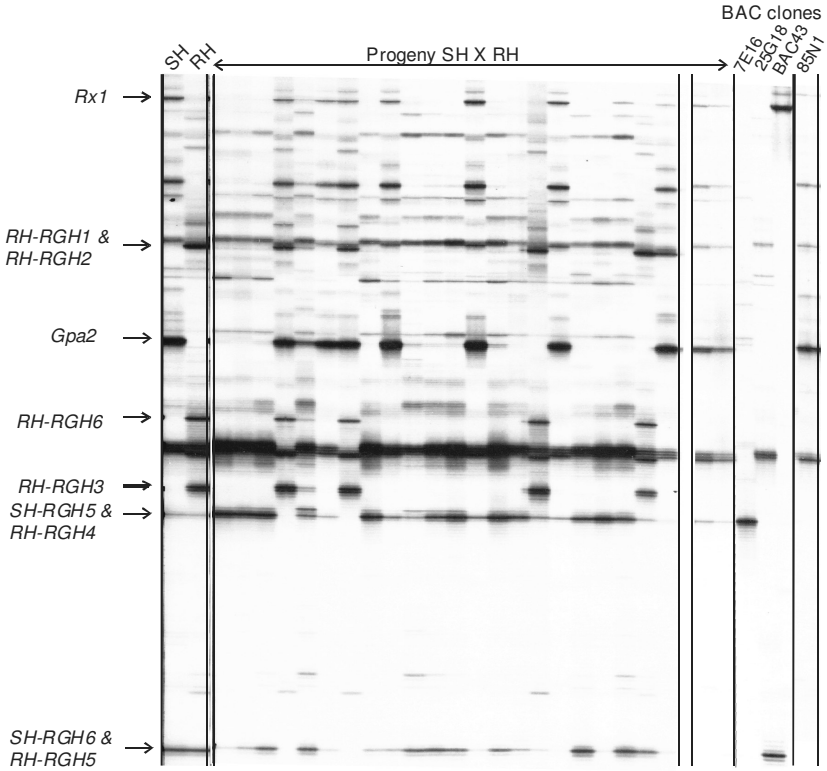


Figure 3 Mapping of RGHs in a SH × RH cross and in a number of BAC clones using a modified AFLP method on a pool of RGH sequences. All the indicated RGHs map on chromosome XII and are linked to *Gpa2* with a LOD score between 88.2 and 12.6. *Rx1* was also identified by the comigrating band on BAC43 from which *Rx1* was originally cloned (Bendahmane et al., 1999). Likewise, we were also able to verify the expected bands for *Gpa2* and *SH-RGH1* by comparing the pattern of the mapping population with the bands in BAC 85N1. Furthermore, we could relate the bands present on BAC clones 7E16 and 25G18 with bands in the progeny. The majority of the remaining, unassigned bands are artefacts of the PCR procedure.

Table 4 Segregation of the RGHs in a progeny of 136 individuals. RGHs with the same mark (++ or --) are in coupling.

SH		RH	
RGH	**segregation**	**RGH**	**segregation**
Gpa2	++	*RH-RGH1*	++
Rx1	++	*RH-RGH2*	++
SH-RGH5	--	*RH-RGH3*	++
SH-RGH6	--	*RH-RGH6*	++
		RH-RGH4	--
		RH-RGH5	--

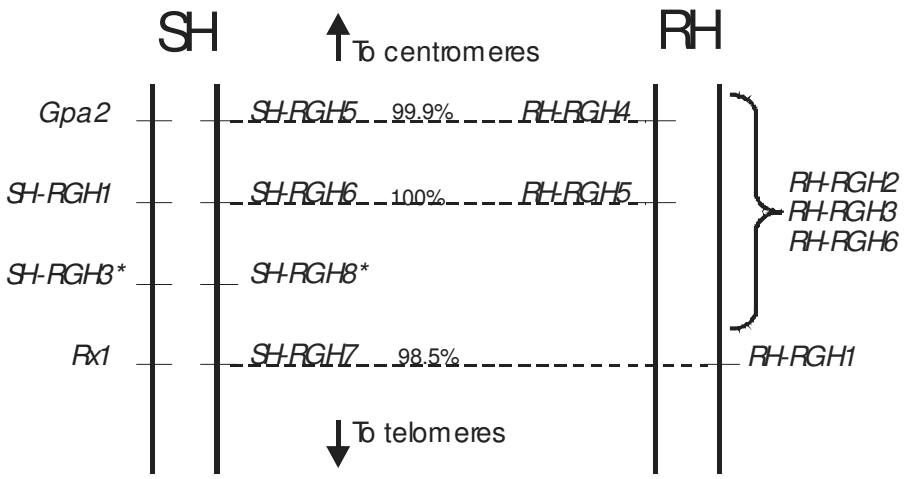


Figure 4. Schematic drawing of the *Gpa2/Rx1* homologues on the four chromosomes of SH and RH. The relative order and orientation of a number of RGHs has been postulated as follows (see also results). *SH-RGH5* is the only RGH present on a BAC clone positioned at the top of the cluster. *RH-RGH4* has a 99.9% identity to *SH-RGH5* and therefore we assume that this RGH is also positioned at the top of the cluster in RH. Because the intron positions of *SH-RGH7* are the same as in *Rx1* we assume that this RGH, like *Rx1*, is at the bottom of the cluster. *RH-RGH1* and *SH-RGH7* have a sequence identity of 98.5 and therefore we assume that *RH-RGH1* is also at the bottom of the cluster in RH. We also assume that the two pseudogenes *SH-RGH3* and *SH-RGH8* that could not be amplified in the PCR (marked with an asterisk) occupy the same position in the cluster. Finally, we assume, based on their 100% sequence identity, that *RH-RGH5* has the same position in the cluster as *SH-RGH6*. No relative order for *RH-RGH2, RH-RGH3* and *RH-RGH6* could be postulated.

Physical mapping of the *Gpa2/Rx1* cluster in SH

The BAC clones containing RGHs were used to construct a physical map of the *Gpa2/Rx1* cluster for both the *Solanum tuberosum* spp. *tuberosum* and the *S. tuberosum* spp. *andigena* haplotypes of SH. Genetic mapping revealed that all these RGHs are located on chromosome XII and hence, the BAC inserts correspond with genomic fragments of this chromosomal region of the potato genome. *SH-RGH5* and *SH-RGH6* comigrate with markers that are in repulsion with *Gpa2* implicating that they are located on the homologous chromosome. Based on this information we could asign the origin of the BAC clones to any of the two haplotypes: BAC 43, BAC 85N51 and BAC 34F16 were derived from the homologous chromosome in *S. tuberosum* ssp. *andigena* introgression segment and BAC clones 25G18, 7E16 and 36G3 were derived from *S. tuberosum* ssp. *tuberosum*. The BAC clones derived from *S. tuberosum* ssp. *andigena* formed a closed contig and could be easily aligned with the original physical map based on the presence and absence of *Gpa2*, *SH-RGH1* and *Rx1* and the markers 73L, 111R, 111L, 221R, 45L, 77R and 77L (Kanyuka et al., 1999). In order to make a contig of the *S. tuberosum* ssp. *tuberosum* haplotype, the BAC ends were sequenced and primers were designed to perform PCR on the other BAC clones of this haplotype. Additionally, BACs 25G18 and 36G3 harbour the same RGHs. These data enabled us to construct a closed contig of these BAC clones (Fig. 5). CHEF electrophoresis of the BAC inserts (data not shown) showed that BAC 7E16 is approximately 120 kb. Both 36G3 and 25G18 are approximately 90 kb in size resulting in a physical map of *Solanum tuberosum* ssp. *tuberosum* of about 200 kb containing *SH-RGH5, 6* and *7*.

Moreover, we were also interested in comparing the genomic organisation of the *Gpa2/Rx1* cluster derived from the S. *tuberosum* spp. *tuberosum* haplotype and the one derived from the *S. tuberosum* ssp. *andigena* introgression segment. A segment of 187 kb of the *S. tuberosum* ssp. *andigena* haplotype has previously been sequenced for cloning *Rx1* and *Gpa2* (Van der Vossen et al., 2000). This sequence information was used to align the BAC end sequences of the BAC clones derived from *S. tuberosum* ssp. *tuberosum* with the contig of the *S. tuberosum* ssp. *andigena* haplotype. The approximate 80% sequence identity for the BAC ends of the left arms of BAC clones 7E16, 25G18 and 36G3 with regions between *Gpa2* and *SH-RGH1* led to the orientation of the BAC contig as proposed in Figure 5. The order of the homologues present on BAC clones 25G18 and 36G3 is not clear. However,

comparison of intron positions between the RGHs showed that *SH-RGH7* has similar intron positions as *Rx1* and *SH-RGH6* has not (Bakker et al., unpublished data). Therefore we presented the order of the homologues as depicted in Figure 5.

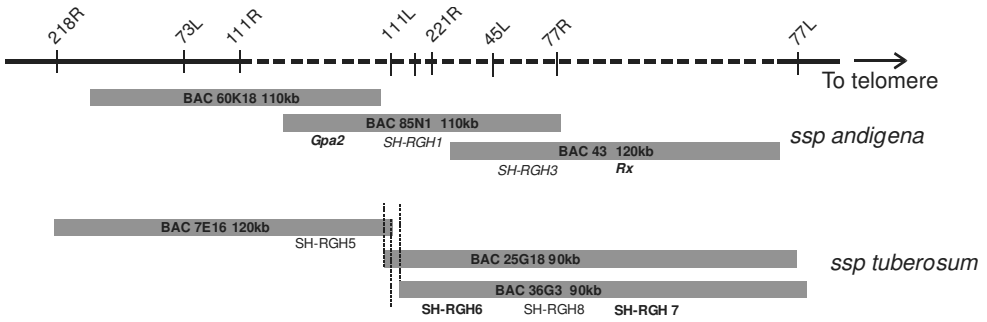


Figure 5. Physical map of the *Gpa2/Rx1* cluster in the diploid potato clone SH. The region (187 kb) around the resistance genes *Gpa2, Rx1* and *SH-RGH1* and the pseudogene *SH-RGH3* on the haplotype derived from *S. tuberosum* ssp *andigena* has been sequenced. On the *S. tuberosum* ssp. *tuberosum* haplotype also four *Gpa2/Rx1* homologues are present: *SH-RGH5, 6,7* and *8*. Like *SH-RGH3,* the homologue *SH-RGH8* could not be identified by PCR, but was detected by Southern blot analysis. The BAC clones are represented in grey rectangles and RGHs are indicated alongside the BACs on which they were detected. RGHs are indicated in bold when intron positions are known. Vertical lines indicate the marker positions used to align the BAC contig from the *S. tuberosum* ssp. *tuberosum* haplotype with the *S. tuberosum* ssp. *andigena* introgression segment. Vertical dotted lines indicate overlap between BAC ends. Dotted horizontal line in the *S. tuberosum* ssp. *andigena* chromosome indicates the position of the sequenced region of 187 kb.

Southern blot analysis

To confirm the specificity of the *Gpa2*/*Rx1*-specific primer set, southern blot analysis was carried out at high stringency conditions on *Hind*III digested DNA of the BAC clones from SH harbouring the RGHs (data not shown). Hybridisation with a 400 bp fragment derived from the 5'-end of the LRR domain of *Gpa2* resulted in a single band for 7E16 and four bands for 25G18 and 36G3. *SH-RGH7* has a *Hind*III site in the probe region which accounts for one of the two additional bands detected in 25G18 and 36G3. The other additional band indicates the presence of an extra RGH in the *Gpa2*/*Rx1* cluster in the haplotype derived from *Solanum tuberosum* ssp*. tuberosum.* Similar to the pseudogene *SH-RGH3*, this RGH is not amplified with the cluster specific primer pair. Based on this information we assume Together with the information on the position of BAC 25G18 and 36G3, we assume that this additional RGH (designated *SH-RGH8*) is also a pseudogene and

that it is most likely positioned between the homologues *SH-RGH6* and *SH-RGH7* (Fig. 4 and Fig. 5). It is noted that the sequenced homologous region in the haplotype of *S. tuberosum* ssp. *andigena* contains exactly the same number of homologues. In addition, a 185 bp fragment was used to determine the number of RGHs with an NBS domain. This probe reached 73 bp into the NBS domain and 70 bp into the LRR domain and gave identical results as the LRR probe. These data show that all the *Gpa2*/*Rx1* homologues identified in the PCR based method also possess at least part of an NBS domain similar to *Gpa2* and *Rx1*.

The total number of *Gpa2*/*Rx1* homologues in the potato genome was also determined by southern blotting on genomic DNA of SH and RH. Hybridisation with the LRR and NBS probe confirmed our previous data and resulted in the detection of eight bands for SH and six for RH. All the bands for SH were also present in the BAC clones harbouring *Gpa2/Rx1* homologues. The observation that Southern blotting resulted in the same number of RGHs as the PCR approach indicates that no additional, slightly modified RGHs are present in RH.

Discussion

In this paper, we describe the identification and mapping of nine resistance gene homologues (RGHs) of a single complex locus in four homologous chromosomes using a cluster-specific primer combination based on the LRR domain of *Gpa2* and *Rx1*. Eight of the nine RGHs could be mapped on chromosome XII in a single step procedure using a modified AFLP method. Physical mapping revealed that the remaining homologue *(SH-RGH7)* was also located in the *Gpa2/Rx1* cluster on chromosome XII. The segregation of all six RH-RGHs in the mapping population shows that the two haplotypes of RH are heterozygous at the *Gpa2/Rx1*. This is conformed by the fact that all six RH-RGHs have different sequences. Only two pairs of RGHs are (virtually) identical: SH-RGH5 and RH-RGH4 (99.9%) and SH-RGH6 and RH-RGH5 (100%). In both cases these identical RGHs are derived from different genotypes.

Gpa2 and *Rx1* are highly homologous. At the amino acid level, they have a homology of 88% and at the nucleotide level their identity is even 93% (Van der Vossen et al., 2000). *SH-RGH1*, an RGH present on the same haplotype as *Gpa2* and *Rx1*, has also a similar sequence identity to the LRR domains of *Gpa2* and *Rx1,* respectively 93.6 and 93.5%. A complete open reading frame was detected for this putative *R* gene based on the 187 kb sequence of this region (Van der Vossen et al., 2000). This indicates that *SH-RGH1* could be a functional homologue of *Gpa2* and *Rx1* with unknown specificity. The nine RGHs identified in this study are all closely related to *Gpa2* and *Rx1* with sequence identities ranging from 93 to 95%. Southern analysis showed that all these RGHs have at least part of an NBS domain.

Our results indicate that all homologues are located in the *Gpa2/Rx1* cluster on chromosome XII. Surprisingly, in *Solanum aucaule* a functional *Gpa2*/*Rx1* homologue with the same specificity as *Rx1* is found (Bendahmane et al., 2000). This gene (*Rx2*), however, is genetically linked to RFLP marker Gp21 on chromosome 5 (Ritter et al., 1991); (Bendahmane et al., 2000). Sequence identity between *Rx1* and *Rx2* is so high (97,9% for the complete genes and 99.4% for the LRR domains) that another *Rx1*/*Gpa2*-like cluster on chromosome V, if present in SH or RH, would certainly have been identified in this study. An explanation for this remarkable phenomenon that highly homologous genes are located on different chromosomes in two closely related species could be a recent translocation event after the speciation of *S. tuberosum* and *S. aucaule*. However, the synteny between the more distantly related species *S. tuberosum* and *Lycopersicon esculentum* is very high (Grube et al., 2000) indicating that such translocation events are rare within the genus *Solanum*.

In this paper, a PCR based method has been used to identify *Gpa2/Rx1* homologues. Remarkably, some of the amplified RGHs gave conflicting results, among others, with regard to physically mapping. For example, PCR analyses of nearly completely overlapping BAC clones (25G18 and 36G3) resulted in totally different RGHs. Sequence alignments of these RGHs with other RGHs (two by two) revealed that they were the result of a chimaeric PCR product derived from two distinct RGHs. These artificial RGHs consisted of two extraordinary stretches of several hundred nucleotides each. One stretch was 100%

identical to one RGH and the other stretch was 100% identical to another, completely different RGH. These results were obtained with BAC DNA and genomic DNA as well. Fortunately, chimaeric PCR products were typically less frequent than genuine RGHs and occurred in most cases only once in a series of reamplified colonies. Although PCR techniques are commonly used to identify and map RGHs, this phenomenon has to our knowledge not been described before in the literature.

Despite the importance of potato as a food crop and its vulnerability to various pests and diseases only few *R* genes have been cloned. To facilitate cloning genes of interest an Ultra High Dense genetic map comprising 10 000 AFLP markers has been constructed (www.dpw.wag-ur.nl/uhd/ (Van Os et al.)). This has been accomplished by using the mapping population of SH and RH. However, not for all potato species that harbour interesting *R* genes a dense genetic map and BAC library will become available. The production of these tools is still laborious and costly, and the possibility to dissect *R* gene clusters with specific primer combination in different species is a promising alternative. Comparative analysis has shown that the genomes of members of the *Solanaceae* family have a large synteny (Grube et al., 2000). The results described in this paper indicate that characterising *R* genes from other potato (sub)species with PCR-based approaches may be feasible.

Material and Methods

Plant material and DNA extraction

A mapping population for *Solanum tuberosum* ssp. *tuberosum* was available from the diploid potato clones SH83-92-488 × RH89-039-16 (Rouppe van der Voort et al., 1997). The female parent SH83-92-488 contains an introgression segment originating from the wild accession *Solanum tuberosum* ssp. *andigena* CPC1673 on which the *Globodera pallida* resistance gene *Gpa2* and the PVX resistance gene *Rx1* are located. The male parent RH89-039-16 has been selected for its fertility and the production of vigorous offspring. SH83-92-488 will be referred to as SH and RH89-039-16 will be referred to as RH. The mapping population F_1 SH × RH consists of 136 vigorous F1 genotypes. Genomic DNA was extracted from frozen leaf tissue of *in vitro* plants as described (Van der Beek et al., 1992).

BAC library and DNA extraction
The construction of the BAC library from the diploid potato clone SH83-92-488 has been described (Rouppe van der Voort et al., 1999). The library has been extended with 30,000 clones and now comprises 90,000 clones. Pooling and preparation for screening was done as described (Kanyuka et al., 1999) and resulted in 255 plate pools. DNA extracted from these BAC clones was used as template for PCR and sequence analysis. For this, clones were cultured overnight at 37°C in 500 ml LB medium supplemented with 12.5 mg/ml chloramphenicol for selection. Plasmid DNA was isolated using the "very low copy plasmid DNA purification protocol" of the plasmid midi kit according to manufacturers instructions (Quiagen, Hilden, Germany). For each BAC, 1 µl of DNA was digested with *Hind*III to check the concentration and the purity.

PCR analysis
A cluster-specific primer pair was designed based on conserved DNA sequences in the LRR domains of *Gpa2*, *Rx1* and *SH-RGH1* (previously designated as *RGC1*(Van der Vossen et al., 2000)). The forward primer (LRR-F: ttg gtg tcg taa cag tga gg) starts at position +1533 of *Gpa2* and the reverse primer (LRR-R: ctg gct agt cct cag aac ac) at position +3192 of *Gpa2* (UTR). A PCR was performed with this primer pair using the Expand High Fidelity PCR System according to manufacturer instructions (Roche, Mannheim, Germany). The following PCR cycle file was applied: 3 min 94°C followed by 30 s 94°C, 30 s 55°C, 90 s 72°C for 10 cycles, 30 s 94°C, 30 s 55°C, 90 s 72°C with an extension of 5 s/cycle for 25 cycles and finally 5 min elongation at 72°C.

Cloning and digestion analysis of the PCR products
PCR products were cloned into the PCR2.1TOPO-vector for transformation of *E. coli* TOP10 cells according to the manufacturer instructions (Invitrogen, San Diego, CA, USA). For each PCR product, 12 positive clones were cultivated overnight at 37°C in 2 ml liquid LB medium with 100 µg/ml ampicillin for selection. Each clone was reamplified with the same LRR primer pair and the resulting PCR products were digested with *Taq*I. PCR was performed using an adjusted PCR buffer containing 100 mM TrisHCl pH 8.3, 500 mM KCl, 25 mM MgCl, 10% Triton X-100 to avoid additional cleaning steps of the PCR products prior to digestion. The DNA fragments were separated on a 4% agarose gel consisting of 1% ultra pure agarose (Life Technologies, Breda, The Netherlands) and 3% NuSieve ® GTG ® agarose (FMC, Philadelphia, PA, USA) in 1× TAE buffer at 120V. Clones were selected for further analysis based on differences in the digestion patterns.

Sequence analysis
Sequence analysis of the PCR products was carried out by Greenomics, PRI, Wageningen, The Netherlands. BAC-end sequencing was performed using approximately 1 µg template DNA in a cycle sequencing reaction using either 100 ng sp6 or t7 primer, 8 µl Big Dye terminator mix (PerkinElmer, Wellesley, MA, USA) in a total volume of 20 µl. The PCR protocol consisted of 25 cycles of 30 s 96°C, 15 s 50°C and 4 min 60°C. After ethanol precipitation at room temperature for 10 min and recovery, the labelled DNA was dissolved in 3 µl formamide. DNA was then heated for 2 min at 96°C and directly cooled on ice. Approximately 1.5 µl was then loaded on a 6% TBE (pH 8.3) polyacrylamide gel.

Sequence electrophoresis was carried out on either an ABI 373XL or ABI 377 sequencer. DNA sequence analysis and comparisons were carried out using DNAstar™ software.

CHEF gel electrophoresis
BAC insert sizes were determined with CHEF electrophoresis on a 1% agarose gel (Seakem ® Gold, FMC, Philadelphia, PA, USA) in 0.5×TBE buffer at 4°C using a BIO-RAD CHEF DR II system (Bio-Rad Laboratories, Hercules, CA, USA) at 200 V with a pulse time of 5 to 15 s for 18 h.

Southern analysis
Approximately 20 ng BAC DNA and 3 µg genomic DNA was digested with *Hind*III. The digested DNA was separated on a 1% agarose (Agarose NA, Pharmacia, Peapack, NJ, USA) gel in 1×TAE buffer at 50V. The gels were blotted overnight on hybridisation filters using capillary forces (Sambrook et al., 1989). Southern analysis was performed based on the DIG Application Manual for filter hybridisation (Roche, Basel, Switzerland). An LRR based probe (position +1533 to position +1936 (403 bp)) and an NBS and LRR based probe (position +1367 to position +1552 (185 bp)), were amplified from *Gpa2* BAC DNA. Primer pairs used are PLRR-F (ttg gtg tcg taa cag tga gg) and PLRR-R (gtt ctc tgt agg ctc atg ac) at an annealing temperature of 60°C, and PNBSLRR-F (gtg gaa tgc atg atg tga cc) and PNBSLRR-R (ctc act gtt acg aca cca ag) at an annealing temperature of 55°C. The results were visualised on an autoradiogram after 5 min up to 2 hours exposure.

Genetic mapping
Mapping was performed using a modified method (Fig. 1) based on the AFLP technique (Vos et al., 1995). The template used was the PCR product generated from gDNA using the iPM4 primers as described (Kanyuka et al., 1999). The PCR product was digested with *Taq*I and a *Taq*I adapter as described (Vos et al., 1995) was ligated to the digestion products. This was followed by fragment amplification on 1:25 diluted template using a labelled *Taq*I primer (gat gag tcc tga ccg a (Vos et al., 1995)) and the original iPM4R primer. The AFLP thermal cycle conditions were applied as described (Vos et al., 1995). Separation of labelled fragments and autoradiography was done as described (Van Eck et al., 1995).
Linkage analysis of pair wise recombination frequencies between segregating RGHs and markers were performed using JOINMAP 1.4 (Stam, 1993).

Physical mapping
BAC-end sequences were employed to design PCR primers to study overlap between BAC clones. The primers designed for each BAC-end are listed in Table 1. PCR conditions used in the amplifications are equal for all primer pairs and are as follows: 3 min 94°C, followed by 30 s 94°C, 30 s 60°C, 90 s 72°C for 35 cycles and 5 min 72°C.
For the alignment of the contigs in SH, we used primers as described (Kanyuka et al., 1999) and 187 kb sequence information derived from 4 overlapping BAC clones harbouring *Gpa2*, *Rx1, SH-RGH1* and *SH-RGH3* (previously *RGC3*) (Van der Vossen et al., 2000).

Acknowledgements This research is financially supported by the Netherlands Technology Foundation (STW) and the European Union. (FAIR5-PL97-3565)

References

Ballvora, A., Ercolano, M.R., Weiss, J., Meksem, K., Bormann, C.A., Oberhagemann, P., Salamini, F. and Gebhardt, C. (2002) The *R1* Gene for Potato Resistance to Late Blight (*Phytophthora Infestans*) Belongs to the Leucine Zipper/Nbs/Lrr Class of Plant Resistance Genes. Plant J 30, 361

Bendahmane, A., Kanyuka, K. and Baulcombe, D.C. (1997) High-Resolution Genetical and Physical Mapping of the *Rx* Gene for Extreme Resistance to Potato Virus X in Tetraploid Potato. Theor Appl Genet 95, 153

Bendahmane, A., Kanyuka, K. and Baulcombe, D.C. (1999) The *Rx* Gene from Potato Controls Separate Virus Resistance and Cell Death Responses. The Plant Cell 11, 781

Bendahmane, A., Querci, M., Kanyuka, K. and Baulcombe, D.C. (2000) *Agrobacterium* Transient Expression System as a Tool for the Isolation of Disease Resistance Genes: Application to the *Rx2* Locus in Potato. Plant J 21, 73

Flor, .H.H. (1971) Current Status of the Gene-for-Gene Concept. Ann Rev Phytopathol 9: 275

Gebhardt, C. and Valkonen J.P.T (2001) Organization of Genes Controlling Disease Resistance in the Potato Genome. Ann Rev Phytopathol 39, 79

Grube, R.C., Radwanski, E.R. and Jahn M. (2000) Comparative Genetics of Disease Resistance within the *Solanaceae*. Genetics 155, 873

Kanyuka, K., Bendahmane, A., Rouppe van der Voort, J.N.A.M., Van der Vossen, E.A.G. and Baulcombe, D.C. (1999) Mapping of Intra-Locus Duplications and Introgressed DNA: Aids to Map-Based Cloning of Genes from Complex Genomes Illustrated by Physical Analysis of the *Rx* Locus in Tetraploid Potato. Theor Appl Genet 98, 679

Lawrence, G.J., Finnegan, E.J., Ayliffe, M.A., and Ellis, J.G., (1995) The *L6* Gene for Flax Rust Resistance Is Related to the *Arabidopsis* Bacterial-Resistance Gene *Rps2* and the Tobacco Viral Resistance Gene-N. Plant Cell 7, 1195

Milligan, S.B., Bodeau, J., Yaghoobi, J., Kaloshian, I., Zabel, P. and Williamson, V. (1998) The Root Knot Nematode Resistance Gene *Mi* from Tomato Is a Member of the Leucine Zipper, Nucleotide Binding, Leucine-Rich Repeat Family of Plant Genes. The Plant Cell 10: 1307-1319

Mindrinos, M., Katagiri, F., Yu, G.L. and Ausubel, F.M. (1994) The *A. thaliana* Disease Resistance Gene *Rps2* Encodes a Protein Containing a Nucleotide-Binding Site and Leucine-Rich Repeats. Cell 78, 1089

Ritter, E., Debener, T., Barone, A., Salamini, F. and Gebhardt, C. (1991) Rflp Mapping on Potato Chromosomes of 2 Genes-Controlling Extreme Resistance to Potato Virus-X (Pvx). Mol Gen Genet 227, 81

Rouppe van der Voort, J., Wolters, P., Folkertsma, R., Hutten, R., Van Zandvoort, P., Vinke, H., Kanyuka, K., Bendahmane, A., Jacobsen, E., Janssen, R. and Bakker, J. (1997) Mapping of the Cyst Nematode Resistance Locus *Gpa2* in Potato Using a Strategy Based on Comigrating Aflp Markers. Theor Appl Genet 95, 874

Rouppe van der Voort, J., Kanyuka, K., Van der Vossen, E., Bendahmane, A., Mooijman, P., Klein-Lankhorst, R., Stiekema, W., Baulcombe, D. and Bakker, J. (1999) Tight Physical Linkage of the Nematode Resistance Gene *Gpa2* and the Virus Resistance Gene *Rx* on a Single Segment Introgressed from the Wild Species *Solanum tuberosum* Subsp. *andigena* Cpc 1673 into Cultivated Potato. Mol Plant-Microbe Interact 12, 197

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Moleculal Cloning: A Loboratory Manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York

Stam, P. (1993) Construction of Integrated Genetic-Linkage Maps by Means of a New Computer Package - Joinmap. Plant J 3, 739

Takken, F.L.W. and Joosten, M. (2000) Plant Resistance Genes: Their Structure, Function and Evolution. Eur J Plant Pathol 106, 699

Van der Beek, J.G., Verkerk, R., Zabel, P. and Lindhout, P. (1992) Mapping Strategy for Resistance Genes in Tomato Based on Rflps between Cultivars: *Cf9* (Resistance to *Cladosporium fulvum*) on Chromosome 1. Theor Appl Genet 84, 106

Van der Vossen, E., Rouppe van der Voort, J., Kanyuka, K., Bendahmane, A., Sandbrink, H., Baulcombe, D., Bakker, J., Stiekema, W. and Klein-Lankhorst, R. (2000) Homologues of a Single Resistance-Gene Cluster in Potato Confer Resistance to Distinct Pathogens: A Virus and a Nematode. Plant J 23, 567

Van Eck, H.J., Rouppe van der Voort, J., Draaistra, J., Van Zandvoort, P., Van Enckevort, E., Segers, B., Peleman, J., Jacobsen, E., Helder, J. and Bakker, J. (1995) The Inheritance and Chromosomal Localization of Aflp Markers in a Non-Inbred Potato Offspring. Mol Breed 1, 397

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Vandelee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995) Aflp - a New Technique for DNA-Fingerprinting. Nuc Acid Res 23, 4407

Vos, P., Simons, G., Jesse, T., Wijbrandi, J., Heinen, L., Hogers, R., Frijters, A., Groenendijk, J., Diergaarde, P., Reijans, M., Fierens-Osterenk, J., De Both, M., Peleman, J., Liharska, T., Hontelez, J. and Zabeau, M. (1998) The Tomato *Mi-1* Gene Confers Resistance to Both Root-Knot Nematodes and Potato Aphids. Nature Biotech 16, 1365

Whitham, S., Dineshkumar, S.P., Choi, D., Hehl, R., Corr, C. and Baker, B. (1994) The Product of the Tobacco Mosaic-Virus Resistance Gene-*N* - Similarity to Toll and the Interleukin-1 Receptor. Cell 78, 1101

Chapter 3

Functional orthologues in distantly related *Solanum* species point out the long lifespan of the *Rx* gene

Patrick Butterbach, Erin Bakker, Edwin van der Vossen, Marjon Arens, Erik Slootweg, Jan Roosien, Geert Smant, Jaap Bakker and Aska Goverse

---

to be submitted

Abstract:

A survey of 51 accessions across 35 species from the genus *Solanum* resulted in the identification of two novel functional orthologues of the disease resistance gene *Rx1* in the wild potato species *S. lepthophyes/sparsipilum* and *S. albicans*, *Rx3* and *Rx4* respectively, conferring extreme resistance to potato virus X (PVX). Functional analysis of *Rx3* and *Rx4* resulted in a specific hypersensitive response in an agroinfiltration assay on leaves of *N. benthamiana* in the presence of the elicitor of *Rx1*-mediated resistance and an extreme resistance response was observed in a gene-for-gene-specific manner upon PVX infection of the corresponding *S. lepthophyes/sparsipilum* and *S. albicans* genotypes. Our findings of *Rx3* and *Rx4,* together with *Rx1* from *S. tuberosum* ssp. *andigena* and *Rx2* from *S. acaule*, represents the first example of functional orthologues from the class of NBS-LRR resistance genes, in four different plant species. Sequence comparison between the functional *Rx* orthologues *Rx1-4* and their paralogues enabled us to design *Rx* specific primers for the screening of an additional set of about 5000 genotypes derived from 1000 accessions across 200 *Solanum* species. This resulted in the identification of 14 additional distantly related *Solanum* species harbouring *Rx* candidate genes, which shows that the *Rx* orthologues are of ancient origin derived from one ancestral gene and that *Rx* mediated PVX recognition arose before the groups of *Tuberosa* (Peru), *Acaule/Demissum* diverged. The remarkable high similarity (>97%) of the LRR domain between the *Rx* orthologues suggest that strong evolutionary constraints affect this part of the protein in order to maintain the *Rx* recognition specificity.

Introduction

Plants have developed a unique genetic surveillance system to detect and to neutralize foreign invaders. This system is based on single dominant resistance genes (*R* genes) encoding proteins that upon recognition of the corresponding avirulence gene (*avr* gene) product from the pathogen activate a defense response in the plant to prevent further spreading of the invader. A single genetic factor that recognizes, directly or indirectly, a genetic factor in the pathogen is known as the gene-for-gene concept (Flor, 1971).

The evolution of gene-for-gene interactions is often compared with an

evolutionary 'arms race' between a parasite and a host plant (Dawkins and Krebs, 1979). The 'arms-race'-theory assumes that the parasite is always capable to overcome resistance by developing a new way to circumvent recognition by the host. To counter this, the host population constantly has to evolve new recognition specificities in resistance genes to match the virulent pathotypes. Driven by the emergence of new virulence genes, defeated *R* gene alleles rapidly decline in natural host plant populations. Evidence to support the arms race theory is found in the organization of *R* genes in multigene families. Through gene duplication events *R* gene homologues occur at close physical proximity to the parent sequence and generate clusters of closely related *R* gene sequences in the genome (Baumgarten et al., 2003). In addition, the majority of *R* genes are comprised of arrays of hypervariable potential ligand-binding sites (Caicedo et al., 2004). A combination of different mechanisms that create variation (*e.g.* point mutations), brings the *R* genes under diversifying selection resulting in the continuous generation of novel recognition specificities (Parniske et al., 1997). The consequence is the evolution of *R* genes in a so-called birth-and-death process (Michelmore and Meyers, 1998). Analysis of sets of *R* gene homologues indeed revealed a pattern of sequence divergence for multiple clusters of *R* genes in *Arabidopsis thaliana* (Mondragón-Palomino et al., 2002).

More recent, the 'trench warfare'-model is proposed to explain the dynamics of *R* genes and *Avr* genes as repeated advances and retreats of resistance and virulence alleles (Frank, 1992; Stahl *et al.*, 1999). The trench warfare model predicts that resistance alleles and recognition specificities are maintained as balanced polymorphisms in natural populations over long periods of time. The origin of *R* gene alleles with a long life-span could eventually predate host speciation events resulting in recognition specificities simultaneously operating in distantly related species. Evidence for the trench warfare model is found at the simple locus of *Rpm1* in *Arabidopsis thaliana*, an NBS-LRR class gene conferring resistance to *Pseudomonas syringae*. A homologue of the *Rpm1* gene was isolated from *A. lyrata*, which is also resistant to *P. syringae* (Stahl *et al.*, 1999). Sequence comparison with homologues from resistant and susceptible *A. thaliana* accessions suggests a long-term maintenance of resistant and susceptible haplotypes of *Rpm1*. Another example of ancient *R* gene specificities are *Pto* from *Lycopersicon pimpinellifolium* and *LhirPto* from *L. hirsutum* (Riely and Martin, 2001). It has been shown that these highly similar orthologues encode for a protein kinase that in both cases elicits a hypersensitive response

when co-expressed with the avirulence factor from *P. syringae*. For the *Cladosporium fulvum* resistance genes *Cf-4* and *Cf-9* from *Lycopersicon pimpinellifolium,* several functional orthologues were found in five different wild species of tomato (Kruijt *et al.*, 2005). The nearly identical orthologues are believed to be derived from a common ancestor within the genus *Lycopersicon*.

In potato, about 30 *R* genes against nematodes, viruses and fungi have been mapped in various *Solanum* species (Gebhardt and Valkonen, 2001) of which many are located in clusters on the genome. Two *R* genes from a single cluster in *S. tuberosum* ssp. *andigena* confer resistance to taxonomically unrelated pathogens. *Rx1* and *Gpa2* are both located in a single cluster of about 110 kb on chromosome XII in potato and share an overall homology in the amino acid sequence of about 87%. However, *Rx*1 results in extreme resistance to PVX, whereas *Gpa2* gives a much slower and milder resistance response to the potato cyst nematode *Globodera pallida* (Van der Vossen et al., 2000). The cluster containing *Rx1* and *Gpa2* has been introgressed into cultivated potato from *S. tuberosum* ssp. *andigena* (Bendahmane et al., 1999) and recently, the functional orthologue *Rx2* was isolated from *S. acaule*, a wild relative of cultivated potato, using the PVX coat protein as an elicitor in a candidate gene approach (Bendahmane et al., 2000).

Here, we present a survey of 51 accessions from 35 wild species across *Solanum* that were screened for the presence of functional alleles of *Rx*. This resulted in the identification of two novel genes *Rx3* from *S. lepthophyes/sparsipilum* (CGN20619) and *Rx4* from *S. albicans* (CGN 18029). Together with *Rx1* from *S. tuberosum* ssp. *andigena* and *Rx2* from *S. acaule*, they form the first example of *R* genes that belong to the major class of NBS-LRR resistance genes, for which a set of four functional alleles from different species is known. Allelic mining of about 1000 accessions comprising approximately 200 *Solanum* species resulted in the identification of the homologous gene Rx_{edn} from *S. edinense* (PI 607474), a non-functional allele that still harbours the *Rx* recognition specificity, and an additional set of *Rx* candidate genes in 14 *Solanum* species. The presence of identical recognition specificities (orthologues) of *Rx* in a set of 19 distinct *Solanum* species provides more insight in the distribution and evolution of the virus resistance gene *Rx* in *Solanaceae*. Our findings on the molecular evolution of *Rx* homologues are discussed in the context of the two prevailing models for the co-evolution of resistance and virulence alleles.

Results

Selection of accessions harbouring a functional *Rx* orthologue

To identify functional alleles of the potato disease resistance gene *Rx1* in *Solanum*, 51 accessions representing 35 wild *Solanum* species were selected from the germplasm collection of the Centre for Genetic Resources (CGN), Wageningen, the Netherlands. The selection was based on available phenotypic data on PVX resistance and on species diversity. For each accession, five genotypes were screened for the presence of a functional *Rx* orthologue.

In a detached leaf assay and a greenhouse plant assay we employed agroinfiltration with constructs harbouring the coat protein CP106 (avirulent) and CP105 (virulent) from PVX (Bendahmane *et al.*, 1995). An infiltration with an *A. tumefaciens* culture without any construct was used as a negative control. A resistant genotype from the diploid potato clone of *S. tuberosum* ssp. *andigena* (SH) (Van der Voort *et al.*, 1999) harbouring the *Rx1* gene (Bendahmane et al., 1999) as well as a susceptible genotype of the diploid potato clone *S. tuberosum* ssp. *tuberosum* (RH) (Van der Voort *et al.*, 1999) were included as controls. Transient expression of CP106 resulted in a specific hypersensitive response (HR) in the resistant genotype SH at the infiltration spot 5 days post infiltration (Fig. 1), while in the susceptible genotype RH no such HR was observed. Combining detached leaf assay and greenhouse assay infiltrations with the avirulent coat protein construct (Table 1) resulted in the detection of an HR for two genotypes derived from the accessions CGN20619 (*S. lepthophyes/sparsipilum*) and CGN18029 (*S. albicans*), respectively (Fig. 1). Infiltration with the virulent coat protein construct did not result in an HR in any of the plants, suggesting that the observed HR was gene-for-gene specific and mediated by a functional *Rx* homologue.

Table 1.

Genotype screening with the PVX elicitors CP105 and CP106 on detached leafs and on greenhouse grown plants using agroinfiltration assay. SH (resistant) and RH (susceptible) were included as a control for PVX recognition by Rx1.

Yellow	= PVX immunity observed
Blue	= recognition CP105 and CP106 (did not result in recognition by Rx homologues)
Red	= specific recognition CP106 (resulted in cloning of funtional orthologue)

				PVX resistance	PVX resistance observed in other accessions	detached leaf atta (HR/ genotypes)		greenhouse assay (repetitions/ HR observed)				specificity HR
group	species	accession	**CGN**	(CGN)		control	cp106	genotype	control	cpa05	cp106	
Tuberosa (Peru)	*S. tuberosum* ssp. *andigena*	genotype	**SH**	I		0/1	1/1	SH	6/0	6/0	6/6	cp106
Tuberosa (Peru)	*S. tuberosum* ssp. *tuberosum*	genotype	**RH**	S		0/1	0/1	RH	6/0	6/0	6/0	no HR
Conicibaccata	*S. agrimonifolium*	viv	18285	nd		0/5	1/5	1	2/0	2/0.5	2/0	no HR
S.acaule/S.demissum	*S. albicans*	cgn	18029	I		0/5	1/5	5	4/0	4/0	4/4	cp106
Tuberosa (Peru)	*S. andigena/sucrense*	cgn	21355	nd	+	0/5	3/5		nt	nt	nt	no HR
	S. berthaultii	cgn	18118	I	+	0/5	0/5		nt	nt	nt	no HR
	S. berthaultii	viv	20644	S		0/1	0/1		nt	nt	nt	no HR
	S. boliviense	cgn	21316	S		0/5	0/5		nt	nt	nt	no HR
	S. brachistotrichum	viv	17681	nd		0/1	0/1		nt	nt	nt	no HR
	S. brachycarpum	viv	17721	nd		0/1	0/1		nt	nt	nt	no HR
	S. brevicaule	cgn	18030	I		0/5	1/5		nt	nt	nt	no HR
Tuberosa (Peru)	*S. bukasovii*	cgn	17588	nd		0/5	1/5	2	2/0	2/0	2/1	no HR
mexican diploid	*S. bulbocastanum*	cgn	17693	nd		0/5	2/5	2	2/0	2/0	2/0.5	no HR
	S. capsicibaccatum	cgn	18254	nd		0/5	0/5		nt	nt	nt	no HR
H_Bol+Arg	*S. chacoense*	cgn	17898	nd	+	0/5	3/5	1	2/2	2/2	2/2	aspecific HR
H_Bol+Arg	*S. chacoense*	viv	18248	nd	+	0/1	1/1	1	4/4	4/3.5	4/4	aspecific HR
	S. chaparense	viv	18060	nd		0/1	0/1		nt	nt	nt	no HR
	S. circaeifolium ssp. *circaeifolium*	cgn	18133	nd		0/5	3/5	5	2/0	2/0.5	2/0.5	no HR
	S. commersonii	cgn	18328	nd		0/5	0/5	1	2/0	2/0.5	2/0	no HR
S.acaule/S.demissum	*S. demissum*	viv	18313	nd		0/1	0/1		nt	nt	nt	no HR
Etuberosum	*S. etuberosum*	cgn	17714	I?		0/5	4/5	3	2/0	2/0	2/0	no HR
H_brevicaule	*S. gourlayi*	cgn	17592	IS		0/5	3/5	4	2/0	2/0	2/0	no HR
H_brevicaule	*S. gourlayi* ssp. *pachytrichum*	cgn	18102	nd		0/5	1/5		nt	nt	nt	no HR
H_brevicaule	*S. gourlayi* ssp. *vidaurrei*	cgn	18040	I		0/5	0/5		nt	nt	nt	no HR
Tuberosa (Peru)	*S. hondelmannii*	cgn	18182	IS		0/5	2/5	1	2/2	2/1	2/2	aspecific HR
H_brevicaule	*S. incamayoense*	cgn	21320	nd		0/5	3/5	4	4/4	3/4	4/4	aspecific HR

	S. infundibuliforme	cgn	18041	I		0/5	0/5		nt	nt	nt	no HR
No_Rgts	*S. megistacrolobum*	cgn	17725	nd	+	0/5	2/5	3	4/1.5	4/1.5	4/2	aspecific HR
H_brevicaule	*S. microdontum gigantophyllum*	viv	18046	nd		0/1	1/1	1	2/0.5	2/0	2/0	no HR
H_brevicaule	*S. microdontum gigantophyllum*	cgn	22372	nd		0/5	3/5	1	4/0	4/0	4/0	no HR
	S. multidissectum	cgn	17840	nd		0/5	0/5		nt	nt	nt	no HR
	S. neorossii	cgn	18051	nd		0/5	0/5		nt	nt	nt	no HR
H_brevicaule	*S. okadae*	viv	18108	nd		0/1	0/1		nt	nt	nt	no HR
H_brevicaule	*S. okadae*	cgn	18109	nd		0/5	1/5	1	4/2.5	4/3	4/4	aspecific HR
	S. oplocense	cgn	18088	nd		0/5	4/5	1	4/0	4/2	4/4	cp105 + cp106
	S. pampasense	cgn	17738	nd		0/5	5/5	4	2/0	2/1	2/0	no HR
mexican diploid	*S. papita*	viv	17830	S		0/1	0/1		nt	nt	nt	no HR
H_brevicaule	*S. raphanifolium*	viv	17753	nd		0/1	1/1	4	2/0	2/0	2/0	no HR
	S. sanctae-rosae	cgn	20564	nd		0/5	1/5		nt	nt	nt	no HR
Tuberosa (Peru)	*S. sandemanii*	cgn	17600	nd		0/5	3/5	1	2/0	2/0.5	2/0	no HR
Tuberosa (Peru)	*S. sparsipilum*	viv	18154	S		0/1	1/1		nt	nt	nt	no HR
Tuberosa (Peru)	*S. sparsipilum*	viv	18230	VS		0/1	0/1		nt	nt	nt	no HR
Tuberosa (Peru)	*S. sparsipilum*	cgn	17756	nd		0/5	1/5		nt	nt	nt	no HR
Hybrids	*S. sparsipilum/leptophyes*	cgn	20619	IS		0/5	3/5	3	6/1.5	6/1	6/6	weak aspecific; cp106
H_brevicaule	*S. spegazzinii*	cgn	17602	nd	+	0/5	5/5	3	2/0	2/0	2/0	no HR
mexican polyploid	*S. stoloniferum*	cgn	17606	IS		0/5	4/5	3	2/0	2/0	2/0	no HR
Tuberosa (Peru)	*S. sucrense*	viv	18205	nd	+	0/1	0/1		nt	nt	nt	no HR
Tuberosa (Peru)	*S. sucrense*	cgn	20563	nd	+	0/5	3/5	1	2/0	2/0.5	2/1	weak cp105 + c106
H_Bol+Arg	*S. tarijense*	cgn	17760	nd	+	0/5	4/5	1	2/0	2/2	2/2	aspecific HR
H_Bol+Arg	*S. tarijense/berthaultii*	cgn	20636	IS		0/5	2/5		nt	nt	nt	no HR
	S. vernei	cgn	17836	nd	+	0/5	0/5		nt	nt	nt	no HR
S. verrucosum	*S. verrucosum*	viv	17768	nd		0/1	1/1	1	2/0	2/0	2/0	no HR
H_brevicaule	*S. virgultorum*	cgn	17775	nd		0/5	3/5	1	2/0	2/0	2/0	no HR

To confirm PVX resistance in the accessions CGN20619 and CGN18029, a virus resistance assay was performed. Per treatment three *in vitro* plants were inoculated either with the virulent PVX strain HB or with the avirulent strain UK3 (Kavanagh *et al.*, 1992). Systemic spreading of PVX was assessed and the accumulation of virus in the apex was analysed using ELISA. In the susceptible control plant RH, systemic spreading of the virus was detected for both the avirulent and virulent PVX strains, but in the resistant control plant SH virus was only detected in plants inoculated with the virulent strain HB (Table 2). A similar result was observed for the accessions CGN20619 and CGN18029, in which accumulation of virus was detected only in plants inoculated with the virulent strain HB. These data are consistent with the CGN PVX immunity score and the specific HR induced in leaves of CGN20619 and CGN18029 upon expression of the avirulent elicitor CP106 and indicates the presence of a functional orthologue of the *Rx* gene in these accessions from *S. lepthophyes/sparsipilum* and *S. albicans*.

Table 2: Virus resistance assay with selected genotypes of *Solanum* species. Susceptibility was indicated by systemic spreading of the virus, which was detected in the apex by ELISA. The results of the ELISA are shown with the measured OD including the standard deviation. Resistance (-) and susceptibility (+) are detected by an OD lower or higher than 0.04 in the ELISA, respectively.

plant species	PVX HB	OD at 405nm			PVX UK3	OD at 405nm		
S. tuberosum ssp *tuberosum*	+	0.25	+/-	0.02	+	0.32	+/-	0.01
S. tuberosum ssp. *andigena*	+	0.28	+/-	0.04	-	0.02	+/-	0.01
S.lepthophyes/ sparsipilum	+	0.19	+/-	0.04	-	0.01	+/-	0.00
S. albicans	+	0.21	+/-	0.01	-	0.01	+/-	0.01

Identification of the functional *Rx* orthologues *Rx3* from *S. lepthophyes/sparsipilum* and *Rx4* from *S. albicans*

To identify functional orthologues from *Rx1*, homologous genomic fragments were amplified from the selected genotypes from accessions CGN20619 and CGN18029 with cluster-specific primers flanking the start and stop codon sites of *Rx1* (Bendahmane et al., 2000). The PCR products were cloned into pDONR207 in *E. coli* and 39 single clones were screened for each genotype with a second pair of *Rx1* cluster-specific primers based on conserved regions of the LRR domains from *Rx1, Rx2,* and *Gpa2* (Bakker *et al.*, 2003). The amplified PCR product was subsequently digested with *TaqI* to assess the diversity of the resistance gene homologues on the basis of differences in their restriction patterns (Fig. 2). This procedure resulted in the identification of 8 different homologues for CGN20619 and 12 homologues for CGN18029. None of the restriction patterns of the homologues was identical to the *TaqI* restriction pattern of the *Rx1* gene (Bakker *et al.*, 2003).

To determine which of the homologues harbours the same recognition specificity as *Rx1*, the predicted open reading frame of each homologue was cloned into the overexpression vector pk2GW7 and transformed to *A. tumefaciens* for transient expression in plants. A series of co-infiltration assays was performed on leaves of six weeks old plants of *N. benthamiana*. Each homologue was combined with either *A. tumefaciens* without a vector, with CP105 or with CP106. Analogous experiments with *Rx1* and *Rx2* have shown that a combined transient overexpression of a functional *R* gene and its elicitor CP106 in *N. benthamiana* results in a specific and visible HR on the leaf (Bendahmane et al., 1999; Bendahmane et al., 2000). Infiltration of the homologues isolated from CGN20619 resulted in the identification of one homologue, hereafter named *Rx3*, showing a specific HR only in combination with CP106 (Table 3, Fig. 3). Likewise, among the 12 homologues isolated from CGN18029, one showed a specific HR in combination with CP106 (Table 3, Fig. 3) and will from now on be referred to as *Rx4*. No response was observed in leaf tissue infiltrated with *Rx3* or *Rx4* in combination with CP105 or the mock treatment. The transient expression of all other homologues, either alone or in combination with CP105 or CP106, did not show an HR except for one homologue from CGN20619, which led to a constitutive HR (Table 3).

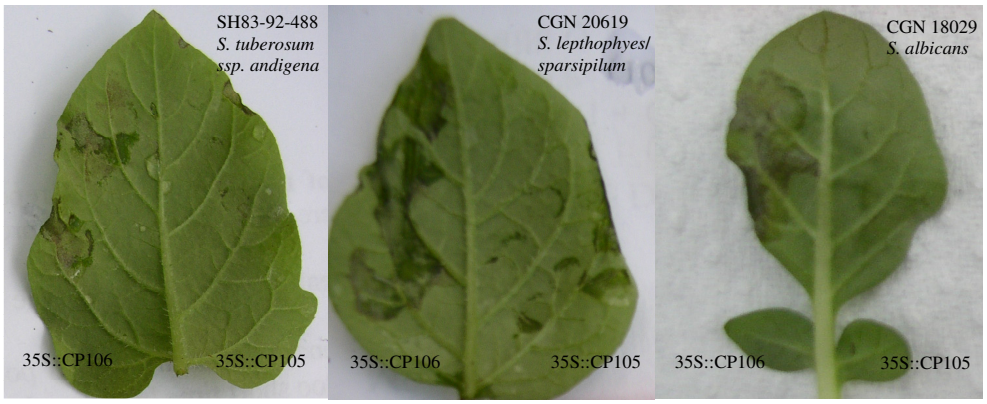


Figure 1: Agroinfiltration of potato accessions with PVX Elicitor constructs 35S::CP105 and 35S::CP106. The pictures show the leaves after 5 dpi. A necrotic lesion is observed on the left sides infiltrated with the elicitor of *Rx*.

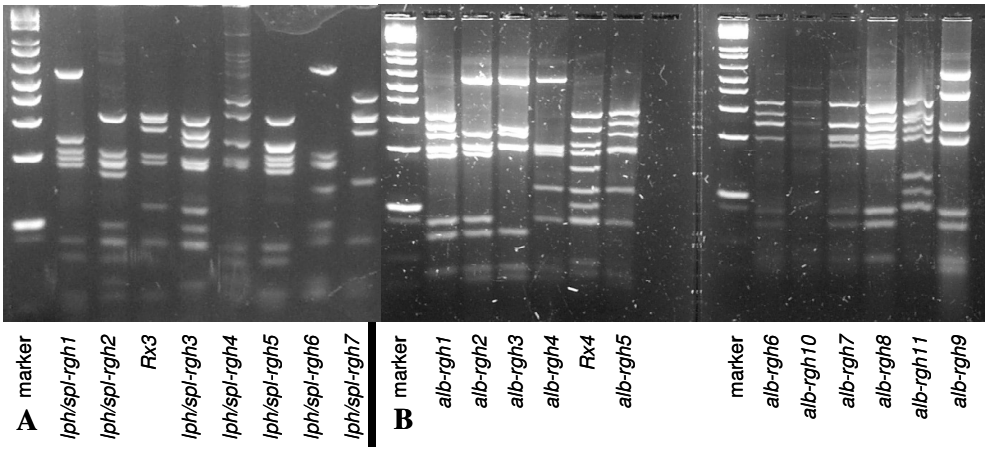


Figure 2: Digestion pattern of *Rx1* homologues from CGN 20619 (A) and CGN 18029 (B) digested with TaqI. A 1kb+ Ladder was used as a marker (marker) for restriction fragment length comparison.

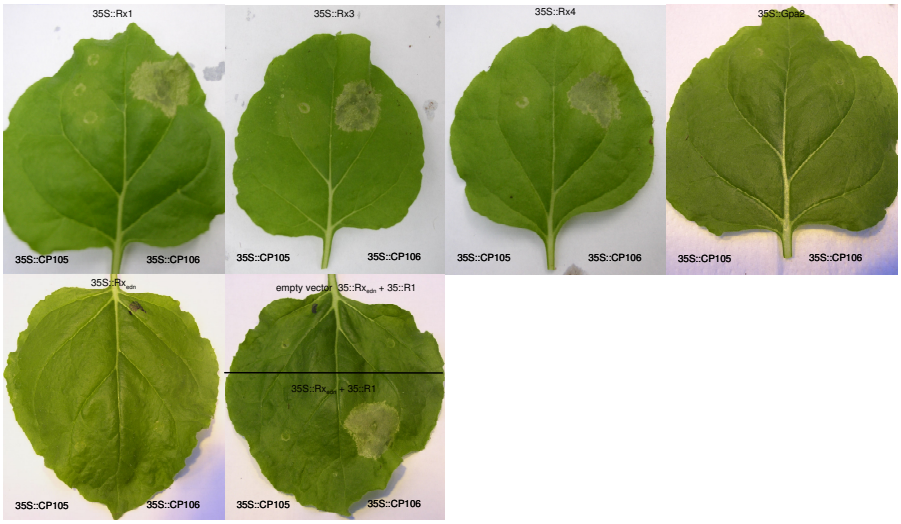


Figure 3: Expression of Rx orthologues, Gpa2 and Avr/vir proteins in *N. benthamiana*. *A. tumefaciens* cultures carrying 35S::CP105 (vir/left) or 35S::CP106 (avr/right) were co-infiltrated with either 35S::Rx1, 35S::Rx2, 35S::Rx4, 35S::Gpa2 (top) or 35S::Rx_{edn} (bottom left) into *N. benthamiana*. For the complementation experiment 35S::CP105 or 35S::CP106 were coinfiltrated with 35S:: Rx_{edn} and 35S::R1 together (bottom right). The leaves are shown 3 days post-infiltration. Cell death developed only in the leaf parts co-infiltrated with the combinations 35S::CP106 and a functional Rx orthologue.

Table 3: Results of the co-expression assay in *N. benthamiana*. *Rx* orthologues isolated from *S. tuberosum* ssp. *andigena*, *S. lepthophyes/sparsipilum*, *S. albicans* and *S. edinense* were coexpressed with either empty vector, CP105 or CP106. Hypersensititive response (HR) was observed after 3 days post-infiltration.

	empty vector	35S::CP105	35S::CP106
35S::Rx1	-	-	HR
35S::Gpa2	-	-	-
35S::lph/spl-rgh1	-	-	-
35S::lph/spl-rgh2	-	-	-
35S::Rx3	-	-	HR
35S::lph/spl-rgh3	-	-	-
35S::lph/spl-rgh4	-	-	-
35S::lph/spl-rgh5	-	-	-
35S::lph/spl-rgh6	HR	HR	HR
35S::lph/spl-rgh7	-	-	-
35S::alb-rgh1	-	-	-
35S::alb-rgh2	-	-	-
35S::alb-rgh3	-	-	-
35S::alb-rgh4	-	-	-
35S::Rx4	-	-	HR
35S::alb-rgh5	-	-	-
35S::alb-rgh6	-	-	-
35S::alb-rgh7	-	-	-
35S::alb-rgh8	-	-	-
35S::alb-rgh9	-	-	-
35S::alb-rgh10	-	-	-
35S::alb-rgh11	-	-	-
35S::Rx_{edn}	-	-	-
35S::Rx_{edn}+35S::R1	-	-	HR

Gene structure of *Rx3* and *Rx4*

Sequencing of *Rx3* and *Rx4* revealed that their lengths from start to stop are 3142bp and 3128bp, respectively. This is in the same range as *Rx1* (3159bp) and *Rx2* (3162bp). In a nucleotide sequence alignment with the sequence of the *Rx1* gene, it was shown that *Rx3* and *Rx4* have a homology with *Rx1* of *98%* and 97% respectively. Using the sequences of *Rx1*, *Rx2*, *Rx3*, *Rx4*, and other homologous genes found in *S. lepthophyes/sparsipilum* and *S. albicans* (data not shown) unique PCR primers were designed to specifically amplify the cDNA sequence of *Rx3* and *Rx4*. Sequence alignments showed that the genes include 3 exons and 2 introns analogous to *Rx1* and *Rx2* (Fig. 4). In *Rx3* Intron I has a length of 234 bp starting at +2712 and ending at +2948. Intron II consists of 95 bp starting at +3075 and ending at +3192. Compared to *Rx1* and *Rx2*, *Rx3* is lacking 22 nucleotides at the N-terminus of Intron II, including the homologous acceptor-site into exon III of *Rx1* and *Rx2*. Instead, an acceptor site 7bp further downstream leads directly into the stopcodon TAA at +3194 (Fig. 4). In *Rx4* Intron I has a length of 234bp starting at +2712 and ending at +2944. Intron II consists of 94bp starting at +3075 and ending at +3185. Due to a deletion that includes the homologous acceptor site position, Intron I ends into a different acceptor site 4bp further upstream. This results in a shortened reading frame ending into the stopcodon TGA at +3017 before Intron II (Fig. 4).

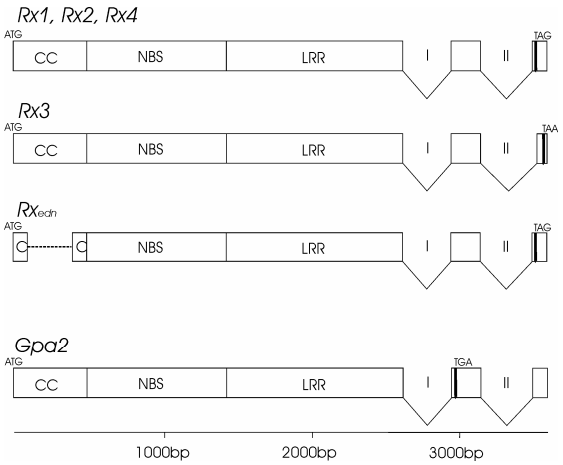


Figure 4. Schematic structural overview of the functional Rx homologues, Rx_{edn} and Gpa2 with their different domains (LZ=leucine zipper, NBS=nucleotide binding site, LRR=leucine rich repeat) The start- (ATG) and stopcodons (TAG, TGA, TAA) are indicated. The horizontal dotted lines indicate sequence gaps.

Table 4. Comparison of functional *Rx* orthologues, Rx_{edn} and *Gpa2* regarding structure of nucleotide and protein sequence. The length is the number of nucleotides in the homologous ORF of *Rx1* including the introns. Displayed are only the amino acid changes on translated protein sequence due to nucleotide substitution in an alignment with *Rx1*. High similarity in nucleotide sequence between the homologues points to a common ancestor gene. In the functional Rx homologues insertions correspond to the reading frame. Decreased changes in the LRR-domain between the functional *Rx* homologues and Rx_{edn} compared to *Gpa2* support the evidence of the LRR domain bearing pathogen-recognition specificity.

	Rx1	*Rx2*	*Rx3*	*Rx4*	Rx_{edn}	*Gpa2*
Origin	*S. andigena*	*S. acaule*	*S.leptophyes/ sparsipilum*	*S. albicans*	*S. edinense*	*S. andigena*
length	3159	3162	3146	3128	2864	3186
translated protein	937	938	937	909	838	912
sequence identity		98%	98%	97%	89%	94%
identity translated Protein		96%	98%	95%	86%	88%
insertions/deletions		+3/0	+3/-23	+3/-35	+5/-300	+20/-57
substitutions		56	20	71	58	266
amino acid changes		40	14	38	21	104
substitutions in domain (aa-change)						
NBS		43 (25)	10(5)	41(25)	43(17)	50(32)
LRR		5(3)	4(3)	7(3)	7(3)	83(57)

Rx3 and Rx4 share a high similarity in their protein sequence

Rx1, *Rx2* and *Gpa2* encode for proteins of 937, 938 and 912 amino acids, respectively (Bendahmane et al., 1999; Bendahmane et al., 2000; Van der Vossen et al., 2000). The predicted open reading frames of Rx3 and Rx4 comprise 936 and 913 amino acids, respectively. Alignment of the two putative amino acid sequences with those of Rx1, Rx2 and Gpa2 resulted in a consensus sequence of 942 amino acids (Fig. 5), to which will be referred to later by pointing out specific amino acid positions. A comparison of the protein sequences of Rx3 and Rx4 reveals a similarity to Rx1 of 98% and 92%, respectively (Table 3). When comparing the translated amino acid sequences of the virus resistance proteins Rx1, Rx2, Rx3 and Rx4 to each other the similarity is always over 90%, whereas similarity to the nematode resistance protein Gpa2 ranges from 84% to 89% (Table 3). The sequences of Rx3 and Rx4 show a high structural similarity to Rx1 and all its known homologues consisting of a CC domain, an NBS domain and an LRR domain.

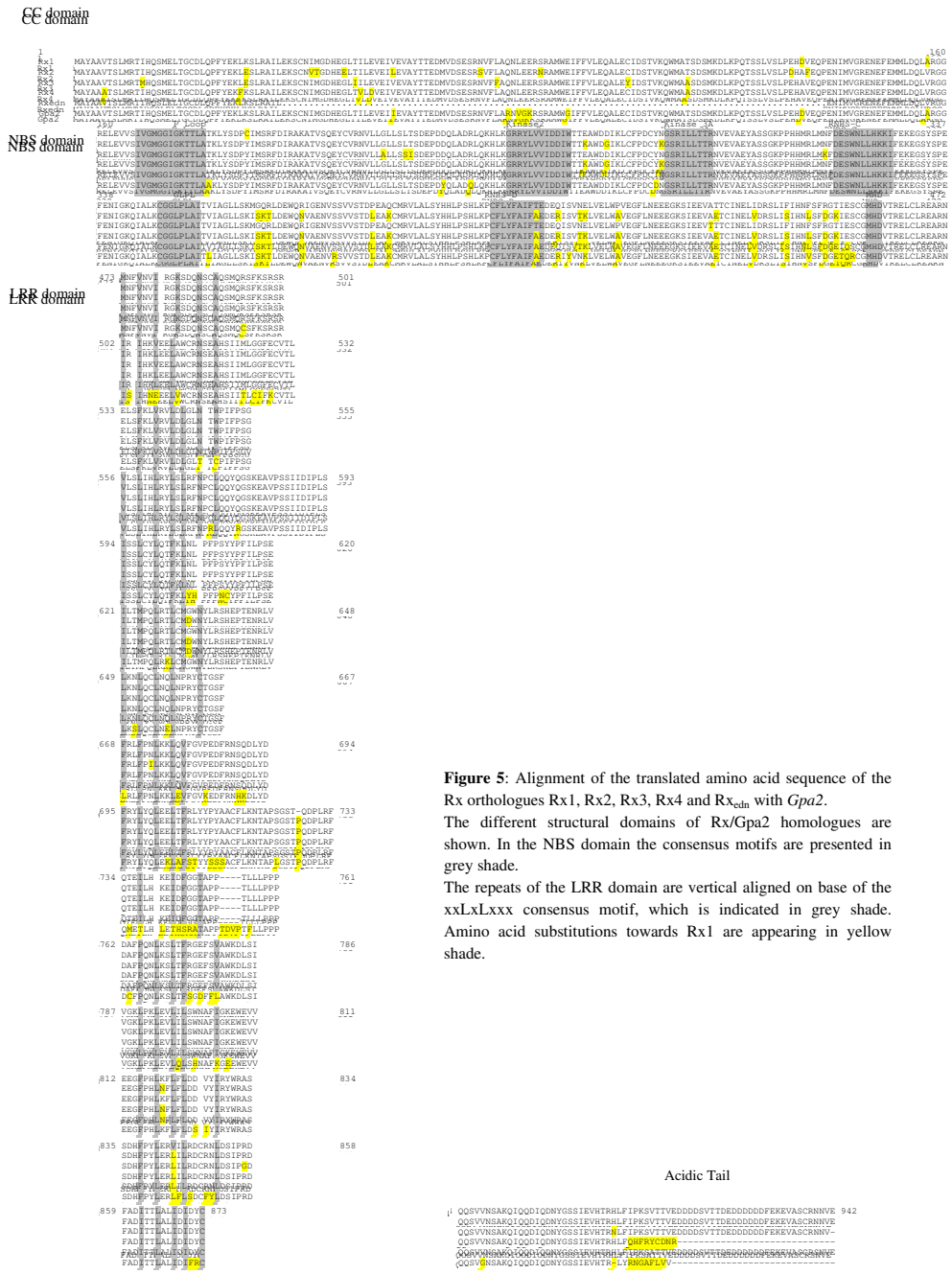


Figure 5: Alignment of the translated amino acid sequence of the Rx orthologues Rx1, Rx2, Rx3, Rx4 and Rx_{edn} with *Gpa2*.
The different structural domains of Rx/Gpa2 homologues are shown. In the NBS domain the consensus motifs are presented in grey shade.
The repeats of the LRR domain are vertical aligned on base of the xxLxLxxx consensus motif, which is indicated in grey shade. Amino acid substitutions towards Rx1 are appearing in yellow shade.

The number of repeats in the LRRs can affect the specificity of resistance genes (Dixon *et al.*, 1998). Similar to Rx1 and Rx2, the predicted proteins of Rx3 and Rx4 both have fifteen of these leucine rich repeats. Within the LRR domain, the solvent exposed amino acids are likely to be involved in interaction (directly or indirectly) with the coat protein of the virus, and therefore these solvent exposed amino acids are predicted to be under strong evolutionary constraints. Comparing the LRR domain of the different *Rx* orthologues (Fig. 5) revealed that the LRR-domain of all four Rx orthologues is identical in 394 out of 401 amino acid positions (98%). In contrast, Gpa2 is identical with all four Rx orthologues in only 332 out of 401 amino acid positions (83%) within the LRR domain.

Table 5. Similarity table of the translated amino acid sequences from Rx1, Rx2, Rx3, Rx4, Rx_{edn} and Gpa2. The comparison of the functional Rx orthologues to each other reveals a similarity ranging from 92%-98%. The similarity to Gpa2 and Rx_{edn} ranges from 84%-89%.

	Rx1	Rx2	Rx3	Rx4	Rx_{edn}	Gpa2
Rx1	100	96	98	92	86	85
Rx2		100	95	94	88	85
Rx3			100	92	85	84
Rx4				100	84	89
Rx_{edn}					100	86
Gpa2						100

The ocurrence of aminoacid insertions and substitutions in the LRR in Rx2, Rx3 and Rx4 and Gpa2 compared to Rx1 was analyzed. Rx3 and Rx4 share with Rx2 and Gpa2 an identical insertion in the LRR-domain resulting in an additional proline at position +727 of the consensus sequence (Fig 5). One substitution is situated at +844 in Rx2, Rx3 and Rx4. Rx2 and Rx4 share three amino acid substitutions at +507, +633 and +819. For Rx3 only two singular amino acid exchanges were detected at +673 and +857. Table 3 shows that Rx2, Rx3 and Rx4 have each three amino acid substitutions in the LRR-domain, while 14-40 amino acid substitutions occur in their whole ORF. In Gpa2 57 out of the total 104 amino acid changes occur within the LRR. Gpa2 has an up to 7 fold increased amino acid substitution rate compared to Rx2, Rx3 and Rx4.

In Figure 5, the LRR-domain is sectioned by aligning the internal structure of the LRR consensus motif showing each repeat of the LRR-domain separately. The proposed LRR-scaffold structure for the translated protein is not affected by amino acid substitutions;

the xxLxLxxx motif is conserved throughout the LRR of all functional Rx orthologues and Gpa2. With one exception, there are no amino acid changes situated within the solvent exposed parts of the LRR domain in the functional Rx orthologues. Rx3 has an aminoacid change from an arginine to a glycine at +857 situated in an exposed part of the LRR according to the alignment.

In R proteins, the NBS domain plays a role in the downstream signalling and activation of the defense response in the plant (van der Biezen and Jones, 1998; Takken et al., 2006). Within the class of NBS-LRR proteins the C-terminal part of the NBS-domain is highly conserved (Aravind et al., 1999); (van der Biezen and Jones, 1998). In the alignment of the protein sequences (Fig. 5) the NBS-domains of Rx2 and Rx4 differ in 15 amino acids from that of Rx1, whereas the number of changes from Gpa2 is twice that number. Rx2 and Rx4 have an identical NBS domain protein sequence. The putative NBS domain in Rx3 shows the highest similarity to Rx1 with only 6 polymorphisms. Several motifs are known within the NBS domain, including the P-Loop, Kinase 2, Kinase 3A, RNBS-C, GLPL, RNBS-D and MHD (Aravind *et al.*, 1999). Figure 6 shows that these motifs are conserved in all Rx homologues, except for the RNBS-D motif in which an arginine is substituted for a threonine in case of Rx2, Rx4 and Gpa2. In the NBS domain, Rx2 and Rx4 share more similarity with Gpa2, while Rx1 and Rx3 are more similar to each other.

The acidic tail structure is dispensable for the function of Rx

Most differences between the Rx orthologues are in the C-terminal end of the sequence coding for an acidic tail motif in Rx1 and Rx2 composed of SVTT(V/D)EDDDD (Bendahmane et al., 1999; Bendahmane et al., 2000). Rx3 also includes an acidic tail, but due to the altered acceptor position at the end of Intron II, Rx3 is lacking the last glutamic acid at position +942 (Fig. 5). Interestingly, Rx4 is lacking the entire acidic repeat structure due to a stopcodon at position +914 but is still able to confer extreme resistance against PVX in a gene-for-gene specific manner as described above. This suggests that this C-terminal tail is not essential for disease signalling or pathogen recognition.

With respect to these findings the relevance of the acidic repeat for the function of the Rx mediated resistance was further investigated. In Rx1 a stop codon was introduced at the N-terminus resulting in a translated amino acid sequence lacking the acidic repeat structure. To determine if the expressed protein from the modified *Rx1* construct has the

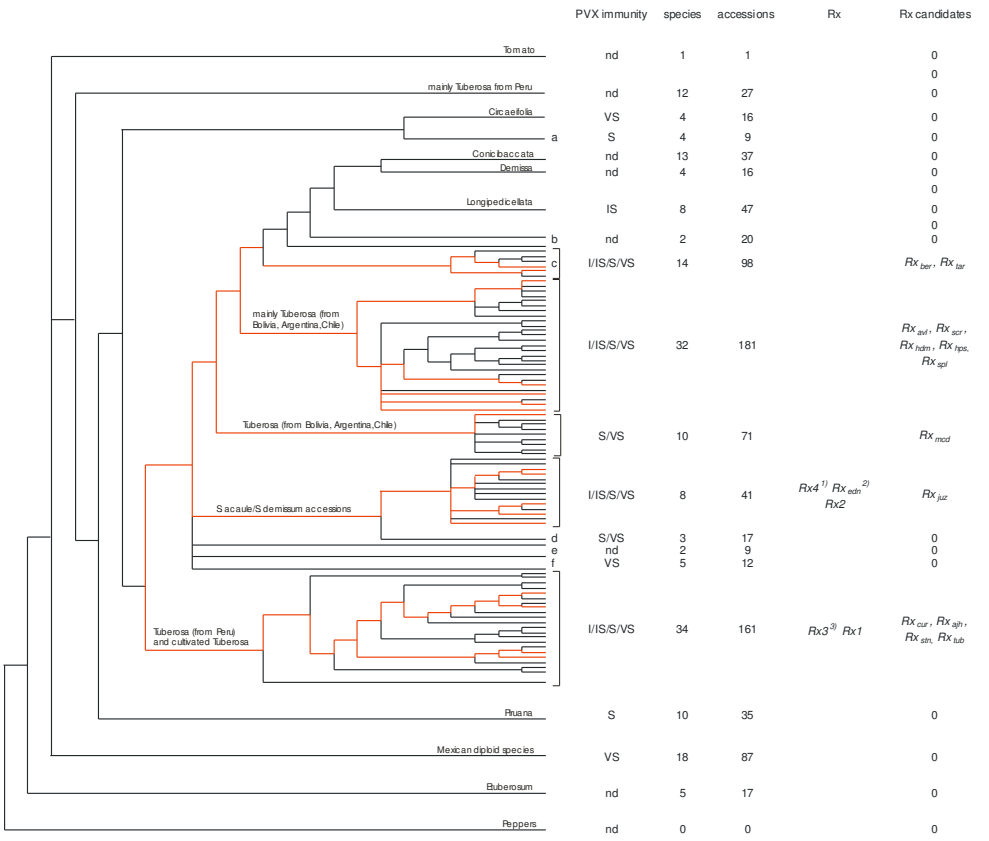
same recognition specificity as *Rx1*, it was cloned into the overexpression vector pBINPLUS and transformed to *A. tumefaciens*. In the coexpression assay of the modified Rx1 construct together with CP106, CP105 or mock in *N. benthamiana* it was shown that the recognition specificity to CP106 and the HR intensity on the leaf was unchanged compared to Rx1 (data not shown). These results support our observation that the presence of the acidic repeat structure in Rx is dispensable for the recognition and the defense signalling function.

Identification and distribution of functional Rx orthologues across *Solanum*

To get a complete overview of the distribution and lifespan of functional *Rx* orthologues across the section *Petota*, 4876 genotypes comprising 1000 accessions derived from 200 species have been screened with a primer pair designed on a specific stretch of sequences that is conserved within the LRR of *Rx1*, *Rx2*, *Rx3* and *Rx4*, but not in other close homologues. This resulted in an amplification product for 237 genotypes comprising 89 accessions from 35 taxa. Sequence analysis revealed that the most variable part of the LRR obtained from genotypes of 14 *Solanum* species was virtually identical to the corresponding region of the functional orthologues *Rx1,2,3* and *4* (~ 98%), whereas the other LRR encoding sequences showed a lower degree of homology to *Rx* like any other *Rx* homologue present in the *Solanum* species tested (~ 94%).

Next, the distribution of (candidate) functional *Rx* orthologues across *Solanum* was determined by placing the 18 corresponding *Solanum* species in a tentative taxonomic tree of this genus based on morphology data, chloroplast DNA sequences and AFLP data (Hawkes, 1990; Van den Berg, 2001); (Spooner and Castillo T, 1997; Kardolus et al., 1998), including tomato and pepper (Fig. 6). This resulted in detection of Rx (candidate) genes in taxa which belong to two groups of *Tuberosa* from Bolivia/Argentina/Chile (*S. avilesii*, *S. hondelmannii*, *S. hoopesii, S. microdontum*, *S. sparsipilum, S. sucrense*) and *Tuberosa* from Peru + cultivated species (*S. ajanhuiri*, *S. curtilobum*, *S. stenotonum*, *S. tuberosum*), a group with various species including *S. berthaulti* and *S. chacoense* (*S. berthaultii*, *S. tarijense*) and to the group of *S. acaule/S.demissum* (*S. juzepczukii*). Two taxa, *S. edinense* from the group of *S. acaule/S. demissum* and S. *huancabambense* from the group *Tuberosa* from Peru + cultivated species, are more far related to the other taxa within their groups. Finally, *Rx* (candidate) genes were detected in *S. fernandezianum* (group

Etuberosum), *S. hougasii* (Mexican Diploid), *S. cardiophyllum* ssp. *ehrenbergii* (Mexican Diploid) which belong to other taxonomic groups of *Solanum*.



a contains *S. sogarandinum, S. scabrifolium* and *S. huancabambense*
b contains *S. verrucosum* and *S. macropilosum*
c contains *S. gandarillasii, S. arnezii, S. setulosistylum, S. ruiz-lealii, S. tarijense, S. berthaultii* and *S. chacoense*
d contains *S. megistacrolobum, S. megistacrolobum* ssp *toralapanum* and *S. sanctae-rosae*
e contains *S. medians* and *S. sandemanii*
f contains *S. violaceimarmoratum, S. buesii, S. santolallae, S. laxissimum* and the hybrid *S. violaceimarmoratum/S. yungasense*
[1]) isolated from *S. albicans* [2]) isolated from *S. edinense* [3])isolated from *S. lepthophyes/sparsipilum*

Figure 7: Schematic representation of the taxonomic relationships of the genus *Solanum* section *Petota* after Jacobs et al. in prepararation, including tomatoes and peppers as derived from Jansky et al., 2006.

For each clade, PVX immunity as was found in some accessions by CGN using disease assays is indicated as I (immune), IS (partly immune), S (susceptible), VS (very susceptible) or nd (not done). In addition, for each clade it is indicated how many species and accessions were screened for presence of candidate *Rx* orthologues and which functional *Rx* orthologues and candidates *Rx* orthologues were found for each clade.

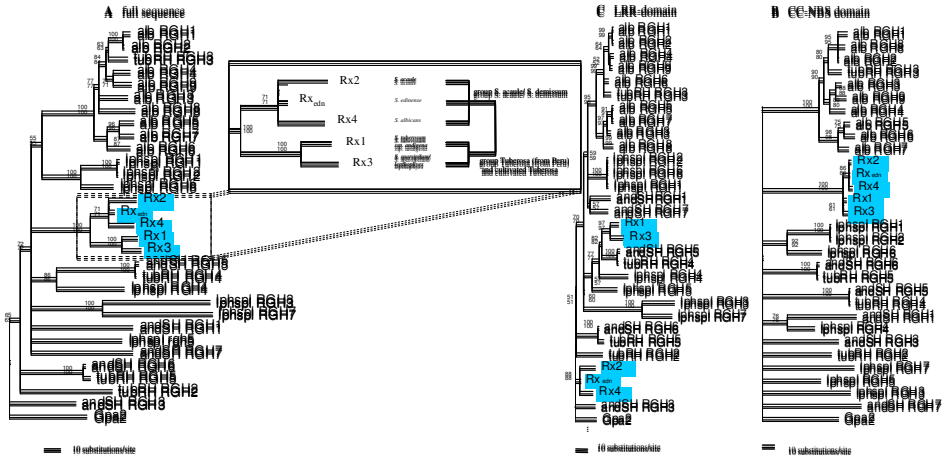


Figure 7. UPGMA distance dendrograms of a nucleotide sequence alignment from the functional *Rx* orthologues, Rx_{edn} and *Gpa2* together with further 7 homologues from *S.lepthophyes/sparsipilum* (*lph rgh1-7*), 9 homologues from *S. albicans* (*alb rgh1-9*), 5 homologues from *S. tuberosum ssp. andigena* (*andSH RGH 3/5/6/7*) and 4 homologues from *S. tuberosum* ssp. *tuberosum* (*tubRH-rgh2/3/4/5*). The position of the functional *Rx* orthologues and Rx_{edn} is indicated in blue. **A**. distance dendrogram of the full length sequence alignment. The clade of functional Rx orthologues and Rx_{edn} is compared with the phylogenetic relationships between the corresponding Solanum species indicated as a schematic tree (Box). The taxonomic scheme is a synthesis of morphological data cpDNA and AFLP based data (Hawkes, J.G. 1990; van den Berg 2001), , (M. Jacobs et al. 2006). **B**. distance dendrogram of the CC and NBS domain sequence alignment **C**. distance dendrogram of the LRR domain sequence alignment

The LRR domain of the non-functional homologue Rx_{edn} still harbours the *Rx* recognition specificity

To find a functional *Rx* orthologue in the most distantly related *Solanum* species, the latter five taxa were selected for further analysis. Therefore, one of the positive genotypes from each of the taxa was used for the isolation and sequencing of the candidate *Rx* orthologues as described for *Rx3* and *Rx4*. Sequence analysis of the complete set of isolated homologues revealed that 4 out of these 5 candidate *Rx* orthologues do not show an increased similarity to *Rx1*, *Rx2*, *Rx3* or *Rx4* than any other *Rx* homologue. The candidate gene Rx_{edn} isolated from *S. edinense* (PI 607474), however, shows a high similarity to the functional *Rx* orthologues (Table 3). The translated amino acid sequence of the LRR domain from Rx_{edn} is identical to that of Rx2, giving strong support of Rx_{edn} harbouring a functional Rx

recognition specificity. However, the functional analysis using the *A. tumefaciens* co-infiltration assay on *N. benthamiana* revealed that in case of Rx_{edn} no response was observed when coinfiltrated with cp105, cp106 or empty vector (Table 2, Fig. 3).

The reason for this could be a deletion of a stretch of 100 amino acids in the CC-domain of the translated protein (Fig. 5). Interestingly, an additionally coinfiltrated CC-domain construct from Rx1 restored the ability to give a specific HR in combination with cp106 in the assay complementing the Rx_{edn} construct (Table 2, Fig. 3). From this we conclude that in *S. edinense* the *Rx* homologue Rx_{edn} still harbours an Rx recognition specificity but lacks the function of the CC domain due to the deletion. This was confirmed in a virus resistance test on greenhouse grown plants from *S. edinense* which were toothpick inoculated with either the virulent PVX amplicon pgR105 or the avirulent PVX amplicon pgR106. For both amplicons, systemic spreading of the virus was observed in the secondary leaves of the inoculated plants showing that *S. edinense* is indeed susceptible to PVX (data not shown).

Rx mediated PVX recognition arose before *Tuberosa* and *Acaulia* diverged

A distance UPGMA dendrogram was built for the complete genes, the conserved CC-NBS domain and the variable LRR domain using the nucleotide sequence alignment of *Rx1*, *Rx2, Rx3*, *Rx4,* Rx_{edn} and *Gpa2* and their corresponding homologues (Fig. 7). Topology of the distance dendrogram obtained for the complete homologues revealed one clade with two subclades containing *Rx1* plus *Rx3* and *Rx2, Rx4* plus Rx_{edn}, respectively (Fig. 7A). Interestingly, the distance dendrogram based on the CC-NBS domains shows the same relationships between the *Rx* orthologues but in two distinct clades (Fig. 7B). In the distance dendrogram based on the LRR domain, it is shown that the *Rx* orthologues harbouring the PVX recognition specificity are grouping in a single clade (Fig. 7C). *Gpa2* is always placed outside the *Rx* clade.

The clade in the distance tree containing the complete functional *Rx* orthologues was aligned with a schematic taxonomy tree of the corresponding *Solanum* species and the two sections resembling the groups S. *acaule/S. demissum* and *Tuberosa* (Peru/cultivated) (Fig. 7A). The comparison of the topology of both trees reveals that the observed division into *Rx* subclades in the dendrogram is corresponding with the division into sections resembling the groups *Tuberosa* and *Acaule/Demissum* of the taxonomy scheme. The

sequence similarities of the functional *Rx* orthologues are corresponding with the phylogenetic relationship of the *Solanum* species from which they are isolated. From this we can conclude that the identified *Rx* orthologues are derived from one ancestral gene and Rx mediated PVX recognition arose before the groups *Tuberosa* (Peru) and *Acaule/Demissum* diverged.

These findings are further supported by the survey across *Solanum* section *Petota* using the primer-based screening, in which both phylogenetic groups are found containing genotypes with an amplification product indicating the presence of *Rx* (Fig. 6). Among them were genotypes from *S. curtilobum*, *S. juzepczukii*, *S. vernei* and *S. sucrense*. It has been found previously, that these species exhibit extreme resistance to PVX analogous to *Rx1* (Querci *et al.*, 1995). The results from the primer-based screening for *Rx* in *Solanum* indicate that functional *Rx* orthologues are also present in several taxa from more distantly related groups within *Solanum* section *Petota* as mentioned above. However, for genotypes from taxa belonging to other phylogenetic groups no amplification product was observed. This indicates, that *Rx* is absent in the groups *Circaeifolia*, *Conicibaccata*, *Demissa*, *Longipedicellata*, *S. verrucosum*, *Piurana* and, after the analysis of the genotypes from *S. hougasii*, *S. ehrenbergii* and *S. fernandezianum*, in the groups of Mexican diploid species and *Etuberosum* as well.

Discussion

In this paper, we present the identification and molecular analysis of two novel functional orthologues of the disease resistance gene *Rx1* from *Solanum tuberosum* ssp. *andigena*. In addition, one orthologue has been identified that resembles the *Rx* specificity, but is not functional due to a deletion in the CC domain. Allelic mining of about 5000 genotypes derived from 1000 accessions comprising 200 *Solanum* species resulted in the identification of *Rx3*, *Rx4* and the truncated Rx_{edn} in the wild potato species *S. lepthophyes/sparsipilum, S. albicans* and *S. edinense*, respectively. Sequence analysis showed that these three genes share high similarity with *Rx1* (98 %, 97% and 86% at the nucleotide level). Functional analysis of *Rx3* and *Rx4* resulted in a specific hypersensitive response in an agroinfiltration

assay on leaves of *N. benthamiana* in the presence of the coat protein from the avirulent PVX strain, which is the elicitor of *Rx1*-mediated resistance, while Rx_{edn} only showed a HR when complemented (*in trans*) with the CC domain from *Rx1*. The resistance gene function of *Rx3* and *Rx4* was shown by the induction of an extreme resistance response in a gene-for-gene manner upon PVX infection of the corresponding genotypes from *S. lepthophyes/sparsipilum* and *S. albicans*. Together with the *Rx2* gene, a functional homologue of *Rx1* which was derived from *S. acaule* (Bendahmane et al., 2000), Rx_{edn}, *Rx1, Rx3* and *Rx4* form a unique dataset of four *R* gene orthologues from the NBS-LRR class of resistance genes harbouring the same recognition specificity and derived from five distinct *Solanum* species.

Several studies on homologues which belong to the class of NBS-LRR genes (Meyers *et al.*, 1998; Mondragón-Palomino *et al.*, 2002; Rose *et al.*, 2004) show that the NBS domain is highly conserved showing low rates of nucleotide changes, while the LRR domain is found to be under diversifying selection. In a survey on *RPP13* homologues in *Arabidopsis thaliana* (Rose *et al.*, 2004) high sequence polymorphism was found in the LRR domain resulting in elevated amino acid substitution rates as a result of possible positive selection towards divergent recognition specificities. An analysis of 11 *Cf*-gene homologues from tomato (Parniske *et al.*, 1997) showed, that particularly some solvent exposed residues within the LRR are hypervariable. In this study, however, strong sequence conservation was shown for the LRR domain of the functional *Rx* orthologues *Rx1*, *Rx2*, *Rx3, Rx4* and Rx_{edn} harbouring identical recognition specificities. Comparison of the genes with the closely related gene *Gpa2* with a putative distinct recognition specificity reveals high variability in the LRR (Van der Vossen et al., 2000). The extreme conservation of the LRR domain between the *Rx* orthologues implicates strong evolutionary constraints imposed by the recognition specificity towards the elicitor from PVX. These data are consistent with a gain-of-function study for *Rx1* showing that mutations at certain positions in the LRR domain, which are strictly conserved in all 4 functional *Rx* orthologues, result in constitutive gain-of-function phenotypes (Bendahmane *et al.*, 2002). Similar results were obtained in an analysis of the *R* genes *Cf4* and *Cf9*, which belong to another *R* gene class including extracellular *R* genes, comparing functional alleles isolated from distinct related wild species from tomato (Kruijt *et al.*, 2005).

Through duplication and clustering of homologous genes the chance of being subject to sequence exchange is increased (Mondragón-Palomino *et al.*, 2002). Sequence exchange by crossing-over events and gene conversions have been found to be an important source of variation resulting in chimaeric gene copies within resistance gene clusters (Meyers et al., 1998; Van der Hoorn et al., 2000; Kuang et al., 2004). Here, such an event of a sequence exchange is reported in a functional gene showing that parts of the NBS domain can be exchanged without changing its function. While the NBS domain of the four functional *Rx* orthologues is mainly conserved, the comparative analysis with 23 corresponding paralogues using UPGMA dendrograms of nucleotide sequences points out an event of sequence exchange separating *Rx1*/*Rx3* and *Rx2*/*Rx4*. With the occurrence of such sequence exchanges, different functional domains get uncoupled by means of their evolutionary constraints. The sequence interval bearing the recognition specificity (LRR-domain) is under constraint matching pathogen recognition, while the sequence interval bearing the downstream-signalling function (NBS-domain) is under a different selection pressure by matching the resistosome.

Rx1 is able to recognize all known strains of PVX with the exception of the resistance breaker strain HB. It has been shown that the susceptibility to HB is based on changes in two amino acid positions of the coat protein (Goulden et al., 1993). Hence, co-expression of avirulent and virulent coat proteins together with the candidate *Rx* homologue in a transient assay on *N. benthamiana* contributes strongly to the identification of functional *Rx* alleles. Transient expression of candidate *Rx* homologues in *N. benthamiana* resulted for some *R* gene homologues in an autoactivated hypersensitive response in the absence of the elicitors. The constitutive overexpression of the genes under control of the CaMV 35S promoter in such transient expression system might differ from the regulation with their endogenous promoter in the *Solanum* genotype from which they were isolated. In addition, a number of *Solanum* genotypes were showing an HR upon transient expression of both the virulent and avirulent PVX coat protein, whereas a gene-for-gene specific recognition was confirmed only for the two genotypes from *S. lepthophyes/sparsipilum* and *S. albicans* harbouring the functional *Rx* orthologues *Rx3* and *Rx4*, respectively. A possible explanation could be physiological stress imposed to the plant tissue by cutting the leaves and syringe disruption, to which different plant species might have a variable tolerance (Mittler and Lam, 1997). Furthermore, a number of aspecific reactions upon infiltration

with *A. tumefaciens* carrying various constructs in potato species have been listed (Van der Hoorn et al., 2000; Kuta and Tripathi, 2005; Wroblewski et al., 2005).

Although PVX nowadays has a worldwide distribution, numerous distinct strains have been collected in the Andean region of Peru and Bolivia suggesting that this might be the center of origin for PVX (Fribourg, 1980; Kavanagh *et al.*, 1992). This overlaps with the center of biodiversity of *Solanum* section *Petota* in the Andes of South America, the area where the species harbouring *Rx* (candidate) genes are most abundant. While *S. acaule* and *S. tuberosum* ssp. *andigena* are quite spread from Peru to Argentina, *S. sparsipilum* and *S. lephtophyes* from which S. *lepthophyes/sparsipilum* was derived are distributed over Peru and Bolivia and *S. albicans* is present in northern Peru (Hawkes, 1990). Remarkably, *S. edinense* harbouring the non-functional gene Rx_{edn} is found in Mexico (Hawkes, 1990). A permanent or episodic absence of corresponding PVX strains in the remote habitate of *S. edinense* and subsequent absent selection pressure on *Rx* possibly could have caused the emergence of this non-functional CC-deletion mutant.

The isolation of *Rx3* from *S. lepthophyes/sparsipilum* which belongs to the group *Tuberosa* (Peru) like *Rx1* and *Rx4* from *S. albicans* plus Rx_{edn} from *S. edinense* which belong to the group *Acaule/Demissum* like *Rx2* shows that functional conservation is not only fixed across species, but traced back in time before the divergence of two taxonomic distinct groups within *Solanum* represented by several distinct related species. The presence of functional *Rx* orthologues in potato species from taxonomically distinct groups point at the ancient origin of the Rx recognition specificity and as such, these data support the trench warfare hypothesis for disease resistance genes in potato. So far, studies on the evolution of NBS-LRR genes across species have been done using *R* gene homologues lacking the information on functionality (Mondragón-Palomino *et al.*, 2002; Ballvora *et al.*, 1996). A study on functional alleles of the resistance gene *Pto* from distinct related species of tomato was conducted resulting in the classification of *Pto* as an ancient conserved recognition specificity (Riely and Martin, 2001). The *Pto* gene, however, belongs to a distinct class of *R* genes encoding a serine-threonine protein kinase (Frederick *et al.*, 1998).

Interestingly, in the discussion about the origin of cultivated potato *S. tuberosum* ssp. *andigena*, based on morphological and cytological studies the species *S. lepthophyes* and *S. sparsipilum* are mentioned as likely parental ancestors (Grun, 1979; Hawkes, 1990). As cultivation of potato arose approximately 10,000 years ago in Peru (Hawkes, 1990) in

case of such a line of inheritance the division of *Rx* from *Tuberosa* (Peru) into *Rx1* and *Rx3* has to be prior to this event. Within the group *Acaule/Demissum* again, chromosomal data supports *S. albicans* (hexaploid) being derived from a parental *S. acaule* (tetraploid) and an unknown diploid progenitor (Hosaka and Spooner, 1992). In the context of the found *Rx* alleles within these species *Rx2* and *Rx4* share a common ancestor gene, which was present in a *S. acaule* population. The pentaploid *S. edinense* belongs to the group *Acaule/Demissum* as well. The timescale in which the 5 species evolved within the groups as well as of the splitting into group *Tuberosa* (Peru+cultivated species) and *Acaule/Demissum* remain yet unknown.

Results from the sequence homology based screening in *Solanum* section *Petota* indicate that further taxa with (potential) functional *Rx* orthologues are present within *Tuberosa* (Peru+cultivated species), the closely related *Tuberosa* from Chile/Bolivia/Argentinia and *Acaule/Demissum*. No presence of a functional *Rx* recognition specificity was detected in the groups *Etuberosum* and the Mexican diploid species. *Etuberosum* is often presented as the phylogenetic outgroup within *Solanum* section *Petota* and the group of Mexican diploid species is considered to be the ancestral sister group for most of the South American groups including *Tuberosa* and *Acaule/Demissum* (Hawkes, 1990). The absence of functional *Rx* orthologues in the groups Mexican Diploid and *Etuberosum* therefore points to an origin of the Rx recognition specificity after the segregation from these groups but before the *Solanum* groups *Tuberosa* (Peru) and *Acaule/Demissum* diverged.

While the majority of the wild potato species is self-incompatible favouring polymorphism among populations, through mostly clonal propagation over tubers, the recombination within cultivated potato is far more limited. For *S. acaule* a high rate of self pollination together with a restricted gene flow between subpopulations is known as well (Hawkes, 1990). In case of the *Cf9* alleles from tomato (Kruijt *et al.*, 2005), increased variation between *R* gene alleles isolated from self-incompatible *Lycopersicon* accessions was found. As the dataset of functional *Rx* orthologues is based on one accession per species, isolation of additional functional *Rx* orthologues within species will provide more insight in the different rates of intraspecific variation between *Rx1* and *Rx2* versus *Rx3* and *Rx4*.

Material and Methods

Plant material

Seeds of 51 *Solanum* accessions were obtained from the Dutch-German Potato Collection, Centre for Genetic Resources (CGN) Wageningen, The Netherlands. Seedlings were grown *in vitro* in 9 cm petridishes on MS20 medium containing 4.7g Murashige&Skoog, 20g sucrose and 8g agarose per liter at pH 5.7 and exposed to 16h light per day at 25˚C. Next, explants were transferred and maintained in jars *in vitro* under similar conditions. The PVX resistant and susceptible potato cultivars SH83-92-488 and RH89-039-16 have been previously described and will be referred to as SH and RH, respectively.

gDNA extraction and cDNA synthesis

Genomic DNA was obtained from the plants by an extraction procedure using the DNeasy Plant Mini kit according to manufacturer's instructions (Invitrogen, San Diego, CA, USA). Total RNA was isolated from potato genotypes using the RNeasy Mini kit (Qiagen) according to the manufacturer's instructions. First strand cDNA was generated using the SuperScript III Reverse Transcriptase kit (Invitrogen, San Diego, CA, USA).

PCR analysis

Rx1/Gpa2 homologues were amplified from genomic DNA template using primer ggggACAAgTTTgTACAAAAAAgCAggCTTTTgTTCATTTTCATACTgAgAg and ggggACCACTTTgTACAAgAAAgCTgggTggCTAgTCCTCAgACCAAC flanking the full open reading frame of *Rx1* (Bendahmane et. al 2000). The primers were adjusted to allow insertion into a Gateway-specific cloning cassette. PCR was performed with the EHF-PCR-kit, a proofreading Polymerase kit (Invitrogen, San Diego, CA, USA). PCR was performed using 10 cycles of 30s 94˚C, 30s 46˚C, 120s 68˚C followed by 25 cycles of 30s 94˚C, 30s 46˚C, 120s 68˚C with an extension of 5s/cycle and finally 5min elongation at 72˚C. cDNA analysis was carried out by using the primers gAAgAATTgACATTTCgTTTATATTATC and CTggCTAgTCCTCAgAACAC amplifying *Rx3* and *Rx4* from cDNA templates derived from *S. lepthophyes/sparsipilum* and *S. albicans*, respectively.

Screening the *Solanum* gDNA collection

A collection of gDNA samples from 4876 genotypes from about 1000 accessions derived from 200 *Solanum* species was kindly provided by B. Vosman at PRI, Wageningen, the Netherlands. The gDNA samples were used as a template in a PCR-based screening approach using the functional *Rx* SNP-specific primer gAAgAATTgACATTTCgTTTATATTATC and a *Rx/Gpa2* cluster specific primer CTggCTAgTCCTCAgAACAC resulting in the amplification of specific fragment of the C-terminal part of the LRR domain. The absence/presence of an amplification product was monitored on agarose gels and recorded.

Cloning and digestion analysis of the PCR products

PCR-products were gel-purified using the Gel purification kit according to the manufacturer's instructions (Invitrogen, San Diego, CA, USA). Purified PCR products were then cloned into the Gateway™ entry vector pDONR_207 (Invitrogen, San Diego, CA, USA) for transformation into *E.coli* TOP10 cells according to the manufacturer's

instructions (Invitrogen, San Diego, CA, USA). For each PCR product 39 positive clones were selected and analysed in a PCR with cluster-specific primers based on the LRR domain followed by a TaqI digestion as described (Bakker *et al.,* 2003). The DNA fragments were separated on a 4% agarose gel stained with 0.4x gelstar (Cambrex Bio Science, Rockland, ME, USA). Clones were selected for further analysis based on differences in their digestion pattern.

Sequence analysis

Sequences of the resistance gene homologues *RGC1* (designated *andSH-RGH1*), *RGC3* (designated *andSH-RGH3)* and the resistance genes *Gpa2* (Van der Vossen *et al.*, 2000), *Rx1* (Bendahmane *et al.*, 1999) and *Rx2* (Bendahmane et al., 2000) are obtained from Genbank (accession numbers AF266747, AF266746, AF195939, AJ011801, AJ249448, respectively).

Sequencing of the PCR products was carried out by Greenomics, PRI, Wageningen, The Netherlands or Baseclear, Leiden, The Netherlands. Sequence assembling was done using ContigExpress from the Vector NTI suite software package (InforMax Inc., Bethesda, USA). Sequence Cluster analysis was performed with PAUP V4.0b10 (Swofford, 1999). Sequence exchanges were investigated using the program Geneconv with default settings (Sawyer, 1989).

Subcloning

Isolated plasmids of the selected constructs were used in an LR-reaction in order to recombine the inserts into the destination vector pK2GW7 for overexpression in plants (MIB, Gent, Belgium). Transformation into *E.coli* TOP10 for selection and maintenance followed by plasmid isolation and transformation into *A. tumefaciens* LBA 4044 were performed according to the manufacturer's instructions (Invitrogen, San Diego, CA, USA).

The PVX coat protein (CP) was amplified from the PVX amplicons pgR106 (Jones *et al.*, 1999) containing cDNA of the *Rx1*-avirulent PVX strain UK3 and pgR105 containing cDNA of the *Rx1*-resistance breaking strain HB (Goulden et al., 1993) using the primers 5UK3cp 5'-TCCATGGGCGGTGG AGTCATGAGCGCACCAGCTAGCACAACACAGCC and 3uk3CP 5'-AGGTACCTGCGGTTATGGTG GTGGTAGAGTGACAACAGC for CP106 and 5HBcp 5'-TCCATGGGCGGTGGAGTCATGACTACG CCAGCCAACACCACTC and 3hbCP 5'-AGGTACCTGCGGTTATGGTGGGGGTAGTGAGATAAC AGC for CP105. The products were cloned into the NcoI-KpnI sites of pUCAP between the CaMV 35S promoter and the Tnos terminator. For agro-infiltration assays and *Agrobacterium tumefaciens* mediated plant transformation, the expression cassettes containing the constructs were cloned into the AscI and PacI sites of the binary vector pBINPLUS (Van Engelen *et al.*, 1995).

For expression under the control of the double enhanced CaMV 35S promoter and Tnos terminator**,** *Rx1* was amplified from the binary plasmid pBINRx1 (Van der Vossen et al., 2000) using the primers 5GpRxbn (5'-TTT TTG GAT CCA TGG CTT ATG CTG CTG TTA CTT CCC) and Rxrev (5'-GAT AGC GTC GAC CAC CTT AAC TAC TCG CTG CA) and cloned into the NcoI-SalI sites of pUCAP (Van Engelen *et al.*, 1995) removing the kappa signal peptide (Schouten *et al.*, 1997), resulting in pUCAPRx1.

Agroinfiltration assays in Nicotiana benthamiana
Bacteria were grown at 28˚C in YEB Medium (5g beef extract (Sigma-Aldrich, St. Louis, MO, USA), 1g yeast extract (Gibco BRL Life technologies, Paisley, Scotland, UK), 5g peptone (Gibco BRL Life technologies, Paisley, Scotland, UK), 5g sucrose (Duchefa, Haarlem, the Netherlands) and 2ml (1M) $MgSO_4$ per liter) with the appropriate antibiotics. Reaching an OD of 0.9-1.5 they were harvested by centrifugation and resuspended to a final OD=0.1 in MMAi Medium (5g Murashige&Skoog salt (Duchefa, Haarlem, the Netherlands), 1,95g MES (Sigma-Aldrich, St. Louis, MO, USA), 20g sucrose (Duchefa, Haarlem, the Netherlands) per liter with a pH adjusted to 5.6. The resuspensions were incubated for 1h at 20˚C to induce virulence as described (Van der Hoorn *et al.*, 2000). Leaves of 6-week-old-seedlings of *N. benthamiana* were infiltrated with the various constructs by *A. tumefaciens* mediated transient assay.

Agroinfiltration assays in *Solanum* species
Detached leaves from *in vitro* grown genotypes from 51 accessions (5 genotypes per accession) derived from 35 *Solanum* species were infiltrated with pBINPLUS35S::CP106 encoding for the elicitor of *Rx1*, *viz.* the coat protein CP106 derived from the avirulent strain PVX_{UK3}. Plants in the greenhouse were grown *in duplo* and 2 leaves from each plant were infiltrated at separate spots with *A. tumefaciens* carrying the constructs pBINPLUS35S::106 and pBINPLUS35S::105 encoding CP106 and CP105, respectively. Bacterial cultures were prepared as described above.

Virus resistance assay
Three week old plantlets of *in vitro* grown *Solanum* species were inoculated *in triplo* with homogenates containing PVX strains UK3 and HB by rubbing two local leaves with carborundum powder. After incubation for 10 days at 25˚C and 16h light leaves of the apex were homogenized in 2ml of 50mM phosphate buffer, pH7. To determine the virus concentration 100µl of the centrifuged supernatant was used in a DAS-ELISA. Plates were coated with a 1:1000 dilution of a polyclonal antibody against PVX to bind the antigen and a polyclonal antibody against PVX conjugated with alkaline phosphatase was used for detection (Valkonen *et al.*, 2000).

Toothpick inoculation
To test the systemic spreading of PVX in the wild potato species *S. edinense,* three lower leaves of a single plant were inoculated in the greenhouse by piercing the leaf 5 times next to a vein with a toothpick carrying *A. tumefaciens* cells (pMOG101) as described by Takken et al. (Takken et al., 2000). *A. tumefaciens* harbouring the constructs pBINPLUS 35S:GFP as a control and the PVX amplicons pgR105 and pgR106 expressing the virulent and avirulent PVX strains HB and UK3, respectively, were plated on selective YEB medium containing kanamycin (50ug/ml) and rifampicin (25 ug/ml). Twelve days after inoculation, 2 inoculated leaves and 2 systemic leaves from the top were collected per plant and homogenized in 3 ml 10 mM phosphate buffer (pH7.0). To test for the presence and systemic spreading of the virus, 50 ul homogenate was inoculated on one lower leaf of *N. benthamiana* by rubbing with 500 mesh carborundum. The appearance of chlorotic infection spots and crinckling of the apical leaves was monitored until 10 days post infection and used as a marker for successful spreading of the virus.

Acknowledgements
We like to thank D. Baulcombe (Sainsbury Laboratory, Norwich, UK) for providing the PVX amplicons pGR105 and pGR106. Furthermore, we like to thank R. Hoekstra (CGN, Wageningen, the Netherlands) and V. Vleeshouwers (Laboratory of Plant Breeding, Wageningen University, the Netherlands) for providing seeds and *in vitro* plant material from wild potato species, respectively and B. Vosman and M. Jacobs for providing the *Solanum* DNA samples and fruitful discussions. This work was supported by a grant from the Bosch Foundation, EU project APOPHYS QLRT-2001-01849 and the Centre of Biosystems Genomics (CBSG), Wageningen, the Netherlands.

References

Aravind, L., Koonin, E.V., and Dixit, V.M. (1999). The domains of death: Evolution of the apoptosis machinery. Trends in Biochemical Sciences **24,** 47.

Bakker, E., Butterbach, P., Rouppe Van der Voort, J., Van der Vossen, E., Van Vliet, J., Bakker, J., and Goverse, A. (2003). Genetic and physical mapping of homologues of the virus resistance gene Rx1 and the cyst nematode resistance gene Gpa2 in potato. Theoretical and Applied Genetics **106,** 1524.

Ballvora, A., Salamini, F., Gebhardt, C., and Leister, D. (1996). A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nature Genetics **14,** 421.

Baumgarten, A., Cannon, S., Spangler, R., and May, G. (2003). Genome-level evolution of resistance genes in Arabidopsis thaliana. Genetics **165,** 309.

Bendahmane, A., Kanyuka, K., and Baulcombe, D.C. (1999). The Rx gene from potato controls separate virus resistance and cell death responses. Plant Cell **11,** 781.

Bendahmane, A., Kohm, B.A., Dedi, C., and Baulcombe, D.C. (1995). The coat protein of potato virus X is a strain-specific elicitor of Rx1-mediated virus resistance in potato. Plant Journal **8,** 933.

Bendahmane, A., Querci, M., Kanyuka, K., and Baulcombe, D.C. (2000). Agrobacterium transient expression system as a tool for the isolation of disease resistance genes: application to the *Rx2* locus in potato. The Plant Journal **21,** 73.

Bendahmane, A., Farnham, G., Moffett, P., and Baulcombe, D.C. (2002). Constitutive gain-of-function mutants in a nucleotide binding site-leucine rich repeat protein encoded at the Rx locus of potato. Plant Journal **32,** 195-204.

Caicedo, A.L., Olsen, K.M., Purugganan, M.D., Stinchcombe, J.R., and Schmitt, J. (2004). Epistatic interaction between Arabidopsis FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. Proceedings of the National Academy of Sciences of the United States of America **101,** 15670.

Dawkins, R., and Krebs, J.R. (1979). Arms races between and within species. Proceedings of the Royal Society of London - Biological Sciences **205,** 489.

Dixon, M.S., Hatzixanthis, K., Jones, D.A., Harrison, K., and Jones, J.D.G. (1998). The tomato Cf-5 disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. Plant Cell **10,** 1915.

Flor, H.H. (1971). Current status of the gene-for-gene concept. [Plants, breeding, disease resistance]. Annu Rev Phytopathol **9,** 275.

Frank, S.A. (1992). Models of plant-pathogen coevolution. Trends in Genetics **8,** 213.

Frederick, R.D., Thilmony, R.L., Sessa, G., and Martin, G.B. (1998). Recognition specificity for the bacterial avirulence protein AvrPto is determined by Thr-204 in the activation loop of the tomato Pto Kinase. Molecular Cell **2,** 241.

Fribourg, C.E. (1980). History and distribution of potato viruses in Latin America. FITOPATOLOGIA **15,** 13.

Gebhardt, C., and Valkonen, J.P.T. (2001). Organization of genes controlling disease resistance in the potato genome. Annual Review of Phytopathology **39,** 79.

Goulden, M.G., Kohm, B.A., Cruz, S.S., Kavanagh, T.A., and Baulcombe, D.C. (1993a). A feature of the coat protein of potato virus X affects both induced virus resistance in potato and viral fitness. Virology **197,** 293.

Grun, P. (1979). Evolution of the cultivated potato: a cytoplasmic analysis. Linn Soc Symp Ser**,** 655.

Hawkes, J.G. (1990). The potato : evolution, biodiversity and genetic resources. (London: Belhaven, 1990. viii, 259 p. : ill.).

Hosaka, K., and Spooner, D.M. (1992). RFLP analysis of the wild potato species, Solanum acaule Bitter (Solanum sect. Petota). Theor Appl Genet **84,** 851-858.

Jones, L., Hamilton, A.J., Voinnet, O., Thomas, C.L., Maule, A.J., and Baulcombe, D.C. (1999). RNA-DNA interactions and DNA methylation in post-transcriptional gene silencing. Plant Cell **11,** 2291.

Kardolus, J.P., Van Den Berg, R.G., and Van Eck, H.J. (1998). The potential of aflps in biosystematics: A first application in Solanum taxonomy (Solanaceae). Plant Systematics and Evolution **210,** 87.

Kavanagh, T., Goulden, M., Santa Cruz, S., Chapman, S., Barker, I., and Baulcombe, D. (1992). Molecular analysis of a resistance-breaking strain of potato virus X. Virology **189,** 609.

Kruijt, M., Kip, D.J., Joosten, M.H.A.J., Brandwagt, B.F., and De Wit, P.J.G.M. (2005). The Cf-4 and Cf-9 resistance genes against Cladosporium fulvum are conserved in wild tomato species. Molecular Plant-Microbe Interactions **18,** 1011.

Kuang, H., Woo, S.S., Meyers, B.C., Nevo, E., and Michelmore, R.W. (2004). Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. Plant Cell **16,** 2870.

Kuta, D.D., and Tripathi, L. (2005). Agrobacterium-induced hypersensitive necrotic reaction in plant cells: A resistance response against Agrobacterium-mediated DNA transfer. African Journal of Biotechnology **4,** 752.

Meyers, B.C., Shen, K.A., Rohani, P., Michelmore, R.W., and Gaut, B.S. (1998). Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell **10,** 1833.

Michelmore, R.W., and Meyers, B.C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Research **8,** 1113.

Mittler, R., and Lam, E. (1997). Characterization of nuclease activities and DNA fragmentation induced upon hypersensitive response cell death and mechanical stress. Plant Molecular Biology **34,** 209.

Mondragón-Palomino, M., Meyers, B.C., Michelmore, R.W., and Gaut, B.S. (2002). Patterns of positive selection in the complete NBS-LRR gene family of Arabidopsis thaliana. Genome Research **12,** 1305.

Parniske, M., Golstein, C., Thomas, C.M., Wulff, B.B.H., Hammond-Kosack, K.E., Jones, D.A., Harrison, K., and Jones, J.D.G. (1997). Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato. Cell **91,** 821.

Querci, M., Baulcombe, D.C., Goldbach, R.W., and Salazar, L.F. (1995). Analysis of the resistance-breaking determinants of potato virus X (PVX) strain HB on different potato genotypes expressing extreme resistance to PVX. Phytopathology **85,** 1003.

Riely, B.K., and Martin, G.B. (2001). Ancient origin of pathogen recognition specificity conferred by the tomato disease resistance gene Pto. Proceedings of the National Academy of Sciences of the United States of America **98,** 2059.

Rose, L.E., Langley, C.H., Michelmore, R.W., Bittner-Eddy, P.D., Holub, E.B., and Beynon, J.L. (2004). The Maintenance of Extreme Amino Acid Diversity at the Disease Resistance Gene, RPP13, in Arabidopsis thaliana. Genetics **166,** 1517.

Sawyer, S. (1989). Statistical tests for detecting gene conversion. Molecular Biology and Evolution **6,** 526.

Schouten, A., Roosien, J., De Boer, J.M., Wilmink, A., Rosso, M.N., Bosch, D., Stiekema, W.J., Gommers, F.J., Bakker, J., and Schots, A. (1997). Improving scFv antibody expression levels in the plant cytosol. FEBS Letters **415,** 235.

Spooner, D.M., and Castillo T, R. (1997). Reexamination of series relationships of South American wild potatoes (Solanaceae: Solanum sect. Petota): Evidence from chloroplast DNA restriction site variation. American Journal of Botany **84,** 671.

Stahl, E.A., Kreitman, M., Bergelson, J., Dwyer, G., and Mauricio, R. (1999). Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis. Nature **400,** 667.

Swofford, D.D. (1999). PAUP* Version4.0b10.

Takken, F.L., Albrecht, M., and Tameling, W.I. (2006). Resistance proteins: molecular switches of plant defence. Current Opinion in Plant Biology **9,** 383.

Takken, F.L.W., Luderer, R., Gabriels, S.H.E.J., Westerink, N., Lu, R., De Wit, P.J.G.M., and Joosten, M.H.A.J. (2000). A functional cloning strategy, based on a binary PVX-expression vector, to isolate HR-inducing cDNAs of plant pathogens. Plant Journal **24,** 275.

Valkonen, J.P.T., Mäki-Valkama, T., Pehu, T., Santala, A., Koivu, K., Lehto, K., and Pehu, E. (2000). High level of resistance to potato virus Y by expressing P1 sequence in antisense orientation in transgenic potato. Molecular Breeding **6,** 95.

Van den Berg, R.G. (2001). Solanaceae V : Advances in taxonomy and utilization. (Nijmegen: Nijmegen University Press).

Van der Biezen, E.A., and Jones, J.D. (1998). The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. Current biology : CB **8,** 226.

Van der Hoorn, R.A.L., Laurent, F., Roth, R., and De Wit, P.J.G.M. (2000). Agroinfiltration is a versatile tool that facilitates comparative analyses of Avr9/Cf-

9-induced and Avr4/Cf-4-induced necrosis. Molecular Plant-Microbe Interactions **13,** 439.

Van der Voort, J.R., Bakker, J., Kanyuka, K., Bendahmane, A., Mooijman, P., Baulcombe, D., Van der Vossen, E., Klein-Lankhorst, R., and Stiekema, W. (1999). Tight physical linkage of the nematode resistance gene Gpa2 and the virus resistance gene Rx on a single segment introgressed from the wild species Solanum tuberosum subsp. andigena CPC 1673 into cultivated potato. Molecular Plant-Microbe Interactions **12,** 197.

Van der Vossen, E.A.G., Rouppe van der Voort, J.N.A.M., Kanyuka, K., Bendahmane, A., Sandbrink, H., Baulcombe, D.C., Bakker, J., Stiekema, W.J., and Klein-Lankhorst, R.M. (2000). Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: A virus and a nematode. Plant Journal **23,** 567.

Van Engelen, F.A., Molthoff, J.W., Conner, A.J., Nap, J.P., Pereira, A., and Stiekema, W.J. (1995). pBINPLUS: An improved plant transformation vector based on pBIN19. Transgenic Research **4,** 288.

Wroblewski, T., Tomczak, A., and Michelmore, R. (2005). Optimization of Agrobacterium-mediated transient assays of gene expression in lettuce, tomato and Arabidopsis. Plant Biotechnology Journal **3,** 259.

Chapter 4

Structural diversity and evolutionary relationships of *Rx1/Gpa2* homologues in *Solanum*

Patrick Butterbach, Erin Bakker, Pjotr Prins, Erik Slootweg, Jan Roosien, Jaap Bakker and Aska Goverse

---

to be submitted

Abtract

To gain more insight in the structural and evolutionary relationship of the resistance gene cluster in potato that harbours the highly similar potato virus X resistance gene *Rx1* and the potato cyst nematode resistance gene *Gpa2*, sequence analysis was carried out on the open reading frames (ORFs) of 75 *Rx1/Gpa2* homologues derived from 10 *Solanum* species with a nucleotide sequence similarity of 69 to 98%. This analysis includes the four functional orthologous virus resistance genes *Rx1-4*, the truncated Rx_{edn} that harbours *Rx* resistance specificity and the nematode resistance gene *Gpa2*. The gene structure was determined based on a subset of cDNA sequences and overall sequence conservation allowing the identification of intron splice sites and putative open reading frames for the complete set of genes. About one third of the homologues were found to encode for a protein harbouring complete CC-, NBS- and LRR- domains analogous to *Rx* and *Gpa2,* indicating the presence of functional *R* genes with unknown specificities. High variability was found at the C-terminus harbouring an acidic tail in 24 homologues including the functional orthologues *Rx1*, *Rx2* and *Rx3*. We identified a local structural variability in the sequence region encoding LRR 10, dividing the homologues in 3 structural subclasses. The occurrence of similar sequence exchange tracks in homologues isolated from different *Solanum* species indicates that interallelic exchanges took place before speciation. Analysis for positive selection revealed very significant positive selection on codon sites within the LRR domain, but also within the CC- and NBS- domain.

Introduction

The innate immune system of plants consists of large arrays of resistance genes (*R*-genes) conferring resistance to a wide range of pests and pathogens. Recent views on the plant immune system describe R proteins as part of a second and third phase of host defence acting after the PAMP-triggered immunity has been overcome by the pathogen (reviewed by Jones and Dangl, 2006). Numerous *R* genes have been found in different plant species (Dangl and Jones, 2001; Hulbert et al., 2001) and the genome of *Arabidopsis thaliana* Col-0 has been estimated to encode for ~150 resistance proteins (Meyers et al., 2003). Based on their structural domains, *R* genes can be divided into distinct classes. *R* genes encoding for

a leucine rich repeat (LRR) domain and a nucleotide binding (NBS) site form the major structural class in plants. Two subfamilies within the NBS-LRR class are known, defined by the presence of a toll/interleukin-1 receptor (TIR) or a coiled-coil (CC) motif rich region in the N-terminus. TIR-NBS-LRR and CC-NBS-LRR differ in the NBS structure, which suggest distinct downstream signalling (reviewed by McHale et al., 2006). Interestingly, TIR-NBS-LRR genes are not found in higher monocots like cereals (Pan et al., 2000; Bai et al., 2002).

The different domains of R proteins resemble functional modules fulfilling different tasks like pathogen recognition and triggering of a disease resistance response. Several lines of evidence point at a role for the LRR domain in the determination of *R* gene specificity. Sequence homology with the LRR domain of the porcine ribonuclease inhibitor suggested that the three-dimensional structure of the LRR domain of *R* genes is horse shoe shaped with a series of parallel β sheets at the inside consisting of conserved structural amino acid residues forming the back bone and variable solvent exposed residues that play a role in protein-protein interactions (Jones, 1997). Comparative analysis of *R* gene sequences revealed that the LRR domain is the most variable part of the R protein suggesting a role in resistance specificity. This is supported by the observation that the solvent exposed amino acid residues of the LRR β-strand/β-turn motifs are hypervariable and subject to diversifying selection (Parniske et al., 1997). Furthermore, combining the LRR domain of the flax resistance gene *L2* with the TIR-NBS regions of the *L6* and *L10* alleles resulted in a chimaeric gene product with *L2* specificity (Ellis et al., 1999), which demonstrates that the LRR domain is the main specificity determinant of the R protein.

Although some *R* genes like the flax *L* gene or the *Arabidopsis* gene *Rpm1* are known to reside in simple loci, most of the *R* genes are members of large gene families that are found in complex loci harbouring several tandemly repeated *R* gene homologues (Michelmore and Meyers, 1998; Leister, 2004); reviewed by (Hulbert et al., 2001). It has been suggested, that the clustered organization facilitates events of sequence exchange among *R* genes (Parniske et al., 1997; Kuang et al., 2004; Meyers et al., 2005; Mondragon-Palomino and Gaut, 2005). Frequent unequal crossing-over and gene conversion could result in the reshuffling of functional domains thereby accelerating the evolution of novel *R* gene specificities (Hulbert et al., 2001). The same process, however, tends to homogenize the paralogues, which does not comply with the findings at the *N*, *Dm* and *Cf* loci where

orthologues are more similar than paralogues (Meyers et al., 1998; Dodds et al., 2001b; Kuang et al., 2004; Parniske et al., 1997). Therefore, Michelmore and Meyers (1998) suggest that *R* genes mainly evolve by divergent evolution and a birth-and-death process.

Functional proteins are expected to be constrained at most of the amino acids and during most evolutionary time, complicating the detection of adaptive evolution by substitution ratios averaging over the whole sequence. Whether a change in the DNA gets fixated in a population and becomes an evolutionary 'success' is determined by random genetic drift, as well as purifying or positive selection in populations. When purifying or positive selection is witnessed, for example in a host-pathogen interaction at the protein level, tests can be developed for adaptive evolution. Powerful tests for adaptive evolution that allow detection of positive selection are derived from the comparison of the relative rates of synonymous and nonsynonymous substitutions using maximum likelihood methods in a phylogenetic framework (Yang, 2005). Subsequently, positive selection restricted to a few amino acid positions and occurring in an episodic fashion have been detected for different *R* gene datasets (reviewed by McDowell and Simon, 2006; Kuang et al., 2004; Mondragón-Palomino et al., 2002).

A number of *R* gene clusters like the *P* locus in flax (Dodds et al., 2006), the *Cf* locus in tomato (Parniske et al., 1997) and the *RPP13* locus in *Arabidopsis thaliana* (Rose et al., 2004) are known to contain race specific resistance genes against different pathotypes of the same pathogen species. In contrast, the *Gpa2/Rx1* locus on chrXII of potato harbours two *R* genes that confer resistance to two completely unrelated pathogens (Van der Vossen et al., 2000; Bendahmane et al., 1999). *Gpa2* confers resistance to the potato cyst nematode *Globodera pallida*, while *Rx1* recognizes the coat protein of the potato virus X. The ratio between non-synonymous and synonymous amino acid substitutions (K_a/K_s) suggested that the LRRs of *Gpa2* and *Rx1* are subject to positive selection (Van der Vossen et al., 2000). These data suggest that the LRR domain of *Gpa2* and *Rx1* is involved in determining nematode and virus recognition specificity, respectively.

To gain more insight in the structural and evolutionary relationships of the *Rx* cluster in potato, the ORFs of 75 *Rx1/Gpa2* homologues derived from 10 *Solanum* species have been analyzed, including the four functional orthologous virus resistance genes *Rx1-4*, Rx_{edn} (Bendahmane et al., 1999; Bendahmane et al., 2000; Chapter 3, this thesis) and the nematode resistance gene *Gpa2* (Van der Vossen et al., 2000). The structural diversity of

the homologues was examined based on the intron-exon composition, the putative amino acid sequence and the presence or absence of known structural motifs in the NBS and LRR – domain. Structural relationships between the homologues were investigated by similarity dendrograms. We report an overall high conservation among sequences from both paralogues and orthologues possibly generated by the frequent events of sequence exchange as observed within this dataset. Using a maximum-likelihood approach for testing codon substitution models strong positive selection was detected at distinct sequence positions in the CC-, NBS- and LRR domain. High overall conservation of the homologues further allowed for the identification of unique amino acid residues in the LRR domain characteristic for *Rx* and *Gpa2* and possibly involved in pathogen recognition.

Results

gDNA alignment

A set of full length open reading frames (ORFs) of 75 resistance gene homologues derived from 10 different *Solanum* species was aligned, including the functional PVX resistance genes *Rx1, Rx2, Rx3, Rx4*, the truncated resistance gene homologue Rx_{edn}, which harbours *Rx* recognition specificity and the nematode resistance gene *Gpa2* (Bendahmane et al., 1999; Bendahmane et al., 2000; Van der Vossen et al., 2000; Chapter 3 this thesis) The alignment revealed an overall sequence similarity of 69 to 98% compared to *Rx1* (Table 1).

Table 1. Overview showing the nucleotide sequence length of the ORF, Introns, and deduced amino acid sequence length of 75 Rx resistance gene homologues derived from 10 *Solanum* species. Introns absent due to a deletion event or missing donor/acceptor sites are indicated as deleted and -, respectively. Joined Introns as in *andSH-RGH5* and *tubRH-RGH6* are marked with ← . Similarities of nucleotide and amino acid sequences are indicated in percentage. The presence of CC, NBS and LRR domain is indicated by a + and in cases of shorter amino acid lengths the number of amino acids is shown. The presence of an acidic tail with one (+) or two (++) SVTT(V/D)EDDDD motifs is indicated and the corresponding amino acid length is shown.

Ploidy	Sequence name	N length	IntronI	IntronII	cDNA isolated	aa length	n similarity to Rx1	aa similarity to Rx1	CC(aa) 160aa	NBS(aa) 312	LRR(aa) 392-403	acidic tail	tail aa length
	Rx1 (and)	3159	234	111	+	937	100%	100%	+	+	+	++	69
	Rx2 (acl)	3162	234	111	+	938	97%	96%	+	+	+	++	69
	Rx3 (lph/spl)	3146	234	95	+	937	98%	98%	+	+	+	++	68
	Rx4 (alb)	3128	230	110	+	909	97%	95%	+	+	+	-	40
	Rx_{edn}	2864	234	113		838	88%	86%	60	+	+	++	69
	Gpa2 (and)	3186	237	112	+	912	94%	88%	+	+	+	-	39
2	*andSH-RGH1*	3146	230	deletion		943	93%	87%	+	+	+	-	31
	andSH-RGH3	2861	234	deletion		156	90%	83%	156	-	-	-	
	andSH-RGH5	3079	201	92		209	93%	97%	+	49	-	-	
	andSH-RGH6	2845	243	←	+	873	85%	85%	+	+	362	-	39
	andSH-RGH7	2845	230	112	+	930	93%	86%	+	+	+	+	59
2	*tubRH-RGH2*	3155	240	112		933	93%	87%	+	+	+	+	58
	tubRH-RGH3	3091	230	113		256	92%	91%	+	96	-	-	
	tubRH-RGH4	3079	205	92		209	92%	97%	+	49	-	-	
	tubRH-RGH5	2845	244	←		873	84%	97%	+	+	362	-	39
2	*lph/spl-RGH1*	3133	230	108		929	94%	88%	+	+	+	+	60
	lph/spl-RGH2	3132	230	111		928	94%	88%	+	+	+	+	59
	lph/spl-RGH3	3153	230	95		209	91%	75%	163	46	-	-	
	lph/spl-RGH4	3038	230	110		878	91%	84%	+	+	366	-	40
	lph/spl-RGH5	3112	235	112		770	93%	90%	+	+	298	-	
	lph/spl-RGH6	3139	230	113		803	93%	89%	+	+	331	-	
	lph/spl-RGH7	3191	234	112		921	92%	84%	163	+	+	-	47
6	*alb-RGH1*	3125	230	113		787	93%	88%	+	310	317	-	
	alb-RGH2	3125	230	113		910	93%	87%	+	310	+	-	39
	alb-RGH3	3066	230	112		304	91%	91%	+	144	-	-	
	alb-RGH4	3130	230	113		106	93%	93%	106	-	-	-	
	alb-RGH5	3033	233	-		928	93%	87%	+	+	+	+	55
	alb-RGH6	2967	234	deletion		304	91%	90%	+	144	-	-	
	alb-RGH7	3131	230	113		913	93%	87%	+	+	+	-	39
	alb-RGH8	3138	230	112		99	93%	89%	99	-	-	-	
	alb-RGH9	3131	230	113		912	93%	87%	+	+	+	-	39
2	*chc-RGH1*	2965	201	112		853	88%	88%	+	+	381	-	
	chc-RGH2	3140	230	113		914	93%	88%	+	+	+	-	41
	chc-RGH3	3145	234	112		932	93%	86%	161	+	+	+	58
	chc-RGH4	3139	230	112		931	93%	84%	161	+	+	+	57
	chc-RGH5	3140	230	113		914	93%	88%	+	+	+	-	41
	chc-RGH6	3131	230	113		912	93%	87%	+	+	+	-	39
	chc-RGH7	3140	230	113		914	93%	88%	+	+	+	-	41
	chc-RGH8	2286	230	110		206	69%	83%	+	46	-	-	

Ploidy	Sequence name	N length	IntronI	IntronII	cDNA isolated	aa length	n similarity to Rx1	aa similarity to Rx1	CC(aa) 160aa	NBS(aa) 312aa	LRR(aa) 392-403	acidic tail	tail aa length
5	*edn-Rgh1*	2286	230	110		206	69%	82%	+	46	-	-	
	edn-Rgh2	3091	230	113		256	92%	91%	+	96	-	-	
	edn-Rgh3	3151	240	109		933	93%	88%	+	+	+	+	58
	edn-Rgh4	3138	234	119		901	93%	89%	+	+	+	-	34
	edn-Rgh5	3125	230	113		228	93%	95%	+	68	-	-	
	edn-Rgh6	3141	230	111		913	91%	87%	+	+	+	-	39
	edn-Rgh7	3140	234	110		908	94%	88%	+	+	+	-	44
	edn-Rgh8	2286	230	110		206	69%	82%	+	46	-	-	
2	*ehr-RGH1*	2286	230	110		206	69%	82%	+	46	-	-	
	ehr-RGH2	2285	230	110		206	69%	82%	+	46	-	-	
	ehr-RGH3	2726	275	-		100	87%	76%	100	-	-	-	
	ehr-RGH4	3138	230	111		914	93%	87%	+	+	+	-	
	ehr-RGH5	2286	230	110		206	69%	82%	+	46	-	-	
	ehr-RGH6	3127	230	111		636	93%	88%	+	+	164	-	
2	*frn-RGH1*	3125	230	110		897	93%	88%	22	+	+	-	31
	frn-RGH2	3011	211	111		873	90%	86%	+	+	366	-	35
	frn-RGH3	2286	230	110		22	68%	70%	+	-	-	-	
	frn-RGH4	3146	234	113		932	94%	89%	+	+	+	++	68
	frn-RGH5	3131	230	-		912	93%	88%	+	+	+	-	39
	frn-RGH6	3148	230	92		425	91%	91%	161	264	-	-	
	frn-RGH7	3147	236	111		22	93%	70%	22	-	-	-	41
2	*hou-RGH1*	2840	230	113		814	84%	77%	60	+	+	-	41
	hou-RGH2	3159	230	111		841	91%	85%	+	308	370	-	
	hou-RGH3	3089	234	111		448	93%	90%	46	296	-	-	
	hou-RGH4	2286	230	110		206	69%	82%	+	46	-	-	
	hou-RGH5	2286	230	110		206	69%	82%	+	46	-	-	
2	*hcb-RGH1*	2817	deletion	deletion		115	81%	76%	115	-	-	-	
	hcb-RGH2	3148	230	112		209	91%	75%	163	46	-	-	
	hcb-RGH3	3150	230	95		209	91%	75%	163	46	-	-	
	hcb-RGH4	3064	230	118		903	92%	86%	+	+	374	+	57
	hcb-RGH5	3105	208	112		908	93%	88%	+	+	+	-	41
	hcb-RGH6	2810	deletion	deletion		892	93%	88%	+	+	333	-	30
	hcb-RGH7	1673	230	113		565	92%	90%	+	+	83	-	
	hcb-RGH8	2285	-	110		89	69%	72%	89	-	-	-	
	hcb-RGH9	3059	230	113		586	92%	93%	+	+	114	-	
	hcb-RGH10	3059	230	113		586	91%	91%	+	+	114	-	

Gene structure

The structure of the homologous resistance genes *Gpa2, Rx1*, *Rx2, Rx3* and *Rx4* is characterized by the presence of two introns located at the 3'end of the gene (Bendahmane et al., 1999; Bendahmane et al., 2000; Van der Vossen et al., 2000; Chapter 3, this thesis). To determine the intron and exon structure of the additional *Rx1/Gpa2 homologues*, a PCR screening was carried out on cDNA from the diploid potato clone SH using a cluster specific primer set based on the sequence of the LRR region. Restriction analysis and sequencing resulted in the identification of cDNA fragments of *andSH-RGH6* and *andSH-RGH7* indicating expression of the two putative resistance genes derived from the susceptible haplotype of SH.

Comparison of the genomic sequences of *andSH-RGH6* and *andSH-RGH7* with their cDNA sequences revealed their putative open reading frames. When comparing the intron splice sites of *andSH-RGH6* and *andSH-RGH7* with those of *Gpa2*, *Rx1*, *Rx2*, *Rx3* and *Rx4*, based on their consensus alignment the intron splice sites are very conserved. Therefore, we assumed that the nucleotide alignment of genomic sequences of all 75 homologues will determine the most likely donor and acceptor sites for the other homologues. At the intron splice sites the alignment was unambiguous and the donor and acceptor sites could be predicted (Fig. 1, Table 1). A majority of 68 genes have splice sites at the same positions as *Rx1*, *Rx2*, *Rx3*, *Rx4* or *Gpa2* resulting in the presence of two introns. There are, however, exceptions resulting in unique gene structures. In *andSH-RGH1* only intron I is spliced, because the donor site of intron II is not present. *andSH-RGH6* and *tubRH-RGH5*, which are virtually identical, have a 158 bp deletion spanning the acceptor site of intron I and a 71 bp deletion spanning the donor site of intron II resulting in the fusion of intron I and intron II. Both intron I and intron II are missing in *hcb-RGH1* and *hcb-RGH6* due to a deletion of the N-terminal region. *ehr-RGH3* is missing intron II due to a deletion in that vicinity and *andSH-RGH3* only has the donor site of intron I in common with the other homologues.

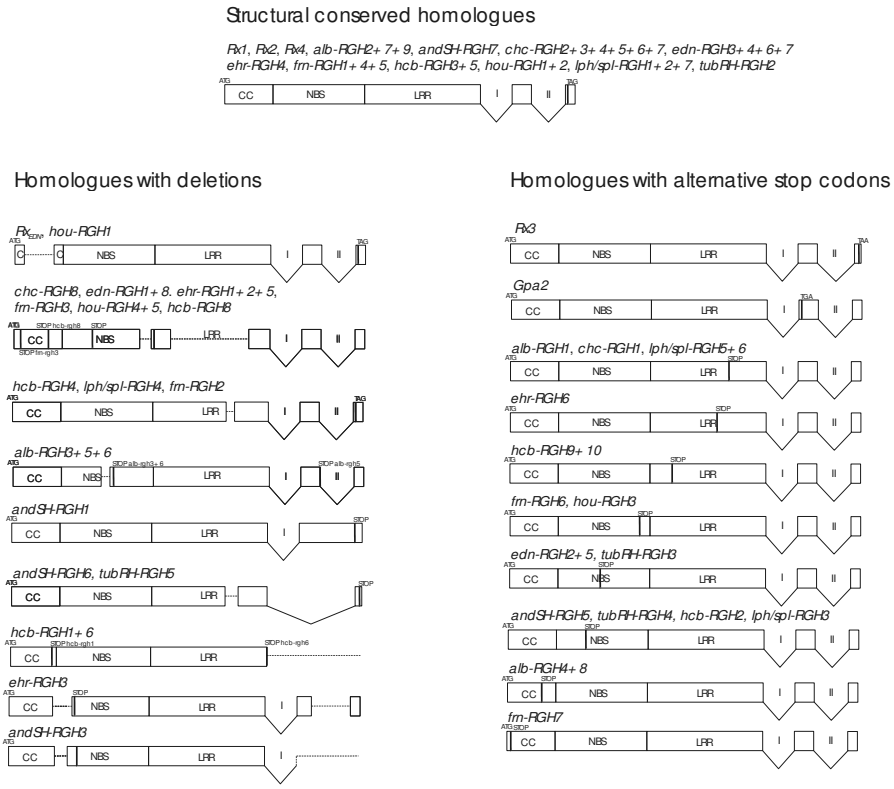


Figure 1. Schematic structural overview of the 75 *Rx* resistance gene homologues derived from *S. tuberosum* ssp. *andigena* (SH) *S. tuberosum* ssp. *tuberosum* (RH), *S. albicans (*alb), *S. acaule* (acl), *S. chacoense* (chc), *S. edinense* (edn), *S. ehrenbergii* (ehr), *S. fernandezianum* (frn), *S. hougasii* (hou), *S. huancabamabense* (hcb) and *S. lephtophyes*/*sparsipilum* (lph/spl) encoding for proteins consisting of different domains (CC=coiled-coil, NBS=nucleotide binding site, LRR=leucine rich repeat). The start- (ATG) and stopcodons (TAG, TGA, TAA or STOP) are indicated as well as the position of intron I and II. The horizontal dotted lines indicate deletions in the nucleotide sequence.

Protein structure diversity

The genes *Rx1*, *Rx2*, *Rx3*, *Rx4* and *Gpa2* encode for a predicted polypeptide of 937, 938, 937, 909 and 912 amino acids, respectively (Bendahmane et al., 1999; Bendahmane et al., 2000; Van der Vossen et al., 2000; Chapter 3, this thesis). The amino acid sequences of all other homologues were predicted by translation of the putative open reading frames based on the isolated or predicted cDNAs and encode polypeptides ranging from 22aa to 943aa (Table 1). Alignment of the 75 putative amino acid sequences resulted in a consensus sequence of 956 amino acids. They exhibit a high degree of homology with a similarity ranging from 70 to 99% compared to Rx1 (Table 1).

Proteins with a CC, NBS and LRR domain

From the 47 genes encoding for a protein with a CC, NBS and LRR domain, a majority of 42 homologues including *Rx1*, *Rx2*, *Rx3*, *Rx4* and *Gpa2* are encoding for an LRR domain with fourteen or fifteen leucine rich repeats (Table 1). *Gpa2* and *chc-RGH3* encode for an LRR domain with fifteen repeats, which are not interrupted by a linker structure in LRR 10. In LRR10 they share a methionine and a leucine preserving the intact LxxLxL motif (Fig. 2). In Rx1, Rx2, Rx3, Rx4, Rx_{edn} and 28 homologues the LxxLxL motif within LRR10 of the consensus alignment is absent due to amino acid substitutions at +748 and +753 (Fig. 2) and they therefore harbour fourteen LRR repeats separated by a linker region dividing the LRR domain in two stretches of nine and five repeats. Another five homologues have one or more LRRs deleted at the C-terminus. Chc-RGH1 and Hou-RGH2 lack one LRR repeat due to stopcodons resulting in two groups of nine and four LRRs. In Alb-RGH1, Lph/Spl-RGH5 and Lph/Spl-RGH6 the last four LRR repeats are not translated as a result of stopcodons resulting in proteins with nine and one repeat. A third group of five homologues is lacking LRR10 because of deletions occurring in this region. A 126 bp deletion in *andSH-RGH6* and *tubRH-RGH5*, a 115 bp deletion in *lph/spl-RGH4* and *frn-RGH2* and an 81 bp deletion in *hcb-RGH4* causes the loss of LRR 10 resulting again in an LRR-domain with continuous fourteen repeats.

Truncated proteins with predicted stop codons in the CC or NBS domain

The remaining 28 homologues have stopcodons within the CC or NBS domain resulting in a shorter predicted protein sequence (Fig. 1). Twenty homologues have truncated NBS domains due to stopcodons within the NBS encoding sequence. From this group, 14 homologues encode for the P-Loop, in TubRH-RGH3 and Edn-RGH2 also for the Kinase-2 motif and in Alb-RGH6 and Alb-RGH3 the Kinase3A motif is also present. Frn-RGH6 and Hou-RGH3 harbour all NBS motifs except MHD. The occurrence of full-length or truncated protein sequences and the presence or absence of the acidic tail motif was observed in homologues derived from all 10 *Solanum* species.

Increased variability in the C terminus by an acidic tail

24 *Rx/Gpa2* homologues encoding full length CC-NBS-LRR proteins exhibit increased variability of the amino acid sequence in the C terminus beyond position +924 in the

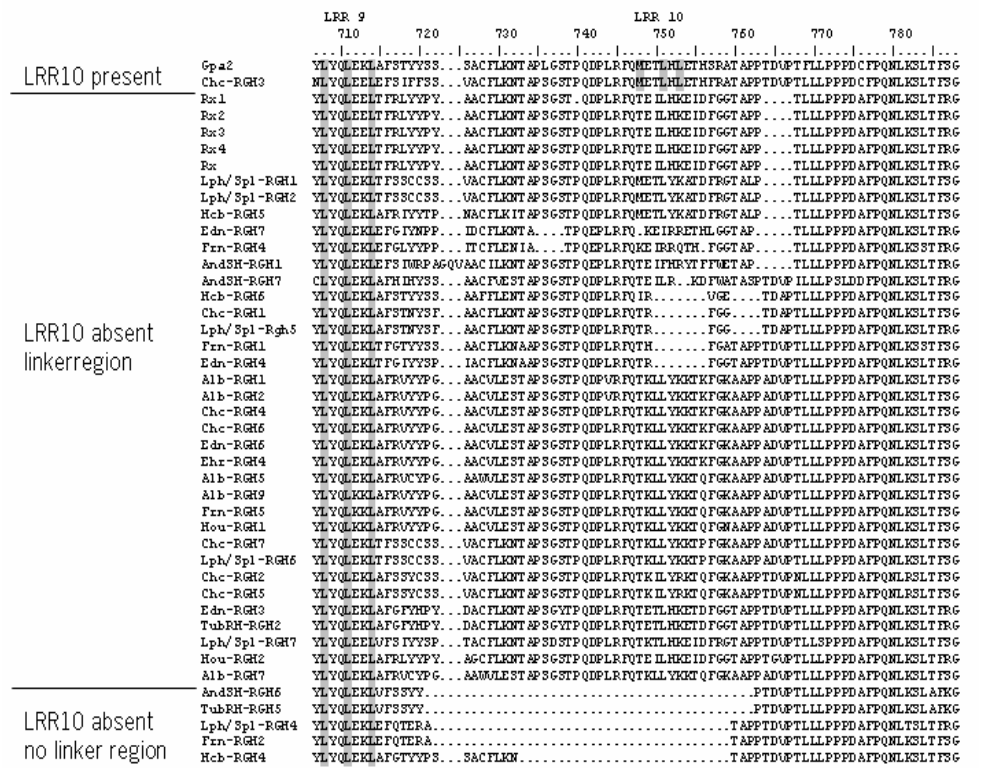


Figure 2. Protein alignment of the amino acid sequences in the region of LRR10 from Rx1/Gpa2 homologues encoding full length proteins. The amino acids encoding for an intact leucine rich repeat are shown in grey shade. The homologues are classified in three categories based on the structural features of LRR 10 as described on the left.

consensus alignment. The variability is generated by frame shifts due to three different splicing variants in intron I and various InDels. From this group sixteen homologues including Rx1, Rx2, Rx3 and Rx_{edn} sharing splicing variant 1 (frame 0) possess one or two repeats of the SVTT(V/D)EDDDD motif forming an acidic tail at the C-terminus of the protein. Lph/Spl-RGH7 and Edn-RGH7 also have splicing variant 1, but lack the acidic tail motif due to a short insertion at +3095 in the consensus alignment causing a frameshift. Furthermore, nine homologues including Rx4 share splicing variant 2 (frame +2) and therefore lack the acidic tail. Instead this subgroup has a C-terminus with two aspargines followed by one to three arginines. Additional in both splicing subgroups frequent small in-frame InDels lead to various amino acid insertions. Due to the intron fusion in *tubRH-RGH5* and *andSH-RGH6*, the deletion of introns in *hcb-RGH6* and *andSH-RGH1* and a

small deletion in *frn-RGH1,* a different structure is present at their C-terminus than those from the previous described subgroups. Gpa2 finally has a unique C-terminal protein sequence due to an insertion event causing a frame shift (Van der Vossen et al., 2000). Interestingly, with one exception in Frn-RGH4, only in the functional Rx/Gpa2 homologues Rx1, Rx2, Rx3 and in Rx_{edn} two motifs are present (Chapter 3, this thesis). It has been shown, that the acidic tail is not necessary for Rx1 and Rx4 to confer resistance to PVX (Chapter 3, this thesis).

The consensus motifs of the NBS domain are virtually identical

47 sequences encode for an NBS domain with the P-Loop, Kinase2, Kinase 3A, RNBS C, GLPL, RNBS D and MHD motifs. The protein sequence alignment shows, that these motifs are very conserved in all homologues. No other amino acids appear in the GLPL motif and MHD motif. Lph/Spl-RGH4, Frn-RGH5 and Ehr-RGH6 are the only homologues that have one amino acid exchange in the P-Loop (T177A, M171T and I174S, respectively). By transiently expressing chimeric constructs from *Rx1* and *Gpa2* it has been shown, that the CC-NBS domain is exchangeable between *Gpa2* and *Rx1* without loosing the resistance function against PVX and *G. pallida* (Slootweg et al., in preparation). Further chimeric constructs from *Rx1* with the CC-NBS domain from its paralogues *and-RGH1, 6* and *7* gave a hypersensitive response when coexpressed with the elicitor CP106 in *N. benthamiana* leaves (Slootweg et al., in preparation). In this context the conservation of motifs throughout the dataset indicates that all these homologues might have a similar mode of action conferred by the NBS domain.

Evolutionary relationships

In SH (Bakker et al., 2003) both haplotypes were found to harbour four homologues. The overview with isolated homologues per genotype from the 10 different *Solanum* species (Table 1) indicates a number of one to four homologues present per haplotype. The genetic and physical mapping of several homologues from the diploid varieties SH and RH suggested the occurrence of orthologous relationships between *tubRH-RGH4* and *andSH-RGH5* as well as between *tubRH-RGH5* and *andSH-RGH6* due to sequence similarities of 99,9 and 99,4%, respectively (Bakker et al., 2003). Based on the sequence distance dendrograms the similarity for whole genes and parts of genes indicate the presence of

various orthologous relationships (Fig. 3). Beside *tubRH-RGH4*/*andSH-RGH5* (99,8%) and *tubRH-RGH5*/*andSH-RGH6* (99,7%) the dendrogram analysis indicate orthologous groups for *lph/spl-RGH1*/*lph/spl-RGH2*, *hcb-RGH3*/*lph/spl-RGH3*, *edn-RGH3/tubRH-RGH2* and *chc-RGH7*/*lph/spl-RGH6* exhibiting nucleotide sequence similarities of 99,8%, 99,7%, 99,5% and 99,6%, respectively. Analogous to these pairs, a group of homologues comprising *edn-RGH1+8*, *ehr-RGH1+2+5*, *hou-RGH4+5*, *chc-RGH8* and *hcb-RGH8* exhibits an extreme similarity to each other ranging from 99,3 to 99,9% suggesting an orthologous relationship as well. The sequence cluster analysis using distance dendrograms (Fig. 3) further indicate that homologues isolated from different species are more similar to each other than to their paralogues from the same species.

Patterns of sequence exchange events

Comparison of the sequences indicates that numerous sequence exchanges have occurred among homologues. Using the statistical algorithms RDP, GENECONV, MaxChi, SiScan and Chimaera from the RDP software package (Martin and Rybicki, 2000) the dataset of 75 *Rx/Gpa2* homologues was analyzed for the occurrence of gene exchange events. All predicted events with a statistical significance of $p<0.05$ in one or more of the algorithms were sorted for exchange track positions, recombination break points and exchange track frequency within the dataset (Fig. 4). 229 sequence exchange tracks have been detected within 43 breakpoints in the sequence alignment. The length of the exchange tracks range from 99 bp to 2252 bp. While a majority of 193 intervals were found in one to five homologues, 36 intervals appeared more frequently and are found in six up to 50 homologues. Most of these more frequent intervals span either the CC domain (breakpoints 488, 541, 560), the combined CC-NBS domain (breakpoints 1266, 1405, 1462, 1527) or the N terminal part of the LRR domain starting before LRR10 (breakpoints 2207, 2229, 2342) (Fig. 4). Interestingly, the sequence part encoding for repeat 11-15 in the LRR domain additionally reveals 29 intervals with low frequencies among the homologues, which are situated within a relative high number of nine recombination breakpoints. The LRR region is sectioned into single leucine rich repeats by one or two breakpoints in between. However, intervals affecting the first part of the LRR domain (repeat 1-10) are rare and exhibit fewer recombination breakpoints.

In an interval within the CC-NBS domain (breakpoints 180-541) a sequence exchange

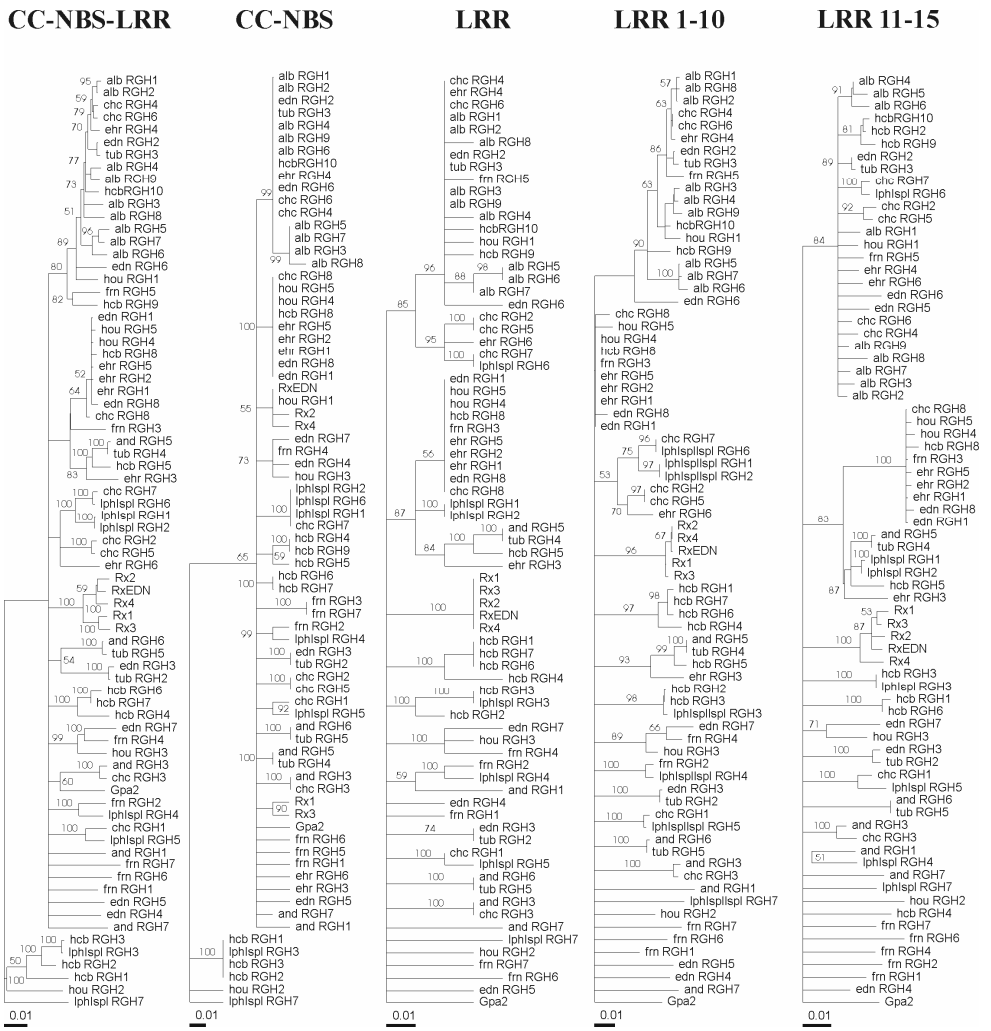


Figure3. UPGMA distance dendrograms of the nucleotide sequence alignment of 75 *Rx* homologues obtained from 10 *Solanum* species.
CC-NBS-LRR. distance dendrogram of the full length ORF alignment. (position 1-3308bp)
CC-NBS. distance dendrogram of the region encoding CC-NBS. (position 1-1462bp)
LRR. distance dendrogram of the region encoding LRR and tail. (position 1463-3308bp)
LRR1-10. distance dendrogram of the region encoding LRR 1-10. (position1463-2342bp)
LRR11-15. distance dendrogram of the region comprising LRR 11-15 + tail. (position 2343-3308bp)
The scale of the trees for 0.01 changes per nucleotide position nucleotide and bootstrap values are indicated.

event was detected for *Rx1*, *Rx2*, *Rx4* and *Gpa2*. A sequence exchange for the NBS domain region comprising the ARC subdomain (breakpoint 944-1405) was detected for *Rx2*, *Rx4* and *Gpa2*. *Rx1*, *Rx3* and Rx_{edn} share two intervals of sequence exchange starting in the ARC domain region (breakpoints 1266, 1374) and ending after the second leucine rich repeat (breakpoints 1546, 1625).

Sequence exchange has been proposed as a major mechanism shaping the diversity among resistance gene homologues in plants (Michelmore and Meyers, 1998; Meyers et al., 1999; Kuang et al., 2004). In our study, all the sequence exchange tracks were detected in homologues from at least two or up to 10 species. The mosaic structure of exchange intervals present in homologues from different species suggests a long complex history of sequence exchanges that have taken place before the speciation of the 10 *Solanum* species.

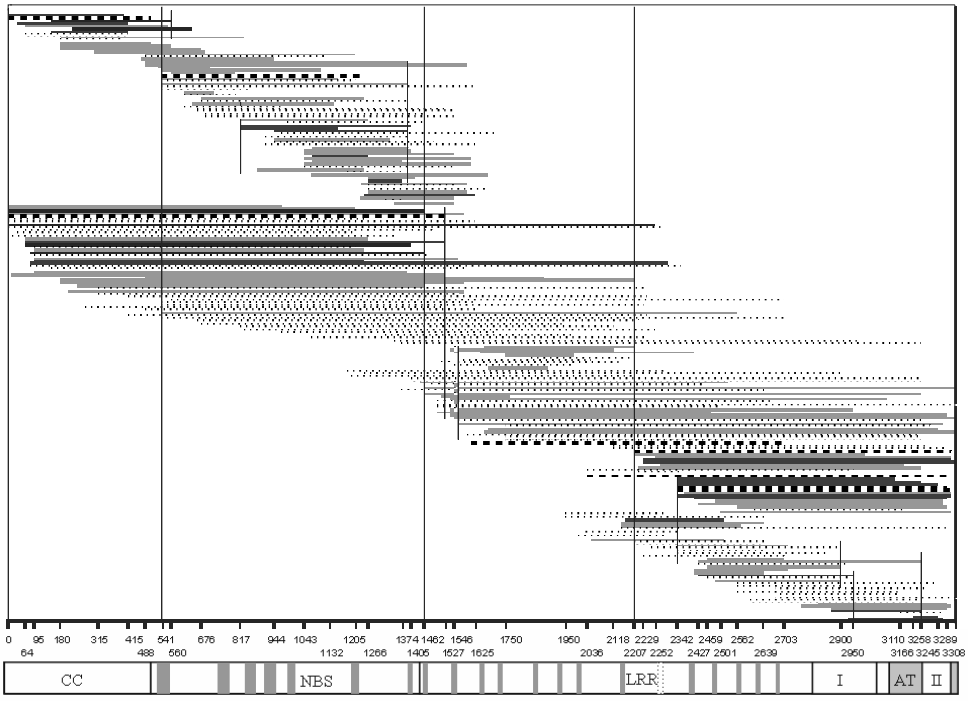


Figure 4. Stretches of nucleotide sequences being subject to sequence exchanges in one or more of 75 *Rx/Gpa2* homologues detected using RDP software package with RDP, GENECONV, MaxChi, Chimaera and SiScan. Conversion tracks are shown as horizontal bars; breakpoints are indicated as vertical bars linked to the nucleotide sequence and protein structure of Rx. The color of the conversion tracks indicates the abundance in the dataset (with pointed 1 homologue, light grey 2-5, dark grey 6-10, striped 11-20 and black >20 homologues present). Below the structural domains (CC, NBS, LRR, Intron I+II, AcidicTail) and motifs (grey bars) are indicated.

Positive selection

The *Rx/Gpa2* homologues were analyzed under the models M0:M3 and M7:M8 within the PAML software package to examine positive selection acting on single codon sites in the CC-NBS domain (Table 2) and the LRR domain (Table 3). Sequences with complex InDel structures or very short deduced amino acid sequence for the corresponding domains were omitted from the analyses resulting in datasets of 67 homologues for the CC-NBS domain analysis and 41 homologues for the LRR domain analysis. Due to decreased homology and the complex InDel structure in the C-terminal tail region we omitted this part from the analysis as well. 56 positively selected sites have been found in the CC-NBS domain and 81 sites were found to be under positive selection within the LRR domain (Table 4).

Positive selected sites in the CC-domain were found to be abundant in a region between amino acid positions +107 to +117 of the protein consensus alignment (Table 4, Fig. 5). As expected from their conservation within a diverse range of species (Van der Biezen and Jones, 1998a; Moffett et al., 2002) the NBS motifs do not contain positive selected sites with the exception of amino acid I306 in the RNBS-C motif. In the ARC domain two patches with positive selected sites were found at positions 443-455 and 401-412 (Table 4, Fig. 5).

In the LRR domain positive selection takes place mostly on codon positions following and in close vicinity to the leucine rich repeat motifs, namely after LRR 3, 4, 5, 6 and LRR 11, 12, 13, 14, 15 (Fig. 6). Under M8, no positive selected sites were found around LRR 1 and LRR 7, whereas positive selection was detected under model M3 (data not shown). LRR 10 contains 10 positively selected sites, 6 within the leucine rich repeat motif and 4 in the ß-strand/turn.

Amino acid residues specific for Rx and Gpa2

The LRR domain is involved in determining the recognition specificity of R proteins (Ellis et al., 1999; Dodds et al., 2001a; Hwang and Williamson, 2003; Rairdan and Moffett, 2006). For *Rx1*, it has been shown through sequence swaps experiments that the recognition specificity resides in the LRR C-terminal half (Rairdan and Moffett, 2006). The overall high similarity between the resistance gene homologues and the extreme conservation of the LRR domains of the functional Rx orthologues and Rx_{edn} allowed pinpointing amino acid positions potentially involved in the recognition specificities of Rx and Gpa2. Certain

Table 2. Log likelihood values and parameter estimates under models of variable ω ratios among sites for 67 CC-NBS sequences. lnL is the log likelihood value as calculated by PAML. 2ΔL is the likelihood ratio test value - here much larger than the χ^2 critical value. ω is the dN/dS ratio as estimated by maximum likelihood. P is the inferred proportion of selected sites in a class.

model	ML estimates of model parameters	lnL	LRT	2ΔL	χ^2 crit. value	degrees of freedom
M0 (one ratio)	ω= 0.89883	-8917				
M3 (discrete)	P0=0.73814 , P1=0.20981, P2=0.05204, ω0=0.30059, ω1=2.46276, ω2=8.47106	-8498	M0-M3	838	13.28	4
M7 (β)	P=0.27958 q=0.33376	-8711				
M8 (β+ω)	P0=0.87598, P=0.51193, q=0.56343 P1=0.12402, ω=5.26932	-8509	M7-M8	404	9.21	2

Table 3. Log likelihood values and parameter estimates under models of variable ω ratios among sites for 41 LRR sequences.

model	ML estimates of model parameters	lnL	LRT	2ΔL	χ^2 crit. value	degrees of freedom
M0 (one ratio)	ω=2.19500	-8692				
M3 (discrete)	P0=0.62463 , P1=0.24451, P2=0.13086, ω0=0.33466, ω1=3.71460, ω2=15.23772	-7943	M0-M3	1498	13.28	4
M7 (β)	P= 0.03007 q= 0.01748	-8426				
M8 (β+ω)	P0=0.80778, P=0.02522, q=0.02144, (P1= 0.19222), ω= 8.69415	-7978	M7-M8	896	9.21	2

Table 4. Amino acid positions under positive selection detected by PAML software package. n is the number of sequences analyzed. The length is the number of codon sites from the alignment consensus

domain	n	length	positive selected sites under M8 (amino acid positions for Rx1, - mark insertions from the alignment not present in Rx1)
CC-NBS	67	476	L 31**, K 32**, G 46*, G 50**, V 59**, N 75 , V 95**, V 107, - 112, T 113**, S 114**, D 115, - 117a, S 125, L 130*, D 134**, V 135**, R 145**, A 157**, T 181**, C 188*, S 202, Y 205**, C 206**, L 211**, L 212, G 213**, L 217**, T 218, D 224**, Y 263, L 269, R 273*, D 296**, H 303, I 306**, M 345**, G 346**, Q 347**, R 348**, R 354**, I 355**, P 367**, Q 370**, Q 401*, S 403*, N 405, E 406*, P 412**, T 430, F 443*, F 447**, R450**, T 452**, E 454**, S 455
LRR	41	416	H 505*, K 506**, A 511, L 524**, G 525**, G 526**, - 526a, - 526b**, E 528**, C 529**, V 530*, L 546, T 548**, W 549**, I 551, N 570**, L 573**, Q 574, Q 577**, G 578, K 580**, S 585, S 586, N 606**, P 610**, S 611**, - 611a**, - 611b**, Y 612**, T 629**, C 631**, G 633**, R 638, Q 678**, P 683**, S 689**, L 692, R 696, Y 697, E 703, T 705**, R 707**, L 708**, Y 709**, Y 710**, P 711**, Y 712**, A 713**, K 718, N 719, T 734**, E 735**, I 736**, L 737**, H 738**, K 739*, E 740**, D 742**, G 744**, G 745**, R 769**, S 773**, I 792, W 795**, A 797**, I 799**, K 801*, D 820**, V 821**, Y 822**, S 829**, D 843**, R 845**, N 846**, R 852**, D 866**, Y 867**, Q 869**, V 873**

amino acid residues are unique for the four functional orthologues of Rx (Table 5). There are in total 6 amino acid residues that exclusively appear in functional Rx orthologues, all of them are located in the LRR domain. From these candidate positions for a putative function two are situated in the solvent exposed residues of LRR 2 and in LRR 14, respectively. Another three of these residues are located in positions that were found to be under positive selection in the consensus alignment of the 75 homologues, and are located after the LRRs 3, 12 and 14.

Similar to the functional Rx orthologues, the functional nematode resistance gene *Gpa2* also exhibits a number of unique amino acid residues compared with all other Rx/Gpa2 homologues (Table 5). There are 11 unique amino acid residues situated in the LRR domain of Gpa2. Three residues of these unique residues were found to be under positive selection and are located after the LRRs 5, 8 and 14. Three unique amino acid residues are situated in the solvent exposed residues of LRR 7, 10 and 14.

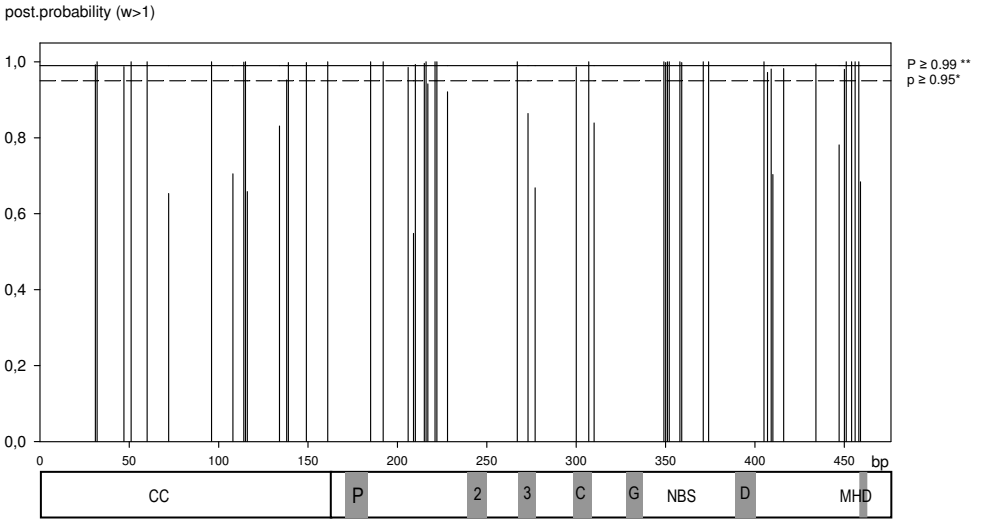


Figure 5. Distribution of codon positions being under diversifying selection within the CC-NBS domains. Positively selected codon positions detected by M8 are indicated as black bars, additionally positions detected under M3 but not under M8 are indicated as grey bars. The positions of structural motifs P-Loop (P), Kinase 2 (2), Kinase 3A (3), RNBS-C (C), GLPL (G) and MHD (MHD) are indicated.

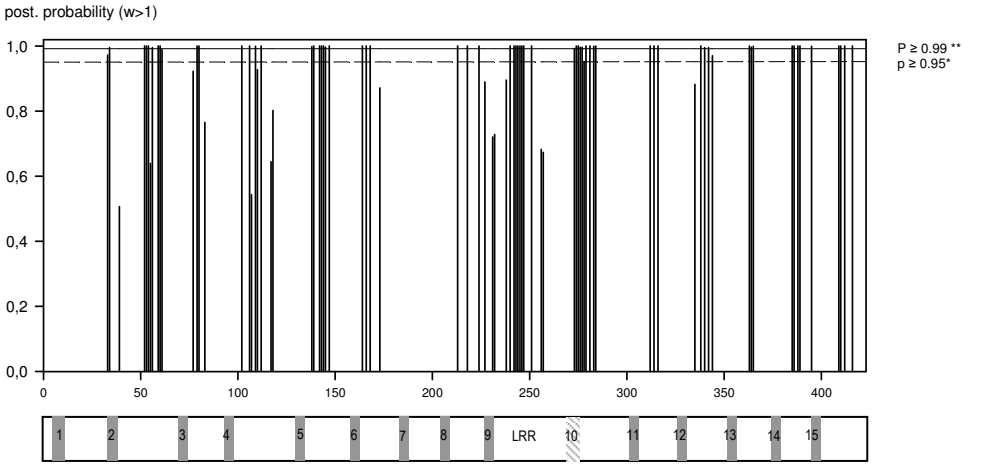


Figure 6. Distribution of codon positions being under diversifying selection within the LRR domain. Positively selected codon positions detected by M8 are indicated as black bars, additionally positions detected under M3 but not under M8 are indicated as grey bars. The positions of the 15 LRR repeat motifs are indicated (1-15). Additionally LRR10, present only in a subset, is marked by dashed lines.

Table 5. The protein sequence alignment of the 75 resistance gene homologues revealed unique amino acid residues for functional Rx orthologues and Gpa2. Amino acid positions also found to be under positive selection by PAML are underlined.

	Number of unique residues	Unique amino acid residues (positions for Rx1 and Gpa2, respectively)
Rx (Rx1, Rx2, Rx3, Rx4, Rx_{edn})	11	R 494, V(L) 507, W 549, I 792, I 840, R 845
Gpa2 (Gpa2)	11	R 572, C 612, E 657, L 668, K 683, L 723, S 744, F 756, S 825, F 850, F 871

Discussion

The structural relationship of the full-length ORF of 75 resistance gene homologues from 10 different *Solanum* species has been investigated to obtain more insight in the role of the different functional domains of *Rx/Gpa2* homologues. In addition, comparison of the nucleotide and amino acid sequences revealed patterns that reflect the mechanisms underlying the evolution of *R* genes.

A multiple sequence alignment from all 75 homologues showed that intron splice-site positions are very conserved among them and through *in silico* splicing the identification of putative open reading frames of all the homologues was possible. A majority of the resistance gene homologues have open reading frames in a comparable size and structure as *Rx1, Rx2, Rx3, Rx4*, Rx_{edn} and *Gpa2*, which indicates that these homologues may encode for functional R proteins of which the recognition specificity is unknown. The *Rx/Gpa2* homologues resemble the so called Type I resistance gene homologues, for which also a high frequency of sequence exchanges was observed (Kuang et al., 2004).

While many TIR-NBS-LRR proteins have an additional larger and variable C-terminal domain past the LRR domain (Bendahmane et al., 2000; Mondragón-Palomino et al., 2002; Meyers et al., 2003) CC-NBS-LRR proteins typically tail into relatively conserved short peptides. Remarkably, in Rx1, Rx2, Rx3, and Rx_{edn} a characteristic C-terminal acidic tail consisting of two direct sequence repeats has been found (Bendahmane et al., 1999; Bendahmane et al., 2000; Chapter 3, this thesis). However, it has been shown that the acidic tail is dispensable for Rx functioning as is the case in the functional orthologue Rx4 from *S. albicans* (Chapter 3, this thesis). Frn-RGH4 is the first known homologue of Rx having the two sequence repeats of the acidic tail, while lacking the Rx resistance function. While a majority of the found ORFs including the functional resistance

gene Gpa2 are not encoding for an acidic tail, further 11 homologues have a single sequence repeat. The role of the acidic tail remains unrevealed, although a possible function through modifying the strength of the resistance response is suggested (Bendahmane et al., 1999).

Within different *R* gene homologue datasets the LRR domain has been shown to contain variable numbers of leucine rich repeats in a module-like construction, which may result in a horseshoe-shaped structure analogous to those of the porcine ribonuclease inhibitor LRR protein (Kobe and Deisenhofer, 1993; Jones, 1997). Similarly, projection of the amino acid sequence of the *A. thaliana* resistance gene RPS5 on the crystal structure of the bovine decorin LRR protein resulted in the prediction of a concave horseshoe structure (McHale et al., 2006). Compared to human or animal LRR proteins, the leucine rich repeat motifs of *R* genes including *Rx1* and *Gpa2* (Jones, 1997; Bendahmane et al., 1999; Van der Vossen et al., 2000) are imperfect which may influence the symmetry of the horseshoe. In our dataset the different homologues are encoding for an LRR domain with a relatively conserved number of 14 to 15 LRR-repeats. An interesting structural variability was found in the region encoding LRR 10, dividing the homologues in 3 structural subclasses. They differ in their structure having either 15 leucine rich repeats, 14 leucine rich repeats divided by a linker region or 14 leucine rich repeats without linker region. These features will have an impact on the overall topology of the LRR domain tertiary structure. A central linker region may break up the horseshoe structure into two separate subdomains. Within the dataset of 75 *Rx/Gpa2* homologues, the majority of homologues including the functional *Rx* orthologues belong to the subclass with a linker.

Sequence exchange by crossing-over events and gene conversion have been found to be an important source of variation (Meyers et al., 1998; Kuang et al., 2004). The analysis here revealed that the 75 *Rx/Gpa2* homologues from different species have been subject to frequent events of sequence exchanges. Sequence exchanges were not exclusive to certain domains, although we found differences in abundance and track lengths. While some sequence exchanges have the potential to create new structures with advantageous functions it is also true that functions can be disrupted. In multi-domain proteins such as the CC-NBS-LRR class the expectation is therefore to find sequence exchange tracks at specific sites that separate domains with independent functions. Breakpoints are indeed detected at the edges of predicted protein domain structures and support their classification.

A conversion track spanning the second part of the NBS region and the N-terminus of the LRR domain was observed for all functional *Rx* orthologues, *Gpa2* and several other homologues. In domain swap experiments with *Rx1* and *Gpa2* it was shown that independent shuffling of ARC2 and LRR domains resulted in autoactive chimeras (Rairdan and Moffett, 2006; Slootweg et al., in preparation). Such limited compatibility could select for the co-exchange of the ARC2 and N-terminal LRR subdomains observed in various Rx/Gpa2 homologues.

A modular exchange pattern of leucine rich repeats is observed for the dataset of the Rx/Gpa2 homologues. This is very obvious in the C-terminal part of the LRR domain from repeat 10 to 15. The occurrence of frequent sequence exchanges of short, single blocks encoding one or more leucine rich repeats has been previously reported from the RGC2 dataset from *L. sativa* (Kuang et al., 2004). The modular shuffling of single repeats in the LRR domain is an attribute of Type I genes as well (Kuang et al., 2004) thereby creating potentially new recognition specificities. Indeed, sequence swaps experiments with *Rx1* and *Gpa2* indicate that the recognition specificity of *Rx* resides in the LRR C-terminal half (Rairdan and Moffett, 2006). Similar results through sequences swaps have been obtained from the resistance genes *Mla1* and *Mla6* in barley (Shen et al., 2003).

Here we report the occurrence of sequence tracks having been subject to exchange present in homologues isolated from different *Solanum* species indicating interallelic exchanges before the speciation. Similar to previously found allelic pairs within the diploid genotypes SH and RH, several homologues from different species were found to be nearly identical in nucleotide sequence indicating phylogenetic orthologous relationship. The overall structural similarity and the identification of putative orthologous alleles suggest that the amplified *Rx/Gpa2* homologues could be located at similar positions on chromosome XII as the *Rx/Gpa2* homologues in SH and RH (Bakker et al., 2003). However, *Rx2* from *S. acaule* has been shown to map on chromosome V (Ritter et al., 1991). Furthermore, frequent sequence exchange between paralogues prevents the correct identification of phylogenetic relationships within the homologue dataset. Obscured orthologous relationships as a result of frequent sequence exchange has been identified as a feature of Type I genes (Kuang et al., 2004).

Positive selection pressure has been documented for many *R* gene datasets and has mainly been found in the LRR domain. Using PAML M0-M3 and M7-M8 models we find

significant positive selection for the *Rx/Gpa2* homologues across both CC-NBS and LRR with a high level of confidence. The heterogenous distribution of positively selected sites was clear from the comparison of the proportion of sites under selection within CC-NBS and LRR regions. As tests based on the ω ratio are most likely conservative (Yang, 2006) and we chose the most conservative methods (using M7-M8 and splitting CC-NBS from LRR analysis) we expect the number of sites considered significant for positive selection to be a conservative estimate.

The division of the LRR domain in an N-terminal and C-terminal part, as discussed above, is also supported by the pattern of positive selection acting on the codon sites in LRR 2-6 and 11-15. The linker region in LRR 10 contains sites under strong positive selection reflecting the structural variability. Comparison of Rx1, Rx2, Rx3, Rx4 and Rx_{edn} resulted in the identification of high sequence conservation in the LRR domain implicating strong evolutionary constraints imposed by PVX for functional *Rx* resistance in *Solanum*. However, as expected, analysis of the functional *Rx* orthologues together with a wider set of homologues reveals very significant positive selection on various codon sites within the LRR domain, reflecting the selection imposed by other pathogen derived effector molecules.

A genome-wide analysis of the NBS-LRR genes of *A. thaliana* revealed distinct patterns of positive selection for the CC/TIR, NBS and LRR domains for different phylogenetic groups (Mondragón-Palomino et al., 2002). The topology of positive selected sites along the LRR domain of the *Rx/Gpa2* homologues is very similar to group 14, which contains together with *RPP8* 11 CC-NBS-LRR genes. Analogous to the *Rx/Gpa2* homologues, positive selected sites are situated in small groups after the leucine rich repeat motifs. However the frequency and pattern of positive selected sites in the CC-NBS domain in the *Rx* homologue dataset was not observed in any group of the NBS-LRR genes from *A. thaliana*. A reason for this could be the origin of the datasets. Our dataset contains a higher number of homologues with high overall similarity and is derived from different species. The Arabisopsis NBS-LRR dataset comprises a wide array of different NBS-LRR subclasses (Mondragón-Palomino et al., 2002; Kuang et al., 2004) with a medium number of entries per group and is derived from a single species.

Based on the mapping data of the *Rx* locus in SH (Bakker et al., 2003) and the number of amplified Rx/Gpa2 homologues, we have a tentative indication for a haplotype

structure consisting of one to four homologues per chromosome in *Solanum*. In the PVX-susceptible diploid potato clone RH, respectively, two and four *Rx/Gpa2* homologues were identified on both haplotypes (Bakker et al., 2003). At higher ploidy levels the number of Rx/Gpa2 homologues per chromosome seems to decrease; ten homologues were amplified from *S. albicans* (hexaploid) and 9 homologues from *S. edinense* (pentaploid). This could be an experimental artifact but could also have a biological origin. With the increasing chance of harbouring identical orthologues in genotypes of higher ploidy, such alleles with redundant functions become dispensable increasing the chance on deletions. However, due to limitations of a PCR based isolation strategy, *Rx/Gpa2* homologues with variation in the ORF flanking untranslated regions could remain undetected. Further research directing the *Rx/Gpa2* locus for different haplotypes is necessary to address this question.

Similar to *R* genes from the *P* locus in flax (Dodds et al., 2001b), the *Rx/Gpa2* homologues show a high degree of sequence identity. Here, it was shown that the LRR domain of the functional *Rx* orthologues is under strong negative selection pressure imposed by the recognition function, while positive selection was detected by analyzing all *Rx/Gpa2* homologues. Combining both results, we identified potential amino acids involved in the pathogen recognition function for *Rx* and *Gpa2*. For *R* proteins a protein-protein interaction involving the LRR domain, either direct to a pathogen effector (Deslandes et al 2003, Dodds 2006, Jia 2000) or indirect in a guard-complex (Van der Biezen and Jones, 1998b; Van der Hoorn et al., 2002; reviewed by Innes, 2004) has been proposed. A study on constructs of *Rx1* leading to recognition of the virulent PVX strain suggests the possibility of a direct interaction of LRR-CP for Rx (Farnham and Baulcombe, 2006). Previously it has been shown that the susceptibility to the virulent strain of PVX is based on changes in two amino acid positions of the coat protein (Goulden et al., 1993). Proposing the pathogen escapes a direct interaction through mutation in two amino acid positions, a corresponding small number of ligand-binding positions may be present in Rx. Similarly, recently it was shown that three independant mutation constructs based on single amino acid changes caused the recognition of Rx to be extended to the virulent coat protein from PVX (Farnham and Baulcombe, 2006). Further analysis like functional mutation analysis and 3D modelling of the LRR structure could enlighten the role of the unique polymorphic amino acid residues for Rx and Gpa2.

Material and Methods

cDNA preparation and PCR screening

Total RNA was isolated from potato clone SH83-92-488 (Rouppe Van der Voort et al., 1997) using the TRIzol LS reagent (Invitrogen, San Diego, CA, USA) according to manufacturers instructions. From the total RNA, mRNA was subtracted using the Oligotex mRNA minikit (Quiagen, Hilden, Germany) according to manufacturers instructions. With the Marathon cDNA Amplification Kit (Clonetech, CA, USA) a 3' RACE was carried out as described (Bendahmane et al., 1999). In the first amplification round a *Gpa2/Rx1* LRR specific forward primer (Bakker et al., 2003) was used together with adapter primer AP1, provided with the kit. In this round, the following cycle file was applied: 2 min 94ºC followed by 5 s 94ºC, 4 min 72ºC for 4 cycles, then 5 s 94ºC, 4 min 70ºC for 4 cycles, and finally 5 s 94ºC, 4 min 68ºC for 24 cycles. The second amplification round was done using the LRR-F primer and a *Gpa2/Rx1* LRR specific reverse primer (Bakker et al., 2003) and the following PCR cycle file: 3 min 94ºC followed by 30 s 94ºC, 30 s 55ºC, 90 s 72ºC for 10 cycles, 30 s 94ºC, 30 s 55ºC, 90 s 72ºC with an extension of 5 s/cycle for 25 cycles and finally 5 min elongation at 72ºC.

The resulting PCR product was cloned into the PCR2.1TOPO-vector for transformation of *E. coli* TOP10 cells according to manufacturers instructions (Invitrogen, San Diego, CA, USA). Seventy positive clones were cultivated overnight at 37ºC in 2 ml liquid LB medium supplemented wit 100 µg/ml ampicillin for selection. Each clone was then reamplified and digested with *Taq*I. The DNA fragments were separated on a 4% agarose gel consisting of 1% ultra pure agarose (Life Technologies, Breda, The Netherlands) and 3% NuSieve ® GTG ® agarose (FMC, Philadelphia, PA, USA) in 1× TAE buffer at 120 V. Clones were selected for further analysis based on differences in the digestion patterns. To avoid sequencing of PCR artefacts (Bakker et al., 2003), 3 clones were selected for each pattern represented by more than one clone. Sequencing was carried out by Greenomics, Wageningen, The Netherlands and sequence assembling was done with the software package Vector NTI (Informax, Inc. Bethesda, Maryland, USA).

Sequence analysis

Sequences of the resistance gene homologues *chc-RGH1* to *8*, *edn-RGH1* to *8*, *ehr-RGH1* to *6*, *frn-RGH1* to *7*, *hou-RGH1* to *5* and *hcb-RGH1* to *10* are obtained as described (Chapter 3, this thesis). Sequencing of the PCR products was carried out by Greenomics, PRI, Wageningen, The Netherlands. Sequence assembling was done using ContigExpress from the Vector NTI suite software package (InforMax Inc., Bethesda, USA). The isolation of *andSH-RGH5*, *6* and *7 tubRH-RGH2* to *5, alb-RGH1* to *9, lph/spl-RGH1* to *7*, *Rx3*, *Rx4* and Rx_{edn} has been described previously.
Sequences of the resistance gene homologues *andSH-RGH1, and SH-RGH3* and the resistance genes *Gpa2* (Van der Vossen et al., 2000), *Rx1* (Bendahmane et al., 1999) and *Rx2* (Bendahmane et al., 2000) are obtained from genbank (accession numbers AF266747, AF266746, AF195939, AJ011801, AJ249448, respectively).
Nucleotide and deduced amino acid sequences were aligned with Clustal W (Chenna et al.,

2003) and analyzed by eye using Vector NTI suite 8.0 (InforMax Inc., Bethesda, USA) and BioEdit version 6.0.5 (Hall 1999).
Using the multiple nucleotide alignment, trees were constructed with the program PAUP* v.4.0b10 (Swofford, 1999) implementing the unweighted pair group method with arithmetic mean (UPGMA) and uncorrected distance model. Bootstrap values were calculated for 1000 replicas.

Positive selection analysis
The ratio was estimated with the codeml program of PAML (phylogenetic analysis by maximum likelihood) (Yang, 1997; Yang and Bielawski, 2000). Two models of fitting codon substitution were used to calculate likelihood ratio statistics (LR), twice the log-likelihood between models is compared with the value of a χ^2 distribution with branches-1 degrees of freedom.
Model M7 (β distributed variable selection pressure) has an ω for each site drawn from a β distribution with parameters *p* and *q*. Model M8 (β plus ω > 1) uses the M7 recipe for a fraction p_0 of the sites and assigns another ω to the remaining fraction. M7 and M8 are nested models, so they can be compared using a likelihood ratio test (LRT) which is generally robust to the assumed distribution of ω over sites. When M8 fits the data significantly better than M7 and the ω ratio estimated under model M8 is greater than 1, we need to ask whether it is significantly greater than 1. To do this, we recalculate the log-likelihood value in M8 while fixing ω to be 1 (model M8A from Swanson et al., 2003)) and compare the change in likelihood with a χ^2 distribution with 1 degree of freedom. Likewise we tested the less complicated models M0 (uniform selective pressure among sites) with M3 (variable selective pressure among sites) and found the results to be less conservative than M7/M8 in its estimates.

Next we tested for positive selection by studying variation among sites identifying amino acids under diversifying selection. This variation is tested with an additional LR test between M7 and M8 (Yang and Nielsen, 2000) using the empirical Bayes theorem as implemented in PAML to calculate the posterior probability that a particular amino acid belongs to a particular class (neutral, negative or positive). A particular site that belongs to the class ω>1 with a posterior probability > 95% is most likely under positive selection. This approach makes it possible to detect positive selection and identify sites under positive selection even if the average ω ratio over all sites is less than 1 (Yang, 2006). Meanwhile, for this type of study it is important to note three test characteristics. First, detection of positive selection requires significant differences between M7 and M8 and estimates of ratio that exceed 1. Second, under M8 it is possible to estimate the proportion of sites that are under positive selection, and this proportion is denoted P1. Third, the application of these models requires a topological or phylogenetic assumption. For each sequence group, PAML analyses were applied using the M0 generated phylogenetic tree. The amino acid sequence alignment was executed by ClustalX (v1.83) (Chenna et al., 2003) and pal2nal (v11) was used to relate the sequences back to a nucleotide alignment. pal2nal is a program that converts a multiple sequence alignment of proteins and the corresponding DNA (or mRNA) sequences into a codon-based DNA (nucleotide) alignment. The program automatically assigns the corresponding codon sequence even if the input nucleotide sequence has mismatches with the input protein sequence, or contains UTRs, polyA tails. It

can also deal with frameshifts in the input alignment, which is suitable for the analysis of pseudogenes. The resulting codon-based DNA alignment can further be subjected to the calculation of synonymous and non-synonymous substitution rates (Suyama et al., 2006). The results, including gaps, were directly used by PAML.

For the CC-NBS we compared 67 sequences and for the LRR 41 sequences. The domain analysis was split, as opposed to full-sequence analysis, as it optimized the number of sequences we could align. Using over 40 sequences of this size, with this level of homology, the statistical power of PAML is high enough.
We used PAML version 3.15 compiled with "-O4 -funroll-loops –fomit -frame -pointer -finline-functions" to the gcc 4.0.3 20051201 (Debian 32-bits Linux 4.0.2-5) compiler on an AMD Athlon(tm). All models were run using codon information, freq. F3X4, one ω for all branches and estimated κ (transition/transversion rate) values.

Gene exchange analysis
The multiple nucleotide alignment was analyzed for sequence exchange events using the statistical algorithms RDP, GENECONV (Padidam et al., 1999), MaxChi (Smith, 1992), Chimaera (Posada and Crandall, 2001) and SiScan (Gibbs et al., 2000) from the RDP software package (Martin and Rybicki, 2000). All predicted events with a statistical significance $p<0.05$ in one or more of the algorithms were sorted for exchange track position, recombination break points and exchange track frequency within the dataset.

Acknowledgement
We like to thank R. Hoekstra (CGN, Wageningen, the Netherlands) and V. Vleeshouwers (Laboratory of Plant Breeding, Wageningen University, the Netherlands) for providing seeds and *in vitro* plant material from wild potato species, respectively. This work was supported by a grant from the Bosch Foundation, EU project APOPHYS QLRT-2001-01849 and the Centre of Biosystems Genomics (CBSG), Wageningen, the Netherlands.

References

Bai, J., Pennill, L.A., Ning, J., Lee, S.W., Ramalingam, J., Webb, C.A., Zhao, B., Sun, Q., Nelson, J.C., Leach, J.E., and Hulbert, S.H. (2002). Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. Genome Research **12,** 1871.

Bakker, E., Butterbach, P., Rouppe Van der Voort, J., Van der Vossen, E., Van Vliet, J., Bakker, J., and Goverse, A. (2003). Genetic and physical mapping of homologues of the virus resistance gene *Rx1* and the cyst nematode resistance gene *Gpa2* in potato. Theoretical and Applied Genetics **106,** 1524.

Bendahmane, A., Kanyuka, K., and Baulcombe, D.C. (1999). The *Rx* gene from potato controls separate virus resistance and cell death responses. Plant Cell **11,** 781.

Bendahmane, A., Kanyuka, K., Baulcombe, D.C., Querci, M., and Bendahmane, A. (2000). Agrobacterium transient expression system as a tool for the isolation of disease resistance genes: Application to the *Rx2* locus in potato. Plant Journal **21,** 73.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Research **31,** 3497.

Dangl, J.L., and Jones, J.D.G. (2001). Plant pathogens and integrated defence responses to infection. Nature **411,** 826.

Dodds, P.N., Lawrence, G.J., and Ellis, J.G. (2001a). Six amino acid changes confined to the leucine-rich repeat β-strand/β-turn motif determine the difference between the *P* and *P2* rust resistance specificities in flax. Plant Cell **13,** 163.

Dodds, P.N., Lawrence, G.J., and Ellis, J.G. (2001b). Contrasting modes of evolution acting on the complex *N* locus for rust resistance in flax. Plant Journal **27,** 439.

Dodds, P.N., Lawrence, G.J., Catanzariti, A.M., Teh, T., Wang, C.I.A., Ayliffe, M.A., Kobe, B., and Ellis, J.G. (2006). Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. Proceedings of the National Academy of Sciences of the United States of America **103,** 8888.

Ellis, J.G., Lawrence, G.J., Luck, J.E., and Dodds, P.N. (1999). Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. Plant Cell **11,** 495.

Farnham, G., and Baulcombe, D.C. (2006). Artificial evolution extends the spectrum of viruses that are targeted by a disease-resistance gene from potato. Proceedings of the National Academy of Sciences of the United States of America **103,** 18828.

Gibbs, M.J., Armstrong, J.S., and Gibbs, A.J. (2000). Sister-scanning: A Monte Carlo procedure for assessing signals in rebombinant sequences. Bioinformatics **16,** 573.

Goulden, M.G., Kohm, B.A., Cruz, S.S., Kavanagh, T.A., and Baulcombe, D.C. (1993). A feature of the coat protein of potato virus X affects both induced virus resistance in potato and viral fitness. Virology **197,** 293.

Hulbert, S.H., Webb, C.A., Smith, S.M., and Sun, Q. (2001). Resistance gene complexes: Evolution and utilization. Annual Review of Phytopathology **39,** 285.

Hwang, C.F., and Williamson, V.M. (2003). Leucine-rich repeat-mediated intramolecular interactions in nematode recognition and cell death signaling by the tomato resistance protein Mi. Plant Journal **34,** 585.

Innes, R.W. (2004). Guarding the goods. New insights into the central alarm system of plants. Plant Physiology **135,** 695.

Jones, D.A.D.A. (1997). The role of leucine-rich repeat proteins in plant defences Adv. Bot. Res **24,** 89.

Jones, J.D.G., and Dangl, J.L. (2006). The plant immune system. Nature **444,** 323.

Kobe, B., and Deisenhofer, J. (1993). Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. Nature **366,** 751.

Kuang, H., Woo, S.S., Meyers, B.C., Nevo, E., and Michelmore, R.W. (2004). Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. Plant Cell **16,** 2870.

Leister, D. (2004). Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends in Genetics **20,** 116.

Martin, D., and Rybicki, E. (2000). RDP: Detection of recombination amongst aligned sequences. Bioinformatics **16,** 562.

McDowell, J.M., and Simon, S.A. (2006). Recent insights into *R* gene evolution. Molecular Plant Pathology **7,** 437.

McHale, L., Tan, X., Koehl, P., and Michelmore, R.W. (2006). Plant NBS-LRR proteins: Adaptable guards. Genome Biology **7**.

Meyers, B.C., Kaushik, S., and Nandety, R.S. (2005). Evolving disease resistance genes. Current Opinion in Plant Biology **8,** 129.

Meyers, B.C., Shen, K.A., Rohani, P., Gaut, B.S., and Michelmore, R.W. (1998). Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell **10,** 1833.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. (2003). Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell **15,** 809.

Meyers, B.C., Michelmore, R.W., Sivaramakrishnan, S., Dickerman, A.W., Sobral, B.W., and Young, N.D. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant Journal **20,** 317.

Michelmore, R.W., and Meyers, B.C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Research **8,** 1113.

Moffett, P., Farnham, G., Peart, J., and Baulcombe, D.C. (2002). Interaction between domains of a plant NBS-LRR protein in disease resistance-related cell death. EMBO Journal **21,** 4511.

Mondragon-Palomino, M., and Gaut, B.S. (2005). Gene conversion and the evolution of three leucine-rich repeat gene families in *Arabidopsis thaliana*. Molecular Biology and Evolution **22,** 2444.

Mondragón-Palomino, M., Meyers, B.C., Michelmore, R.W., and Gaut, B.S. (2002). Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. Genome Research **12,** 1305.

Padidam, M., Sawyer, S., and Fauquet, C.M. (1999). Possible emergence of new geminiviruses by frequent recombination. Virology **265,** 218.

Pan, Q., Wendel, J., and Fluhr, R. (2000). Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. Journal of Molecular Evolution **50,** 203.

Parniske, M., Golstein, C., Thomas, C.M., Wulff, B.B.H., Hammond-Kosack, K.E., Jones, D.A., Harrison, K., and Jones, J.D.G. (1997). Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato. Cell **91,** 821.

Posada, D., and Crandall, K.A. (2001). Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. Proceedings of the National Academy of Sciences of the United States of America **98,** 13757.

Rairdan, G.J., and Moffett, P. (2006). Distinct domains in the ARC region of the potato resistance protein Rx mediate LRR binding and inhibition of activation. Plant Cell **18,** 2082.

Ritter, E., Debener, T., Barone, A., Salamini, F., and Gebhardt, C. (1991). RFLP mapping on potato chromosomes of two genes controlling extreme resistance to potato virus X (PVX). Molecular and General Genetics **227,** 81.

Rose, L.E., Langley, C.H., Michelmore, R.W., Bittner-Eddy, P.D., Holub, E.B., and Beynon, J.L. (2004). The Maintenance of Extreme Amino Acid Diversity at the Disease Resistance Gene, RPP13, in Arabidopsis thaliana. Genetics **166,** 1517-1527.

Rouppe van der Voort, J., Wolters, P., Folkertsma, R., Hutten, R., Van Zandvoort, P., Vinke, H., Kanyuka, K., Bendahmane, A., Jacobsen, E., Janssen, R., and Bakker, J. (1997). Mapping of the cyst nematode resistance locus *Gpa2* in potato using a strategy based on comigrating AFLP markers. Theoretical and Applied Genetics **95,** 874.

Shen, Q.H., Zhou, F., Bieri, S., Haizel, T., Shirasu, K., and Schulze-Lefert, P. (2003). Recognition specificity and RAR1/SGT1 dependence in barley *Mla* disease resistance genes to the powdery mildew fungus. Plant Cell **15,** 732.

Smith, J.M. (1992). Analyzing the mosaic structure of genes. Journal of Molecular Evolution **34,** 126.

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Research **34**.

Swanson, W.J., Nielsen, R., and Yang, Q. (2003). Pervasive adaptive evolution in mammalian fertilization proteins. Molecular Biology and Evolution **20,** 18.

Swofford, D.D. (1999). PAUP* Version4.0b10.

Van der Biezen, E.A., and Jones, J.D. (1998a). The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. Current biology : CB **8,** pp. R226.

Van der Biezen, E.A., and Jones, J.D.G. (1998b). Plant disease-resistance proteins and the gene-for-gene concept. Trends in Biochemical Sciences **23,** 454.

Van der Hoorn, R.A.L., De Wit, P.J.G.M., and Joosten, M.H.A.J. (2002). Balancing selection favors guarding resistance proteins. Trends in Plant Science **7,** 67.

Van der Vossen, E.A.G., Rouppe Van der Voort, J.N.A.M., Kanyuka, K., Bendahmane, A., Sandbrink, H., Baulcombe, D.C., Bakker, J., Stiekema, W.J., and Klein-Lankhorst, R.M. (2000). Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: A virus and a nematode. Plant Journal **23,** 567.

Yang, J.P.B.a.Z. (2005). Statistical Methods in Molecular Evolution, chapter 5 Adaptive Protein Evolution. (Springer).

Yang, Z. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences **13,** 555.

Yang, Z. (2006). Computational Molecular Evolution. (Oxford Series in Ecology and Evolution).

Yang, Z., and Bielawski, J.R. (2000). Statistical methods for detecting molecular adaptation. Trends in Ecology and Evolution **15,** 496.

Yang, Z., and Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Molecular Biology and Evolution **17,** 32.

Chapter 5

Comparative sequence analysis of the *Rx1/Gpa2* cluster in distinct *Solanum* species

Patrick Butterbach, Johannes Sandbrink, Willem Stiekema, Marjon Arens, Edwin van der Vossen, Erin Bakker, Jaap Bakker and Aska Goverse

---

to be submitted

Abstract

Comparative sequence analysis was carried out on several BAC clones derived from *S. tuberosum* ssp. *andigena*, *S. tuberosum* ssp. *tuberosum* and *S. lepthophyes/sparsipilum* harbouring 17 *Rx1/Gpa2* homologues including the functional resistance genes *Rx1* and *Gpa2* and three novel *Rx/Gpa2* homologues from *S. lepthophyes/sparsipilum*. The 5'and 3'end UTRs were analysed showing a patch work of sequence homology between the different *Rx/Gpa2* homologues. A contig was assembled for the resistant and susceptible haplotype of SH and one haplotype of RH containing *Rx1/Gpa2* homologues that match syntenic regions of the *R* gene locus. Comparison of these homeologous chromosomal regions with the sequences from *S. lepthophyes/sparsipilum* showed a high intergenic sequence divergence in this region across different species of *Solanum*. In total, 21 genes were identified with high homology to *RMA1* from *A. thaliana* encoding a RING finger motif protein. They were located in vicinity of the *Rx1/Gpa2* homologues suggesting that tandem duplication may have occurred in this *R* gene cluster.

Introduction

In recent years, the speed of automated DNA sequencing has accelerated the availability of whole genomes including those of several plant species (Lin et al., 1999). While complete genomes are available for *Arabidopsis thaliana* and rice (Kaul et al., 2000; Ohyanagi et al., 2006), similar genome initiatives are starting up for several solanaceous crop plants like tomato and potato (Shibata, 2005; Van Os et al., 2006; Lee et al., 2007). Furthermore, sequencing projects for specific genome regions harbouring resistance gene clusters and quality traits in minor crop plants like eggplant and pepper are more and more used as an alternative for comparative genetic linkage to syntenic regions in related major crops like tomato (Doganlar et al., 2002; Lee et al., 2007). Genomic sequencing and annotation enable the identification of putative genes together with their precise locations in the corresponding genome region. With an increasing amount of functionally analysed genes from model plants such as *Arabidopsis thaliana*, comparative sequence analysis using bioinformatic tools for gene annotation enables the selection of functional candidate genes from other plants.

Understanding the genomic structure, function and evolution of disease resistance

genes (*R* genes) is of special interest for plant breeding research as these genes offer a large potential of effective and durable plant protection against pathogens. Many plant *R* genes have been cloned and characterized at the molecular level (Hammond-Kosack and Jones, 1997; Hulbert et al., 2001; Nimchuk et al., 2003; Parker, 2003). The structure of these plant disease resistance genes is generally conserved and at least seven classes have been recognized based on their structural domains (Martin et al., 2003; Gu et al., 2005). Genes encoding similar structural domains in their proteins have been found to be effective against pathogens like viruses, bacteria, fungi, nematodes and insects. One common class of resistance genes encodes proteins with a coiled-coil domain, a nucleotide binding site (NBS) and a leucine-rich (repeat) region (LRR). The NBS domain plays a role in the downstream signalling and activation of the defense response (Van der Biezen and Jones, 1998; Takken et al., 2006), whereas for several R proteins (Ellis et al., 1999; Chin et al., 2001; Dodds et al., 2001; Moffett et al., 2002) it has been shown that the LRR domain is involved in direct or indirect recognition of the pathogen elicitor.

Plant disease resistance genes are located in restricted areas of the genome, which can be organized in a very complex way. Molecular data from at least 10 families of resistance genes including loci from tomato, lettuce, rice, flax and *Arabidopsis* indicate that these loci frequently contain arrays of related genes (Michelmore and Meyers, 1998; reviewed by Hulbert et al., 2001). So far, few studies exist about the variation in structure and function of orthologous *R* gene loci in different genera. Significant variation in copy number and position of homologous resistance genes was found in a comparative study in the Graminae (Leister et al., 1998) suggesting limited conservation of genomic positions for *R* gene homologues, even from closely related species. However, an extensive study tracing close homologues of known functional *R* genes in *Lycopersicon*, *Solanum* and *Capsicum* indicates the existence of several cross-generic clusters of *R* genes (Grube et al., 2000). Pepper homologues of the tomato *R* genes *Sw-5*, *N*, *Pto*, *Prf* and *I2* were found in syntenic regions containing resistance loci in other solanaceous genomes suggesting that the chromosomal locations of *R* genes may be broadly conserved through speciation.

In potato, disease resistance loci against various major pathogens have been identified and mapped throughout the potato genome (reviewed by Gebhardt and Valkonen, 2001). They are often located in so called '*hot spots'* for resistance. The gene *Rx1* and its close homologue *Gpa2* are both located in a single *R* gene cluster of about 110 kb on

chromosome XII in potato and share an overall homology in the amino acid sequence of about 87%. Interestingly, these two *R* genes confer resistance to taxonomically unrelated pathogens, a virus and a nematode. However, *Rx1* results in extreme resistance to potato virus X (PVX) (Bendahmane et al., 1999), whereas *Gpa2* gives a much slower and milder resistance response to the potato cyst nematode *Globodera pallida* (Van der Vossen et al., 2000). Recently, a set of 75 *Rx/Gpa2* homologues was obtained from 10 species of *Solanum* including the resistant diploid potato clone SH (*S. tuberosum* ssp. *andigena*), the susceptible diploid potato clone RH (*S. tuberosum* ssp. *tuberosum*) and 9 wild *Solanum* species including a hybrid of *S. lepthophyes* and *S. sparsipilum* (Butterbach et al., Chapter 4) .

To better understand the genetic architecture and evolution of resistance gene clusters in potato, the complete nucleotide sequence was analysed for a region spanning a 187 kb genomic DNA fragment derived from the *Rx1/Gpa2* locus in the resistant haplotype of SH (*S. tuberosum* ssp. *andigena*). In addition, BAC clones harboring *Rx1/Gpa2* homologues were obtained and sequenced from the susceptible haplotype of SH and RH (*S. tuberosum* ssp. *tuberosum*) and the diploid potato species *S. lepthophyes/sparsipilum*. These data were used for a comparative sequence analysis resulting in new insights in the genomic organisation and evolution of the *Rx/Gpa2* cluster in potato.

Results

BAC sequencing and contig construction

In this study, the sequence of four BAC clones (Van der Vossen et al., 2000; Bakker et al., 2003) assembled in a minimal tiling path of 187,352 bp (AF265664) was used for analysis of the resistant chromosome - an introgression segment from *S. tuberosum* ssp. *andigena* - of the diploid potato clone SH harbouring the resistance genes *Gpa2*, *Rx1* and the closely related homologues *andSH-RGH1* and *andSH-RGH3*.

For the susceptible homeologous chromosome from *S. tuberosum* ssp. *tuberosum*, two BAC clones were obtained harbouring the *Rx/Gpa2* homologues *andSH-RGH5, 6, 7* and *8* (Bakker et al., 2003). The size of the BACs was 120,000 bp (BAC7E16) and 90,000 bp (BAC25G18), respectively, and sequencing resulted in 3 contigs for BAC7E16 and 3 contigs for BAC25G18. No overlap was found between any of the contigs from these two BACs.

For the susceptible diploid potato clone RH (*S. tuberosum* ssp. *tuberosum*), the sequences of two BAC clones – BAC RH153B24 and RH193P22 with an estimated size of 110,000 bp and 120,000 bp, respectively - were obtained harbouring the *Rx/Gpa2* homologues *tubRH-RGH4* and *tubRH-RGH5*.

A BAC library was constructed from the wild diploid potato species *S. lepthophyes/ sparsipilum* that harbours the functional *Rx* orthologue *Rx3* (Chapter 3, this thesis). The BAC library was screened with a primer set specific for the conserved LRR domain of *Gpa2/Rx1* homologues (Bakker et al., 2003). Six positive unique BAC clones were retrieved and sequenced. Table 1 gives an overview of the number of contigs after assemblage and the length of the six BAC clones from *S. lepthophes/sparsipilum*.

Table 1. Total length, number of contigs and ORFs of six BAC clones from *S. lepthophyes/sparsipilum* that harbour *Rx/Gpa2* homologues. Similar to the findings in the *Rx* loci on the SH and RH haplotypes, several zinc-finger protein encoding ORFs as well as transposon elements were found in close proximity indicating a similar region structure. Among the other genes found on the BACs are several retrotransposon elements.

BAC clone	Length in basepairs	contigs	*Rx/Gpa2* homologue	zinc finger protein	other genes with homology to:	highest similarity to:	query length in basepairs	e-value
SH AF265664	187352	1	*Rx1*	*RFP01*	retrotransposon element	putative retrotransposon [*Solanum demissum*]	714	0.0
			Gpa2	*RFP02*	retrotransposon element	putative retrotransposon [*Solanum demissum*]	1320	0.0
			andSH-RGH1	*RFP03*	methyltransferase	sugar transporter MFS 1 [*Medicago tranculata*]	1344	0.0
			andSH-RG3	*RFP04*				
SH 25G18	90000	3	*andSH-RGH6*	*RFP05*	retrotransposon element	putative retrotransposon [*Solanum demissum*]	714	1e-32
			andSH-RGH7	*RFP06*				
			andSH-RGH8	*RFP07*				
				RFP08				
SH 7E16	120000	3	*andSH-RGH5*	*RFP09*	retrotransposon element	putative retrotransposon [*Solanum demissum*]	243	5e-31
				RFP10	sugar transporter	sugar transporter MFS 1 [*Medicago tranculata*]	1344	0.0
					methyltransferase	caffeic acid O-methyltransferase [*Rosa chinensis* var. *spontanea*]	489	2e-42
RH 153B24	110000	6	*tubRH-RGH5*	*RFP11*	retrotransposon element	putative retroelement [*Solanum tuberosum*]	219	2e-13
					RNA polymerase	RNA polymerase II [*Lycopersicon esculentum*]	120	8e-10
					resistance gene	leucine-rich repeat [*Medicago truncatula*]	2583	1e-178
						putative fimbriae-associated protein [*Trichomonas vaginalis* G3]	711	7e-06
RH 192P22	120000	8	*tubRH-RGH4*	*RFP12*	retrotransposon element	putative retroelement [*Solanum tuberosum*]	4302	0.0
				RFP13	retrotransposon element	putative retrotransposon [*Solanum demissum*]	1353	0.0
					retrotransposon element	putative retrotransposon [*Solanum demissum*]	714	0.0
					sugar transporter	sugar transporter MFS 1 [*Medicago tranculata*]	1344	9e-84
					resistance gene	putative late blight resistance protein [*Solanum demissum*]	249	3e-22
					methyltransferase	caffeic acid O-methyltransferase [*Rosa chinensis* var. *spontanea*]	489	6e-42

BAC clone	Length in basepairs	contigs	*Rx/Gpa2* homologue	zinc finger protein	other genes with homology to:	highest similarity to:	query length in basepairs	e-value
ZYxH1	45000	2	*lph/spl-RGH6*	*RFP14*	retrotransposon element	putative retrotransposon [*Solanum demissum*]	2592	0.0
ZYxH2	60000	2	*lph/spl-RGH4*	*RFP21*	retrotransposon element	putative retrotransposon [*Solanum demissum*]	2421	0.0
					Rnase	RNase H family protein [*Solanum demissum*]	1619	2e-34
ZYxH3	50000	3	*lph/spl-RGH9*	*RFP18*	retrotransposon element	putative retrotransposon [*Solanum demissum*]	744	2e-52
			lph/spl-RGH10		Rnase	RNase H family protein [*Solanum demissum*]	1626	2e-34
ZYxH4	50000	2	*lph/spl-RGH5*	*RFP15*	RNA polymerase	RNA polymerase II [*Lycopersicon esculentum*]	3000	1e-74
					leucine rich repeat	leucine-rich repeat [*Medicago truncatula*]	2583	1e-175
					retrotransposon element	putative retroelement [*Solanum tuberosum*]	4032	0.0
ZYxH5	40000	3	*lph/spl-RGH11*	*RFP19*	resistance gene	putative late blight resistance protein [*Solanum demissum*]	255	2e-18
				RFP20	resistance gene	putative late blight resistance protein [*Solanum demissum*]	300	4e-37
ZYxH6	45000	6	*lph/spl-RGH7*	*RFP16*	retrotransposon element	putative retrotransposon [*Solanum demissum*]	837	3e-67
				RFP17	retrotransposon element	putative retrotransposon [*Oryza sativa*]	1515	1e-129

Gene prediction and annotation

Gene prediction scans using BLASTX were carried out on the BAC clone sequences to find the *Rx/Gpa2* homologues including their 5' and 3' UTR regions and other putative genes. Blast homology searches were used to establish putative functions of the predicted proteins. In total, 17 *Rx1/Gpa2* homologues were found on the BAC clone sequences (Table 1). In accordance with previous findings, the functional resistance genes *Rx1* and *Gpa2* and their close homologues *andSH-RGH1, 3, 5, 6, 7* and *8* were retrieved from the corresponding BAC clones from SH. Several *Rx1/Gpa2* homologues (*lph/spl-RGH4, 5, 6* and *7*) previously isolated from the genotype of *S. lepthophyes/sparsipilum* (Chapter 3, this thesis) were found on sequences of various BAC clones, except for *Rx3*. Three additional *Rx1/Gpa2* homologues were found *viz. lph/spl-RGH9*, *lph/spl-RGH10* and *lph/spl-RGH11* showing 95% sequence similarity to *Rx1*. They resemble the gene structure of the major group of *Rx1* homologues (Chapter 4, this thesis) with an ORF encoding a putative R protein consisting of a CC-NBS-LRR domain.

The sequences of the untranslated regions at the 5' and 3´end UTR regions of the 14 homologues including *Rx1* and *Gpa2* comprising 5000 bp and up to 2360 bp, respectively, were aligned and analyzed (Fig. 1). *Rx1*, *Gpa2, andSH-RGH6, andSH-RGH7* and *andSH-RGH8* show a very high homology of 93.6 -96.7% in their sequence upstream of the conserved ATG codon. The homology extends to 2572 nucleotides in front of the conserved ATG for *Rx1* and 2614 in front of the *Gpa2* gene. The 5' end UTR region contains microsatellite repeats variable for different homologues. At two positions the *Gpa2* 5' end region contains a longer microsatellite repeat than *Rx1*. At position -2442 a complex microsatellite region is located in which *Gpa2* has $(TA)_{12}$, whereas *Rx1* has $(TA)_5$. At position -1300 *Gpa2* has $(TA)_{20}$, where *Rx1* has $(TA)_5$. At position -207 *Rx1* has a stretch of 5 extra thymidines, whereas *Gpa2* has two extra thymidines at -6. The remaining homologous sequence upstream of the conserved ATG is 97.6 % identical (60 mutations out of 2567).

The 5' end UTR of *andSH-RGH1, lph/spl-RGH 5, 6* and *7* have an insertion of up to 1756 bp in the region upstream of position -281. This insert is flanked by (imperfect) inverted repeats of 400 bp. The insert sequence has no homology to any known sequence in the public databases. The 5' region of *andSH-RGH1* flanking this 1756 bp insert sequence has again high homology to the 5' regions of *Gpa2* and *Rx1*. While the 5' UTR sequence of

Rx1, *Gpa2*, *andSH-RGH1* and *andSH-RGH8* exhibit a high homology across a region encompassing

2500-5000 bp, the 5' UTR of most of the other homologues reveals much shorter homologous regions. The homology of the region 5' of *andSH-RGH3* and *lph/spl-RGH9* to the other 5' regions stops at position –653, while the sequence continues to have high homology among those two.

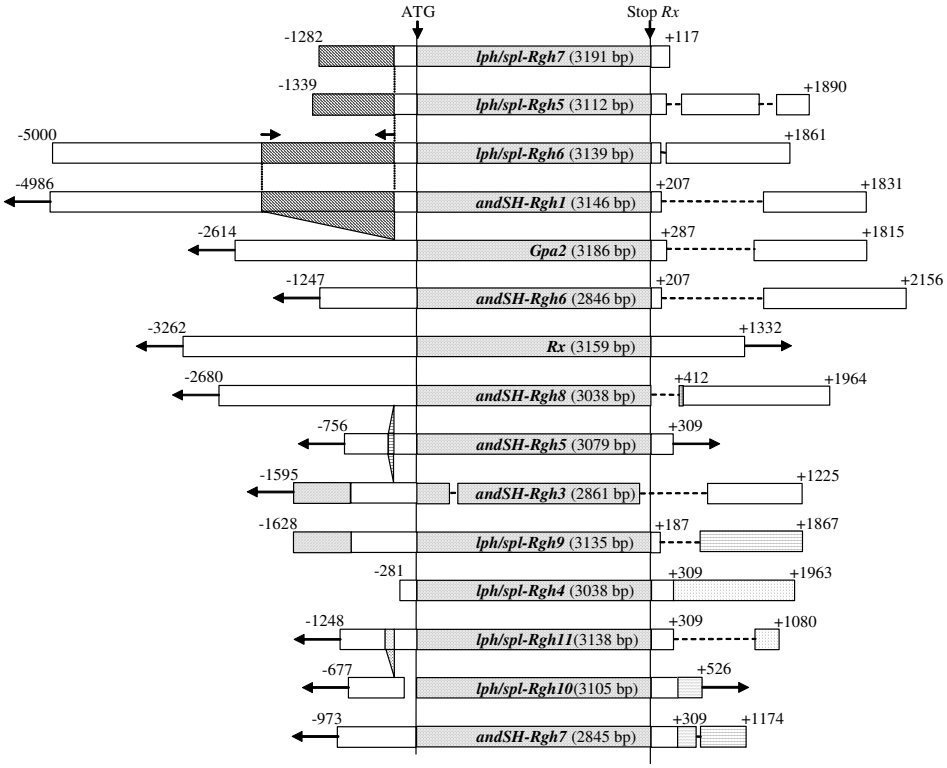


Figure 1. Comparison of the genome structure of the *Rx1/Gpa2* homologue untranslated 5' and 3' regions with the major indels. The sequence homologous to the genomic DNA of *Rx1* containing the cDNA sequences (including introns), between start (ATG) and stopcodon (stop *Rx1*) is shown in grey. UTR sequence parts with homology to that of *Rx1* are shown as white boxes, sequence parts with no homology to *Rx1* UTRs are patterned. Deletions are marked as broken lines, insertions compared to *Rx1* in the 5' UTR are shown as patterned rectangulars.

The structural variation at the 3' end of the *Rx/Gpa2* homologues is far more extensive than the variation observed at the 5' region (Fig. 1). Structural rearrangements 3' to the coding sequence can be observed in all copies resulting in a mosaic-like patchwork of homologous regions dissected in different positions by deletions and insertions with different or without any homology to other 3'end UTR sequences. DNA sequence homology between *Gpa2* and *Rx1* was observed until 287 nucleotides downstream of the stopcodon of *Rx1*. However, the coding sequence of *Gpa2* differs from *Rx1* due to an 11 bp insertion behind the first intron of *Gpa2* resulting in a different C-terminal composition (Van der Vossen et al., 2000). *Gpa2*, *andSH-RGH1* and *andSH-RGH6* share a deletion of about 1300 bp after position +206, +206 and +281. Sequence homology between *andSHRGH3* and the other homologues stops at the beginning of exon 2, but the 3' region continues to be highly homologous at a stretch of 1224 bp. However, this homology does not start directly 3' of the gene, but respectively 327 and 575 bp from the start of the second exon of *Rx1*.

The 3'end UTR of several homologues contains regions that share close homology to only one other UTR, while it is unrelated to the 3'UTRs of other homologues across species. The 3' UTRs of *lph/spl-RGH4* and *lph/spl-RGH11* share homology in a stretch of 771bp (Fig. 1), whereas the 3' UTR of *andSH-RGH7* contains a 400 bp stretch homologous to the 3' UTR sequence of *lph/spl-RGH10* followed by a 671 bp long region with homology to the 3' UTR of *lph/spl-RGH9*.

In total, 21 genes were identified showing a very high similarity of 80.6 to 96.4% to each other and the highest similarity (51.3%) to *RMA1* from *Arabidopsis* (Matsuda et al., 2001), which encodes a RING finger protein (RFP) and to a gene (68.3%) from *Capsicum annum* encoding a RING finger protein. RING zinc-finger protein encoding genes were found across the BAC sequences from SH, RH and *S. lepthophyes/sparsipilum*. The zinc finger motif of seven cysteine residues and one histidine (Saurin et al., 1996) is present in all 21 genes. The identity with RMA1 at the protein level is 40% and pertains to the RING finger and the C-terminal tail. In all 21 putative RFPs, the C-terminal tail contains 2 hydrophilic amino acid residues in the last 16 residues, which makes it unlikely that this region is anchored to a membrane as was hypothesized for RMA1 by Matsuda & Nakano (1998).

Some predicted genes derived from RH and *S. lepthophyes/sparsipilum* showed

homology to resistance related proteins like a putative late blight protein from *S. demissum* and an LRR from *Medicago tranculata*. Furthermore, 14 retrotransposon elements were found in the BAC sequences with close homology to retrotransposon elements from *S. tuberosum* and *S. demissum*. Transposons are frequently observed within resistance gene loci (Ballvora et al., 2007; Kuang et al., 2005) and they may be involved in the diversification of members of *R* gene families located in complex loci (Richter and Ronald, 2000).

Several other predicted putative genes revealed homology to a caffeic acid O-methyltransferase, sugar transporters, Rnase H and RNA polymerase II subunits (Table 1).

Comparative sequence analysis

The complete sequence information of the *Rx1/Gpa2* locus of the resistant haplotype from SH (*S. tuberosum* ssp. *andigena)* was used to search for syntenic regions present in the BAC clone sequences derived from the homeologous susceptible haplotype of SH (*S. tuberosum* ssp. *tuberosum*), the unlinked BAC clone sequences from RH (*S. tuberosum* ssp. *tuberosum*) and *S. lepthophyes/sparsipilum*. However, no clear overall sequence homology was observed upon comparison of the different BAC clones indicating high intergenic sequence variation, which did not allow the identification of putative syntenic regions by sequence alignment. To investigate the lack of sequence homology at the *Rx1/Gpa2* cluster across species of *Solanum*, the sequences encoding the *Rx1/Gpa2* locus of the resistant and susceptible haplotypes and an 80000 bp stretch from BAC RH153B24 for RH were compared using dot-plots (Fig. 2). For the SH resistant and susceptible haplotypes, only the regions harbouring the coding sequence and the corresponding untranslated regions of the *Rx/Gpa2* homologues revealed homology to each other. For the sequences obtained for the susceptible haplotype from SH and the haplotype from RH - both derived from *S. tuberosum* ssp. *tuberosum* - the dot-plot showed a high sequence similarity for a region encompassing 45000 bp. No homologous stretches were detected in the sequences derived from S. lephtophyes/sparsipilum (data not shown). These data indicate that the level of intergenic sequence variation increases substantially in more distantly related *Solanum* species.

Therefore, common microsatellites or simple sequence repeats (SSR) were identified in the sequences (Table 2) and used to link the different chromosomal regions.

Three microsatellite markers from the *S. tuberosum* ssp. *andigena* haplotype were found to match with the sequences from the *S. tuberosum* ssp. *tuberosum* haplotype of SH and from the two BAC sequences from the *S. tuberosum* ssp. *tuberosum* haplotype of RH. In addition, the position and order of markers flanking *Rx1/Gpa2* homologues and other predicted genes matched those from different BAC clone sequences. None of these SSR markers were detected in the BAC sequences from *S. lepthophyes/sparsipilum*.

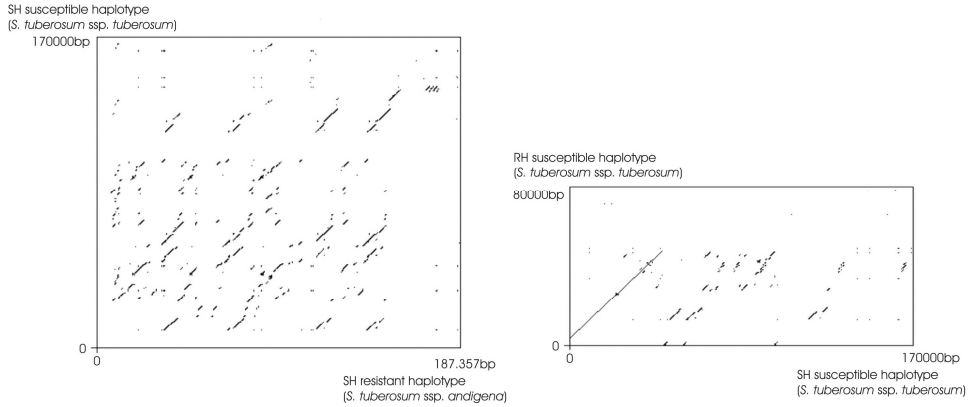


Figure 2. Dot-Plot comparisons done using DOTTUP (EMBOSS software package) for the *Rx2/Gpa2* locus of the resistant with the susceptible haplotype of SH and the susceptible haplotype of SH with an 80000 bp contig from BAC153B24 of RH. Similarity between the haplotypes of SH derived from *S. tuberosum* ssp. *andigena* and *S. tuberosum* ssp. *tuberosum* was only observed in the regions harbouring the *Rx/Gpa2* homologues coding sequence and the corresponding close adjacent untranslated regions. In the SH haplotype and the RH haplotype both derived from *S. tuberosum* ssp. *tuberosum* sequence the dot-plot reveals high sequence similarity for a region encompassing 45000 bp, which indicates increased sequence conservation compared to the corresponding region *S. tuberosum* ssp. *andigena*.

Table 2. Common microsatellites or simple sequence repeats (SSR) were identified in the sequences (Table 1) and used to link the different chromosomal regions. Three microsatelite markers from the *S. tuberosum* ssp. *andigena* haplotype were found to match with the sequences from the *S. tuberosum* ssp. *tuberosum* haplotype of SH and from the two BAC sequences from the *S. tuberosum* ssp. *tuberosum* haplotype of RH. Positions on the BAC clone sequences and *Rx1/Gpa2* region for the resistant haplotype from SH are indicated.

	SH AF265664	SH 25G18	SH 7E16	RH 153B24	RH 192P22
SSR microsatellite marker					
GGTCCTATTACCCCCCCCCCCCAAACTTATTTT	+20094	+38433	-	+44784	-
ATCTATCTATCTCTCTCTCTCTCTCTATATATATAT	+84804	-	+13940	-	+23281
CCCTAATTATTATTATTATTATTTTTTT	+155010	-	+57330	-	

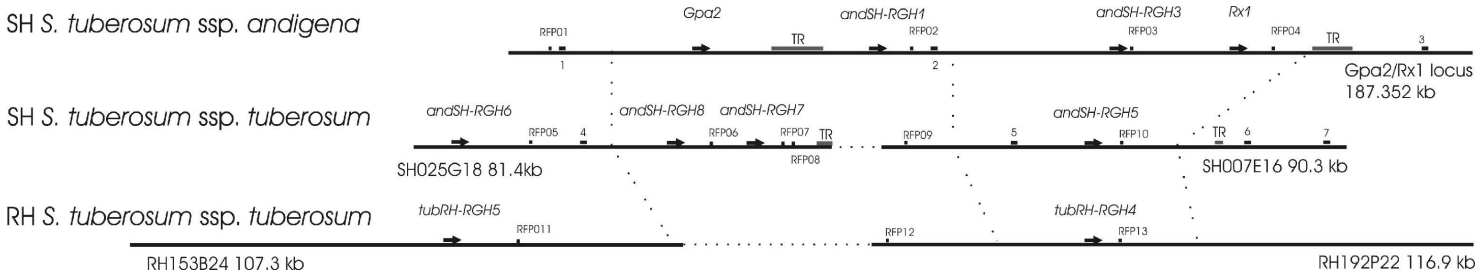


Figure 3. Tentative representation of the *Rx/Gpa2* locus in 3 haplotypes derived from the diploid potato clones SH and RH. The region around the resistance genes *Gpa2* and *Rx1* in SH on the haplotype derived from *S. tuberosum* ssp. *andigena* has been published recently (Bakker, Butterbach et al. 2003). 2 BAC clones derived from the second haplotype in SH (Bakker, Butterbach et al. 2003) and 2 BAC clones of RH from the potato BAC genome library were found to harbour *Rx/Gpa2* homologues. Unique SSR marker sequences found on different haplotypes allowed for a tentative assembly of the BAC clone derive sequences indicating syntenic regions relative to the full sequenced loci from *S. tuberosum* ssp. *andigena*. Based on their homology for several homologues a putative allelic relationship has been previously suggested (Bakker, Butterbach et al. 2003). According to this scheme, the homologues *andSH-RGH6*, *tubRH-RGH5* and *andSH-RGH5*, *tubRH-RGH4* match syntenic regions on 2 different haplotypes of *S. tuberosum* ssp. *tuberosum* which give support for them being alleles. Transposons (TR), RING encoding genes (RFP1-13), retrotransposon elements (1,4,6), RNA polymerase encoding gene (3), sugar transporter MFS 1 (2, 5), caffeic acid O-methyltransferase gene (7) are indicated.

The resulting positions and orientation of the *Rx1/Gpa2* homologues and predicted genes present on the haplotypes of SH and RH are shown in Figure 3. A different distribution of the *Rx1/Gpa2* homologues was observed for the homeologous chromosomal regions of SH, whereas the organisation of the *Rx1/Gpa2* homologues was more similar between the *S. tuberosum* ssp. *tuberosum* haplotypes of SH and RH. The observed syntheny between microsatelites and the allelic relationship of *andSH-RGH6/tubRH-RGH5* and *andSH-RGH5/tubRH-RGH4* (Bakker et al., 2003) indicate that the BAC clones from RH are derived from the same haplotype.

With the exception of *Gpa2* and *lph/spl-RGH10*, it was observed that the *Rx1/Gpa2* homologues were situated in close vicinity upstream of an *RFP* gene in this region of SH, RH and *S. lepthophyes/sparsipilum* with 6 out of the 13 *RFP* genes from RH and SH, which are located at a similar distance to the *Rx1/Gpa2* homologues (between 5 and 6 kb). A cluster analysis was performed to reveal the sequence relationships of the 21 *RFP* homologues from distinct *Solanum* species (Fig. 4). Interestingly, comparison of the tree topologies of the *RFP* genes and *Rx1/Gpa2* homologues resulted in identification of two pairs of *Rx/Gpa2* homologues and their corresponding *RFP* gene. This suggests an allelic relationship and a common origin of these *Rx/Gpa2* homologues and *RFP* genes derived from the *S. tuberosum* ssp. *tuberosum* haplotype of RH and SH. However, for the other *Rx1/Gpa2* homologues and closely situated *RFP* genes no such correlation was found.

Discussion

Here, we describe the comparative analysis of chromosomal sequences obtained from different *Solanum* species that contain *Rx1/Gpa2* homologues, including the functional resistance genes *Rx1* and *Gpa2*. The complete sequence of 187,352 kb spanning the *Rx1/Gpa2* cluster on chromosome XII of the resistant haplotype from the diploid potato clone SH (*S. tuberosum* ssp. *andigena*) was used to align the sequences derived from the susceptible haplotypes of SH and RH (*S. tuberosum* ssp. *tuberosum*) and BAC clone sequences from *S. lepthophyes/sparsipilum.* This resulted in the detection of an overall large sequence divergence in this region across different species of *Solanum*. However, we

were able to anchor the sequences from *S. tuberosum* ssp. *tuberosum* and *S. tuberosum* ssp. *andigena*, but not the BAC clone sequences from *S. lepthophyes/sparsipilum*. This demonstrates that intergenic sequence variation increases with the corresponding species distance in *Solanum.*

In contrast to the intergenic sequence divergence on the *Rx1/Gpa2* loci derived from more closely related *Solanum* species overall high sequence conservation was found for the coding sequence of *Rx1/Gpa2* homologues from even distantly related taxa within *Solanum* section *Petota* (Chapter 4, this thesis). More recently, two studies addressing the divergence of *R* gene loci across species came to rather contrasting results presenting substantially divergence in numbers and location of *R* genes in Graminae (Leister et al., 1998) versus the presence of syntenic cross-generic *R* gene clusters in tomato, potato and pepper (Grube et al., 2000). This could be due to the different approaches used as Grube et al. (2000) was focussing on known functional *R* genes like *Sw-5*, *N*, *Pto*, *Prf* and *I2*, while Leister et al. (1998) conducted a survey using sequence defined classes of *R* gene homologues. The approach and results described in this study resemble more those of Grube et al. (2000) aiming for homeologous regions of the *Rx1/Gpa2* locus harbouring the functional resistance genes *Rx1* and *Gpa2* in other haplotypes and across species.

The very high similarity (>87%) of *Rx2* to the copies of the *Rx/Gpa2* cluster is striking, as *Rx2* was isolated from *Solanum acaule* and *Rx/Gpa2* homologues from *Solanum tuberosum* ssp. *andigena*. The mapping data of *Rx2* presented in the papers of Ritter et al. (1991) and Bendahmane *et al.* (2000) locates *Rx2* to chromosome 5 while *Rx* and *Gpa2* are on chromosome 12. This would indicate that resistance gene clusters do not only evolve by intergenic and intra-allelic recombination within a cluster, but possibly also through recombination between clusters on different chromosomes. Hence, it would be very interesting to study the organization and evolution of both the *Rx2* region and of homeologous haplotypes of the *Rx/Gpa2* region in other potato genotypes.

The DNA sequence homology (95%) in the regions 5' of the starting ATG between several homologues including *Rx1* and *Gpa2* is very striking. As *Gpa2* encodes for a resistance gene against a nematode and *Rx1* against a virus, one would expect that the putative promoters of the genes might have significant differences as nematodes are soil-borne and *Rx1* is air-borne. However, the only obvious difference between the two 5' regions is the presence of two long microsatellite regions in the 5' region of *Gpa2* on the

position where *Rx1* has two short microsatellites. It is known that microsatellites evolve quicker than ordinary sequences and possibly the quick evolution of the microsatellites in the 5' region of *Gpa2* has played a role in a possible functional evolution of the two genes. The biological relevance of these observations remains to be proven.

An interesting result is the identification of a gene family of at least 21 members with high homology to RMA1 (a RING finger motif protein) from *A. thaliana* and *RDP* from *C. annum* in vicinity of the *Rx1/Gpa2* homologues located on chromosomal regions obtained from different *Solanum* species. Numbers and relative position of *Rx1/Gpa2* homologues and *RFP* on the contig, but also on the unlinked BAC sequences from *S. lepthophyes/sparsipilum* (data not shown) suggest that they may be tandemly duplicated on the *Rx1/Gpa2* locus. Co-localisation with the functional resistance genes *Rx1* and *Gpa2* might indicate that RDPs plays a role in resistance.

Matsuda & Nakano (1998) showed by Southern blotting that RMA1 is present in the genome of *Arabidopsis* as a single copy gene. Searches with the most conserved amino acid stretches in the *Arabidopsis* database only yielded three hits of which one was RMA1. RING finger motifs have been found to interact with small GTPases. There is functional evidence for RING-finger proteins being involved in resistance protein associated functions in plants. Durrant *et al.* (2000) reported that the expression of a RING-H2 zinc finger protein was strongly induced after Avr9/Cf-9 eliciting in tobacco cell cultures, indicating that this RING-H2 zinc finger protein might be involved in signal transduction. The homologous proteins RIN2 and RIN3 encode RING-finger type ubiquitin ligases acting as a positive regulator of the hypersensitive response mediated by RPM1 and RPS2, which confer resistance to *Pseudomonas syringae* in *A. thaliana* (Kawasaki et al., 2005). RIN2 and RIN3 interact with the N-terminal region of RPM1, including the CC motif. Similarly to *Rx1*, it has been shown for RPM1 that the R protein is interacting in a resistosome complex with RAR1 and HSP90 indicating partially redundancy in downstream signalling function (Moffett et al., 2002; Belkhadir et al., 2004; Schulze-Lefert, 2004). Whether the RING finger protein encoding genes found at the *Rx1/Gpa2* cluster of potato are involved in the resistance response of *Rx1* or *Gpa2* remains to be proven.

Material and Methods

BAC contigs preparation for SH and RH

The construction of the BAC libraries from the resistant haplotype in SH and PCR screening of these BAC libraries have been described by Rouppe van der Voort *et al.* (1999) and Kanyuka *et al.*(1999). A physical map of the region is published by Van der Vossen et al. (2000); (Bakker et al., 2003). Homology search using *Rx1/Gpa2* cluster specific primers; Butterbach et al., Chapter 3) retrieved the BAC clone sequences from BAC and BAC harbouring andRH-RGH5 and andRH-RGH4 (kindly provided by CBSG, Wageningen, the Netherlands).

BAC library and BAC contigs of S. lepthophyes/sparsipilum

The resistant genotype of *S. lepthophyes/sparsipilum* CGN20619 harbouring *Rx3* (Chapter 3, this thesis) was used as DNA source for the construction of a BAC library. High molecular weight DNA preparation and BAC library construction were carried out as described by Rouppe van der Voort et al. (1999). The BAC library was stored in 20 pools of about 1000 clones. For the screening plasmid DNA was isolated from each pool of clones using the standard alkaline lysis protocol and PCR was carried out with cluster-specific primers based on the the LRR domains, followed by a TaqI digestion, as described (Bakker et al., 2003) to identify positive pools. Bacteria corresponding to positive pools were diluted and plated on LB agar plates containing chloramphenicol (12µg ml-1). 6 individual single positive BAC clones, ZYxH1, ZyYH2,ZYxH3, ZYxH4, ZYxH5, ZYxH6 were identified.

CHEF gel electrophoresis

BAC insert sizes were determined with CHEF electrophoresis on a 1% agarose gel (Seakem ® Gold, FMC, Philadelphia, PA, USA) in 0.5×TBE buffer at 4°C using a BIO-RAD CHEF DR II system (Bio-Rad Laboratories, Hercules, CA, USA) at 200 V with a pulse time of 5 to 15 s for 18h.

Template preparation, PCR reactions and sequence gel electrophoresis

Large-scale plasmid preparation from 5 ml cultures of shotgun clones was performed using a Qiagen 9600 robot using the Qiaprep 96 Turbo miniprep kit (Qiagen) according to the manufacturers protocol. Approximately 500 ng template DNA was used in a cycle sequencing reaction using either 100 ng M13 forward or reverse primer, 8 µl Big Dye terminator mix (PerkinElmer) in a total volume of 20 µl. The PCR protocol consisted of 25 cycles of 30 sec at 96 °C , 15 sec at 50°C and 4 min at 60°C. After ethanol precipitation at RT for 10 min and recovery, the labelled DNA was dissolved in 3 µl formamide. DNA was then denatured for 2 min at 96°C and immediately cooled on ice. Approximately 1.5 µl was then loaded on a 6% TBE polyacrylamide gel pH 8.3. Sequence electrophoresis was carried out on either an ABI 373XL or ABI 377 sequencer.

Sequence assembly and analysis

Shotgun sequences were assembled and edited using the Staden software package (Bonfield *et al.* 1995) on a Sun_Solaris platform. Consensus BAC sequence was established from both plus and minus strand directed sequences covering the total insert length with an

average redundancy of 6 for BACs from SH and 34 for BACs from *S. lepthophyes/sparsipilum*. Sequence quality and assembly integrity was verified by referencing restriction enzyme data and overlapping sequence data from neighbouring BAC clones.

pBELOBAC 11 vector sequences were removed from the BAC sequence with the module Seqman of the DNAstar suite (www.dnastar.com). The four BACs from the resistant haplotype of SH were assembled into one sequence with the same program. Quality of this sequence was checked by comparing the 25,908 bp of overlapping sequence.

Sequence analysis

Positions of putative genes were predicted using GENSCAN (Burge and Karlin, 1997). Homology searches were carried out using the BLAST program (Altschul et al., 1997).

The isolation of *andSH-RGH5*, *6* and *7, tubRH-RGH2* to *5, lph/spl-RGH1* to *7*, *Rx3* has been described previously. Sequences of the resistance gene homologues *andSH-RGH1, and SH-RGH3* and the resistance genes *Gpa2* (Van der Vossen et al., 2000), *Rx1* (Bendahmane et al., 1999) and *Rx2* (Bendahmane et al., 2000) are obtained from genbank (accession numbers AF266747, AF266746, AF195939, AJ011801, AJ249448, respectively).

Nucleotide and deduced amino acid sequences were aligned with Clustal W (Chenna et al., 2003) and analyzed by eye using Vector NTI suite 8.0 (InforMax Inc., Bethesda, USA) and BioEdit version 6.0.5 (Hall 1999).

Using the multiple nucleotide alignment, trees were constructed with the program PAUP* v.4.0b10 (Swofford, 1999) implementing the unweighted pair group method with arithmetic mean (UPGMA) and uncorrected distance model. Bootstrap values were calculated for 1000 replicas.

BAC clone sequences were analyzed for microsatellites using webTROLL at http://wsmartins.net/webtroll/troll.html (Castelo et al., 2002). Dot-plot comparsion was carried out using DOTTUP from the EMBOSS software package (Rice et al., 2000) with a windowsize of 20 basepairs.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research **25,** 3389.

Bakker, E., Butterbach, P., Rouppe Van Der Voort, J., Van Der Vossen, E., Van Vliet, J., Bakker, J., and Goverse, A. (2003). Genetic and physical mapping of homologues of the virus resistance gene *Rx1* and the cyst nematode resistance gene *Gpa2* in potato. Theoretical and Applied Genetics **106,** 1524.

Ballvora, A., Jocker, A., Viehover, P., Ishihara, H., Paal, J., Meksem, K., Bruggmann, R., Schoof, H., Weisshaar, B., and Gebhardt, C. (2007). Comparative sequence analysis of *Solanum* and *Arabidopsis* in a hot spot for pathogen resistance on potato chromosome V reveals a patchwork of conserved and rapidly evolving genome segments. BMC Genomics **8,** 112.

Belkhadir, Y., Subramaniam, R., and Dangl, J.L. (2004). Plant disease resistance protein signalling: NBS-LRR proteins and their partners. Current Opinion in Plant Biology **7,** 391.

Bendahmane, A., Kanyuka, K., and Baulcombe, D.C. (1999). The *Rx* gene from potato controls separate virus resistance and cell death responses. Plant Cell **11,** 781.

Bendahmane, A., Kanyuka, K., Baulcombe, D.C., Querci, M., and Bendahmane, A. (2000). Agrobacterium transient expression system as a tool for the isolation of disease resistance genes: Application to the *Rx2* locus in potato. Plant Journal **21,** 73.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. Journal of Molecular Biology **268,** 78.

Castelo, A.T., Martins, W., and Gao, G.R. (2002). TROLL - Tandem repeat occurrence locator. Bioinformatics **18,** 634.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Research **31,** 3497.

Chin, D.B., Arroyo-Garcia, R., Ochoa, O.E., Kesseli, R.V., Lavelle, D.O., and Michelmore, R.W. (2001). Recombination and spontaneous mutation at the major cluster of resistance genes in lettuce (*Lactuca sativa*). Genetics **157,** 831.

Dodds, P.N., Lawrence, G.J., and Ellis, J.G. (2001). Contrasting modes of evolution acting on the complex *N* locus for rust resistance in flax. Plant Journal **27,** 439.

Doganlar, S., Frary, A., Daunay, M.C., Lester, R.N., and Tanksley, S.D. (2002). A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the *Solanaceae*. Genetics **161,** 1697.

Ellis, J.G., Lawrence, G.J., Luck, J.E., and Dodds, P.N. (1999). Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. Plant Cell **11,** 495.

Gebhardt, C., and Valkonen, J.P.T. (2001). Organization of genes controlling disease resistance in the potato genome. Annual Review of Phytopathology **39,** 79.

Grube, R.C., Radwanski, E.R., and Jahn, M. (2000). Comparative genetics of disease resistance within the solanaceae. Genetics **155,** 873.

Gu, K., Yang, B., Tian, D., Wu, L., Wang, D., Sreekala, C., Yang, F., Chu, Z., Wang, G.L., White, F.F., and Yin, Z. (2005). *R* gene expression induced by a type-III effector triggers disease resistance in rice. Nature **435,** 1122.

Hammond-Kosack, K.E., and Jones, J.D.G. (1997). Plant disease resistance genes. Annual Review of Plant Biology **48,** 575.

Hulbert, S.H., Webb, C.A., Smith, S.M., and Sun, Q. (2001). Resistance gene complexes: Evolution and utilization. Annual Review of Phytopathology **39,** 285.

Kanyuka, K., Bendahmane, A., Rouppe Van Der Voort, J.N.A.M., Van Der Vossen, E.A.G., and Baulcombe, D.C. (1999). Mapping of intra-locus duplications and introgressed DNA: Aids to map-based cloning of genes from complex genomes illustrated by physical analysis of the *Rx* locus in tetraploid potato. Theoretical and Applied Genetics **98,** 679.

Kaul, S., Koo, H.L., Jenkins, J., Rizzo, M., Rooney, T., Tallon, L.J., Feldblyum, T., Nierman, W., Benito, M.I., Lin, X., Town, C.D., Venter, J.C., Fraser, C.M., Tabata, S., Nakamura, Y., Kaneko, T., Sato, S., Asamizu, E., Kato, T., Kotani, H., Sasamoto, S., Ecker, J.R., Theologis, A., Federspiel, N.A., Palm, C.J., Osborne, B.I., Shinn, P., Conway, A.B., Vysotskaia, V.S., Dewar, K., Conn, L., Lenz, C.A., Kim, C.J., Hansen, N.F., Liu, S.X., Buehler, E., Altafi, H., Sakano, H., Dunn, P., Lam, B., Pham, P.K., Chao, Q., Nguyen, M., Yu, G., Chen, H., Southwick, A., Jeong Mi, L., Miranda, M., Toriumi, M.J., Davis, R.W., Wambutt, R., Murphy, G., Du?sterho?ft, A., Stiekema, W., Pohl, T., Entian, K.D., Terryn, N., Volckaert, G., Salanoubat, M., Choisne, N., Rieger, M., Ansorge, W., Unseld, M., Fartmann, B., Valle, G., Artiguenave, F., Weissenbach, J., Quetier, F., Wilson, R.K., De la Bastide, M., Sekhon, M., Huang, E., Spiegel, L., Gnoj, L., Pepin, K., Murray, J., Johnson, D., Habermann, K., Dedhia, N., Parnell, L., Preston, R., Hillier, L., Chen, E., Marra, M., Martienssen, R., McCombie, W.R., Mayer, K., White, O., Bevan, M., Lemcke, K., Creasy, T.H., Bielke, C., Haas, B., Haase, D., Maiti, R., Rudd, S., Peterson, J., Schoof, H., and Frishman, D. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408,** 796.

Kawasaki, T., Nam, J., Boyes, D.C., Holt Iii, B.F., Hubert, D.A., Wiig, A., and Dangl, J.L. (2005). A duplicated pair of *Arabidopsis* RING-finger E3 ligases contribute to the RPM1- and RPS2-mediated hypersensitive response. Plant Journal **44,** 258.

Kuang, H., Wei, F., Marano, M.R., Wirtz, U., Wang, X., Liu, J., Shum, W.P., Zaborsky, J., Tallon, L.J., Rensink, W., Lobst, S., Zhang, P., Tornqvist, C.E., Tek, A., Bamberg, J., Helgeson, J., Fry, W., You, F., Luo, M.C., Jiang, J., Robin Buell, C., and Baker, B. (2005). The *R1* resistance gene cluster contains three groups of independently evolving, type I *R1* homologues and shows substantial structural variation among haplotypes of *Solanum demissum*. Plant Journal **44,** 37.

Lee, S., Sung, H.J., and Choi, D. (2007). *Solanaceae* genomics: Current status of tomato (*Solanum lycopersicum*) genome sequencing and its application to pepper (*Capsicum* spp.) genome research. Plant Biotechnology **24,** 11.

Leister, D., Kurth, J., Laurie, D.A., Yano, M., Sasaki, T., Devos, K., Graner, A., and Schulze-Lefert, P. (1998). Rapid reorganization of resistance gene homologues in cereal genomes. Proceedings of the National Academy of Sciences of the United States of America **95,** 370.

Lin, X., Kaul, S., Rounsley, S., Shea, T.P., Benito, M.I., Town, C.D., Fujii, C.Y., Mason, T., Bowman, C.L., Barnstead, M., Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J., Ronning, C.M., Koo, H.L., Moffat, K.S., Cronin, L.A., Shen, M., Pal, G., Van Aken, S., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D., Nierman, W.C., White, O., Elsen, J.A., Salzberg, S.L., Fraser, C.M., and Venter, J.C. (1999). Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature **402,** 761.

Martin, G.B., Bogdanove, A.J., and Sessa, G. (2003). Understanding the Functions of Plant Disease Resistance Proteins. Annual Review of Plant Biology **54,** 23.

Matsuda, N., and Nakano, A. (1998). *RMA1*, an *Arabidopsis thaliana* gene whose cDNA suppresses the yeast sec15 mutation, encodes a novel protein with a RING finger motif and a membrane anchor. Plant and Cell Physiology **39,** 545.

Matsuda, N., Suzuki, T., Tanaka, K., and Nakano, A. (2001). *Rma1*, a novel type of RING finger protein conserved from Arabidopsis to human, is membrane-bound ubiquitin ligase. Journal of Cell Science **114,** 1949.

Michelmore, R.W., and Meyers, B.C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Research **8,** 1113.

Moffett, P., Farnham, G., Peart, J., and Baulcombe, D.C. (2002). Interaction between domains of a plant NBS-LRR protein in disease resistance-related cell death. EMBO Journal **21,** 4511.

Nimchuk, Z., Eulgem, T., Holt Iii, B.F., and Dangl, J.L. (2003). Recognition and Response in the Plant Immune System. Annual Review of Genetics **37,** 579.

Ohyanagi, H., Tanaka, T., Sakai, H., Shigemoto, Y., Yamaguchi, K., Habara, T., Fujii, Y., Antonio, B.A., Nagamura, Y., Imanishi, T., Ikeo, K., Itoh, T., Gojobori, T., and Sasaki, T. (2006). The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. *japonica* genome information. Nucleic acids research. **34**.

Parker, J.E. (2003). Plant recognition of microbial patterns. Trends in Plant Science **8,** 245.

Rice, P., Longden, L., and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics **16,** 276.

Richter, T.E., and Ronald, P.C. (2000). The evolution of disease resistance genes. Plant Molecular Biology **42,** 195.

Ritter, E., Debener, T., Barone, A., Salamini, F., and Gebhardt, C. (1991). RFLP mapping on potato chromosomes of two genes controlling extreme resistance to potato virus X (PVX). Molecular and General Genetics **227,** 81.

Saurin, A.J., Borden, K.L.B., Boddy, M.N., and Freemont, P.S. (1996). Does this have a familiar RING? Trends in Biochemical Sciences **21,** 208.

Schulze-Lefert, P. (2004). Plant Immunity: The Origami of Receptor Activation. Current Biology **14**.

Shibata, D. (2005). Genome sequencing and functional genomics approaches in tomato. Journal of General Plant Pathology **71,** 1.

Swofford, D.D. (1999). PAUP* Version4.0b10.

Takken, F.L., Albrecht, M., and Tameling, W.I. (2006). Resistance proteins: molecular switches of plant defence. Current Opinion in Plant Biology **9,** 383.

Van der Biezen, E.A., and Jones, J.D. (1998). The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. Current biology : CB **8**.

Van der Voort, J.R., Kanyuka, K., Van Der Vossen, E., Bendahmane, A., Mooijman, P., Klein-Lankhorst, R., Stiekema, W., Baulcombe, D., and Bakker, J. (1999). Tight physical linkage of the nematode resistance gene *Gpa2* and the virus resistance gene *Rx* on a single segment introgressed from the wild species Solanum tuberosum subsp. andigena CPC 1673 into cultivated potato. Molecular Plant-Microbe Interactions **12,** 197.

Van der Vossen, E.A.G., Rouppe van der Voort, J.N.A.M., Kanyuka, K., Bendahmane, A., Sandbrink, H., Baulcombe, D.C., Bakker, J., Stiekema, W.J., and Klein-Lankhorst, R.M. (2000). Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: A virus and a nematode. Plant Journal **23,** 567.

Van Os, H., Andrzejewski, S., Bakker, E., Barrena, I., Bryan, G.J., Caromel, B., Ghareeb, B., Isidore, E., De Jong, W., Van Koert, P., Lefebvre, V., Milbourne, D., Ritter, E., Rouppe Van Der Voort, J.N.A.M., Rousselle-Bourgeois, F., Van Vliet, J., Waugh, R., Visser, R.G.F., Bakker, J., and Van Eck, H.J. (2006). Construction of a 10,000-marker ultradense genetic recombination map of potato: Providing a framework for accelerated gene isolation and a genomewide physical map. Genetics **173,** 1075.

Chapter 6

General Discussion

Patrick Butterbach, Erik Slootweg, Kamila Koropacka, Jan Roosien, Laurentiu Spiridon, Andrei Petrescu, Geert Smant, Erin Bakker, Jaap Bakker and Aska Goverse

Has been published as a modified version in the Proceedings of the XIIth International Congress for MPMI, 2007 Sorrento, Italy

Introduction

Plants and animals are constantly under attack by a wide range of pathogens and pests. To defend themselves, most animals have both an innate and an adaptive immune system at their disposal of which the latter continues to somatically generate novel resistance specificities against non-self invaders during the organism's lifetime. In plants, however, such an acquired immunity is absent and instead they depend on meiotically formed innate pathogen-recognition mechanisms for defence (Nimchuk et al., 2003; Jones and Takemoto, 2004). The plant immunity is in principal a cell based surveillance system consisting of receptor-like resistance proteins (R proteins), encoded by a large set of single dominant resistance (*R*) genes.

According to their structural characteristics, the R proteins are divided into at least seven classes (Martin et al., 2003; Gu et al., 2005). A shared feature of the proteins in most of these classes is the leucine-rich repeat structure, which is expected to be involved in the specific recognition. The vast majority of *R* genes encode for proteins belonging to the super family of cytosolic NBS-LRR disease resistance proteins, which encode modular proteins consisting of a C-terminal leucine-rich repeat domain (LRR), and a nucleotide-binding site domain (NBS). At the N-terminus several domain types can be found, most often either a coiled-coil domain (CC) or a Toll/interleukin-1 receptor domain (TIR). The exact function of the different R protein domains is not yet unveiled, but the protein is expected to function as a molecular switch, comparable to proteins with a similar domain architecture that function in animal innate immunity and cell death, like NOD and Apaf-1.

The LRR is expected to keep the protein in an inactive state by shielding the NBS region via intramolecular interactions. Specific recognition via the LRR then triggers a conformational change, releasing the intramolecular interactions, and allowing among others a change in nucleotide-binding status and the binding of further signalling components (Moffett et al., 2002; Hwang and Williamson, 2003; Takken et al., 2006). Eventually the specific recognition leads to a complex resistance response, including the production of anti-pathogenic compounds, the induction of a reactive oxygen burst and a local programmed cell death, the so-called hypersensitive response.

On a genetic level the, direct or indirect, recognition of pathogen derived products by corresponding R proteins in the plant can be described as a gene-for-gene interaction (Flor, 1971). The genetic interaction drives the co-evolution of the *R* gene and the pathogen

avirulence gene. *R* genes are often located in clusters of homologues in the plant genome, and thereby subject to the evolutionary dynamics inherent to clusters. Multiple *R* gene copies arise by duplication events. The presence of homologous sequence in close proximity makes these areas of the genome more prone to unequal crossing-over by mispairing during meiosis. Unequal crossing-over can be a source of variation through for example intragenic sequence exchange, creating new recognition varieties, but on the other hand, if it occurs frequently it leads to gene conversion and a progressing homogenisation of the homologous genes in a cluster.

Progressing insights in the phylogenetics of *R* gene clusters have shown that although intragenic sequence exchanges occur, they are not frequent enough in most cases to cause gene conversion. Interallelic recombination and divergent evolution are the main mechanisms for development of new specificities (Michelmore and Meyers, 1998; Baumgarten et al., 2003; Leister, 2004). From the total number of known *R* gene sequences, only a few can yet be coupled to their cognate pathogen. In most clusters the specificities of the paralogues are completely unknown or they confer resistance to different strains of a single pathogen species like shown for the tomato *Cf* genes and the flax *L* genes (Ellis et al., 1999; Van der Hoorn et al., 2001b; Van der Hoorn et al., 2001a).

In potato, however, a unique cluster (Bakker et al., 2003) of four *R* gene homologues is located on chromosome XII including a truncated gene, a putative *R* gene of which the function is unknown and the two closely related genes *Rx1* (Bendahmane et al., 1999) and *Gpa2* (Van der Voort et al., 1997; Van der Vossen et al., 2000). Rx1 and Gpa are 88 % identical in amino acid sequence and the fast majority of the differences are found in the predicted solvent exposed regions of the LRRs. On the other (susceptible) haplotype, a similar cluster is present containing another four paralogues of which the function is unknown, which share about 90% sequence identity with Rx1 and Gpa2 (Bakker et al., 2003).

Interestingly, the closely related genes *Rx1* and *Gpa2* confer resistance to totally unrelated pathogens a virus and a nematode, respectively. In response to the release of an unknown elicitor from the avirulent nematode population *Globodera pallida* D383, *Gpa2* mediates a late cell death response in root cells surrounding the nematode induced feeding site. *Rx1* on the other hand mediates an extreme resistance response in leaves against Potato Virus X upon specific recognition of the viral coat protein, which under natural

circumstances stops the virus from spreading without causing a visible cell death response. One feature both pathogens have in common is the fact that they extensively manipulate the plant cell at a molecular level to adapt it to their needs. The nematode secretes a whole range of proteins into the cytoplasm via its stylet, and induces in this way the formation of a metabolically hyperactive syncytium, whereas PVX employs the plants transcription and translation machinery to replicate itself. Therefore, the *Rx1/Gpa2* cluster provides an excellent experimental model system to study the molecular dynamics underlying the evolution of *R* gene specificities.

Identification of novel functional orthologues across *Solanum* point out the long lifespan of the *Rx* gene

The ancestry of the cultivated potato is still under debate and a complex origin involving several wild species is suggested (Grun, 1990; Hawkes, 1990). Several lines of evidence point at the cultivated species *Solanum stenotonum* as the putative first cultivated potato species, from which *S. tuberosum* ssp. *andigena* and later *S. tuberosum* ssp. *tuberosum* could have derived (Grun, 1990). Within the large genus *Solanum* the cultivated potato and its wild relatives are classified in the Section *Petota* Dumort., including about 200 species originating from Central and South America (Hawkes, 1990; D'Arcy, 1991). Even far more back in time than the birth of cultivated potato lies the origin of these wild species within *Solanum* section *Petota*. Based on genomic and crossability data of wild potato species, Hawkes (1990) has postulated in his migration hypothesis that the first *Petota* species came from Central America to the South and subsequently experienced radial speciation in the new diverse habitats about 3.5 million years ago, when the Panama isthmus formed a land bridge between the landmasses of North- and South America.

The centre of diversity of potato is in the highlands of the Andes between Bolivia and Peru (Hawkes, 1990). Several wild species in germplasm collections from this region are being used as a resource for isolating new *R* genes (Van der Vossen et al., 2003; Flis et al., 2005; Gebhardt et al., 2004). To date, about 30 *R* genes against nematodes, viruses and fungi have been mapped in *Solanum* species (Gebhardt and Valkonen, 2001) of which many are located in clusters on the genome.

The cluster containing *Rx1* and *Gpa2* has been introgressed into cultivated potato from *S. tuberosum* ssp. *andigena* (Bendahmane et al., 1999). Recently, a functional

orthologue of *Rx1* was found (Bendahmane et al., 2000), which was introgressed from *S. acaule,* a wild relative of cultivated potato (Ritter et al., 1991). *Rx2* was isolated from a library of genes with homology to *Rx1* by employing an *Agrobacterium* transient expression system to identify the candidate gene with the *Rx1* like resistance specificity. *Rx1* and *Rx2* are functionally identical in their recognition of the PVX coat protein. The LRR domain, which is supposed to be the main determinant of *R* gene specificity (Jones, 1997; Ellis et al., 1999), is highly conserved between *Rx1* and *Rx2* orthologues suggesting that this domain is indeed involved in Avr protein recognition. The identification of an ancient recognition specificity for the same *Avr* gene in PVX in two distantly related wild potato species implied that more functional *Rx* alleles could be present within the genus *Solanum*.

In this thesis, we show that a survey of about 51 accessions across 35 species from the genus *Solanum* resulted in the identification of two novel functional orthologues of the disease resistance gene *Rx1* in the wild potato species *S. lepthophyes/sparsipilum* and *S. albicans*, *Rx3* and *Rx4,* respectively, conferring extreme resistance to potato virus X (PVX). Functional analysis of *Rx3* and *Rx4* resulted in a specific hypersensitive response in an agroinfiltration assay on leaves of *N. benthamiana* in the presence of the elicitor of *Rx1*-mediated resistance. An extreme resistance response in a gene-for-gene-manner was observed upon PVX infection of the corresponding genotypes from *S. lepthophyes/sparsipilum* and *S. albicans*. Our finding of *Rx3* and *Rx4,* together with *Rx1* from *S. tuberosum* ssp. *andigena* and *Rx2* from *S. acaule*, represents the first example of functional orthologues from the class of NBS-LRR resistance genes in four different plant species. In addition, the homologous gene Rx_{edn} was identified from *S. edinense* (PI 607474) and although not functional, it was shown that Rx_{edn} harbours the Rx recognition specificity. The remarkable extremely high similarity (>97%) of the LRR domain between the *Rx* orthologues suggest that strong evolutionary constraints affect this locus in order to maintain the Rx recognition specificity.

Sequence comparison between the functional *Rx* orthologues and their paralogues enabled us to design *Rx* specific primers for the screening of an additional set of about 5000 genotypes derived from 1000 accessions across 200 Solanum species. This resulted in the identification of an additional set of 13 different *Solanum* species harbouring *Rx* candidate genes (this thesis, Chapter 3). This shows that the *Rx* orthologues are of ancient origin and

derived from one ancestral gene and that *Rx* mediated PVX recognition arose before the groups of *Tuberosa* (Peru), *Acaule/Demissum* diverged. Analogous to the case of functional orthologues encoding for the pathogen defence related protein kinase *Pto* in tomato species, these findings support trench warfare dynamics acting on the *Rx* recognition specificity in *Solanum*.

Riely et al. (2001) pointed out that the *Pto* locus might be ancient as the genome organisation is conserved between *Lycopersicon esculentum* and *L. pimpinellifolium*. In addition, *Pto* homologues were found to map in syntenic regions in tomato, potato and pepper. However, an ancient origin of a gene or locus is in itself not indicative for trench warfare dynamics. Instead, a long lifespan of *R* genes is determined by a continuous and simultaneous selection for certain pathogen recognition specificities in distinctly related species as shown for the functional *Rx* orthologues (this thesis, Chapter 3). Stahl et al. (1999) estimated that the *Rpm1* polymorphism arose 9.8 million years ago, which was based on comparative analysis of substitution rates between *Rpm1* loci of different *A. thaliana* ecotypes. Similarly, Riely et al. (2001) suggested that *Pto* and *AvrPto* could have coexisted for a few million years as *Pto* mediated resistance arose before the speciation of *L. hirsutum* and *L. pimpinellifolium*.

The geographic distribution of 32 accessions derived from 14 distantly related *Solanum* species harbouring putative functional *Rx* orthologues is shown in Figure 1. For accessions derived from five species no geographic data were available and hence, they are not included in this overview. A majority of 17 accessions was found in the area of Lake Titicaca and Lake Ouro in Bolivia, a region located within the centre of diversity for potato. Accessions from several other species are derived from locations in Argentina and the accession of *S. albicans* harbouring *Rx4* was collected in Peru. The most remarkable observation is the distribution of three *S. edinense* accessions in a remote location in central Mexico. For one accession, we have shown that it contains a non-functional allele that still harbours the *Rx* recognition specificity (this thesis, Chapter 3). Although PVX has a worldwide distribution nowadays, we cannot exclude a permanent or episodic absence of corresponding PVX strains in the habitat of *S. edinense,* which could explain the presence of a non-functional CC-deletion mutant as the result in the lack of selection pressure on the *Rx* gene. Further analysis on the candidate *Rx* orthologues in the genotypes of the other accessions from *S. edinense* will show whether a fully functional *Rx* allele can be found or

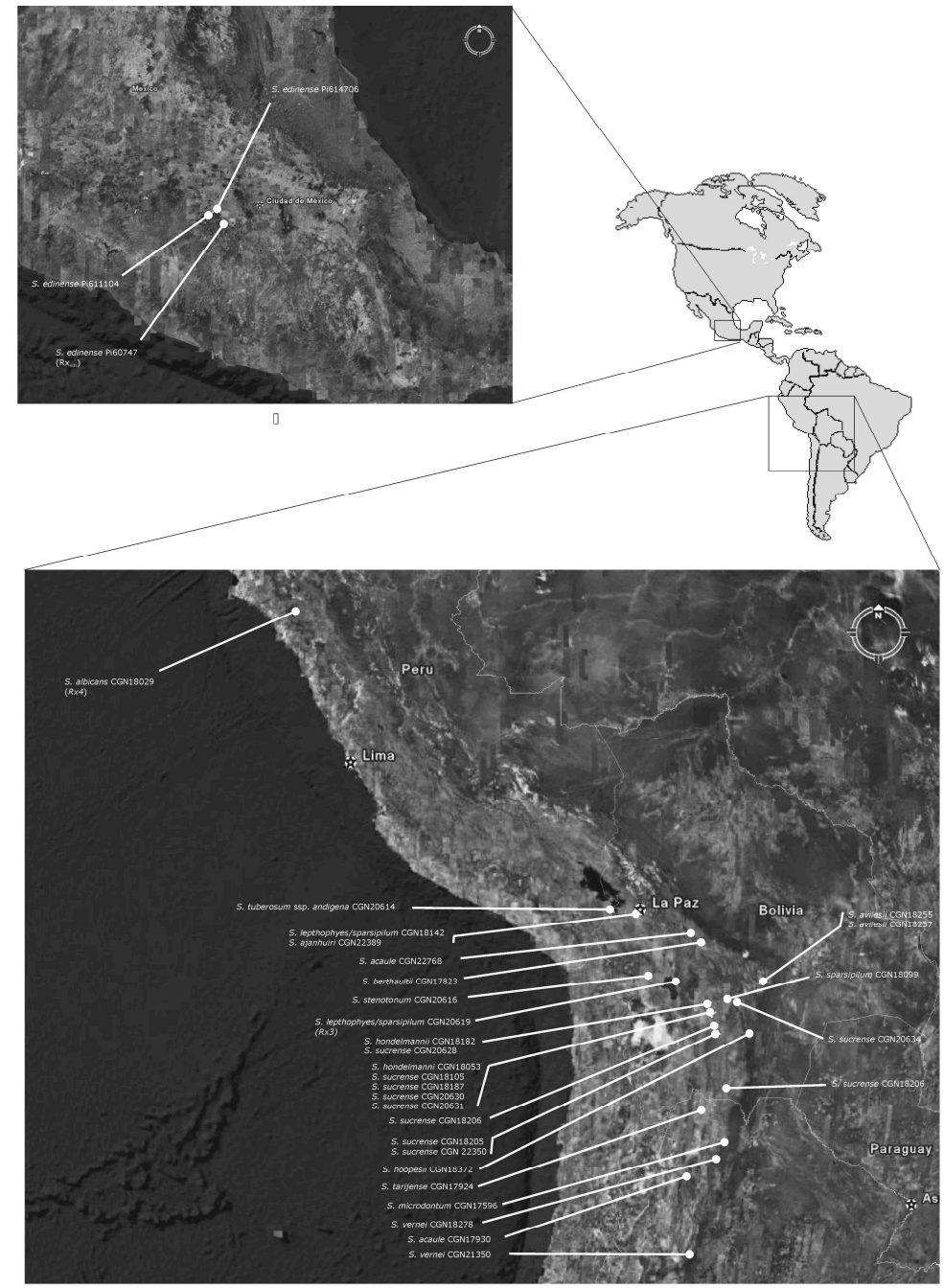


Figure 1. The geographic distribution of 32 accessions from *Solanum* species harbouring the putative functional *Rx* candidate genes, the functional *Rx* orthologues *Rx3*, *Rx4* and Rx_{edn} as described in chapter 3. Accessions were mapped using Google Earth™ on the location information from the accession passport data provided by CGN genebank

whether the non-functional allele is omnipresent in *S. edinense*.

Structural diversity and evolutionary dynamics of *Rx/Gpa2* homologues in *Solanum*

To gain more insight in the structural and evolutionary relationship of the resistance gene cluster in potato that harbours the highly similar potato virus X resistance gene *Rx* and the potato cyst nematode resistance gene *Gpa2*, sequence analysis on the ORFs of 75 highly similar *Rx/Gpa2* homologues derived from 10 *Solanum* species was carried out. This analysis includes the four functional orthologous virus resistance genes *Rx1-4*, the truncated Rx_{edn}, that harbours *Rx* resistance specificity and the nematode resistance gene *Gpa2*. cDNA sequences were obtained for the functional resistance genes *Gpa2* and *Rx1-4* (Van der Vossen *et al,* 2000; Bendahmane *et al.,* 1998*;* Bendahmane *et al.,* 1999*;* this thesis, Chapter 3*)* as well as for two other resistance gene homologues. Overall sequence conservation allowed for the identification of intron splice sites and putative open reading frames for all other genes.

Acquiring such a unique and extensive dataset encompassing related *Rx/Gpa2* homologues from different haplotypes across *Solanum* species facilitates the detection of genetic processes and evolutionary dynamics shaping *R* gene clusters. Recently, the analysis of several hundreds of *RGC2* genes from accessions of *Lactuca sativa* led to the distinction between two types of *R* genes (Kuang et al., 2004). The so-called Type I genes were found to evolve rapidly generating chimaeric structures as a result of frequent sequence exchange resulting in obscure orthologous relationships. Additional characteristic features of this group are diversifying selection and conserved introns. Type II genes on the other hand evolve more slowly, are relatively conserved and recombine infrequently. The structural diversity, evolutionary relationship, patterns of positive selection and sequence exchange observed for the *Rx/Gpa2* homologues (this thesis, Chapter 4) are consistent with a Type I gene classification.

The reason of heterogeneous rates of evolution within *R* gene clusters is not yet clear. Based on the different selection patterns it has been suggested, that Type II genes could harbour indispensable resistance specificities, while Type I genes are functional resistance genes still ”in-progress” (Kuang et al., 2005). About one third of the *Rx/Gpa2* homologues were found to encode for a protein harbouring complete CC-, NBS- and LRR-domains analogous to *Rx* and *Gpa2,* indicating the presence of functional *R* genes with

unknown specificities. In this context, it is anticipated that at least some homologues of the isolated *Rx/Gpa2* homologues could resemble intermediate functional *R* genes that still may undergo changes, whereas others may be more stable like observed for the functional *Rx* orthologues.

Sequence homology with the LRR domain of the porcine ribonuclease inhibitor suggested that the three-dimensional structure of the LRR domain of *R* genes is horse shoe shaped with at the inside a series of parallel β sheets consisting of conserved structural amino acid residues forming the back bone and variable solvent exposed residues that play a role in protein-protein interactions (Jones, 1997). Recently, both the NBS and LRR domains have been modelled for several *R* genes based on sequence similarities to genes with known protein crystal structures supporting the folding of the LRR domain into a horse shoe like structure (Albrecht and Takken, 2006; McHale et al., 2006). Comparing the Rx/Gpa2 homologues a local structural variability was detected in the sequence region encoding LRR 10, dividing the homologues in 3 structural subclasses (this thesis, Chapter 4). To test the hypothesis that this sequence variation affects the predicted horse shoe structure of the LRR domains of the corresponding *Rx/Gpa2* homologues, computational 3D modelling of the full length proteins is currently in progress. A horseshoe structure consisting of 15 leucine rich repeats was obtained for Gpa2, but the majority of the other *Rx/Gpa2* homologues exhibit a linker like structure in place of the LRR10 thereby splitting the LRR in two twisted domains. It will be interesting to find out whether these intriguing differences in the tertiary structure of the Rx1 and Gpa2 protein determines the distinct recognition specificities and resistance responses mediated by these highly homologous R proteins.

Comparative analysis of *R* gene sequences revealed that the LRR domain is the most variable part of the R protein suggesting a role in resistance specificity. This is supported by the observation that the solvent exposed amino acid residues of the LRR β-strand/β-turn motifs are hyper-variable and subject to diversifying selection (Parniske et al., 1997). Furthermore, combining the LRR domain of the flax resistance gene *L2* with the TIR-NBS regions of the *L6* and *L10* alleles resulted in a chimaeric gene product with *L2* specificity (Ellis et al., 1999), which demonstrates that the LRR domain is the main specificity determinant of the R protein. Analysis for diversifying selection revealed very significant positive selection on codon sites within the LRR domain, which is consistent

with a role in pathogen recognition.

However, amino acid residues under diversifying selection were also detected within the CC- and NBS- domain involved in disease signalling. It has been shown for Rx, that intramolecular interactions involving the CC and NBS domain occur (Moffett et al., 2002). Comparison of positive selected sites in the NBS domain on the 3D model of Gpa2 revealed two patches of sites in the ARC domain that lie in two opposing, solvent exposed loop-structures (Andrei Petrescu et al. personal communication). A detailed analysis of the correlation between positive selected sites in the Rx/Gpa2 homologues is in progress to determine the occurrence of coevolution between different R protein structures. By comparison of sequence similarities compared to subsets of the functional *Rx* orthologues and the nematode resistance gene *Gpa2*, we identified potential amino acid positions potentially involved in the pathogen recognition function for *Rx* and *Gpa2*. Comparison with the Gpa2 LRR 3D model shows that these positions orientate on the same site of the horseshoe. Sequence comparison within the functional *Rx* orthologues revealed that strong negative selection has taken place imposed by the recognition specificity in the LRR domain, whereas increased variability was found for *Gpa2* (Bendahmane et al., 2000; Van der Vossen et al., 2000; this thesis, Chapter 3). Moreover, the ratio between non-synonymous and synonymous amino acid substitutions (K_a/K_s) suggested that the LRRs of Gpa2 and Rx1 are subject to positive selection (Van der Vossen et al., 2000). These data suggest that the LRR domain of Gpa2 and Rx1 is involved in determining nematode and virus recognition specificity, respectively.

Intergenic sequence exchange between members of the *Rx1/Gpa2* cluster point out functional constraints of novel *R* genes

Analogous to the immune system of vertebrates, it has been predicted that plants have developed a versatile genetic system to generate novel recognition specificities to sense new variants of a pathogen. Recently, evidence has been obtained that the leucine-rich repeat (LRR) domain, which is found in most resistance (*R*) genes identified to date and is often directly attached to the putative effector domain, is a flexible recognition determinant that enables switching the host plant response from nematode to virus resistance (Slootweg *et al.*, in preparation). The mild inhibition of a nematode-induced multinucleate syncytium could be converted into extreme resistance to Potato Virus X and *visa versa* by exchanging

the LRR domains of the homologous *R* genes *Rx1* and *Gpa2*. These data point out that extensive sequence exchange events as observed for the *Rx/Gpa2* homologues isolated from different *Solanum* species (this thesis, Chapter 4) contributes to the generation of novel recognition specificities.

Functional recombinant R proteins showing the Rx1 recognition specificity were also obtained by intergenic recombination with closely related homologues located at the same cluster on two different haplotypes derived from *S. tuberosum ssp tuberosum* and *S. tuberosum ssp andigena,* respectively. In addition, a shift in the original Rx1 recognition specificity was observed for some domain swaps. Extensive sequence exchange within the LRR domains of *Rx1* and *Gpa2* resulted in the identification of specific regions involved in pathogen recognition, whereas others are shown to play a role in the regulation of the resistance response. This indicates that genetic mechanisms, like unequal crossing-overs or gene conversions, operating on LRRs may alter the specificity of an *R* gene leading to novel resistance specificities, even against a taxonomically unrelated pathogen with a distinct route of invasion, an entirely different mechanism of parasitism and highly deviant symptoms upon infection.

Despite the high sequence homology between *Rx1* and *Gpa2*, intergenic recombination resulted sometimes in loss-of-function mutants or the activation of a constitutive cell death response. Therefore, these data demonstrate on the one hand the compatibility between different members of a single *R* gene cluster, but on the other hand point out the functional constraints for sequence exchange as a mechanism for *R* gene evolution.

References

Albrecht, M., and Takken, F.L.W. (2006). Update on the domain architectures of NLRs and R proteins. Biochemical and Biophysical Research Communications **339,** 459.

Bakker, E., Butterbach, P., Rouppe Van der Voort, J., Van der Vossen, E., Van Vliet, J., Bakker, J., and Goverse, A. (2003). Genetic and physical mapping of homologues of the virus resistance gene *Rx1* and the cyst nematode resistance gene *Gpa2* in potato. Theoretical and Applied Genetics **106,** 1524.

Baumgarten, A., Cannon, S., Spangler, R., and May, G. (2003). Genome-level evolution of resistance genes in *Arabidopsis thaliana*. Genetics **165,** 309.

Bendahmane, A., Kanyuka, K., and Baulcombe, D.C. (1999a). The *Rx* gene from potato controls separate virus resistance and cell death responses. Plant Cell **11,** 781.

Bendahmane, A., Kanyuka, K., Baulcombe, D.C., Querci, M., and Bendahmane, A.

(2000). *Agrobacterium* transient expression system as a tool for the isolation of disease resistance genes: Application to the *Rx2* locus in potato. Plant Journal **21,** 73.

D'Arcy, W.G.W.G. (1991). The *Solanaceae* since 1976 with a review of its biogeography Solanaceae III**,** 75.

Ellis, J.G., Lawrence, G.J., Luck, J.E., and Dodds, P.N. (1999). Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. Plant Cell **11,** 495.

Flis, B., Strzelczyk-Zyta, D., Marczewski, W., Hennig, J., and Gebhardt, C. (2005). The *Ry-fsto* gene from *Solanum stoloniferum* for extreme resistant to Potato virus Y maps to potato chromosome XII and is diagnosed by PCR marker GP122718 in PVY resistant potato cultivars. Molecular Breeding **15,** 95.

Flor, H.H. (1971). Current status of the gene-for-gene concept. Annu. Rev. Phytopathol. **9,** 297.

Gebhardt, C., and Valkonen, J.P.T. (2001). Organization of genes controlling disease resistance in the potato genome. Annual Review of Phytopathology **39,** 79.

Gebhardt, C., Ballvora, A., Walkemeier, B., Oberhagemann, P., and Schüler, K. (2004). Assessing genetic potential in germplasm collections of crop plants by marker-trait association: A case study for potatoes with quantitative variation of resistance to late blight and maturity type. Molecular Breeding **13,** 93.

Grun, P. (1990). The Evolution of Cultivated Potatoes. Economy Botany **44,** 39.

Gu, K., Yang, B., Tian, D., Wu, L., Wang, D., Sreekala, C., Yang, F., Chu, Z., Wang, G.L., White, F.F., and Yin, Z. (2005). *R* gene expression induced by a type-III effector triggers disease resistance in rice. Nature **435,** 1122.

Hawkes, J. (1990). The potato. evolution, biodiversity and genetic resources. Belhaven, London

Hwang, C.F., and Williamson, V.M. (2003). Leucine-rich repeat-mediated intramolecular interactions in nematode recognition and cell death signaling by the tomato resistance protein Mi. Plant Journal **34,** 585.

Jones, D.A., and Takemoto, D. (2004). Plant innate immunity - direct and indirect recognition of general and specific pathogen-associated molecules. Curr. Opin. Immunol. **16,** 48.

Jones, D.A. (1997). The role of leucine-rich repeat proteins in plant defences
Adv. Bot. Res **24,** 89.

Kuang, H., Woo, S.S., Meyers, B.C., Nevo, E., and Michelmore, R.W. (2004). Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. Plant Cell **16,** 2870.

Kuang, H., Wei, F., Marano, M.R., Wirtz, U., Wang, X., Liu, J., Shum, W.P., Zaborsky, J., Tallon, L.J., Rensink, W., Lobst, S., Zhang, P., Tornqvist, C.E., Tek, A., Bamberg, J., Helgeson, J., Fry, W., You, F., Luo, M.C., Jiang, J., Robin Buell, C., and Baker, B. (2005). The *R1* resistance gene cluster contains three groups of independently evolving, typeI *R1* homologues and shows substantial structural variation among haplotypes of *S. demissum*. Plant Journal **44,** 37.

Leister, D. (2004). Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends in Genetics **20,** 116.

Martin, G.B., Bogdanove, A.J., and Sessa, G. (2003). Understanding the Functions of Plant Disease Resistance Proteins. Annual Review of Plant Biology **54,** 23.

McHale, L., Tan, X., Koehl, P., and Michelmore, R.W. (2006). Plant NBS-LRR

proteins: Adaptable guards. Genome Biology **7**.

Michelmore, R.W., and Meyers, B.C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res. **8,** 1113.

Moffett, P., Farnham, G., Peart, J., and Baulcombe, D.C. (2002). Interaction between domains of a plant NBS-LRR protein in disease resistance-related cell death. EMBO Journal **21,** 4511.

Nimchuk, Z., Eulgem, T., Holt Iii, B.F., and Dangl, J.L. (2003). Recognition and Response in the Plant Immune System. Annual Review of Genetics **37,** 579.

Parniske, M., Golstein, C., Thomas, C.M., Wulff, B.B.H., Hammond-Kosack, K.E., Jones, D.A., Harrison, K., and Jones, J.D.G. (1997). Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. Cell **91,** 821.

Riely, B.K., and Martin, G.B. (2001). Ancient origin of pathogen recognition specificity conferred by the tomato disease resistance gene *Pto*. Proceedings of the National Academy of Sciences of the United States of America **98,** 2059.

Ritter, E., Debener, T., Barone, A., Salamini, F., and Gebhardt, C. (1991). RFLP mapping on potato chromosomes of two genes controlling extreme resistance to potato virus X (PVX). Molecular and General Genetics **227,** 81.

Stahl, E.A., Kreitman, M., Bergelson, J., Dwyer, G., and Mauricio, R. (1999). Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. Nature **400,** 667.

Takken, F.L., Albrecht, M., and Tameling, W.I. (2006). Resistance proteins: molecular switches of plant defence. Current Opinion in Plant Biology **9,** 383.

Van der Hoorn, R.A.L., Roth, R., and De Wit, P.J.G.M. (2001a). Identification of distinct specificity determinants in resistance protein *Cf-4* allows construction of a *Cf-9* mutant that confers recognition of AVR4. Plant Cell **13,** 273.

Van der Hoorn, R.A.L., Kruijt, M., Roth, R., Brandwagt, B.F., Joosten, M.H.A.J., and De Wit, P.J.G.M. (2001b). Intragenic recombination generated two distinct *Cf* genes that mediate AVR9 recognition in the natural population of *Lycopersicon pimpinellifolium*. Proceedings of the National Academy of Sciences of the United States of America **98,** 10493.

Van der Vossen, E., Te Lintel Hekkert, B., Wouters, D., Pereira, A., Stiekema, W., Sikkema, A., Gros, J., Stevens, P., Muskens, M., and Allefs, S. (2003). An ancient *R* gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. Plant Journal **36,** 867.

Van der Vossen, E.A.G., Rouppe van der Voort, J.N.A.M., Kanyuka, K., Bendahmane, A., Sandbrink, H., Baulcombe, D.C., Bakker, J., Stiekema, W.J., and Klein-Lankhorst, R.M. (2000). Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: A virus and a nematode. Plant Journal **23,** 567.

Van der Voort, J.R., Wolters, P., Folkertsma, R., Hutten, R., vanZandvoort, P., Vinke, H., Kanyuka, K., Bendahmane, A., Jacobsen, E., Janssen, R., and Bakker, J. (1997). Mapping of the cyst nematode resistance locus *Gpa2* in potato using a strategy based on comigrating AFLP markers. Theor. Appl. Genet. **95,** 874.

Summary

Potato (*Solanum tuberosum* ssp. *tuberosum*) is the fourth most important food crop with an annual yield of about 300 million tons over the world. The history of the domestication of potato shows that disease-causing agents followed the tracks of potato cultivation in temperate climates across continents, resulting in substantial crop losses. Plants including potato have evolved defence mechanisms against pathogens, of which the pathotype-specific system involving resistance genes (*R* genes) is very effective. In search for durable resistance in crop plants, an increasing number of *R* genes has been identified and characterized, providing valuable information on their genomic organisation and evolutionary dynamics. They are often located in *R* gene clusters on complex loci or so called 'hotspots' for resistance in the plant genome.

The evolution of host pathogen-interactions based on gene-for-gene interactions has been described in two prevailing theories. The arms race hypothesis assumes that new *R* gene alleles continuously emerge in a plant population and replace resistance alleles that are overcome by the pathogen. A short lifespan is postulated for a particular *R* gene allele. More recent, the 'trench warfare'-model is proposed to explain the dynamics of *R* genes and *Avr* genes as repeated advances and retreats of resistance and virulence alleles. This frequency dependent selection pressure results in the maintenance of resistance alleles as balanced polymorphisms in natural populations over long periods of time (long lifespan).

The aim of the research described in this thesis was to study the molecular mechanisms underlying the evolution of *R* gene recognition specificities using the virus resistance gene *Rx1* from potato. *Rx1* and its close homologue *Gpa2* are both located in a single *R* gene cluster of about 110 kB on chromosome XII in potato and share an overall homology in the amino acid sequence of about 87%. Interestingly, these two *R* genes confer resistance to taxonomically unrelated pathogens, a virus and a nematode. However, *Rx1* results in extreme resistance to potato virus X (PVX), whereas *Gpa2* gives a much slower and milder resistance response to the potato cyst nematode *Globodera pallida.*

The *Rx1/Gpa2* resistance cluster was analysed on three *S. tuberosum* ssp. *tuberosum* haplotypes derived from the diploid potato clones SH and RH, homeologous to the resistant *S. tuberosum* ssp. *andigena* haplotype of SH, which resulted in the identification of nine additional homologues of *Rx1* and *Gpa2* in this region.

In postulating a long lifespan for *Rx*, the findings of functional orthologues from distant related *Solanum* species give support for the trench warfare hypothesis**.** A survey across 35 species from the genus *Solanum* resulted in the identification of two novel functional orthologues of the disease resistance gene *Rx1* in the wild potato species *S. lepthophyes/sparsipilum* and *S. albicans*, *Rx3* and *Rx4.* By functional assays it was shown, that *Rx3* and *Rx4* confer *Rx*-mediated resistance to PVX in the corresponding *S. lepthophyes/sparsipilum* and *S. albicans* genotypes. Our finding of *Rx3* and *Rx4,* together with *Rx1* from *S. tuberosum* ssp. *andigena* and *Rx2* from *S. acaule*, represents the first example of functional orthologues from the class of NBS-LRR resistance genes in four different plant species.

Subsequent screening of about 5000 genotypes derived from 1000 accessions across 200 *Solanum* species resulted in the identification of 14 additional distantly related *Solanum* species harbouring *Rx* candidate genes, which shows that the *Rx* orthologues are of ancient origin and derived from one ancestral gene. This indicates that *Rx* mediated PVX recognition arose before the groups of *Tuberosa* (Peru) and *Acaule/Demissum* diverged.

To gain more insight in the structural and evolutionary relationship of the resistance gene cluster a set of 75 *Rx1/Gpa2* homologues was obtained from 10 different *Solanum* species, including the four functional orthologous virus resistance genes *Rx1-4*, the truncated Rx_{edn} that harbours *Rx* resistance specificity and the nematode resistance gene *Gpa2*. About one third of the homologues were found to encode for a protein harbouring complete CC-, NBS- and LRR- domains analogous to *Rx* and *Gpa2,* indicating the presence of functional *R* genes with unknown specificities. We identified a local structural variability in the sequence region encoding LRR10, dividing the homologues in 3 structural subclasses. The occurrence of similar sequence exchange tracks in homologues isolated from different *Solanum* species indicates that interallelic exchanges took place before speciation. Analysis for positive selection revealed very significant positive selection on codon sites within the LRR domain, but also within the CC- and NBS- domain.

In addition, the genomic organisation of the *Rx1/Gpa2* cluster was analysed. Therefore, sequence analysis was carried out on several BAC clones derived from *S. tuberosum* ssp. *andigena*, *S. tuberosum* ssp. *tuberosum* and *S. lepthophyes/sparsipilum* harbouring in total 17 *Rx1/Gpa2* homologues, including the functional resistance genes *Rx1* and *Gpa2* from potato. The homeologous chromosomal regions derived from *S. tuberosum* ssp. *andigena* and *S. tuberosum* ssp. *tuberosum* were assembled and aligned based on syntenic regions. However, comparative sequence analysis revealed large intergenic sequence divergence in this region, which was increasing for more distant related Solanum species. The occurrence of a family of conserved genes with homology to RING zink-finger protein encoding genes in vicinity of the *Rx/Gpa2* homologues suggests that tandem duplication of these genes has occurred at this locus.

The major results described in this thesis are discussed and an outlook on structure-function analyses of *Rx1* and *Gpa2* is presented.

Samenvatting
Aardappel (*Solanum tuberosum* ssp. *tuberosum*) is het vierde voedsel gewas wereldwijd met een jaarlijkse opbrengst van ongeveer 400 miljoen ton. De geschiedenis van de domesticatie van aardappel toont aan dat ziekteverwekkers het spoor van de aardappel volgen, wat tot substantiële productie verliezen leidt. Planten, waaronder aardappel, hebben verdedigingsmechanismen ontwikkeld die samen met de potentiële ziekteverwekkers zijn geëvolueerd. Zeer effectief is een systeem dat gebruikt maakt van resistentiegenen (*R* genen). De eiwitten waarvoor de *R* genen coderen zijn in staat om speciefieke eiwitten afkomstig van een pathogeen te herkennen. De laatste eiwitten worden aangeduid als avirulentieproducten en de genen die ervoor coderen als avirulentiegenen (*Avr* genen). Het mechanisme waarin één *R* genproduct (direct of indirect) specifiek interacteert met één *Avr* genproduct wordt ook wel gen-om-gen interactie genoemd. Tijdens de zoektocht naar duurzame resistentie in voedselgewassen is een nog steeds toenemend aantal *R* genen geïdentificeerd en gekarakteriseerd. Dit heeft belangrijke informatie over de genomische organisatie en de evolutionaire dynamiek opgeleverd. *R* genen zijn bijvoorbeeld vaak georganiseerd in complexe clusters in het genoom; zogenaamde 'hotspots' voor resistentie.

De evolutie van de gastheer-pathogeen interactie die gebaseerd is op de gen-om-gen interactie wordt beschreven volgens de twee volgende hypotheses. De eerste hypothese is de wapenwedloophypothese ('arms race hypothesis') die veronderstelt dat nieuwe *R* gen-allelen continue ontstaan in een plantenpopulatie en dat resistentie-allelen die door de pathogeen doorbroken zijn vervangen worden. Deze hypothese postuleert een korte levensverwachting voor een specifiek *R* gen-allel. Meer recentelijk is de loopgraafhypothese ('trench warfare hypothesis') voorgesteld. In deze hypothese wordt verondersteld dat de dynamiek van *R* genen en *Avr* genen te verklaren is door een herhaaldelijke vooruitgang en teruggang van resistentie- en virulentie-allelen. De frequentie van deze dynamiek is afhankelijk van de selectiedruk en resulteert in het behoud van resistentie-allelen als gebalanceerde polymorfismen in een natuurlijke populatie gedurende langere tijd (lange levensverwachting).

Het doel van het onderzoek beschreven in dit proefschrift is het bestuderen van de moleculaire mechanismen die ten grondslag liggen aan de evolutie van de specificiteit van *R* gen herkenning. Hierbij maken we gebruik van het virusresistentiegen *Rx1* in aardappel. *Rx1* en het nauwverwante nematoderesistentiegen *Gpa2* bevinden zich beide in één *R* gen cluster. In dit cluster zijn nog twee resistentiegenhomologen geidentificeerd, waarvan er één getrunceerd is en één een volledig functioneel gen met onbekende specificiteit is. De lengte van dit *R* gen cluster is ongeveer 110 kb en ligt op een introgressiesegment afkomstig van *S. tuberosum* ssp. *andigena* op chromosoom 12 van de diploide aardappel SH. De overeenkomst tussen de eiwitsequenties van *Rx1* en *Gpa2* is ongeveer 87%. Ondanks de grote overeenkomst induceren deze twee *R* genen resistentie tegen twee taxonomisch ongerelateerde ziekteverwekkers. Daarnaast lijkt de werking van deze twee genen ook totaal verschillend te zijn. *Rx1* resulteert in een extreem snelle resistentiereactie tegen het aardappelvirus X (PVX) terwijl *Gpa2* een veel langzamere en mildere resistentie tegen het aardappelcystenaaltje*Globodera pallida* vertoont. In de aardappelsoort *S. acaule* is een gen geidentificeerd dat zeer homoloog, maar niet identiek, is aan *Rx1*. Dit gen, *Rx2*, heeft exact dezelfde specificiteit en werking als *Rx1* en is dus een functionele ortholoog.
Om te beginnen is het *Rx1/Gpa2* resistentiecluster van drie *S. tuberosum* spp. *tuberosum* haplotypen afkomstig van de diploide aardappels SH en RH geanalyseerd. Dit resulteerde in de identificatie van negen additionele homologen van *Rx1* en *Gpa2* in deze regio.

Vervolgens is een studie gedaan naar het voorkomen van functionele *Rx1* orthologen in wilde aardappelsoorten. De ontdekking van functionele orthologen in ver verwante *Solanum* soorten zal een lange levensduur van *Rx* en daarmee het loopgraaf model

ondersteunen. In eerste instantie zijn 35 soorten, die allemaal tot het geslacht *Solanum* behoren getest. Dit resulteerde in de ontdekking van twee nieuwe functionele orthologen van het resistentiegen *Rx1*. Deze twee functionele orthologen zijn ontdekt in een natuurlijke hybride van de wilde aardappelsoorten *S. lepthophyes* en *S. sparsipilum* en in de soort *S. albicans. Ze* hebben de namen *Rx3* en *Rx4* gekregen. Een functionele analyse van *Rx3* en *Rx4* toont aan dat deze twee orthologen resistentie tegen PVX induceren in respectievelijk *S. lepthophys/sparsipilum* en *S. albicans*. De identificatie van *Rx3* en *Rx4* samen met *Rx1* en *Rx2* is het eerste voorbeeld van functionele orthologen uit de klasse van NBS-LRR resistentiegenen in vier verschillende plantensoorten. Hieropvolgend zijn ongeveer 5000 genotypen, afgeleid van 1000 accessies die 200 *Solanum* soorten vertegenwoordigen geanalyseerd. Dit resulteerde in de ontdekking van 14 additionele *Rx* kandidaatgenen vanuit ver gerelateerde *Solanum* soorten. Onder deze kandidaatgenen bevindt zich een gen ($Rx_{edn)}$ dat wel de *Rx* specificiteit bezit, maar niet functioneel is vanwege een deletie in de eerste helft van het gen. Uit deze resultaten kan afgeleid worden dat PVX herkenning door *Rx* orthologen relatief oud is en dat de Rx orthologen waarschijnlijk afgeleid zijn van één voorvaderlijk gen. Dit gen is waarschijnlijk ontstaan voordat de groepen *Tuberosa* (Peru) en *Acaule/Demissum* gedivergeerd zijn.

Om meer inzicht te krijgen in de structurele en evolutionere relaties van dit resistentiegencluster, zijn de sequenties van 75 *Rx1/Gpa2* homologen van 10 verschillende *Solanum* soorten verder geanalyseerd. Deze analyse behelsde ook de vier functionele orthologe virusresistentiegenen *Rx1-4*, het getrunceerde gen Rx_{edn} en het nematoderesistentiegen *Gpa2*. Ongeveer eendere van de homologen codeert voor een eiwit met een compleet CC-, NBS- en LRR- domein, analoog aan *Rx* en *Gpa2*. Dit suggereert de aanwezigheid van functionele *R* genen met onbekende specificiteit. Een structureel variabele sequentie is geïdentificeerd in LRR10. Op basis van deze variabele sequentie kunnen de homologen in 3 structurele onderklassen worden indeeld. De aanwezigheid van sporen van sequentieuitwisselingen in de homologen van verschillende *Solanum* soorten geeft aan dat er uitwisseling tussen allelen heeft plaatsgevonden vóór het ontstaan van de verschillende soorten. Analyse voor positieve selectie toont een significante positieve selectie aan in de codons binnen het LRR domein, het CC domein, en het NBS- domein.

De genomische organisatie van het *Rx1/Gpa2* cluster is verder geanalyseerd met behulp van de sequenties van verschillende BAC klonen afkomstig uit *S. tuberosum* ssp. andigena en *S. tuberosum* ssp. *tuberosum.* Deze BAC klonen werden geassembleerd en aligned op basis van syntenische regio's. De analyse toonde aan dat de sequenties in deze regio een grote onderlinge afwijking laten zien, die toe neemt naarmate de soorten waaruit deze sequenties afkomstig minder nauw aan elkaar verwant zijn. De aanwezigheid van een familie van geconserveerde genen met homologie aan RING zincvingergenen in de omgeving van de *Rx/Gpa2* homologen suggereert dat tandem duplicaties van deze genen is voorgekomen op dit locus.

De resultaten zijn beschreven in dit proefschrift en worden bediscussieerd. Daarnaast wordt een vooruitblik op de structurele en functionele analyse van *Rx1* en *Gpa2* gepresenteerd.

Zusammenfassung

Die Kartoffel (*Solanum tuberosum* ssp. *tuberosum*) ist die viertwichtigste Feldfrucht mit einem jährlichen Ertrag von ungefähr 300 Millionen Tonnen weltweit. Die Geschichte der Kultivierung von Kartoffeln zeigt, dass Pflanzenschädlinge der geographischen Ausbreitung des Kartoffelanbaus folgten, was schließlich zu erheblichen Ertragsverlusten vor allem in den gemäßigten Klimaregionen führte. Pflanzen wie die Kartoffel haben Abwehrmechanismen gegen Schädlinge entwickelt. Sehr effektiv ist ein schädlingsspezifisches Abwehrsystem, das auf Resistenzgenen (*R* Genen) beruht. Auf der Suche nach dauerhaften Resistenzen werden zunehmend mehr *R* Gene gefunden und beschrieben. Daraus konnten wertvolle Erkenntnisse zur genetischen Organisation und evolutionären Dynamik gewonnen werden. So befinden sich *R* Gene oft in komplexen Ansammlungen von mehreren Genen (Gen-Clustern), den sogenannten "Hotspots" für Resistenzgene.

Für die Evolution von Wirts-Parasit-Interaktionen, die auf Gen-für-Gen Wechselwirkungen beruhen, gibt es zwei unterschiedliche Theorien. Die ‚arms-race' (Wettrüsten) Hypothese besagt, dass immerwährend neue *R* Gen Allele innerhalb einer Pflanzenpopulation entstehen und die durch den Schädling überwundenen Allele ersetzen. Daraus folgt, dass jedes einzelne *R* Gen Allel für sich gesehen eine kurze Lebensdauer hat. Vor einiger Zeit wurde eine zweite Hypothese, genannt ‚trench warfare' (Grabenkampf) aufgestellt. Diese Hypothese beschreibt die Dynamik von *R* Genen und *Avr* Genen als sich wiederholende Episoden von Ausbreitung und Rückzug der entspechenden Resistenz- und Virulenzallele. Da der Selektionsdruck auf ein einzelnes Allel von seiner Häufigkeit abhängt, bleiben *R* Gen Allele in einem polymorphen Gleichgewicht innerhalb natürlicher Populationen über lange Zeiträume hinweg erhalten (lange Lebensdauer).

Ziel des Forschungsprojektes in dieser Dissertation war es, die molekularen Mechanismen, welche der Evolution von *R* Gen Erkennungspezifitäten zugrunde liegen, am Beispiel des Virusresistenz-Gens *Rx1* der Kartoffel zu analysieren. *Rx1* und das nah verwandte, homologe Gen *Gpa2* befinden sich beide im selben *R* Gen Cluster. Die Größe des Clusters beträgt ungefähr 110 000 Basenpaare und befindet sich auf Chromosom XII der Kartoffel. Die Aminosäuresequenzen von *Rx1* und *Gpa2* sind zu 87% identisch. Interessanterweise kodieren diese beiden *R* Gene für Resistenzen gegen taxonomisch nichtverwandte Schädlinge, einen Virus und einen Fadenwurm (Nematode). Es gibt jedoch Unterschiede bei der Resistenzwirkung. Während *Rx1* eine extrem effektive Resistenz gegen den Kartoffelvirus X (PVX) hervorruft, bildet sich durch *Gpa2* eine langsamere und schwächere Teilresistenz gegen den Nematoden *Globodera pallida* aus.

In drei verschiedenen Haplotypen von *S. tuberosum* ssp. t*uberosum*, welche den diploiden Kartoffelklonen SH und RH entstammen und homolog zum resistenzgebenden Haplotyp von *S. tuberosum* ssp. *andigena* sind, wurde der *Rx1*/*Gpa2* Cluster analysiert. Dabei wurden in dieser Region neun weitere Gene mit Homologie zu *Rx1*/*Gpa2* gefunden.

Die Entdeckung von funktionellen Orthologen aus fern miteinander verwandten *Solanum* Arten (Familie: Nachtschattengewächse) unterstützt die These für eine lange Lebensdauer von *Rx* und damit die ‚trench warfare' Hypothese. Eine Studie, die 35 Arten der Gattung *Solanum* umfasste, führte zur Entdeckung von zwei neuen funktionellen Orthologen des Resistenzgens *Rx1*. Die *Rx3* und *Rx4* genannten Gene wurden in den wildwachsenden, kartoffelverwandten Arten *S. lepthophyes/sparsipilum* and *S. albicans* gefunden. Die funktionelle Analyse zeigt, dass *Rx3* und *Rx4* eine *Rx*-spezifische PVX Resistenz in *S. lepthophyes/sparsipilum* und *S. albicans* hervorrufen. Unsere Ergebnisse für

Rx3 und *Rx4*, zusammen mit den bereits bekannten Genen *Rx1* aus *S. tuberosum* ssp. *andigena* und *Rx2* aus *S. acaule* ist das erste bekannte Beispiel für die Existenz von funktionellen Orthologen der NBS-LRR Strukturklasse aus vier verschiedenen Pflanzenarten.

In einem darauffolgenden Screening wurden ungefähr 5000 Genotypen aus 1000 Akzessionen (gesammelte Pflanzenproben), die wiederum 200 Solanum- Arten umfassen, auf die Anwesenheit von funktionellen *Rx* Orthologen untersucht. In weiteren 14 verwandten *Solanum* Arten konnten *Rx* Genkandidaten gefunden werden. Daraus lässt sich schließen, dass der Ursprung der *Rx* abhängigen Resistenz gegen den PVX Virus sehr alt sein muss und diese funktionellen *Rx* Orthologe von einem gemeinsamen (Ur-)Gen abstammen. Die Ergebnisse zeigen außerdem, dass *Rx* abhängige Resistenz gegen PVX schon vor der Aufspaltung der Gattung *Solanum* in die taxonomischen Gruppierungen *Tuberosa* und *Acaule/Demissum* entstanden ist.

Um einen Einblick in die strukturellen und evolutionären Zusammenhänge des Resistenzgen-Clusters zu bekommen, wurden 75 *Rx1/Gpa2* Homologe, die aus 10 verschiedenen *Solanum* Arten isoliert sind, weitergehend analysiert. Diese Analyse umfasste auch die 4 funktionellen orthologen *Rx* Virus-Resistenzgene Rx1-4, das verkürzte Gen Rx_{edn} , das ebenfalls eine *Rx* Erkennungspezifität aufweist, und das Nematoden-Resistenzgen *Gpa2*. Ungefähr ein Drittel der Homologe kodieren analog zu *Rx1* und *Gpa2* für ein Protein mit kompletter CC-, NBS- und LRR-Strukturdomäne. Dies deutet darauf hin, dass es sich hierbei ebenfalls um funktionelle *R* Gene, jedoch mit unbekannter Erkennungspezifität handelt. Wir entdeckten eine variable Struktur in der Sequenzregion, die für LRR10 kodiert. Die Anwesenheit von Sequenzauswechselungen in den Homologen weist darauf hin, dass genetischer Austausch zwischen Allelen vor der Enstehung dieser Arten stattgefunden hat. Die Analyse zur positiven Selektion zeigte significant positive Selektion in Kodons der LRR-Domäne, aber auch in CC- und NBS-Domäne.

Die genetische Organisation des *Rx1/Gpa2* Clusters wurde außerdem anhand von BAC Klon- Sequenzen aus *S. tuberosum* ssp. *andigena*, *S. tuberosum* ssp. *tuberosum* und *S. lepthophyes/sparsipilum* analysiert. 17 *Rx1/Gpa2* Homologe, darunter die funktionellen *R* gene *Rx1* und *Gpa2*, konnten auf DNA-Abschnitten identifiziert werden. Für *S. tuberosum* ssp. *andigena* und *S. tuberosum* ssp. *tuberosum* wurden homologe Chromosomenregionen identifiziert und, basierend auf syntenischen Abschnitten, ein Sequenzalignment erstellt. Eine vergleichende Sequenzanalyse zeigte jedoch, dass es große Unterschiede zwischen Sequenzabschnitten verschiedener Haplotypen gibt. Diese Unterschiede nehmen weiter zu, je weitläufiger die entsprechenden Arten miteinander verwandt sind. Eine Gruppe von zueinander beinahe identischen Genen wurde gefunden, die Ähnlichkeit mit RING Zink-Finger-Protein kodierenden Genen besitzen. Ihre genetische Anordnung in unmittelbarer Nähe der *Rx1/Gpa2* Homologe unterstützt die Annahme, dass gekoppelte Duplikation dieser Gene in diesem DNA-Abschnitt stattgefunden hat.

Die wichtigsten Ergebnisse dieser Dissertation werden abschließend diskutiert und ein Ausblick auf weitere Analysen zur Struktur und Funktion von *Rx1* und *Gpa2* präsentiert.

Acknowledgements

At this point I would like to thank the people that helped me, directly or indirectly, to accomplish this thesis.

Jaap, I would like to thank you for the trust you put in me and that you gave me the opportunity to work in this excellent research group. I still remember making plans for the first experiments to isolate functional *Rx* orthologues at the start of my PhD thesis, but also some doubts if we will be able to find any. But it was a success and here we stand now with a really nice dataset in our hands!

Aska, you are a great supervisor. I appreciated the scientific debates with you and I learned a lot during the discussions. Most important for me, you always gave me new motivation when I was down, e.g. because an experiment had failed. Thank you for intense supervision during the past five years!

Erin, your thesis was the fundament for my one and everybody knows how important it is to have such a solid fundament. You teached me a lot about *R* genes, molecular technics and bioinformatics. I enjoyed our cooperation!

Geert, thank you for the critical views on manuscripts and science in general.

Many thanks go to all my colleagues in the laboratory of Nematology. I have to say that I liked especially the communicative atmosphere in this place right from the beginning. At this point I will start with my room mates as there are: Aneta; We had a great time there! Jan van Bezooijen; Thank you for the nice time downstairs. And finally Kamila; There was always time for a chat! Erik, *Rx* is our passion. For you all the best with finishing your thesis! Piotr, you got PAML working! Thanks for being my paranymph. Dieuwertje, thank you for the help. Nikkie, I will always remember our Mexican hurricane adventure, that was quite something, greetings also to Dave and Sid! Hein, I enjoyed the dinner partys in your place! Jan R., thanks a lot for the nice time in the lab. I learned a lot from you. Sven and Joost, it was a pleasure to work in the lab with you, we had quite some fun. The EK/WK-Oranje-lab deco was really impressive! Rikus, thank you for the always so competent support in the lab. Sajid, I guess you are the next… All the best for the last steps! Martijn, thank you for the help with the phylogenetic trees.

Lisette and Priscilla, thank you for the support, because sometimes forms for just easy things can be so difficult to fill out... . Lisette, I especially appreciate that you are very

active in teaching the dutch language to the foreign students in the lab, keep up the good work!

I want to thank Edwin van der Vossen and Marjon Arens for the support and lively discussions. Special thanks also go to Mirjam Jacobs, Ben Vosman and Ronald van den Berg for providing information and material on the taxonomy of *Solanum*.

I would like to thank the contact persons from the Bosch Foundation, dr. Ingrid Wünning, Atje Drexler and Michaela Reininger, for their good work. I enjoyed the annual network meetings a lot (in the beautiful locations in southern Bavaria…), they always have a very friendly atmosphere.

Special attention deserves of course my polish friends (and their partners…) from Nematology and around Wageningen. Pikutek, I am glad that you are my paranymph. Thank you and also your husband Richard for the help with finalizing this book! Pisiek, thank you for being Pisiek! Dziękuję bardzo.

Marysia and Piotrek, I was always very welcomed in your home, thank you for the nice Settler and Carcassonne board gaming sessions! Agniezka, I enjoyed our regular scientific debates on evolution, PhD, women and men, or paranormal phenomena or any combinations of these subjects. Also my thanks go to André and Tomek. Ula, nice to have you back here together with Hassan and Hannah.

A warm 'thank you' goes to the people from the NIOO for the regular invitations to birthday dinners in one of the Wageningen restaurants and drinks in the pubs (and I still memorize the crazy session of dutch Trivial Pursuit we once had!).

I would like to thank my parents, who enabled my study, for the support with this thesis and also for the patience with various home experiments during my childhood involving insects and other bugs. And my sister Pia, I wish you all the best with your future career.

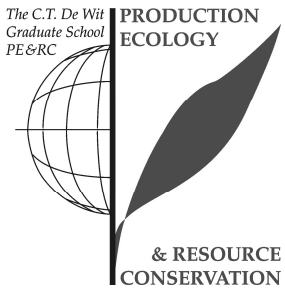
Finally I want to say that finishing a thesis very often comes along with long working evenings and this thesis was not an exception... So, Ania, I thank you for the patience and active support in the past months!

Patrick

Curriculum vitae

Patrick Bernhard Ernst Butterbach was born on the 30th of June 1976 in Trier-Ehrang, Germany. Already during his school time an affinity with science (and even potato!) became apparent through a short research project on the “Vitamin C content in potato cultivars”. In June 1994, he obtained his graduation diploma at the St. Willibrord Gymnasium, Bitburg, awarded for the best results in Science by the national funds of the chemistry industry. After that he was called to fulfill his duty for society by absolving 15months civil service at the Caritasverband Bitburg e.V. He started his study in biology at the University of Kaiserslautern in September 1995, with the specialization on fungal biotechnology. During a 2 months research fellowship at Pharma Research Center, Bayer Life Science, Wuppertal, Germany he got the opportunity to work on bacterial isolates for novel bioactive-compound screenings. In 2000 he began his thesis entitled “Establishing a culture and a bio-assay of *H. schachtii* and isolation of nematicidal metabolites out of higher fungi” at the Fraunhofer institute for biotechnology and metabolite research, Kaiserslautern University, Germany and in 2001 he completed his university education with the diploma in biology. After that he obtained a 9month research scholarship at the Laboratory of Nematology, Plant Sciences, Wageningen UR. There, being granted with a PhD fellowship from the Robert Bosch Foundation he worked on his PhD thesis on the molecular evolution of resistance genes in potato species. The obtained results of these past 5 years are presented in this dissertation.

PE&RC PhD Education Certificate



With the educational activities listed below the PhD candidate has complied with the educational requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

Review of Literature (5.6 credits)

- Evolution of resistance gene clusters: recent case studies on arms race versus trench warfare dynamics for functional resistance genes

Post-Graduate Courses (8.4 credits)

- Basic and advanced statistics; PE&RC (2003
- Summer school "The analysis of natural variation within crop and model plants", Wageningen; EPS (2003)
- Summer school "Functional genomics: theory and hands-on data analysis", Utrecht; EPS (2003)
- Molecular phylogenies: reconstruction and interpretation; EPS (2004)
- Bio-information technology-1; VLAG (2004)

Competence Strengthening / Skills Courses (1.4 credits)

- Techniques for writing and presenting a scientific paper; PE&RC (2004)

Discussion Groups / Local Seminars and Other Scientific Meetings (6.1 credits)

- CBSG potato cluster meetings (2004-2007)
- PE&RC discussion group "In the tracks of evolution"(2005)
- Annual ALW dagen Lunteren, the Netherlands, 2 days (2005)
- 4e Gewasbeschermingsmanifestatie: is het al tijd om te oogsten?!, Ede, the Netherlands (2005)
- WCS day in Utrecht (2006)
- EPS theme day/WCS day in Amsterdam, talk (2007)

PE&RC Annual Meetings, Seminars and Introduction Days (1.2 credits)

- PE&RC day (2003)
- Introduction weekend (2003)

International Symposia, Workshops and Conferences (12.2 credits)

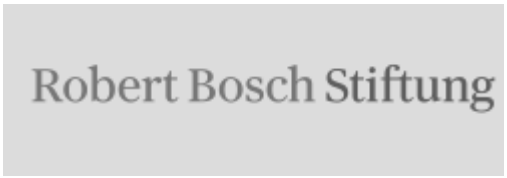
- 12th International congress on molecular plant-microbe interactions, Merida, Mexico (2005)
- 4th Plant genomics European meeting, Amsterdam, the Netherlands (2005)
- 10th International congress of the European society for evolutionary biology, Krakau, Poland (2005)
- 12th *Solanaceae* genome workshop, centre of bio-systems and genomics, Wageningen, the Netherlands
- 4th International meeting with talk at the NEIKER institute, Victoria, Spain (2007)

Courses in which the PhD Candidate has Worked as a Teacher

- Environmental sciences; PSG-Nematology, 14 days

The research presented in this thesis has been carried out at the Laboratory of Nematology, Plant Sciences, Wageningen University.

This research project was funded by a grant from the Robert Bosch Foundation. Further financial support was provided by the EU project APOPHYS QLRT-2001-01849 and the Centre of Biosystems Genomics (CBSG), Wageningen, the Netherlands.



Printed by: Gildeprint B.V., Enschede

About the cover

The cover presents two drawings from the autograph manuscript of Guaman Poma's *Nueva corónica y buen gobierno* (Gl. kgl. S. 2232, 4º). This manuscript is part of the collection in the Royal Library Museum in Copenhagen, Denmark. Recently, it has been included in the UNESCO's "Memory of the World" list.

It was written between the years of 1600 and 1616 by Felipe Guaman Poma de Ayala, an Andean Indian from the region of Huamanga (Ayacucho, Peru). The document is actually an extremely long letter-petition destined for Philip II, king of Spain. However, it never reached this address. In the 1190 pages of the original manuscript, with its 398 drawings, the author proposed in the first part ("new chronicle") to reconstruct the history of humanity in the Andes, from its origins to the last Incas, interweaving it with biblical history and presenting a detailed portrait of the social, political, and religious organization of Tahuantinsuyu (the Inca empire). In the second part ("good government") he denounced the miserable and oppressed condition of the indigenous Andean population under Spanish rule, suggesting to the monarch a series of corrective measures to reverse the perilous trend.

The manuscript contains a lot of information on the daily life of the indigenous population in that era, like in this example of potato cultivation.

Source: 2004. The illustrated Codex (1615/1616) of Felipe Guaman Poma de Ayala (2004). Toward a new Era of Reading. Fund og Forskning 43, pp. 423-436

Front page
Drawing 391(page 1157). June: Time of digging up the potatoes; *Hawkay Kuski Killa*, month of rest after the harvest.
Back page
Drawing 397(page 1175). December: Time of planting potatoes and uqa, tubers; *Qhapaq Inti Raymi Killa*, month of the festivity of the lord sun.