MOLECULAR CLONING AND CHARACTERISATION OF POTENTIAL *FUSARIUM* RESISTANCE GENES IN BANANA (*Musa acuminata* ssp. *malaccensis*)

by

SANTY PERAZA ECHEVERRIA

Plant Biotechnology Program

Science Research Centre

A thesis submitted for the degree of Doctor of Philosophy at

the Queensland University of Technology

2007

Abstract

Banana is the most important fruit crop in the world but ironically one of the crops least studied. This fruit constitutes a major staple food for millions of people in developing countries and also it is considered the highest selling fruit in the world market making this crop a very important export commodity for the producing countries. At the present time, one of the most significant constraints of banana production that causes significant economical losses are fungal diseases. Among these, Panama disease, also known as Fusarium wilt has been the most catastrophic. Panama disease is caused by the soil-borne fungus *Fusarium oxysporum* formae specialis (f.sp) *cubense* (FOC), which infects susceptible bananas through the roots causing a lethal vascular wilt. To date, the race 4 of this pathogen represents the most serious threat to banana production worldwide since most of the commercial cultivars are highly susceptible to this pathogen. Introduction of FOC resistance into commercial cultivars by conventional breeding has been difficult because edible bananas are sterile polyploids without seeds. Genetic transformation of banana, which has already been established in various laboratories around the world has the potential to solve this problem by transferring a FOC race 4 resistance gene into susceptible banana cultivars (eg. Cavendish cultivars). However, a FOC resistant (R) gene has not been isolated. Genes that confer resistance to *Fusarium oxysporum* have been isolated from tomato and melon using a map-based positional cloning approach. The tomato *I2* and melon *Fom-2* genes belong to the non-Toll/interleukin like receptors (TIR) subclass of nucleotide-binding site and leucine-rich repeat (NBS-LRR) R genes. These genes confer resistance only to certain races of *F. oxysporum* in their corresponding plant families limiting their use in other plant families. The fact that these two *Fusarium* resistance genes share the same basic non-TIR-NBS-LRR structure suggests a similar *Fusarium* resistance mechanism is shared between the families Solanaceae and Cucurbitaceae. This observation opens the possibility to find similar *Fusarium* resistance genes in other plant families including the Musaceae.

A remarkable discovery of a population of the wild banana *Musa acuminata* subspecies (ssp.) *malaccensis* segregating for FOC race 4 resistance was made by Dr. Ivan Buddenhagen (University of California, Davis) in Southeast Asia. Research carried out at Queensland Department of Primary Industries (Australia) using this plant material has demonstrated that a single dominant gene is involved in FOC race 4 resistance (Dr. Mike Smith, unpublished results). Tissue-culture plantlets of this FOC race 4 segregating population were kindly provided to the Plant Biotechnology Program (Queensland University of Technology) by Dr. Mike Smith to be used in our research. This population holds the potential to assist in the isolation of a FOC race 4 resistance gene and other potential *Fusarium* resistance genes. The overall aims of this research were to isolate and characterise resistance gene candidates of the NBS-type from *M. acuminata* ssp*. malaccensis* and to identify and characterise potential *Fusarium* resistance genes using a combination of bioinformatics and gene expression analysis.

Chapter 4 describes the isolation by degenerate PCR of five different classes of NBS sequences from banana (*Musa acuminata* ssp *malaccensis*) designated as resistance gene candidates (RGCs). Deduced amino acid sequences of the RGCs revealed the typical motifs present in the majority of known plant NBS-LRR resistance genes. Structural and phylogenetic analyses showed that the banana RGCs are related to non-TIR subclass of NBS sequences. The copy number of each class was estimated by Southern hybridisation and each RGC was found to be in low copy number. The expression of the RGCs was assessed by RT-PCR in leaf and root tissues of plants resistant or susceptible to *Fusarium oxysporum* f. sp. *cubense* (FOC) race 4. Four classes showed a constitutive expression profile whereas no expression was detected for one class in either tissue. Interestingly, a transcriptional polymorphism was found for RGC2 whose expression correlated with resistance to FOC race 4 suggesting a possible role of this gene in resistance to this devastating FOC race. Moreover, RGC2 along with RGC5 showed significant sequence similarity to the *Fusarium* resistance gene *I2* from tomato and were chosen for further characterisation. The NBS

sequences isolated in this study represent a valuable source of information that could be used to assist the cloning of functional R genes in banana.

Chapter 5 describes the isolation and characterisation of the full open reading frame (ORF) of RGC2 and RGC5 cDNAs. The ORFs of these two banana RGCs were predicted to encode proteins that showed the typical structure of non-TIR-NBS-LRR resistance proteins. Homology searches using the entire ORF of RGC2 and RGC5 revealed significant sequence similarity to the *Fusarium* resistance gene *I2* from tomato. Interestingly, the phylogenetic analysis showed that RGC2 and RGC5 were grouped within the same phylogenetic clade, along with the *Fusarium* resistance genes *I2* and *Fom-2*. These findings suggest that the banana RGC2 and RGC5 are potential resistance gene candidates that could be associated with *Fusarium* resistance. The case of RGC2 is more remarkable because its expression was correlated to FOC race 4 resistance (Chapter 4). As a first step to test whether RGC2 has a role in FOC race 4 resistance, different expression constructs were made with the ORF of this sequence. One of the constructs contains a RGC2 putative promoter region that was successfully cloned in this work. These constructs will be used to transform susceptible banana plants that can then be challenged with FOC race 4 to assess whether resistance has been acquired by genetic complementation.

The results of this thesis provide interesting insights about the structure, expression and phylogeny of two potential *Fusarium* resistance genes in banana, and provide a rational starting point for their functional characterisation. The information generated in this thesis may lead to the identification of a *Fusarium* resistance gene in banana in further studies and may also assist the cloning of *Fusarium* resistance genes in other plant species.

Key words: banana, *Musa acuminata* ssp. *malaccensis*, *Fusarium oxysporum* f. sp. *cubense* race 4, Panama disease, disease resistance gene candidates, nucleotide binding site.

LIST OF PUBLICATIONS/PATENTS

Dale, J.L. and **Peraza-Echeverria, S**. Banana resistance genes and uses thereof. International patent application. PCT/AU2004/001300, WO2005/028 651.

Table of Contents

CHAPTER 5: Structural and Phylogenetic Analysis of Two Potential *Fusarium* Resistance Genes from Banana (*Musa acuminata ssp. malaccensis*)

List of Figures

List of Tables

List of Abbreviations

ATP	adenosine triphosphate
Avr	avirulence
BBTV	*Banana bunchy top virus*
bp	base pair (s)
CC	coiled coil
cDNA	complementary DNA
DIG	digoxygenin
DNA	deoxyribonucleic acid
dNTPs	deoxynucleoside triphosphates
EDTA	ethylenediamine tetra acetic acid
ET	ethylene
FOC	*Fusarium oxysporum* formae specialis *cubense*
g	gram (s)
GFP	green fluorescent protein
GTP	guanine triphosphate
GUS	β-glucuronidase
HR	hypersensitive response
hr	hour (s)
JA	Jasmonic acid
kb	kilobase (s)
kDa	kilodalton (s)
LRR	Leucine rich repeat
M	molar
mM	millimolar
mg	milligram (s)
min	minute (s)
ml	millilitre (s)
μg	microgram (s)
μl	microlitre (s)

μM	micromolar
NBS	nucleotide binding site
ng	nanogram (s)
NO	nitric oxide
NPTII	neomycin phosphotransferase
nt	nucleotide (s)
ORF	open reading frame
PCR	polymerase chain reaction
PR	pathogenesis related
R	resistance
RACE	rapid amplification of cDNA ends
RGC	resistance gene candidates
RNA	ribonucleic acid
ROI	reactive oxygen intermediates
rpm	revolutions per minute
RTF	restricted taxonomic functionality
s	second (s)
SA	salicylic acid
SAR	systemic acquired resistance
SDS	sodium dodecyl sulphate
TE	Tris-EDTA
TIR	Toll/Interleukine-1 receptor homology region
u	units
Ubi-1	maize polyubiquitin promoter
UTR	untranslated region

Declaration

The work contained in this thesis has not been previously submitted for a degree or diploma at any other higher education institute. To the best of my knowledge and belief, this thesis contains no material previously published or written by another person except where due reference is made.

Signed: .....

Date: October 2, 2006

Acknowledgements

I would like to thank my principal supervisor Professor James Dale for all his advice, support, guidance and encouragement. Thanks for sharing with me your vision of modern biotechnology and all the scientific discussion seminars we had. I would also like to thank my associate supervisor Associate Professor Chris Collet for all his guidance and valuable suggestions to improve my thesis. Special thanks to Associate professor Robert Harding for all his guidance and support.

Thanks to the staff and students of the Plant Biotechnology Program for not only technical assistance, but also providing a number of valuable friendships. Special thanks to Matthew Webb, Luke Devitte, Ben Dugdale, Clair Bolton, Maiko Kato, Maino and Cuong. Special thanks also to Jennifer Kleidon for taking care of the plants.

Thanks to the Queensland University of Technology, an excellent place to do scientific research. I would also like to thank all the lab support and administration staff. Special thanks to Diana O'Rourke and Jenny Mayes.

Thanks to the Australian culture for showing me a different way to live, think and face new challenges. It is indeed a great nation and a great place to do science.

I would like to thank the "Consejo Nacional de Ciencia y Tecnología" (CONACyT, México) for the PhD Scholarship granted, without it this scientific journey would have never been possible.

Thanks to my soul mate, Vicky, when bad moments came you were always an oasis of comfort, optimism and wisdom that inspired me to keep going. When good times came, you made my life even better. Thanks for sharing with me this wonderful adventure. Finally, I would like to thank my mother, sister, brother and nephews for their support and encouragement.

Dedication

This PhD thesis is dedicated to the memory of my grandparents:

Sebastián and Estílita

Chapter 1

Literature Review

1.1 Importance of banana

Banana is the most important tropical fruit crop in the world. It is a giant monocotyledonous herb that belongs to the Musaceae family and is cultivated in approximately 120 countries (Jones 2000). World banana production is around 100 million tonnes annually (FAO 2005), of which bananas cultivated for export trade account for only 10%. Hence, fruit harvested from bananas are important components of food security in the tropical world, and provide income to the farming community through local and international trade (Crouch et al. 1998). Banana is an attractive perennial crop for farmers in developing countries. The fruit can be produced all year round, thus providing a steady cash income or supply of nutritious food (Jones 2000). It has been estimated that the highest consumption rates are on the island of New Guinea and in the Great Lakes region of East Africa, where banana *per capita* consumption is approximately 200-250 kg per year. This contrasts to North America and Europe where the banana *per capita* consumption is approximately 15-16 kg per year (Jones 2000).

Although bananas are best known as a food crop, almost every part of the plant can be used in one way or another. Indeed in India it is popularly known as “kalpatharu”, meaning “herb with all imaginable uses” (Sharrock 1996). In Central and East Africa, the juice from the ripe banana known as “beer bananas” is drunk fresh or fermented to make a beer with a low alcohol content. Ripe bananas can also be used to feed cattle and pigs and the unripe fruit can be dried and made into a meal which can be used to substitute up to 70-80% of the grain in pig and dairy diets with little change in performance. Bananas are the source of a fibre used extensively in the manufacture of certain papers, particularly where great strength is required. The paper is used for making tea bags and bank notes amongst other things. The fibre has numerous other uses, including textile manufacture, for making

ropes, string and thread, and for the production of various handicrafts. The large leaves of banana make ideal green umbrellas and they are frequently used as disposable “biological” plates. They can also be used for thatching, for wrapping food during cooking, as bowl covers and tablecloths, as temporary mats and for the covering of earth ovens to hold in the heat (Sharrock 1996). All cultivated bananas are propagated vegetatively either from suckers, division of the corm or increasingly by micropropagation *in vitro* (Dale 1999).

1.1.1 Classification, origin and distribution of bananas

The family Musaceae belongs to the order Zingiberales and contains three genera, *Musa*, *Ensete* and *Musella* (Constantine and Rossel 1999). The Musaceae are distributed from West Africa to the Pacific, but are predominantly of Southeast Asian origin (Stover and Simmonds 1997). Characteristics of the Musaceae that differentiate this family from others in the same order are that the leaves and bracts are spirally arranged, male and female or hermaphrodite flowers are separated within one inflorescence and the fruit is a many-seeded berry (Stover and Simmonds 1997). In *Musa*, the bracts and flowers are inserted independently on the inflorescence axis, the bracts are usually female. In *Ensete*, the bracts and flowers, which are integral with each other and with the axis are persistent and the basal flowers are often hermaphrodite (Stover and Simmonds 1997).

The genus *Musa* is divided into four sections according to Horry et al. (1997) (Table 1.1), two of which contain species with a chromosome number of 10 (2n=20) (*Callimusa* and *Australimusa*) while the species in the other two sections (*Eumusa* and *Rhodochlamys*) have a basic chromosome number of 11 (2n=22). The species in the sections *Callimusa* and *Rhodochlamys* are of ornamental interest only, as the characteristic of parthenocarpy is absent and they do not produce edible fruit. The section *Australimusa* contains *Musa textilis* (Abaca) from which Manila hemp is produced and it is within this section that the edible Fe’i bananas found mainly in the Pacific islands have evolved (Horry et al. 1997).

Virtually all banana and plantain cultivars arose from the *Eumusa* group of species. This section is the biggest in the genus and the most geographically widespread, with species being found throughout Southeast Asia from India to the Pacific (Sharrock 1997). The section contains 11 species (Table 1.1). Most cultivars are derived from two species: *Musa acuminata* (A genome) with a genome size ranging from 591 to 615 megabase pairs (Mbp) and *Musa balbisiana* (B genome) with average genome size of 537 Mbp (Lysak et al. 1999; Jones 2000). Edibility of mature fruits of diploid *Musa acuminata* (AA) arose as a consequence of two mutational events, female sterility and parthenocarpy (Sharrock 1997; Jones 2000). Triploid AAA cultivars arose from these diploids, perhaps as a result of crosses between edible diploids and wild *M. acuminata* subspecies, resulting in a wide range of AAA phenotypes (Table 1.2). The diploid and triploid *acuminata* cultivars were taken by man to areas where *M. balbisiana* is native and natural hybridisations resulted in the formation of hybrid progeny with the genomes AB, AAB, ABB, AAAB, etc (Table 1.2). It is thought that the subsequent dispersal of edible bananas outside Asia was brought about solely by man (Stover and Simmonds 1987). Secondary diversification within the major groups of cultivated bananas has been the result of somatic mutations rather than sexual reproduction (Stover and Simmonds 1987).

The history of many popular banana cultivars is relatively simple according to De Langhe (1996). From about the 5th to the 15th century, and perhaps earlier, the Indian Ocean was navigated by traders from Arabia, Persia, India and Indonesia. Banana varieties from Southeast Asia, including Indonesia and India were, by this means, distributed over the coastal regions of the Indian Ocean (De Langhe 1995). From the 16th to the 19th century the Portuguese and the Spaniards carried bananas all over tropical America. Dutch, British, French and German traders also played a role in the distribution of the popular banana cultivars ‘Gros Michel’ and the Cavendish group to West Africa, Latin America and the Caribbean (De Langhe 1995).

Table 1.1 Systematics of the family Musaceae (Horry et al. 1997)

This table is not available online.
Please consult the hardcopy thesis
available from the QUT Library

Table 1.2 Classification of cultivated varieties in the *Eumusa* section (Daniells et al. 2001).

This table is not available online.
Please consult the hardcopy thesis
available from the QUT Library

1.1.2 Pathogens affecting banana production

Commercial banana growing for export has always been a monoculture based on genetically similar members of the Cavendish subgroup (AAA) (Jones 2000). This lack of genetic diversity makes banana vulnerable to a range of fungal, viral, bacterial and nematode pathogens that cause significant economical losses every year to the export industry (Jones 2000). At the global level, the most serious constraint to the banana production is considered to be black Sigatoka leaf spot disease caused by the air-borne fungus *Mycosphaerella fijiensis*. Leaf necrosis caused by this fungus results in yield losses estimated at 30-76% (Agrios 1997; Carlier et al. 2000; Marín et al. 2003). *M. fijiensis* is spread mainly by wind and water. Petroleum oils and fungicides of the benzimidazole, morpholine, and triazole groups are at the basis of most successful control programs (Swennen and Rosales 1994). This method of control, however, is becoming unsatisfactory as up to 40 applications per year are required to control fungal outbreaks leading to considerable risk of environmental damage and there are now reports of resistant strains of the fungus arising (Dale 1999). In addition, smallholder and subsistence farmers are unlikely to have access to fungicides. Another leaf spot disease of banana is yellow Sigatoka, which is caused by the closely related species *Mycosphaerella musicola*. Although this disease has been largely replaced by black Sigatoka in many banana-producing areas, it remains a significant problem at higher altitudes and cooler temperatures (Carlier et al. 2000). There are sources of resistance to both these diseases in *Musa* germplasm and most breeding programmes include resistance to these diseases as essential characteristic for any new cultivar (Dale 1999). Considerable losses are also caused by Panama disease (also known as Fusarium wilt) caused by the soil-borne fungus *Fusarium oxysporum* f. sp. *cubense* (FOC) which affects many important cultivars of banana (see section 1.4). Current control of FOC is only through resistant cultivars and sources of resistance have been identified in *Musa* germplasm (Dale 1999).

Banana plants also suffer from nematodes, which damage the root system, impeding nutrient and water uptake and consequently impact on yields

(Sarah 2000). The most damaging nematodes are *Radopholus similis*, *Pratylenchus spp.*, *Helicotylenchus multicinctus* and *Meloidogyne incognita*. Control has been through the use of nematicides. Again, sources of resistance in *Musa* germplasm have been identified (Dale 1999).

The most important bacterial disease of banana is Moko disease caused by *Ralstonia solanacearum*, which is transmitted from plant to plant by man or tools used for pruning, root to root, and by insects when visiting the flowers (Thwaites et al. 2000). External symptoms can be confused with those of Panama disease on mature plants. However, two key features help distinguish Panama disease from Moko. External symptoms of Panama disease do not usually develop on plants and suckers that are less than about four months old, whereas plants that are affected by Moko disease will wilt and become chlorotic at a very early stage of development. In addition, internal portions of fruit are discoloured brown by Moko disease, but not by Panama disease. (Ploetz and Pegg 2000). Moko can be controlled by early detection and elimination of infected plants, through a rigorous disinfection of tools, and removal of male buds to prevent dissemination (Swennen and Rosales 1994). Bananas are also affected by a number of viral diseases of which banana bunchy top, caused by the nanovirus *Banana bunchy top virus* (BBTV) is the most important (Dale 1987). BBTV is transmitted by the black banana aphid, *Pentaolonia nigronervosa*, and through vegetative planting material. Infected plants do not produce fruit and the disease can rapidly destroy a plantation unless controlled. The only control at present is through the use of virus-free planting material and sources of resistance have not been identified (Dale 1999).

1.1.3 Developing disease resistance in banana through Biotechnology

Breeding for disease resistance involves the identification of resistance genes in traditional cultivars or related wild species and the incorporation of this resistance into commercially acceptable varieties. The latter can be achieved through conventional breeding using hybridisation techniques or through molecular biotechnology using genetic transformation (Hammond-Kosack and Jones 2000). In the case of banana, traditional breeding

approaches are particularly difficult as almost all of the cultivated varieties are sterile and do not produce seeds (Stover and Simmonds 1987; Jones 2000). Resistant high yielding bananas have been bred and supplied to smallholders in the 1990s after nearly 70 years of traditional breeding (Swennen et al. 2002). This extremely slow progress is due to high sterility, poor seed germination rate, the need for interploidy crosses and the long generation cycle. Consequently, a breeding program can supply only a few promising hybrids per year for further evaluation. Only 0.1% of the selected hybrids are high yielding and resistant to some diseases but they have lost other desired characteristics such as pulp texture and shelf life (Swennen et al. 2002). Hence, genetic transformation offers an opportunity for plant breeders to overcome the constraints imposed by the sterility of the most popular cultivars by transferring specific resistant traits without compromising other important agronomic traits such as pulp texture or flavour. As a result of this feature molecular biotechnology offers great hope for improving commercial banana cultivars in the near future. Protocols have been developed that enable banana plants to be regenerated from cell suspensions (Dhed'a et al. 1991) and somatic embryos (Escalant et al. 1994) that makes genetic transformation of the whole plant possible. Reports of the genetic transformation of banana using biolistics (Sagi et al. 1995; Becker et al. 2000) or *Agrobacterium* (May et al. 1995, Khanna et al. 2004) show that it is possible to express foreign proteins in banana such as β–glucuronidase (GUS), green fluorescent protein (GFP), neomycin phosphotransferase II (NPTII), antifungal proteins and others (Sagi et al. 1995; May et al. 1995; Becker et al. 2000; Swennen et al. 2002). Furthermore, different constitutive or tissue specific promoters such as the cauliflower mosaic virus 35S (CaMV35S), maize ubiquitin, banana actin-1, BBTV 1-6 promoters and others have been tested in banana with great success (Sagi et al. 1995; May et al. 1995; Becker et al. 2000; Hermann et al. 2001; Swennen et al. 2002). These results open the possibility that many desirable agronomic traits such as pathogen resistance could be incorporated into the banana genome. Although the transformation technology in banana is ready, to date, no

resistance genes to the most destructive banana diseases have been cloned and transferred to susceptible banana cultivars.

1.2 *Fusarium oxysporum* f. sp. *cubense*: causal agent of Panama disease

Panama disease or Fusarium wilt was the first serious disease to affect bananas produced for the export trade and it ranks as one of the most destructive plant diseases of all time (O'Donnell et al. 1998). By 1960, Panama disease had destroyed an estimated of 40,000 ha of the cultivar 'Gros Michel' (AAA), causing the export industry to convert to cultivars of the Cavendish subgroup (AAA) (Stover 1962). Panama disease is caused by the soil-borne hyphomycete, *Fusarium oxysporum* formae specialis (f. sp.) *cubense* (FOC). It is one of more than 120 formae speciales (special forms) of *F. oxysporum* that cause vascular wilts in flowering plants (Di Pietro et al. 2003). Each formae specialis consists of strains with ability to cause wilt on a unique host or set of plant host species.

1.2.1 Morphology

Fusarium oxysporum species contains pathogenic and saprophytic strains that cannot be distinguished morphologically (Ploetz and Pegg 2000). In culture, colonies are fast-growing on potato dextrose agar at 24^{o}C, with sparse to abundant aerial mycelium, and white, pink, salmon or purple pigmentation. When formed, sporodochia are tan to orange and sclerotia are blue. Some strains of *F. oxysporum* produce strong odours in culture and these have used to classify isolates of FOC (Moore et al. 1991). Microscopic features of the species include the production of micro- and macroconidia on branched and unbranched monophialides. Microconidia are one- or two-celled and oval to kidney-shaped and are borne in false heads. Macroconidia are four- to eight-celled, sickle-shaped, thin-walled and delicate, with foot-shaped basal and attenuated apical cells. Dimensions of the micro- and macroconidia typically are in the range of 5-16 μm x 2.4-3.5 μm and 27-55 μm x 3.3-5.5 μm, respectively. Terminal and intercalary chlamydospores are usually globose and are formed singly (7-11 μm) or in pair in hyphae or conidia. Although chlamydospore production is a diagnostic character for the

species, they are not produced by isolates of FOC in VCG 01214 (Ploetz and Pegg 2000).

1.2.2 Process of vascular infection and symptoms

As a soil inhabitant, *F. oxysporum* can remain dormant for extended periods in the absence of the host, mainly in the form of thick-walled chlamydospores. Once an area becomes infected with *F. oxysporum*, it usually remains so for many years (Agrios 1997). The proximity of roots induces the dormant propagules to germinate and initiate infection. The ultrastructure of the infection process has been well documented by a series of light and electron microscopy studies (Rodriguez-Gálvez and Mendgen 1995) and most recently using green fluorescent protein (GFP) as a marker system (Di Pietro et al. 2001; Lagopodi et al. 2002). After germination, infection hyphae adhere to the host roots and penetrate the epidermis directly. The mycelium then advances intercellularly through the root cortex until it reaches the xylem vessels entering through the pits. At this point, the fungus switches to a highly peculiar mode of infection, during which it remains exclusively within the xylem vessels, using them as avenues to rapidly colonize the host. This is mainly accomplished by the production of microconidia, which are detached and carried upward in the sap stream. The microconidia eventually germinate and the mycelium penetrates the upper wall of the vessels, producing more microconidia in the next vessel. The characteristic wilt symptoms appear as a result of severe water stress, mainly due to clogging of the vessels. Wilting is most likely caused by a combination of pathogen activities, such as the accumulation of fungal mycelium and/or toxin production and host defence responses, including production of gels, gums and tyloses and vessel crushing by proliferation of adjacent parenchyma cells (Beckman 1987). As long as the plant is alive, the vascular wilt fungus remains strictly limited to the xylem tissues and a few surrounding cells. Only when the infected plant is killed by the disease does the fungus invade the parenchymatous tissue and sporulate profusely on the plant surface (Ploetz and Pegg 2000). Thus, successful infection by *F. oxysporum* is a complex phenomenon that requires a series of highly regulated processes: (1) recognition of roots through unknown host signals,

(2) root surface attachment and differentiation of penetration hyphae, (3) penetration of the root cortex and degradation of physical host barriers such as the endodermis in order to reach the vascular tissue, (4) adaptation to the hostile plant environment, including tolerance to plant antifungal compounds, (5) hyphal proliferation and production of microconidia within the xylem vessels, and (6) secretion of virulence determinants such as small peptides or phytotoxins (Di Pietro et al. 2003).

The characteristic internal symptom of Fusarium wilt is reddish to dark brown discoloration of the host's vascular system. The first internal symptoms occur in the roots, which are the initial sites of infection. These symptoms progress to the rhizome and are most pronounced where the stele joins the cortex (Stover 1962). Eventually, the pseudostem is colonized. In the latter organ, symptoms are often evident as faint brown streaks and/or flecks when outer portions of older leaf sheaths are examined (Ploetz 2000). The first external symptoms of Fusarium wilt in banana are a yellowing of the oldest leaves or a longitudinal splitting of the lower portion of the outer leaf sheaths on the pseudostem. This is followed by a wilt and collapse of leaves at the petiole base. In some cases, these leaves remain green. As the disease progresses, younger and younger leaves collapse, until the entire canopy consists of dead or dying leaves. At this stage, a pronounced, red-brown discoloration of the vascular tissue is usually evident if the pseudostem is cut. After the plant dies, they usually remain standing for 1-2 months before they decay and topple (Ploetz and Pegg 2000).

1.2.3 Host range and distribution

FOC affects the following species in the Order Zingiberales: in the family Musaceae, *Musa acuminata* colla, *Musa balbisiana* Colla, *Musa schizocarpa* and *Musa textilis*; and in the family Heliconeaceae, *Heliconia caribaea*, *Heliconia chartaceae*, *Heliconia crassa*, *Heliconia collinsiana*, *Heliconia latispatha*, *Heliconia mariae*, *Heliconia rostrate* and *Heliconia vellerigera*. Additional hosts include hybrids between *M. acuminata* and *M. balbisiana*, and between *M. auminata* and *M. schizocarpa* (Ploetz and Pegg 2000). FOC has been classified into four physiological races based on pathogenicity to

host cultivars in the field. Race 1 was responsible for the epidemics on ‘Gros Michel’ (1890-1960) and also affects ‘Maqueno’ (AAB, Maia Maoli-Popoulu subgroup), ‘Silk’ (AAB), ‘Pome’ (AAB), ‘Pisang Awak’ (ABB) and ‘I.C.2’ (AAAA); race 2 affects ABB cooking bananas, such as ‘Bluggoe’ (ABB) and some AAAA tetraploids; race 3 was reported to affect *Heliconia* spp. and was weakly pathogenic on ‘Gros Michel’ and seedlings of *Musa balbisiana*. Finally, race 4 affects race 1 and race 2 susceptible clones in addition to the Cavendish cultivars (Ploetz and Pegg 2000; Ploetz 2005). Subtropical race 4 (STR4) has been reported to cause serious crop losses in the subtropical regions of Australia (Queensland and New South Wales), South Africa (Natal and Transvaal) and the Canary Islands. Cold winter temperatures in the subtropics are thought to predispose Cavendish cultivars to subtropical race 4. Another genetically distinct form of FOC race 4 that affects Cavendish but in the tropics was designated as tropical race 4 (TR4) (Ploetz 2005). TR4 is found in Taiwan, Australia (Northern Territory), islands in Indonesia (Sumatra, Halmahera,Irian Jaya, Java and Sulawesi), where no predisposing factors have been identified (Ploetz 2005). If race 4 was to become established in the Americas, the world export industries could be severely affected, as there is no widely accepted replacement for the Cavendish cultivars.

1.2.4 Origins and genetic diversity

It is believe that FOC co-evolved with banana in Southeast Asia and was subsequently dispersed to other regions by humans (Stover 1962). The presence in this region of numerous *Musa acuminata subspecies* such as *malaccensis*, *burmannica*, *microcarpa* and *siamea* which are highly resistance to Panama disease indicate that these subspecies had been exposed to selection pressures exerted by FOC (Vakili 1965).

A teleomorph (sexual stage) for *F. oxysporum* has not been found, the pathogen appears to rely solely on asexual reproduction. Genetic variation that occurs in the species is assumed to arise by neutral mutations that are maintained by mitosis (Kistler and Miao 1992). Vegetative or somatic compatibility (VCGs) has been used extensively to characterize worldwide

populations of FOC (Ploetz and Corell 1988; Brake et al. 1990; Moore et al. 1993). Usually, genetic complementation (heterokaryon formation) between nitrate-non-utilizing auxotrophic (*nit*) mutants is used to identify compatible isolates (Ploetz and Pegg 2000). To date, 16 VCGs or VCG complexes have been reported in FOC (Ploetz and Pegg 2000). Boehm et al. (1994) determined the chromosome number and genome size for 118 isolates in 12 VCGs complexes. Two major groups of isolates were distinguished by the karyotype data, those with large genome (40-50 Mbp) and high chromosome numbers (11-14) and those with smaller genomes (32-45 Mbp) and fewer chromosomes (9-12). Interestingly, the VCGs 0124-0125-0128-01220 were found in the former and 0120-01215 in the latter major group. Subsequent phylogenetic studies have recognized the same major groups that were identified by Boehm et al. (1994). Group 1 contained isolates in VCGs 0120-01215, 0121, 0122, 0126, 0129, 01210, 01211, 01213-01216 (TR4) and 01219, whereas group 2 contained isolates in VCGs 0123, 0124-0125-0128-012120, 01212, 01214, 01217 and 01218 (Koening et al. 1997; Bently et al. 1998; O'Donnell et al. 1998). These matched, respectively, the *odoratum* and *inodoratum* groups of Moore et al. (1991), which were determined by whether or not isolates formed volatile aldehydes in culture. These studies have relevant implications for understanding this important plant pathogen, and for the deployment of resistant genotypes (Ploetz 2005).

1.2.5 Control of Panama disease

In general, effective chemical control measures for panama disease do not exist (Ploetz and Pegg 2000). The only methods of controlling panama disease are by quarantine or by planting resistant varieties (Stover and Simmonds 1987). Panama disease susceptible clones can be grown if pathogen-free propagation material is used in uninfected soils. Micropropagated plantlets are the most reliable source of clean material. Since they are also free of bacterial, nematode and other fungal pathogens, plantlets are used whenever possible. It should be noted, however, that plants grown from tissue-cultured plantlets have been shown to be more susceptible to Panama disease than those grown from conventional planting material (Smith et al. 1998). The expense of plantlets, however, may make their use in subsistence agriculture impractical. Clearly, the use of resistant genotypes is the best way to combat this disease in infested soils (Ploetz and Pegg 2000).

1.2.6 Identification of genes that confer resistance to Fusarium wilt.

Two genes that confer resistance to Fusarium wilt have been identified in tomato (*Lycopersicon esculentum*) and melon (*Cucumis melo*) (Simons et al. 1998; Joobeur et al. 2004). In tomato, three races of *F. oxysporum* f.sp. *lycopersici* (FOL) are known, races 1, 2 and 3 (Sela-Buurlage et al. 2001). A map-based positional cloning strategy was used to identify the gene *I2* that confers resistance to FOL race 2. The *I2* gene does not confer resistance to the other two FOL races, race 1 and 3, whose resistance is located at different loci of the tomato genome (Sela-Buurlage et al. 2001). The *I2* gene is a dominant gene that belongs to the NBS-LRR class of disease resistance genes (see section 1.3). In the case of melon, three races (0, 1 and 2) of *F. oxysporum* f. sp. *melonis* are known. Resistance to race 1 and race 2 is conferred by a single dominant gene *Fom-2* and *Fom-1*, respectively. Both genes also confer resistance to race 0 (Zink and Thomas 1990; Schreuder et al. 2000). Recently, the *Fom-2* was isolated by a map-based positional cloning strategy (Joobeur et al. 2004). This gene also belongs to the NBS-LRR class of disease resistance genes (see section 1.3). Introduction of the *I2* gene into a FOL race 2 susceptible tomato genotype using genetic

transformation has demonstrated the success of this technology to develop *Fusarium* resistance in a susceptible genotype (Simons et al. 1998). However, the race specificity that these R genes have shown in their corresponding plant families imposes a serious limitation for their use in other plant families. The fact that both *Fusarium* resistance genes belong to the same class of disease resistance genes suggest that similar genes may be responsible for resistance in other plant families, including the Musaceae. Therefore, cloning these types of genes in the Musaceae could ultimately assist the development of resistance in banana to Panama disease.

1.3 Plant disease resistance genes

In the 1940s, using flax (*Linum usitatissimum*) and its fungal rust pathogen *Melampsora lini*, H.H. Flor studied the inheritance of plant resistance and pathogen virulence (Flor 1971). Based on this work, he developed the classic "gene-for-gene" model that proposes that for resistance to occur, complementary pairs of dominant genes, one in the host and the other in the pathogen, are required. A loss or alteration to either the plant resistance (R) gene or the pathogen avirulence (Avr) gene leads to disease. This simple model holds true for most biotrophic pathogens, including viruses, bacteria, fungi and nematodes (Hammond-Kosack and Jones 1997). This gene-for-gene type resistance is generally interpreted as an elicitor-receptor model: the plant R protein recognizes directly or indirectly the pathogen-derived Avr product. Once this recognition has occurred, defense responses are triggered (Staskawicz et al. 1995). Most R gene-triggered resistance is associated with a rapid defense response, termed the hypersensitive response (HR). The HR results in a localized cell and tissue death at the site of infection, which constrains further spread of the infection (Hammond-Kosack and Jones 1997). This local response often triggers non-specific resistance throughout the plant, a phenomenon known as systemic acquired resistance (SAR). Once triggered, SAR provides resistance to a wide range of pathogens for days. The plant-pathogen interactions are divided into "compatible interactions" in which a susceptible plant becomes diseased upon attack by a virulent pathogen and "incompatible interactions" in which a resistant plant does not develop disease upon attack by an avirulent

pathogen (Hammond-Kosack and Jones 2000). These interactions are illustrated in figure 1.1.

Until 1992, no plant R gene had been cloned and characterized at the molecular level. Since then, R genes from several plant species have been cloned using either transposon-based gene tagging or a map-based positional cloning approach. The majority of R gene products have domains involved in protein recognition such as the leucine-rich repeat (LRR) or coiled-coil (CC) domains, and signal transduction such as the nucleotide-binding site (NBS) or protein kinase (PK) domains. Some R genes are predicted to be cytoplasmic proteins while others span the cell membrane via a transmembrane domain (TM). These domains are combined in several arrangements to give rise to different classes of R proteins. At least six different classes of R-genes have been identified according to structural characteristics: 1) NBS-LRR, 2) LRR-TM-PK, 3) LRR-TM, 4) CC-TM, 5) PK and 6) PK-PK (Dangl and Jones 2001; Brueggeman et al. 2002) (Figure 1.2).

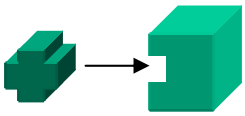
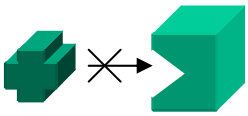
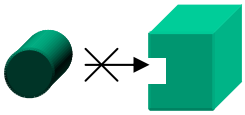
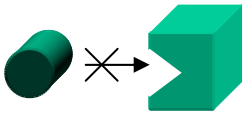
Pathogen genotype	**Host plant genotype**	
	R1	***r1***
Avr1	 **Avr1** **R1 protein** No disease Plant and pathogen are incompatible	 **Avr1** **r1 protein** Disease Plant and pathogen are compatible
avr1	 **avr1** **R1 protein** Disease Plant and pathogen are compatible	 **avr1** **r1 protein** Disease Plant and pathogen are compatible

Figure 1.1 Flor's gene-for-gene model. For resistance (incompatibility) to occur, complementary pairs of dominant genes, one in the host and one in the pathogen, are required. An alteration or loss of the plant resistance gene (R changing to r) or of the pathogen avirulence gene (Avr changing to avr) leads to disease (compatibility)(Hammond-Kosack and Jones 2000). *R1* and *AVR1*, dominant genes from the plant and pathogen respectively. *r1* and *avr1*, recessive genes from the plant and pathogen respectively.

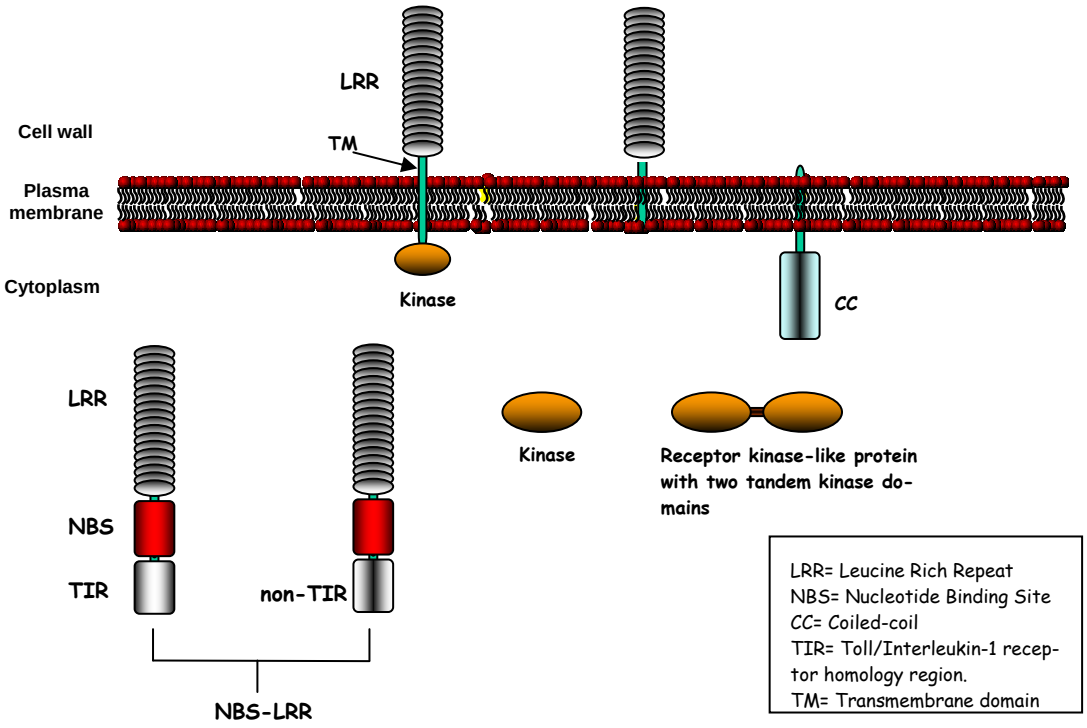


Figure 1.2 Schematic representation of the predicted structure of disease resistance proteins in plants.

1.3.1 Structure and function of NBS-LRR disease resistance genes

The NBS-LRR class of disease resistance genes is by far the largest group of characterized R genes with more than 30 cloned genes to date from different plant species (Table 1.3). The NBS-LRR proteins are predicted to be located in the cytoplasm and they confer resistance to a diverse array of pathogens such as virus, bacteria, fungi, nematodes and pests (Hulbert et al. 2001) (Table 1.3). Three different domains form the basic structure of this class: the TIR or non-TIR domains are found at the N-terminus along with the NBS, and the LRR domain which is located at the C-terminus of the protein (Hammond-Kosack and Jones 2000). Most of the NBS-LRR genes identified to date have been isolated by a map-based positional cloning approach.

Table 1.3 Plant disease NBS-LRR resistance genes cloned from 1994 to 2006.

Host	Pathogen	PLANT (R) PROTEIN STRUCTURE	R PROTEIN NAME	REFERENCE
Flax	*Melampsora lini*	TIR-NBS-LRR	L	Lawrence et al. 1995
Tobacco	*Tobacco mosaic virus*	TIR-NBS-LRR	N	Whitman et al. 1996
Flax	*Melampsora lini*	TIR-NBS-LRR	M	Anderson et al. 1997
Arabidopsis	*Peronospora parasitica*	TIR-NBS-LRR	RPP 5	Parker et al. 1997
Arabidopsis	*Pseudomonas* syringae	TIR-NBS-LRR	RPS4	Gassmann et al. 1999
Flax	*Melampsora lini*	TIR-NBS-LRR	P	Dodds et al. 2000
Arabidopsis	*Ralstonia solanacearum*	TIR-NBS-LRR-WRKY	RRS1-R	Deslandes et al. 2002
Arabidopsis	*Leptosphaeria maculans*	TIR-NBS-LRR	RLM	Staal et al. 2006
Arabidopsis	*Pseudomonas syringae*	non-TIR-NBS-LRR	RPS2	Mindrinos et al. 1994
Tomato	*Pseudomonas syringae*	non-TIR-NBS-LRR	Prf	Salmeron et al. 1996
Arabidopsis	*Pseudomonas syringae*	non-TIR-NBS-LRR	RPM1	Grant et al. 1996
Tomato	*Fusarium oxysporum*	non-TIR-NBS-LRR	I2	Simons et al. 1998
Tomato	*Meloidogyne sp*	non-TIR-NBS-LRR	Mi	Milligan et al. 1998
Tomato	*Macrosiphum euphorbie*	non-TIR-NBS-LRR	Mi	Milligan et al. 1998
Arabidopsis	*Peronospora parasitica*	non-TIR-NBS-LRR	RPP1	Botella et al. 1998
Rice	*Xanthomonas oryzae*	non-TIR-NBS-LRR	Xa1	Yoshimura et al. 1998
Arabidopsis	*Pseudomonas syringae*	non-TIR-NBS-LRR	RPS5	Warren et al. 1998
Maize	*Puccinia sorghi*	non-TIR-NBS-LRR	Rp1-D	Collins et al. 1999
Pepper	*Xanthomonas campestris*	non-TIR-NBS-LRR	Bs2	Tai et al. 1999
Rice	*Magnaporthe grisea*	non-TIR-NBS-LRR	Pib	Wang et al. 1999
Potato	*Potato virus X*	non-TIR-NBS-LRR	Rx2	Bendahmane et al. 1999
Rice	*Magnaporthe grisea*	non-TIR-NBS-LRR	Pi-ta	Bryan et al. 2000
Arabidopsis	*Peronospora parasitica*	non-TIR-NBS-LRR	RPP 13	Bittner-Eddy et al. 2000
Barley	*Blumeria graminis*	non-TIR-NBS-LRR	Mla	Zhou et al. 2001
Tomato	*Tospovirus*	non-TIR-NBS-LRR	Sw-5	Brommonschenkel et al. 2002
Lettuce	*Bremia lactucae*	non-TIR-NBS-LRR	Dm3	Shen et al. 2002
Potato	*Globodera pallida*	non-TIR-NBS-LRR	Gpa 2	Van der Vossen et al. 2000
Potato	*Phytophtora infestans*	non-TIR-NBS-LRR	R1	Ballvora et al. 2002

Tomato	*Globodera rostochiensis*	non-TIR-NBS-LRR	Hero	Ernst et al. 2002
Potato	*Phytophthora infestans*	non-TIR-NBS-LRR	RB	Song et al. 2003
Potato	*Phytophthora infestans*	non-TIR-NBS-LRR	Rpi-blb1	Van der Vossen et al. 2003
Melon	*Fusarium oxysporum*	non-TIR-NBS-LRR	Fom-2	Joobeur et al. 2004
Soybean	*Pseudomonas syringae*	non-TIR-NBS-LRR	Rpg1-b	Ashfield et al. 2004
Potato	*Phytophthora infestans*	non-TIR-NBS-LRR	R3a	Huang et al. 2005
Maize	*Xanthomonas oryzae*	non-TIR-NBS-LRR	Rxo1	Zhao et al. 2005

1.3.1.1 TIR and non-TIR domains

The NBS-LRR class of R genes can be divided into two distinct subclasses based on the presence or absence of an N-terminal domain with homology to the *Drosophila* Toll and human Interleukin-1 receptors (TIR) (Meyers et al. 1999; Pan et al. 2000), the TIR and non-TIR subclasses, respectively. The non-TIR subclass commonly has a predicted coiled-coil (CC) structure, sometimes in the form of a leucine zipper (Meyers et al. 1999; Pan et al. 2000; Hulbert et al. 2001). By analogy to the animal TIR proteins, plant TIR proteins are thought to function in signal transduction. However, the plant TIR domain may contribute to determining R gene specificity as swapping the TIR domains of L^6 and L^7 R genes in flax switches their specificity in rust resistance (Luck et al. 2000). This, together with evidence for diversifying selection in the TIR region of the flax rust R genes (Ellis et al. 1999), may also indicate that pattern recognition operates as a complex. Thus, it appears that both the TIR and LRR domains play a role in pathogen recognition (Luck et al. 2000). In the case of the non-TIR domain, a CC predicted structure is usually found. A typical CC structure shows a heptad repeat where the seven positions are labelled *a* through *g*. Residues *a* and *d* tend to be hydrophobic, and the residues at the *e* and *g* positions are charged or polar (Fluhr et al. 2001). A large subset of dicot and cereal NBS-LRR genes contain CC-like structures in their non-TIR domain with over 95% probability (Pan et al. 2000). In this context, they may serve the function of adapter TIR-like motifs.

1.3.1.2 Nucleotide Binding Site (NBS) domain

The NBS domain comprises three motifs predicted to bind ATP or GTP, and several conserved motifs whose functions are not known (Hammond-Kosack and Jones, 1997). This region has homology to two activators of apoptosis in animal cells: APAF-1 and CED. By analogy to these well-characterized regulators of programmed cell death, the corresponding domain in NBS-LRR proteins might operate as an intramolecular signal transducer (Van der Biezen and Jones, 1998; Aravind et al. 1999). Biochemical evidence has revealed that the NBS of the tomato Mi-1 and I2 non-TIR-NBS-LRR resistance proteins can bind and hydrolase ATP *in vitro* (Tameling et al. 2002) which reinforces the idea that the NBS domain functions in signal transduction. The TIR and non-TIR subclasses of NBS-LRR genes can also be distinguished by the motifs found within the NBS domain or by a single amino acid residue in the final portion of the NBS kinase-2 (Kin-2) motif, which in most cases is an aspartic acid for the TIR subclass and a tryptophan for the non-TIR subclass (Meyers et al. 1999). The non-TIR subclass is widely distributed in both monocotyledonous and dicotyledonous species, whereas the TIR subclass appears to be restricted to dicotyledonous species (Meyers et al. 1999; Pan et al. 2000; Cannon et al. 2002). Using PCR with degenerate primers targeting the conserved GVGKTT (P-loop), GSRIIITTRD or GLPLA motifs of the NBS domain of R genes (Figure 1.3) has resulted in the isolation of numerous NBS-containing genes from a variety of plant species such as soybean (Kanazin et al. 1996; Yu et al. 1996; Graham et al. 2000), potato (Leister et al. 1996), lettuce (Shen et al. 1998), rice and barley (Leister et al. 1998), wheat (Seah et al. 2000), common bean (Rivkin et al. 1999; López et al. 2003) and other plant species (Meyers et al. 1999). Significantly, the genetic position of these sequences is frequently at or near R-gene loci, indicating that these NBS sequences may form part of R-genes (Meyers et al. 1999). In summary, the NBS domain of NBS-LRR genes has been used extensively to identify and to classify these genes. The popular use of this domain stems from a number of reasons: The NBS domain has some conserved amino acid motifs that assist in cloning these genes via PCR amplification and recognizing them in databases; the conserved motifs assist in aligning the sequences for

phylogenetic analyses, and classification of NBS-LRR genes by their NBS region sequences accurately predicts whether they belong to the TIR or non-TIR subclass (Bai et al. 2002).

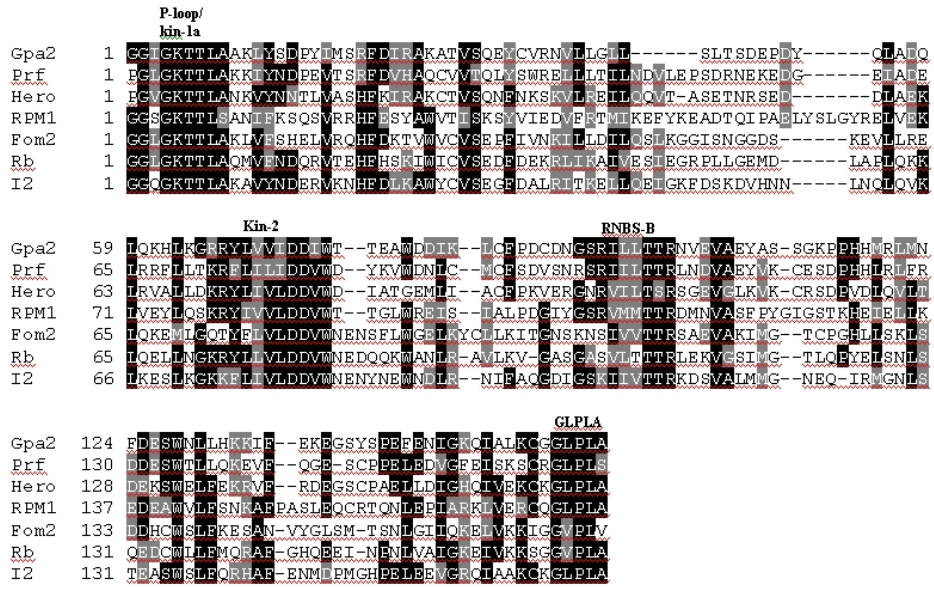


Figure 1.3 Alignment of the NBS domain of multiple NBS-LRR disease resistance proteins. The conserved NBS motifs as determined by Meyers et al. (1999) are indicated. Identical amino acids are shaded in black and conservative substitutions are shaded in grey

1.3.1.3 Leucine Rich Repeat (LRR) domain

The C-terminal comprises a LRR domain with the consensus sequence xxLxLxx (where x is any residue), which is thought to be involved in ligand binding and pathogen recognition (Hammond-Kosack and Jones 1997). Genetic evidence indicates that the β-strand/β-turn of the LRR is a key region in the R protein and appears to determine its pathogen specificity (Hammond-Kosack and Jones 1997; Jones and Jones 1997). Given the crystal structure determined for porcine ribonuclease inhibitor protein, the conserved leucines (L) in the plant R proteins within this consensus are predicted to occupy the hydrophobic protein core, whereas the other residues (x) form a solvent-exposed surface that can participate in binding other proteins (Hammond-Kosack and Jones 1997; Jones and Jones 1997). R gene sequence comparisons reveal that the x residues in this region are

hypervariable. These data suggest the xxLxLxx region creates a surface that has evolved to detect variations in the multitude of pathogen-derived ligands. Parts of the LRR motif in plant R proteins may also participate in relaying downstream signalling through interactions with effector proteins. The large size of the LRR domain in most R proteins could even permit both the recognition and the effector functions to be accommodated by different binding specificities within different LRR subdomains or by interactions with more than one pathogen-derived ligand (Hammond-Kosack and Jones 2000). Direct evidence for interaction of LRR domain with avirulence factors is based on the finding that a single amino acid difference in the LRR domain distinguished susceptible and resistance alleles of the rice *Pi-ta* R-gene that confers resistance to *Magnaporthe grisea* (Bryan et al. 2000). In this case, by using the yeast two-hybrid system, the recombinant LRR domain of the resistance allele could be shown to directly interact with its avirulence factor while the susceptible allele displayed a much weaker interaction (Jia et al. 2000).

1.3.2 Organization of NBS-LRR genes in the plant genome.

In different plants, NBS-LRR loci are found both as isolated genes (singletons) and as tightly linked arrays of related genes (gene clusters) (Holub 2001). In some cases, gene clusters contain copies of NBS-LRR genes from different phylogenetic clades (Hulbert et al. 2001; Leister 2004). The complete *Arabidopsis* sequence of 125 Mb has allowed a comprehensive analysis of the diversity and organization of NBS-LRR R gene sequences in a single plant genome. Annotation has revealed ~150 sequence with homology to the NBS-LRR class of R genes (The *Arabidopsis* genome initiative 2000). R gene homologues are unevenly distributed between chromosomes, with 49 on chromosome I, two on chromosome II, 16 on chromosome III, 28 on chromosome IV, and 55 on chromosome V. Despite the fact that many previously isolated R genes seem to reside in local multigene families, there are 46 singleton *Arabidopsis* R-gene homologues, 25 doublets, seven loci with three copies, and individual loci with four, five, seven, eight and nine NBS-LRR-encoding genes. There are

more TIR-NBS-LRR genes (94) than non-TIR-NB-LRR genes (55) (Meyers et al. 2003). In the case of the *Oryza sativa* L. (var. Nipponbare) genome sequence of 420 Mbp there are ~535 NBS-coding sequences, including 480 non-TIR-NBS-LRR genes (Goff et al. 2002; Zhou et al. 2004). The other NBS-coding sequences are totally different in their structures from the majority, or are simply truncated (Zhou et al. 2004). The 480 non-TIR-NBS-LRR genes identified represent about 1% of all the predicted ORFs in the rice genome (Goff et al. 2002), while the *A. thaliana* genome non-TIR and TIR-NBS-LRR represent 0.43% of the total predicted ORFs. Thus both the absolute number and relative representation of NBS-LRR genes in the rice genome are clearly higher than in the *Arabidopsis* genome. A few genes with a TIR-like domain have been identified in rice, but these did not encode any obvious LRR domain, and were otherwise divergent from NBS-LRR genes (Bai et al. 2002; Zhou et al. 2004). The chromosomal distribution of NBS-coding genes in rice vary from 20 on chromosome 9 to 133 on chromosome 11. 263 NBS-coding genes resided in 44 gene clusters and the average number of genes in a cluster was six. There are 15 clusters with four copies, 13 with five, 4 with six, 4 with seven, 3 with eight, 3 with ten and 1 each with 15 or 17 NBS-coding genes (Zhou et al. 2004). Besides these clusters, there were 40 tightly linked doublets and 17 triplets. Therefore a total of 394 genes resided either in a gene cluster or in tandem array. In all, 125 NBS singletons were dispersed over the entire chromosome (Zhou et al. 2004). The ratio of singletons to the total number of NBS genes in the rice genome (24.1%) was similar to that in *Arabidopsis* (26.8%; Meyers et al. 2003).

1.3.3 Evolution of NBS-LRR genes

1.3.3.1.Diversification

For an increasing number of R genes, including the NBS-LRR genes, evidence of the selection for diversity of residues in the LRR region that are predicted to be solvent exposed, and hence may constitute ligand contact points, has been observed (Parniske et al. 1997; Wang et al. 1998; Meyers et al. 1998; Botella et al. 1998). Protein variation can be assessed by

comparing base-pair changes in nucleotide sequence from numerous variants of the same gene (orthologues or paralogues) that either alter the encoded amino acid (non-synonymous substitutions) or leave the amino acid unaltered (synonymous substitutions). The ratio of non-synonymous (Ka) to synonymous (Ks) amino-acid changes provides a measure of diversifying selection; a Ka/Ks value less than 1 indicates conservation of the sequence, whereas a value greater than 1 indicates positive selection to diversify (Krietman and Akashi 1995). Parniske et al. (1997) were the first to use this comparative method to analyse sequence variation in R genes, in an examination of tandemly repeated genes at the *Cf4/Cf9* locus from different subspecies of tomato. They and others (Meyers et al. 1998; Botella et al. 1998) reached the conclusion that the LRR domain shows much higher levels of diversity, particularly at solvent-exposed faces in the repeats, than other domains within the genes. These changes have occurred in addition to changes in the number of LRR repeats. Modification of the length of the LRR appears to be an important contributor to R-gene diversification. For example, whereas the genes at the *Cf4*/*Cf9* locus of tomato vary principally because of multiple nucleotide substitutions, the related genes at the *Cf2/5* locus have additionally undergone deletion/expansion events involving individual LRR repeat units (Dixon et al. 1998). Furthermore, these events have been restricted to the amino-terminal LRR region of the protein, a region of Cf proteins that determines specificity differences between paralogues (Thomas et al. 1997). In flax *L* alleles, the LRR repeats are more degenerate and the DNA sequences encoding the repeats are probably not sufficiently related for inter-repeat recombination. Nevertheless, examples occur in which blocks of sequence encoding LRR units within flax and *Arabidopsis* NBS-LRR R genes have undergone duplication (Ellis et al. 1999; Noel et al. 1999). These direct repeats are able to undergo unequal exchange events that can give rise to cycles of repeat expansion and reduction. For example, although most *L* alleles in flax contain two direct repeats of 450 base pairs comprising six individual LRR repeat units, functional alleles with either one or four copies of the 450 base pair repeat occur (Ellis et al. 1999). Another example is provided by the *RPP5* locus where more complex arrangements of direct repeats consisting of sets of

four individual LRR units exist (Noel et al. 1999). Exchange events giving rise to paralogues with 8, 13, 21, and 25 LRR units have taken place during the evolution of the locus. The combined effect of point mutations and changes in the number of LRR repeats indicate that variation in the LRR domain might be important for determining the specificity of a given R gene. However, recent evidence from flax indicates that the more highly conserved TIR domain can also determine resistance specificity (Luck et al. 2000).

1.3.3.2 Ancient origins of disease resistance

Phylogenetic analysis of NBS-LRR clearly shows an ancestry dating to the emergence of higher plants (Holub 2001). The coupling of NBS-LRR genes to different defence responses provides an alternative means of containing parasites, and presumably was an important selective factor in divergence of the two main NBS-LRR types (TIR and non-TIR). Interestingly, TIR-NBS-LRR genes have not been identified so far in cereals, even though they represent two-thirds of the NBS-LRR genes in *Arabidopsis*. As examples of this subclass have been found in pine, Meyers et al. (1999) and Pan et al. (2000) speculate that the two main NBS-LRR types are older than the divergence of angiosperms and gymnosperms, which occurred at least 200 million years ago (Figure 1.4). This suggests that TIR-NBS-LRR genes were lost during the evolution of monocotyledons, such as the cereals (Figure 1.4). Stahl et al. (1999) provided the first attempt to estimate the age of a functional R gene using a comparative analysis of DNA sequence variation in regions flanking the *RPM1* locus. They compared variation in sequence among accessions of two *Arabidopsis* species, and concluded that the functional resistance allele and the null deletion allele have coexisted at this locus for ~ 10 million years. This estimate coincides with the predicted divergence of *Brassica* and *Arabidopsis* lineages, in which deletions of *RPM1* seem to have occurred independently (Grant et al. 1998). Vision et al. (2000) has provided estimates for the age of large duplicated regions distributed throughout most of the *Arabidopsis* genome. These regions seem to have remained intact with respect to gene order for the estimated age of each region. It is interesting that all of the known functional NBS-LRR genes are located in regions estimated to be at least 50 million years old. For

instance, three of the single gene loci shown (*RPM1*, *RPS2* and *RPP8*) lie in regions that date to ~100 million years ago, an important period for speciation in the angiosperms.

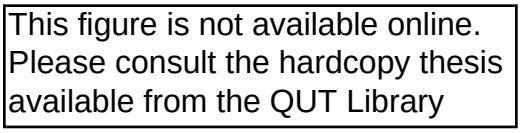


Figure 1.4 Model for the evolution of NBS-LRR disease resistance genes and their homologues. The evolution of NBS-LRR genes involved at least two stages. Stage I was characterized by the presence of a few NBS-LRRs with a broad spectrum of specificity. In stage II, after the monocot/dicot separation, the disease resistance genes evolved by divergent gene duplication followed by gene diversification. During this stage, TIR-NBS-LRR group degenerated in monocot genomes (Pan et al. 2000).

1.3.3.3 Behaviour of NBS-LRR genes in natural plant populations: the 'arms race' and 'trench warfare' models.

In nature, the ongoing battle between plants that develop novel resistance specificities and pathogens that try to circumvent recognition by these plants can be seen as an arms race (Dawkins and Krebs 1979). Such an arms race implies a transient polymorphism of R genes, which means that high disease pressure causes the replacement of old R genes by new ones, resulting in

relatively young R genes and monomorphic R gene loci (Bergelson et al. 2001).

The debate surrounding how polymorphisms are maintained in natural populations revolves around the issue of whether ‘defeated’ R-gene alleles are driven to extinction and are therefore transient in host population, or instead simply become rare until they are recycled, increasing in frequency as the corresponding avirulence re-emerges in the pathogen population. In an attempt to determine whether R genes are transient or recycled in plants Stahl et al. (1999) used a collection of 26 ecotypes of *Arabidopsis* to investigate allelic variation at the single gene locus *RPM1*. The transient polymorphism model, according to which R alleles are replaced in each cycle by new ones, and that Stahl et al. (1999) referred to in a restrictive sense as an ‘arms race’, was rejected on the basis that *RPM1* is an old resistance specificity without alternative functional alleles. The apparent lack of functional alternatives at this locus might actually support a transient model if the contemporary functional allele represents an adaptive optimum. A major constraint in testing a transient model with only one example of a single gene locus is that the crucial evidence, namely extinct alleles, are not available. Conversely, Stahl et al. (1999) provide a strong argument in favour of recycling polymorphism, which they prefer to call ‘trench warfare’, in which advances and retreats of resistance-allele frequency maintain variation for disease resistance as a dynamic polymorphism. This argument is supported by evidence from many other loci of *Arabidopsis* (*RPP1*, *RPP8*, *RPP13*) in which alternative alleles are common, showing that polymorphism has been generated, has accumulated and is apparently maintained at these loci for millions of years. The trench warfare model can explain many recent observations but does not provide an explanation for the generation of new R genes with novel specificities. Therefore, R gene dynamics in a natural plant population probably reflect a combination of trench warfare and an arms race, the latter perhaps being relatively slow (Van der Hoorn et al. 2002).

1.3.4 Expression of NBS-LRR genes.

Resistance genes are typically expressed at low levels, they are usually unaffected by pathogen inoculation and transcripts can be difficult to detect by gel blot analysis (Hulbert et al. 2001). RT-PCR or Northern blot analysis performed on several NBS-LRR genes such as *RPM1*, *Prf, RPP5*, *Mi*, *L6, Rp1-D*, *Pib, RPP8* and *I2* have revealed the presence of low levels of transcripts in unchallenged plants (Grant et al. 1995; Salmeron et al. 1996; Parker et al. 1997; Milligan et al. 1998; Ayliffe et al. 1999; Collins et al. 1999; Wang et al. 1999; Cooley et al. 2000; Mes et al. 2000). On the other hand, transcription of the rice *Xa1* resistance gene appears to increase following pathogen inoculation (Yoshimura et al. 1998). Infection by pathogens has also been demonstrated to affect turnover of Rpm1 protein in *Arabidopsis* and splicing of *N* gene transcripts in tobacco (Dinesh-Kumar and Baker 2000).

1.3.5 NBS-LRR signal transduction

A logical prediction of the gene-for-gene model is that R genes encode receptors that interact physically with products of matching Avr gene, enabling recognition of the pathogen and subsequent elicitation of an array of plant defense responses that eventually lead to resistance (Keen 1990) (Figure 1.5A). The structure and predicted location of R and Avr proteins are usually consistent with this model (Takken et al. 2000). For example, most R proteins carry leucine rich repeats (LRRs), which are thought to form a versatile binding domain that could fulfil the receptor role of the R protein. In addition, membrane-anchored R proteins mediate the perception of Avr factors that are produced in or injected into the host cytoplasm by the pathogen. Although these observations agree with the ligand-receptor model, a direct physical interaction between Avr and R proteins has only been shown for the AvrPto-Pto and AvrPita-Pi-ta pairs (Tang et al. 1996; Jia et al. 2000). In most other cases, in spite of extensive and detailed studies, no evidence for a direct interaction between the two gene products has been found (Van der Hoorn et al. 2002). Lack of evidence for direct Avr-R interactions led to the formulation of new models for Avr perception by resistant plants. One interesting model is that R proteins confer recognition

of Avr factors only when these Avr factors are complexed with their host avirulence targets. This model was initially proposed to explain the role of Prf in AvrPto-Pto signalling (Van der Biezen and Jones 1998) and was later referred to as the guard model (Dangl and Jones 2001) (Figure 1.5B). In this model, Pto is considered to be the avirulence target of AvrPto, which is guarded by the NBS-LRR resistance protein, Prf. In general, three observations support the guard model. First, no direct interaction is found between Avr factors and R proteins, Second, recognition of the Avr factor requires and additional host protein that is specific for each Avr-R pair. Third the structure and predicted function of this host protein suggests that it might be an avirulence target for the pathogen. The available data suggest that resistance based on guarding is prevalent in gene-for-gene interactions (Van der Hoorn et al. 2002).

There are various additional observations that might be explained by the guard model. For example, the dual recognition shown by some NBS-LRR R genes. For example, *RPM1* recognizes two non-homologous avr gene products of *Pseudomonas syringae* (Grant et al. 1995) while the tomato *Mi* gene confers not only nematode resistance but also aphid resistance (Rossi et al. 1998). Furthermore alleles of the *RPP8*/*HRT* gene recognize an oomycete parasite and a virus (Cooley et al. 2000). Similarly, the closely related potato *Rx* and *Gpa2* genes confer virus and nematode resistance, respectively (Van der Vossen et al. 2000). These findings suggest that there are different Avr proteins that recognize the same avirulence target protected by the R protein. Alternatively, it is possible that different Avr-avirulence target interactions are detected by the same R protein (Van del Hoorn et al. 2002). This type of combinatorial interactions may explain how plants are capable of coping with different kind of pathogens with a limited set of R genes.

Pathogen recognition by NBS-LRR proteins causes the rapid activation of appropriate defenses. Activation of the hypersensitive response (HR) triggers a systemic resistance response known as systemic acquired resistance (SAR). This response includes the accumulation of the signal molecule

salicylic acid (SA) throughout the plant and the consequent expression of a characteristic set of defense genes, including pathogenesis-related proteins (PRs). Plants expressing SAR are more resistant to subsequent attack by a variety of otherwise virulent pathogens (Glazebrook 2001). Some defense responses are activated by signal transduction networks that require jasmonic acid (JA) and ethylene (ET) as signal molecules. Different pathogens are limited to different extents by SA-dependent responses and by JA/ET-dependent responses. There appears to be considerable cross-talk between these signal transduction networks, with at least some SA-dependent responses limited by JA/ET-dependent responses and vice versa (Glazebrook 2001). The discovery of genes or mutants allows further dissection of local (HR) and systemic signalling networks and begins to highlight the complex interplay between defense molecules such as SA, JA, ET, nitric oxide (NO) and reactive oxygen intermediates (ROI) (Hammond-Kosack and Parker 2003).

Mutational analyses, almost exclusively conducted using *Arabidopsis,* have led to the identification of genes that are essential for the function of NBS-LRR proteins providing an important first step in the elucidation of defence signalling (Feys and Parker 2000). In *Arabidopsis,* the *ndr1* and *eds1* mutants were defined in screens for loss of race-specific resistance to strains of the bacterium *Pseudomonas syringae* or the oomycete *Peronospora parasitica*. *EDS1* and *NDR1*, which encode a lipase-like protein and a membrane associated protein respectively, are each required for the function of different NBS-LRR genes (Century et al. 1997; Falk et al. 1999). The R genes suppressed by the *ndr1* mutation are not affected by *eds1* mutants, and vice versa. *eds1* suppresses TIR-NBS-LRR R genes, whereas *ndr1* suppresses a subset of non-TIR-NBS-LRR resistance proteins. Although these observations suggest a model in which EDS1 and NDR1 mediate distinct R gene-dependent signalling pathways (Aarts et al. 1998), there are several examples of non-TIR-NB-LRR R proteins which function independently of both EDS1 and NDR1 (Glazebrook 2001). *RAR1*, was identified in mutational screens for suppressors of *Mla12* resistance in barley to the powdery mildew fungus. This gene encodes Cys-and His-rich

(CHORD) Zn^{2+} binding domains that are conserved in sequence and tandem organization in all eukaryotic phyla examined (Shirasu et al. 1999). RAR1 is required by multiple barley *Mla* genes as well as other unlinked powdery mildew resistance loci. Barley *rar1* mutant plants are impaired in whole cell reactive oxygen intermediates accumulation and in the hypersensitive response of attacked host epidermal cells in *Mla12*-specified resistance, suggesting that RAR1 acts early in the plant resistance cascade. In *Arabidopsis*, RAR1 is also an early component of R gene-triggered resistance against avirulent *Peronospora* and *Pseudomonas syringae*, exerting rate-limiting control of defence signal fluxes leading to hypersensitive plant cell death and is used by both TIR- and non-TIR-NBS-LRR proteins, indicating that its recruitment is not conditioned by a particular R protein structural type in contrast to EDS1 and NDR1 (Muskett et al. 2002; Tornero et al. 2002). Other important defense regulators to emerge are components of mitogen-activated protein kinase (MAPK) cascades that constitute functionally conserved eukaryotic signal relay systems in response to various environmental stresses (Asai et al. 2002; Romeis 2001). Importantly, the MAPK kinase kinase, EDR1, negatively regulates SA-inducible defenses (Frye et al. 2001), whereas MAPK4 appears to differentially regulate SA and JA signals (Petersen 2000). These findings strongly implicate MAPK modules in molecular communication between different plant defense pathways. Another key element of systemic signalling is the *Arabidopsis NPR1* gene (non-expressor of PR1), which encodes an ankyrin repeat protein, initially identified as an SA response regulator. The addition of SA to *Arabidopsis* seedlings promotes movement of NPR1 to the nucleus (Kinkema et al. 2000) where it is able to bind several TGA (TGACG DNA motif) class transcription factors, conferring a possible direct route to defense gene induction (Fan et al. 2002). Identification of an apoplastic lipid transfer protein, DIR1, as an inducer of long distance defense signalling in SAR suggests that lipid-derived molecules may have a role (Maldonado et al. 2002). An overview of the local signalling networks controlling activation of local defense responses is presented in figure 1.6.

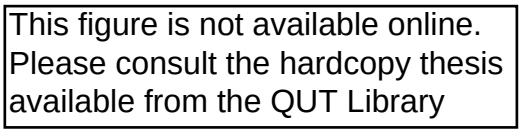


Figure 1.5 Incompatible interaction conferred by a resistant host plant can arise in two ways. (A) The R protein directly recognises the Avr protein itself. This situation is now considered to occur only rarely. (B) The R protein is a guard protein, recognising the modified plant avirulence target caused by the earlier binding of the Avr factor (Hammond-Kosack and Parker 2003).

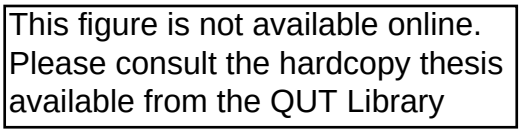


Figure 1.6 Activation of local defense responses mediated by NBS-LRR resistance proteins. Most non-TIR-NBS-LRR resistance proteins require NDR1, whereas TIR-NBS-LRR proteins are dependent on EDS1. A convergence point of the TIR and non-TIR-NBS-LRR proteins is at RAR1/SGT1, both operating upstream of the hypersensitive response (HR) and oxidative burst (OB). Another early defense signal generated is nitric oxide (NO), which can potentiate both the HR and OB. Activation of later potentiating defense responses by TIR-NBS-LRR proteins involves the combined actions of EDS1 and PAD4, EDS5, SA and NPR1. EDR1, MAPK4 and SSI2 can each repress activation of the SA pathway, while various SA-binding proteins (SABP) located in distinct cellular compartments may modulate the local concentrations of available SA signal. The OB can potentiate SA-mediated signalling directly and via the induction of various MAPK cascades, for example, SIPK. NPR1 is required downstream of SA, which also stimulates NPR1 translocation into the nucleus where it interacts with TGA transcription factors and induces the expression of *PR* genes (Hammond-Kosack and Parker 2003).

1.3.6 Engineering pathogen resistance in crop plants using NBS-LRR genes

To control diseases in elite commercial cultivars, plant breeders traditionally have used lengthy breeding programs to introgress new R genes from wild relatives of crop species (Figure 1.7). Currently, the availability of cloned R genes for genetic transformation is opening the possibility of direct transfer into elite lines within a single generation (Figure 1.8) (Hammond-Kosack and Jones 2000). The introduction of R genes by plant transformation also removes the barriers presented when interspecies infertility prevents gene introduction by traditional plant breeding (Hammond-Kosack and Jones 2000; Campbell et al. 2002). For example, in the *Solanaceae* family several NBS-LRR resistance genes have been isolated and transferred from one species to other species of the same family with successful results (Whitham et al. 1996; Tai et al. 1999; Van der Vossen et al. 2003). Attempts to demonstrate function in species outside of the family from which the R gene was isolated have, however, been unsuccessful. For example, the *Arabidopsis RPS2* gene that confers resistance to *Pseudomonas syringae* is non-functional in transgenic tomato and this phenomenon has been referred to as “restricted taxonomic functionality” (RTF) (Tai et al. 1999). The molecular basis of RTF is unknown but might reflect an inability of the R protein to interact with signal transduction components that have diverged in the heterologous host (Hulbert et al. 2001). It remains to be seen whether RTF is a general attribute of R genes. Nevertheless, the transfer of resistance genes even between related species will be a great step forward for plant breeders (Rommens and Kishore 2000). Plant transformation also offers the immediate possibility of introducing simultaneously several different R gene alleles that are effective against a single pathogen species. In theory, this should slow the process of microbe evolution, because the various R genes should be overcome only if all the corresponding Avr gene products mutate simultaneously within a single pathogen isolate (Hammond-Kosack and Jones 2000; McDowell and Woffenden 2003).

The rapid activated and localized defense response that frequently culminates in the hypersensitive response is one of the most prevalent and effective mechanisms deployed by plants to minimize pathogen attack.

Through the combined expression of both an R gene and the complementary Avr gene in a single plant genotype, an engineered “trigger” for HR is possible (Hammond-Kosack and Jones 2000). However, if both components are expressed continuously in a single transgenic plant, the HR induced is devastating, destroying not only the pathogen but also the entire plant. Therefore, the expression of either one component or both must be tightly regulated (Hammond-Kosack and Jones 2000). The desired resistance phenotype may be obtained by a pathogen-inducible promoter (a two-component system) (Figure 1.9A). An ideal pathogen-inducible promoter would be activated rapidly in response to a wide range of pathogens and therefore be effective in providing broad-spectrum resistance. In reality, pathogens have different infection biologies (biotrophs, hemibiotrophs and necrotrophs) (Gurr and Rushton 2005a) and it might be that a pathogen-inducible promoter will only be activated by a subset of possible interactions. The promoter must also be inactive under disease-free conditions to ensure that there are no spurious defense responses triggered by leaky expression of the transgene (McDowell and Woffenden 2003). Furthermore, the promoter should not be autoactivatable by the transgene. This could lead to an uncontrolled spread of gene expression; so-called ‘runaway cell death’ (Gurr and Rushton 2005b). Alternatively, the desired resistance phenotype may be obtained by a limited restoration of R gene function through the somatic excision of a transposable element from an R gene, in combination with constitutive Avr expression; this approach is known as the genetically engineered acquired resistance (GEAR) system (Figure 1.9B). These two approaches have the advantage that the entire multifactorial defense response would be activated, thereby potentially achieving broad-spectrum pathogen control (Hammond-Kosack and Jones 2000).

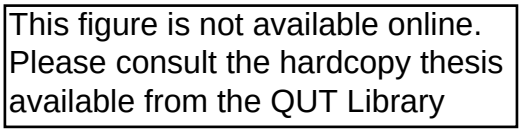


Figure 1.7 Traditional breeding method. In a traditional breeding program, as much as 0.4% of the genome complement from each donor parent can reside in the seventh backcross generation along with the R gene of interest (originally from parent 1) (Hammond-Kosack and Jones 2000).

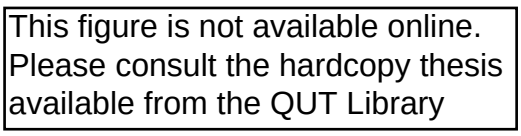


Figure 1.8 Transgenic method. In a transgenic approach, multiple R genes from several initial sources are first assembled into a single Ti plasmid. After T-DNA integration into the plant genome, these R genes cosegregate in all subsequent breeding steps, greatly simplifying the subsequent backcrossing program for introducing multiple new traits into a cultivar. When the transgenic transformation approach is used, the entire sequence of the introduced DNA is known, whereas in traditional breeding program, neither the total extent of the DNA introgressed nor its sequence identity is known. LB and RB represent the left and right T-DNA borders, respectively. NOS T represents the nopaline synthase terminator; the neomycin phosphotransferase gene (NPTII) that confers resistance to kanamycin and the cauliflower mosaic virus 35S (CaMV35S) promoter (caMV35S pro) are indicated (Hammond-Kosack and Jones 2000).

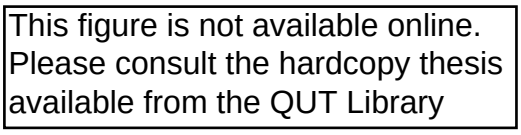

This figure is not available online.
Please consult the hardcopy thesis
available from the QUT Library

Figure 1.9 Two plant biotechnology approaches have been devised that enable plants to activate the entire multifactorial defense response and thereby achieve broad-spectrum resistance. (A) The two-component sensor system includes in one plant the sensor, an Avr gene under the control of a pathogen-inducible promoter, and an effector, an R gene. The promoter fused to the Avr gene is activated by nonspecific elicitors from the attacking pathogen. The Avr gene product then interacts with the resistance gene product, leading to activation of the defense response (HR). (B) Genetically engineered acquired resistance (GEAR) operates through limited restoration of the R gene function, in combination with constitutive Avr expression. R gene function is regulated by inserting a transposable element in the R gene coding sequence, which results in low frequency of somatic excision of this transposon coincident with plant cell division. In the plant cells where R protein function is restored, recognition of the cognate Avr protein occurs, which triggers activation of the plant defense responses. Subsequent defense signals emanating from the cell responding to R-Avr induce resistance responses in the surrounding plant cells, in which the R gene is still nonfunctional. The GEAR technology creates a plant that is a genetic mosaic for cells with and without restored R protein function; in most cells, however, defense responses are active and give the plant improved protection against pathogen attack (Hammond-Kosack and Jones 2000).

1.4 References

Agrios, G.N. (1997). *Plant Pathology*. 4th ed. Academic Press. USA. p. 342-346.

Aarts, N., Metz, M., Holub, E., Staskawic, B., Daniels, M. and Parker, J. (1998). Different requirements for *EDS1* and *NDR1* by disease resistance genes define at least two *R* gene-mediated signalling pathways in *Arabidopsis*. *Proceedings of the National Academy of Science (USA)* 95: 10306-10311.

Anderson, P., Lawrence, G., Morrish, B., Finnegan, E. and Ellis, J. (1997). Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeate coding region. *Plant Cell* 9: 641-651.

Aravind. L., Dixit, V. and Koonin, E. (1999). The domains of the death: evolution of the apoptosis machinery. *Trends in Biochemical Science* 24: 47-53.

Asai, T., Tena, G., Plotnikova, J., Willmann, M., Chiu, W., Gomez-Gomez, L., Boller, T., Ausubel, M. and Sheen, J. (2002). MAP kinase signalling cascade in *Arabidopsis* innate immunity *Nature* 415: 977-983.

Ayliffe, M.A., Frost, D., Finnegan, E.J., Lawrence, G.J., Anderson, P.A. and Ellis, J.G. (1999). Analysis of alternative transcripts of flax *L6* rust resistance gene. *Plant Journal* 17: 287-292.

Ashfield, T., Ong, L., Nobuta, K., Schneider, C., and Innes, R. (2004). Convergent evolution of disease resistance gene specificity in two flowering plant families. *Plant Cell* 16: 309-318.

Bai, J., Pennill, L.A., Ning, J., Lee, S.W., Ramalingam, J., Leach, J.E. and Hulbert, S.H. (2002). Diversity in nucleotide binding site-leucine rich repeat genes in cereals. *Genome Research* 12: 1871-1884.

Ballvora, A., Ercolano, M., Weib, J., Meksem, K., Bormann, C., Oberhagemann, P., Salamini, F. and Gebhardt, C. (2002). The *R1* gene for potato resistance to late blight (*Phytophtora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance gene. *Plant Journal* 30: 361-371.

Becker, D., Dugdale, B., Smith, M., Harding, R. and Dale, J. (2000). Genetic transformation of Cavendish banana (*Musa* spp. AAA group) cv 'Grand Nain' via microprojectile bombardment. *Plant Cell Reports* 19: 229-234.

Beckman, C.H. (1987). *The nature of wilt disease of plants*. APS press. USA. pp. 175.

Bendahmane, A., Kanyuka, K. and Baulcombe, D. (1999). The *Rx* gene from tomato controls separate virus resistance and cell death responses. *Plant Cell* 11: 781-791.

Bentley, S., Pegg, K.G., Moore, N.Y., Davis, R.D. and Buddenhagen, I.W. (1998). Genetic variation among vegetative compatibility groups of *Fusarium oxysporum* f. sp. *cubense* analysed by DNA fingerprinting. *Phytopathology* 88: 1283-1293.

Bergelson, J., Kreitman, M., Sthal, E.A and Tian, D. (2001). Evolutionary dynamics of plant R genes. *Science* 292: 2281-2285.

Bittner-Eddy, P., Crute, I., Holub, E. and Beynon, J. (2000). *RPP13* is a simple locus in *Arabidopsis thaliana* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. *Plant Journal* 21: 177-188.

Botella, M., Pareker, J., Frost, L., Bittner-Eddy, P. and Beynon, J. (1998). Three genes of the *Arabidopsis RPP1* complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. *Plant Cell* 10:1847-1860.

Brommonschenkel, S., Fray, A. and Tanksley, S. (2000). The broad-spectrum tospovirus resistance gene *Sw-5* of tomato is a homolog of the root-knot nematode resistance gene Mi. *Molecular Plant-Microbe Interactions* 13: 1130-1138.

Brueggeman, R., Rostoks, N., Kudrna, D., Kilian, A., Han, F., Chen, J., Druka, A., Steffenson, B. and Kleinhofs, A. (2002). The barley stem rust-resistance gene Rpg1 is a novel disease resistance gene with homology to receptor kinases. *Proceedings of the National Academy of Science (USA)* 99: 9328-9333.

Bryan, G., Wu, K., Farall, L., Jia, Y., Hershey, H., McAdams, S., Faulk, K., Donaldson, G., Tarchini, R. and Valent, B. (2000). A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. *Plant Cell* 12: 2033-2045.

Campbell, M.A., Fitzgerald, H.A. and Ronald, P.C. (2002). Engineering pathogen resistance in crop plants. *Transgenic Research* 11: 599-613.

Carlier, J., Fouré, E., Gauhl, F., Jones, D.R., Lepoivre, P., Mourichon, X., Pasberg-Gauhl, C. and Romero, R.A. (2000). Fungal diseases of the foliage. pp. 37-141. In *Diseases of Banana, Abacá and Enset.* Jones, D.R. (Ed.). CABI. UK.

Century, K., Shapiro, A., Repetti, P., Dahlbeck, D., Holub, E. and Staskawicz, J. (1997). NDR1, a pathogen-induced component required for *Arabidopsis* disease resistance. *Science* 278: 1963-1965.

Collins, N., Drake, J., Ayliffe, M., Sun, Q., Ellis, J., Hulbert, S. and Pryor, T. (1999). Molecular characterization of the Maize *Rp1-D* rust resistance haplotype and its mutants. *Plant Cell* 11: 1365-1376.

Constantine, D. and Rossel, G. (1999). The Musaceae. An annotated list of the species of *Ensete*, *Musa* and *Musella*. http://www.users.globalnet.co.uk/~drc/

Cooley, M.B., Pathirana, S., Wu, H.J., Kachroo, P. and Klessig, D.F. (2000). Members of Arabidopsis *HRT/RPP8* family of resistance genes confer resistance to both viral and oomycete pathogens. *Plant Cell* 12: 663-676.

Crouch, J.H., Vuylsteke, D. and Ortiz, R. (1998). Perspectives on the application of biotechnology to assist the genetic enhancement of plantain and banana (*Musa* spp.). *Electronic Journal of Biotechnology* 1: 1-12.

Dale, J.L. (1987). Banana bunchy top: an economically important tropical plant virus disease. *Advances in Virus Research* 33: 301-325.

Dale, J.L. (1999). Banana biotechnology: already a reality. *AgBiotechNet* 1: 1-6.

Dangl, L. and Jones, J. (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411: 826-833.

Daniells, J., Jenny, C., Karamura, D. and Tomekpe, Kodjo. (2001). Musalogue: A catalogue of *Musa* germplasm. INIBAP. France. pp. X-XI

Dawkins, R. and Krebs, J. (1979). Arms races between and within species. *Proceedings of the Royal Society of London 205: 489-511.*

De Langhe, E.A.L. (1995). Banana and plantain: the earliest fruit crops? INIBAP: Annual report 1995. INIBAP. France. pp. 6-8.

Deslandes, L., Olivier, J., Theulieres, F., Hirsch, J., Feng, D.X., Bittner-Eddy, P., Beynon, J. and Marco, Y. (2002). Resistance to *Ralstonia solanacearum* in *Arabidopsis thaliana* is conferred by the recessive *RRS1-R* gene, a member of a novel family of resistance genes. *Proceedings of the National Academy of Science (USA)* 99: 2404-2409.

Dhed'a, D., Dumortier, F., Panis, B., Vuylsteke, D. and De langhe, E. (1991). Plant regeneration in cell suspension cultures of cooking banana cv. Bluggoe (*Musa* spp. ABB group). *Fruits* 46: 125-135.

Dinesh-Kumar, S.P. and Baker, B.J. (2000). Alternatively spliced *N* resistance gene transcripts: their possible role in tobacco mosaic virus resistance. *Proceedings of the National Academy of Science (USA)* 97: 1908-1913.

Di Pietro, A., Garcia-Maceira, F.I., Meglecz, E. and Roncero, M.I.G. (2001). A MAP kinase of the vascular wilt fungus *Fusarium oxysporum* is essential for root penetration and pathogenesis. *Molecular Microbiology* 39: 1140-1152.

Di Pietro, A., Madrid, M.P., Caracuel, Z., Delgado-Jarana, J. and Roncero, M.I.G. (2003). *Fusarium oxysporum*: exploring the molecular arsenal of vascular wilt fungus. *Molecular Plant Pathology* 4: 315-325.

Dixon, M., Hatzixanthis, K., Jones, D., Harrison, K. and Jones, J. (1998). The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. *Plant Cell* 10: 1915-1925.

Dodds, P., Lawrence, G. and Ellis, J. (2000). Six amino acid changes confined to the leucine-rich repeat B-strand/B-turn motif determine the difference between the *P* and *P2* rust resistance specificities in flax. *Plant Cell* 13: 163-178.

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: 495-506.

Ernst, K., Kumar, A., Kriseleit, D., Kloos, D., Phillips, M. and Ganal, M. (2002). The broad-spectrum potato cys nematode resistance gene (*Hero*) from tomato is the only member of a large gene family of NBS-LRR genes with an unusual amino acid repeat in the LRR region. *Plant Journal* 31: 127-136.

Escalant, J.V., Teisson, C. and Cote, F. (1994): Amplified somatic embryogenesis from male flowers of triploid banana and plantain cultivars (*Musa* spp.). *In vitro Plant Cellular and Developmental Biology* 30: 181-186.

Falk, A., Feys, B., Frost, L., Jones, J., Daniels, M. and Parker, J. (1999). EDS1, an essential component of *R* gene-mediated disease resistance in *Arabidopsis* has homology to eukaryotic lipases. *Proceedings of the National Academy of Science (USA)* 96: 3292-3297.

Fan, W.H. and Dong, X.N. (2002). In vivo interaction between NPR1 and transcription factor TGA2 leads to salicylic acid-mediated gene activation in *Arabidopsis. Plant Cell* 14: 1377-1389.

FAO, (2005). http://apps.fao.org/lim500/nph-wrap.pi?production.crops.primary&Domain=SUA&

Feys, B.J. and Parker, J.E. (2000). Interplay of signalling pathways in plant disease resistance. *Trends in Genetics* 16: 449-455.

Flor, H.H. (1971). Current status of the gene-for-gene concept. *Annual Review of Phytopathology* 9: 275-296.

Fluhr, R. (2001). Sentinels of disease. Plant resistance genes. *Plant Physiology* 127: 1367-1374.

Frye, C.A., Tang, D. and Innes, R.W. (2001). Negative regulation of defense responses in plants by a conserved MAPKK kinase. *Proceedings of the National Academy of Science (USA)* 98: 373-378.

Gassmann, W., Hinsch, M. and Staskawicz, B. (1999). The *Arabidopsis RPS4* bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease resistance genes. *Plant Journal* 20: 265-277.

Glazebrook, J. (2001). Genes controlling expression of defence responses in *Arabidopsis*-2001 status. *Current Opinion in Plant Biology* 4: 301-308.

Goff, S. et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). *Science* 296: 92-100.

Graham, M., Mareck, L., Lohnes, D., Cregan, P. and Shoemaker, R. (2000). Expression and genome organization of resistance gene analogues in soybean. *Genome* 43: 86-93.

Grant, M., Godard, L., Straube, E., Ashfield, T., Leward, J., Sattler, A., Innes, R. and Dangl, J. (1995). Structure of the *Arabidopsis RPM1* enabling dual specifity disease resistance. *Science* 269: 843-846.

Grant, M. (1998). Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis*. *Proceedings of the National Academy of Science (USA)* 95: 15843-15848.

Gurr, S.J. and Rushton, P.J. (2005a). Engineering plants with increased disease resistance: what are we going to express? *Trends in Biotechnology* 23: 275-282.

Gurr, S.J. and Rushton, P.J. (2005b). Engineering plants with increased disease resistance: how are we going to express it? *Trends in Biotechnology* 23: 283-290.

Hammond-Kosack, K. and Jones, J. (1997). Plant resistance genes. *Annual Review of Plant Physiology and Plant Molecular Biology* 48: 575-607.

Hammond-Kosack, K. and Jones, D.G. (2000). Responses to plant pathogens. p.p. 1102-1156. In *Biochemistry and Molecular Biology of Plants*. (Buchanan, B., Gruissem, W. and Jones R. eds.). American Society of Plant Physiologists. USA.

Hammond-Kosack, K. and Parker, J. (2003) Deciphering plant-pathogen communication: fresh perspectives for molecular resistance breeding. *Current Opinion in Biotechnology*, 14, 177-193.

Hermann, SR., Harding, RM. and Dale, J.L. (2001). The banana *Actin 1* promoter drives near-constitutive transgene expression in vegetative tissues of banana (*Musa* ssp). *Plant Cell Reports* 20: 525-530.

Holub, E.B. (2001). The arms race is ancient history in *Arabidopsis*, the wildflower. *Nature reviews/ Genetics* 2: 516-527.

Horry, J-P., Ortiz, R., Arnaud, E., Crouch, J.H., Ferris, R.S.B., Jones, D.R., Mateo, N., Picq, C. and Vuylsteke, D. (1997). Banana and Plantain. pp. 67-81. *In biodiversity in trust. Conservation and use of plant genetic resources in CGIAR centres.* Fuccillo, D., Sears, L. and Stapleton, P. (Eds). Cambridge University Press. UK.

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: 251-261.

Hulbert, S., Webb, C., Smith, S. and Sun, Q. (2001). Resistance gene complexes: evolution and utilization. *Annual Review of Phytopathology* 39: 285-312.

Jia, Y., McAdams, S., Bryan, G., Hershey, P. and Valent, B. (2000). Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO Journal* 19: 4004-4014.

Joobeur, T., King, J., Nolin, S., Thomas, C., and Dean, R. (2004). The *Fusarium* wilt resistance locus *Fom-2* of melon contains a single resistance gene with complex features. *Plant Journal* 39: 283-297.

Jones, D.A. and Jones, J.D.G. (1997). The role of leucine-rich repeat proteins in plant defenses. *Advances in Botanical Research* 24: 89-167.

Jones, D.R. (2000). Introduction of Banana, Abacá and Enset. pp. 1-36. In *Diseases of Banana, Abacá and Enset.* Jones, D.R. (Ed.). CABI. UK.

Kanazin, V., Mareck, L. and Shoemaker, R. (1996). Resistance gene analogs are conserved and clustered in soybean. *Proceedings of the National Academy of Science (USA)* 93: 11746-11750.

Keen, N.T. (1990). Gene-for-gene complementarity in plant-pathogen interactions. *Annual Review of Genetics* 24: 447-463.

Khanna, H., Becker, D., Kleidon, J. and Dale, J. (2004). Centrifugation assisted *Agrobacterium*-mediated transformation (CAAT) of embryogenic

cell suspensions of banana (*Musa* spp. Cavendish AAA and Lady finger AAB). *Molecular Breeding* 14: 239-252.

Kinkema, M., Fan, W.H. and Dong, X.N. (2000). Nuclear localization of NPR1 is required for activation of PR gene expression. *Plant Cell* 12: 2339-2350.

Krietman, M. and Akashi, H. (1995). Molecular evidence for natural selection. *Annual Review of Ecology and Systematics* 26: 403-422.

Lagopodi, A.L., Ram, A.F., Lamers, G.E., Punt, P.J., Van den Hondel, C.A., Lugtenberg, B.J. and Bloemberg, G.V. (2002). Novel aspects of tomato root colonization and infection by *Fusarium oxysporum* f.sp. *radicis-licopersici* revealed by confocal laser scanning microscopic analysis using the green fluorescent protein as a marker. *Molecular Plant-Microbe Interactions* 15: 172-179.

Lagudah, E., Moullet, O. and Appels, R. (1997). Map-based cloning of a resistance gene sequence encoding a nucleotide-binding domain and leucine rich region at the *Cre3* nematode resistance locus of wheat. *Genome* 40: 659-665.

Lawrence, G.J., Finnegan, E.J., Ayliffe, M. and Ellis, J. (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: 1195-1206.

Leister, D., Ballvora, A., Salamini, F. and Gebhardt, C. (1996). A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. *Nature Genetics* 14: 421-429.

Leister, D., Kurth, J., Laurie, D., Yano, M., Sasaki, T., Devos, K., Graner, A. and Schulze-Lefert, P. (1998). Rapid organization of resistance gene homologues in cereal genomes. *Proceedings of the National Academy of Science (USA)* 95: 370-375.

Leister, D. (2004). Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends in Genetics* 20: 116-122.

López, C., Acosta, I., Jara, C., Pedraza, F., Gaitan-Solis, E., Gallego, G., Beebe, S. and Tohme, J. (2003). Identifying resistance gene analogs associated with resistance to different pathogens in common bean. *Phytopathology* 93: 88-95.

Luck, J.E. (2000). Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination. *Plant Cell* 12: 1367-1377.

Lysak, M.A., Dolezelova, M., Horry, J.P., Swennen, R. and Dolezel, J. (1999). Flow cytometric analysis of nuclear DNA content in *Musa*. *Theoretical and Applied Genetics* 98: 1344-1350.

Maldonado, A.M., Doerner, P., Dixon, R.A., Lamb, C.J. and Cameron, R.K. (2002). A putative lipid transfer protein involved in systemic resistance signalling in *Arabidopsis*. *Nature* 419: 399-403.

Marín, D.H., Romero, R.A., Guzmán, M. and Sutton, T.B. (2003). Black sigatoka: An increasing threat to banana cultivation. *Plant Disease* 87: 208-222.

May, G., Afza, R., Mason, H., Wiecko, A., Novak, F. and Arntzen, C. (1995). Generation of transgenic banana (*Musa acuminata*) plants via Agrobacterium-mediated transformation. *Bio/Technology* 13: 486-492.

McDowell, J.M. and Woffenden, B.J. (2003). Plant disease resistance genes: recent insights and potential applications. *Trends in Biotechnology* 21: 178-183.

Mes, J., Van Doorn, A., Wijbrandi, J., Simons, G., Cornelissen, B. and Haring, M. (2000). Expression of the *Fusarium* resistance gene *I-2* colocalizes with the site of fungal containment. *Plant Journal* 23: 183-193.

Meyers, B., Shen, K., Rohani, P., Gaut, B. and Michelmore, R. (1998). Receptor –like genes in the mayor resistance locus in lettuce are subject of divergent selection. *Plant Cell* 11: 1833-1846.

Meyers B, Dickerman A, Michelmore R, Sivaramakrishnan S, Sobral, B. and Young, N. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant Journal* 20: 317-332.

Meyers, B., Kozik, A., Griego, A., Kuang, H. and Michelmore, R.W. (2003). Genome-wide analysis of NBS-LRR encoding genes in *Arabidopsis*. *Plant Cell* 15: 809-834.

Milligan, S., Bodeau, J., Yaghoobi, J., Kaloshian, I., Zabel, P. and Williamson, V. (1998). The root knot resistance gene *M*i from tomato is a member of the leucine zipper, nucleotide binding site, leucine-rich repeat family of plant genes. *Plant Cell* 10: 1307-1319.

Mindrinos, M., Katagiri, F., Yu, G. and Ausubel, F. (1994). The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell* 78: 1089-1099.

Moore, N.Y., Hargreaves, P.A., Pegg, K.G. and Irwin, J.A.G. (1991). Characterization of strains of *Fusarium oxysporum* f. sp. *cubense* by production of volatiles. *Australian Journal of Botany* 39: 161-166.

Muskett, P., Kahn, K., Austin, M., Moisan, L., Sadanandom, A., Shirasu, K., Jones, J. and Parker, J. (2002). *Arabidopsis* RAR1 exerts rate-limiting control of *R* genes-mediated defenses against multiple pathogens. *Plant Cell* 14: 979-992.

Noel, L., Moores, T.L., Van der Biezen, E.A., Parniske, M., Daniels, M.J., Parker, J.E. and Jones, J.D.G. (1999). Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus in *Arabidopsis*. *Plant Cell* 11: 2099-2111

O'Donnell, K., Kistler, H.C., Cigelnik, E. and Ploetz, R.C. (1998). Multiple evolutionary origins of the fungus causing Panama disease of banana: Concordant evidence from nuclear and mitochondrial gene genealogies. *Proceedings of the National Academy of Science (USA)* 95: 2044-2049.

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: 203-213.

Parker, J., Coleman, M., Szabo, V., Frost, V., Schmidt, R., Van der Biezen, E., Moores, T., Dean, C., Daniels, M. and Jones, J. (1997). The *Arabidopsis* downy mildew resistance gene *RPP5* shares similarity to the toll interleukine-1 receptors with N and L6. *Plant Cell* 9: 879-894.

Parniske, M., Hammond-Kosack, K., Golstein, C., Thomas, C., Jones, D., Harrison, K., Wulff, B.B. Jones, J. (1997). Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4*/*9* locus of tomato. *Cell* 91: 821-832.

Petersen, M., Brodersen, P., Naested, H., Andreasson, E., Lindhart, U., Johansen, B., Nielsen, H.B., Lacy, M., Austin, M.J. and Parker, J.E. (2000). *Arabidopsis* MAP kinase 4 negatively regulates systemic acquired resistance. *Cell* 103: 1111-1120.

Ploetz, R.C. (2000). Panama disease: A classical and destructive disease of banana. Online. *Plant Health Progress* doi: 10.1094/PHP-2000-1204-01-HM. www.planthealthprogress.org.

Ploetz, R.C. and Pegg, K.G. (2000). Fungal diseases of the root, corm and pseudostem. pp. 143-172. In *Diseases of Banana, Abacá and Enset*. (Jones, D.R., ed). CABI. UK.

Ploetz, R.C. (2005). Panama disease: An old nemesis rears its ugly head (part 2, the Cavendish era and beyond). *Plant Health Progress* doi: 10.1094/PHP-2005-1221-01-RV. www.planthealthprogress.org.

Rivkin, M., Vallejos, C. and McClean, P. (1999). Disease-resistance related sequences in common bean. *Genome* 42: 41-47.

Rodriguez-Gálvez, E. and Mendgen, K. (1995). The infection process of *Fusarium oxysporum* in cotton root tips. *Protoplasma* 189: 61-72.

Romeis, T. (2001). Protein kinases in the plant defense response. *Current Opinion in Plant Biology* 4: 407-414.

Rommens, C. and Kishore, G. (2000). Exploiting the full potential of disease-resistance genes for agricultural use. *Current Opinion in Biotechnology* 11: 120-125.

Rossi, M., Goggin, F.L., Milligan, S.B., Kaloshian, I., Ullman, D.E. and Williamson, V.M. (1998). The nematode resistance gene *Mi* of tomato confers resistance against the potato aphid. *Proceedings of the National Academy of Science (USA)* 95: 9750-9754.

Sagi, L., Panis, B., Remy, S., Schoofs, H., De Smet, K., Swennen, R. and Cammue, P. (1995). Genetic transformation of banana and plantain (*Musa* spp) via particle bombardment. *Bio/Techniques* 13: 481-485.

Salmeron, J., Oldroyd, E., Rommens, C., Scofield, S., Kim, H., Lavelle, D., Dahlbeck, D. and Staskawicz, B. (1996). Tomato *Prf* is a member of a leucine-rich repeat class of plant disease resistance gene and lies embedded within the *Pto* kinase gene cluster. *Cell* 86:123-133.

Sarah, J.L. (2000). Nematode pathogens. pp. 295-323. In *Diseases of Banana, Abacá and Enset.* Jones, D.R. (Ed.). CABI. UK.

Schreuder, W., Lamprecht, S.C. and Holz, G. (2000). Race determination and vegetative compatibility grouping of *Fusarium oxysporum* f.sp. *melonis* from South Africa. *Plant Disease* 84: 231-234.

Seah, S., Spielmeyer, W., Jahier, J., Sivasithamparam, K. and Lagudah, S. (2000). Resistance gene analogs within and introgressed chromosomal segment derived from *Triticum ventricosum* that confers resistance to nematode and rust pathogens in wheat. *Molecular Plant Microbe Interactions* 13: 334-341.

Sela-Buurlage, M., Budai-Hadrian, O., Pan, Q., Carmel-Goren, L., Vunsch, R., Zamir, D., and Fluhr, R. (2001). Genome-wide dissection of *Fusarium* resistance in tomato reveals multiple complex loci. *Molecular Genetics and Genomics* 265:1104-1111.

Shen, K., Meyers, B., Islam-Faridi, M., Chin, D., Stelly, D. and Michelmore, R. (1998). Resistance gene candidates identified by PCR with degenerate oligonucleotide primers Map to clusters of resistance genes in lettuce. *Molecular Plant Microbe Interactions* 11: 815-823.

Simons, G., Groenendijk, J., Wijbrandi, J., Reijans, M., Groenen, J., Diergaarde, P., Van der Lee, T., Bleeker, M., Onstenk, J., Both, M., Haring, M., Mes, J., Cornelisse, B., Zabeau, M. and Vos, P. (1998).

Dissection of the *Fusarium I2* gene cluster in tomato reveals six homologs and one active gene copy. *Plant Cell* 10: 1055-1068.

Sharrock, S. (1996). Uses of banana. INIBAP: Annual report 1996. INIBAP. France. pp. 42-44.

Sharrock, S. (1997). The banana and its relatives. INIBAP: Annual report 1997. INIBAP. France. pp. 52-55.

Shen, K.A., Chin, D.B., Arroyo-García, R., Ochoa, O.E., Lavelle, D.O., Wroblewski, T., Meyers, B.C. and Michelmore, R.W. (2002). *Dm3* is one member of a large constitutively expressed family of nucleotide binding site leucine-rich repeat encoding genes. *Molecular Plant Microbe Interactions* 15: 251-261.

Shirasu, K., Lahaye, T., Tan, M., Zhou, F., Azevedo, C. and Schulze-Lefert, P. (1999). A novel class of eukaryotic zinc-binding proteins is required for disease resistance signalling in barley and the development in *C. elegans*. *Cell* 99: 355-366.

Smith, M.K., Whiley, A.W., Searle, C., Langdon, P.W., Schaffer, B. and Pegg, K.G. (1998). Micropropagated bananas are more susceptible to Fusarium wilt than plants grown from conventional material. *Australian Journal of Agricultural Research* 49: 1133-1139.

Song, J., Bradeen, J.M., Naess, S.K., Raasch, J.A., Wielgus, S.M., Haberlach, G.T., Liu, J., Kuang, H., Austin-Phillips, S., Buell, C.R., Helgeson, J.P. and Jiang, J. (2003). Gene *RB* cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late blight. *Proceedings of the National Academy of Science (USA)* 100: 9128-9133.

Staal, J., Kaliff, M., Bohman, S. and Dixelius, C. (2006). Transgressive segregation reveals two *Arabidopsis* TIR-NBS-LRR genes effective against *Leptosphaeria maculans*, causal agent of blackleg disease. *Plant Journal* 46: 218-230.

Sthal, E.A., Dwyer, G., Mauricio, R., Kreitman, M. and Bergelson, J. (1999). Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* 400: 667-671.

Staskawicz, B.J., Ausubel, F.M., Baker, B.J., Ellis, J.G. and Jones, J.D. (1995). Molecular genetics of plant disease resistance. *Science* 268: 661-667.

Stover, R.H. (1962). *Fusarial wilt (Panama disease) of bananas and other Musa species.* Commonwealth Mycological Institute. UK. pp. 117.

Stover, R.H. and Simmonds, N.W. (1987). *Bananas*. 3th ed. Longman Scientific & Technical. USA.

Swennen, R. and Rosales, F.E. (1994). Bananas. Encyclopedia of Agricultural Science, Vol. 1. Academic Press. pp. 215-232.

Swennen, R., Arinaitwe, G., Cammue, B.P.A., Francois, I., Panis, B., Remy, S., Sagi, L., Santos, E., Strosse, H. and Van den Houwe, I. (2002). Transgenic approaches for resistance to *Mycosphaerella* leaf spot disease in *Musa* spp. pp. 209-238. In *Mycosphaerella* leaf spot disease of bananas: present status and outlook. (Jacome, L., Lepoivre, P., Marin, D. Ortiz, R. Romero, R. and Escalant, J.V., eds). Proceedings of the 2nd international workshop on *Mycosphaerella* leaf spot diseases held in San José, Costa Rica, 20-23 May 2002.

Tai, T., Dahlbeck, D., Clarck, E., Gajiwala, P., Pasion, R., Whalen, M., Stall, R. and Staskawicz, J. (1999). Expression of the *Bs2* pepper gene confers resistance to bacteria spot disease in tomato. *Proceedings of the National Academy of Science (USA)* 96: 14153-14158.

Takken, F.L.W. and Joosten, M.H.A.J. (2000). Plant resistance genes: their structure, function and evolution. *European Journal of Plant Pathology* 106: 699-713.

Tameling, W., Elzinga, S., Darmin, P., Vossen, J., Takken, F., Haring, M. and Cornelissen, B. (2002). The tomato R gene products I2 and Mi-1 are functional ATP binding proteins with ATPase activity. *Plant Cell* 14: 2929-2939.

Tang, X., Frederick, R.D., Zhou, J., Halterman, D.A., Jia, Y., Martin, G.B. (1996). Initiation of plant disease resistance by physical interaction of AvrPto and Pto kinase. *Science* 274: 2060-2063.

The *Arabidopsis* genome initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815.

Thwaites, R., Eden-Green, S.J. and Black, R. (2000). Diseases caused by bacteria. Pp. 213-239. In *Diseases of Banana, Abacá and Enset.* Jones, D.R. (Ed.). CABI. UK.

Thomas, C., Jones, D., Parniske, M., Harrison, K. and Balint-Kurti, P. (1997). Characterization of the tomato *Cf-4* gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in Cf-4 and Cf-9. *Plant Cell* 9: 2209-2224.

Tornero, P., Merritt, P., Sadanandom, A., Shirasu, K., Innes, R. and Dangl, J. (2002). RAR1 and NDR1 contribute quantitatively to disease resistance in *Arabidopsis*, and their relative contributions are dependent on the *R* gene assayed. *Plant Cell* 14: 1005-1015.

Van der Biezen, E. and Jones, D. (1998). Plant disease-resistance proteins and the gene-for-gene concept. *Trend in Biochemestry Science* 23: 454-456.

Van der Hoorn, R., De Wit, P. and Joosten, M. (2002). Balancing selection favors guarding resistance proteins. *Trends in Plant Science* 7: 67-71.

Van der Vossen, E., Rouppe, J., Kanyuka, K., Bendahmane, A. and Sandbrink, H. (2000). Homologue of a single resistance-gene cluster in potato confers resistance to distinct pathogens: a virus and a nematode. *Plant Journal* 23: 567-576.

Van der Vossen, E., Sikkema, A., Hekkert, B.T.L., Gros, J., Stevens, P., Muskens, M., Wouters, D., Pereira, A., Stiekema, W. 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: 867-882.

Vision, T.J., Brown, D.G. and Tanskley, S.D. (2000). The origins of genomic duplication in *Arabidopsis*. *Science* 290: 2114-2117.

Wang, Z.X., Yano, M., Yamanouchi, U., Iwamoto, M. and Monna, L. (1999). The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine rich repeat class of plant disease resistance genes. *Plant Journal* 19: 55-64.

Warren, R., Henk, A., Mowery, P., Holub, E. and Innes, R. (1998). A mutation within the leucine-rich repeat domain of the *Arabidopsis* disease resistance gene *RPS5* partially suppresses multiple bacterial and downy mildew resistance genes. *Plant Cell* 10: 1439-1452.

Whitham, S., McCormick, S. and Baker, B. (1996). The *N* gene of tobacco confers resistance to tobacco mosaic virus in transgenic tomato. *Proceedings of the National Academy of Science (USA)* 93: 8776-8781.

Yoshimura, S., Yamanouchi, U., Katayose, Y. and Toki, S. (1998). Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. *Proceedings of the National Academy of Science (USA)* 95: 1663-1668.

Yu, Y., Buss, G. and Maroof, M. (1996). Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. *Proceedings of the National Academy of Science (USA)* 93: 11751-11756.

Zhao, B., Lin, X., Poland, J., Trick, H., Leach, J. and Hulbert, S. (2005). A maize resistance gene functions against bacterial streak disease in rice. *Proceedings of the National Academy of Science (USA)* 102: 15383-15388.

Zhou, F., Kurth, J., Wei, F., Elliott, C., Vale, G., Yahiaoui, N., Keller, B., Somerville, S., Wise, R. and Schulze-Lefert, P. (2001). Cell-autonomous expression of barley Mla1 confers race-specific resistance to the powdery

mildew fungus via a Rar1-independent signalling pathway. *Plant Cell* 13: 337-350.

Zhou, T., Wang, Y., Chen, J.Q., Araki, H., Jing, Z., Jiang, K., Shen, J. and Tian, D. (2004). Genome-wide identification of NBS genes in *japonica* rice reveals significant expansion of divergent non-TIR-NBS-LRR genes. *Molecular Genetics and Genomics* 271: 402-415.

Zink, F.W. and Thomas, C.E. (1990). Genetics of resistance in muskmelon to *Fusarium oxysporum* f. sp. *melonis* race 0, 1 and 2 in muskmelon line MR-I. *Phytopathology* 80: 1230-1232.

Chapter 2

Aims of the Study

2.1 Aims

Fifty years ago the banana export industry based on the cultivar ‘Gros Michel’ was wiped out by the soil-borne fungus *Fusarium oxysporum* f. sp. *cubense* race 1 (Ploetz and Pegg 2000). The plantations were replaced by natural resistant triploids of the Cavendish group and the cultivar ‘Gros Michel’ was rapidly abandoned. Over the past few years the FOC race 4 has emerged as the most serious threat to the banana production worldwide since cultivars of the Cavendish group are highly susceptible to this race. Unlike fifty years ago, today there are no naturally resistant substitutes that possess the fruit attributes (flavour, aroma and pulp texture) that the world market demands and that make current commercial cultivars so attractive for consumers. Efforts to develop FOC race 4 resistant varieties through conventional breeding have resulted in the development of a promising tetraploid hybrid called FHIA-01® (‘Goldfinger’) (Rowe and Rosales 2000). Although this new variety is resistant to FOC race 4, it does not have the attractive flavour attributes of the Cavendish cultivars and thus the broad acceptance of this variety in the world banana market is unlikely. The technology of genetic engineering in banana is already established in numerous laboratories around the world (Sagi et al. 1995; May et al. 1995; Becker et al. 2000; Khanna et al. 2004). This technology holds the promise to introduce FOC resistance (R) genes in the current commercial susceptible cultivars by genetic transformation without compromising the valuable traits of the fruit and also holds the promise to bring back on the market the once popular banana cultivar, ‘Gros Michel’. In order to develop FOC resistance in banana through the use of R genes and genetic engineering, the corresponding R gene (s) need to be identified, which is one of the major goals of the Plant Biotechnology Program (Queensland University of Technology, Australia). So far, cloning of genes that confer resistance to *F.*

oxysporum has been achieved only in tomato and melon. In tomato, the *I2* gene confers resistance to *F. oxysporum* f. sp. *licopercisi* (FOL) race 2 (Simons et al. 1998); and in melon, the *Fom-2* gene confers resistance to *F. oxysporum* f.sp. *melonis* (FOM) races 0 and 1 (Joobeur et al. 2004). Interestingly, both genes belong to the same non-TIR subclass of the largest class of plant disease resistance genes, the NBS-LRR. The fact that both genes belong to same class of disease resistance genes suggests a similar *Fusarium* resistance mechanism shared by two different plant families (Solanaceae and Cucurbitaceae), such type of *Fusarium* R genes may be extended in other plant families including the Musaceae. The *I2* and *Fom-2* genes represent a valuable resource to develop *Fusarium* resistance in tomato and melon, respectively. However, the race specificity of these R genes imposes a serious limitation for their use in other crops such as banana. Consequently, other potential sources of FOC resistance must be considered. One potential source of *Fusarium* resistance genes for developing resistance in banana is the wild, progenitor banana species, *Musa acuminata* ssp. *malaccensis* which is a diploid subspecies that produces small fruits with seeds that have no commercial value. This subspecies is, however, highly resistant to most banana pathogens including FOC race 4. In field trials using a population of healthy *Musa acuminata* ssp. *malaccensis* originating from a FOC tropical race 4-infected site on Sumatra, the population segregated for resistance in an Australian FOC subtropical race 4-infected site in an Mendelian ratio of 3:1 suggesting a single dominant gene was involved in conferring resistance to the fungal pathogen (Smith and Hamill 1999). Tissue-culture plantlets from this segregating population were kindly provided to the Plant Biotechnology Program (Queensland University of Technology) by Dr. Mike Smith (Queensland Department of Primary Industries and Fisheries, Nambour, Australia) to be used in our research. This population holds the potential to assist in the isolation of a FOC race 4 resistance gene and other potential *Fusarium* resistance genes in banana.

This project aims to characterise resistance gene candidates (RGCs) of the NBS-type from the wild banana *Musa acuminata* ssp. *malaccensis* and to

identify potential *Fusarium* resistance genes using a combination of bioinformatics and gene expression analysis.

Specifically, the aims of this project are:

(i) To isolate and characterise the structure, phylogeny and expression of disease resistance gene candidates of the NBS-type from *Musa acuminata* ssp. *malaccensis*.

(ii) To isolate and characterise the structure and phylogeny of potential *Fusarium* resistance genes.

2.2 References

Becker, D., Dugdale, B., Smith, M., Harding, R. and Dale, J. (2000). Genetic transformation of Cavendish banana (*Musa* spp. AAA group) cv 'Grand Nain' via microprojectile bombardment. *Plant Cell Reports* 19: 229-234.

Joobeur, T., King, J., Nolin, S., Thomas, C., and Dean, R. (2004). The *Fusarium* wilt resistance locus *Fom-2* of melon contains a single resistance gene with complex features. *The Plant Journal* 39: 283-297.

Khanna, H., Becker, D., Kleidon, J. and Dale, J. (2004). Centrifugation assisted *Agrobacterium*-mediated transformation (CAAT) of embryogenic cell suspensions of banana (*Musa* spp. Cavendish AAA and Lady finger AAB). *Molecular Breeding* 14: 239-252.

May, G., Afza, R., Mason, H., Wiecko, A., Novak, F. and Arntzen, C. (1995). Generation of transgenic banana (*Musa acuminata*) plants via Agrobacterium-mediated transformation. *Bio/Technology* 13: 486-492.

Ploetz, R. and Pegg, K. (2000). Fungal disease of the root, corm and pseudosteam. pp. 143-171. *In Diseases of banana, abaca and enset*. Jones, D.R. (Ed.). CABI. Wallingford, UK.

Rowe, P.R. and Rosales, F.E. (2000). Conventional banana breeding in Honduras. pp. 435-449. *In Diseases of banana, abaca and enset*. Jones, D.R. (Ed.). CABI. Wallingford, UK.

Sagi, L., Panis, B., Remy, S., Schoofs, H., De Smet, K., Swennen, R. and Cammue, P. (1995). Genetic transformation of banana and plantain (*Musa* spp) via particle bombardment. *Bio/Techniques* 13: 481-485.

Simons, G., Groenendijk, J., Wijbrandi, J., Reijans, M., Groenen, J., Diergaarde, P., Van der Lee, T., Bleeker, M., Onstenk, J., Both, M., Haring, M., Mes, J., Cornelisse, B., Zabeau, M., and Vos, P. (1998). Dissection of the Fusarium *I2* gene cluster in tomato reveals six homologs and one active gene copy. *Plant Cell* 10: 1055-1068.

Smith MK, Hamill SD (1999) Banana tissue culture for clean, sustainable production. In: Final Report (FR96013), Horticultural Research & Development Corporation, Gordon, NSW, Australia.

CHAPTER 3
General Materials and Methods

3.1 Plant material

Musa acuminata ssp. *malaccensis* (genotype AA) plantlets, resistant (accessions 850 and 852) or susceptible (accessions 845 and 846) to FOC subtropical race 4 (Smith et al. 1998) were kindly provided by Dr. Mike Smith (Queensland Department of Primary Industries, Nambour, Australia). They were grown in pots in an incubator chamber and used as the source for the harvested tissues, leaves and roots of 4-month-old plants. Harvested tissues were frozen in liquid nitrogen and stored at –80^{o}C until DNA or RNA extraction.

3.2 Nucleic acid extraction

3.2.1 DNA extraction

Genomic DNA was extracted from 2 g of leaf tissue using the protocol of Dellaporta et al. (1983). Leaves frozen in liquid nitrogen were ground into a fine powder using a mortar and pestle. Immediately, the powder was transferred to a 50 ml Falcon tube with 15 ml of extraction buffer (100 mM Tris-HCl pH 8, 50 mM EDTA pH 8.0, 500 mM NaCl, 10 mM mercaptoethanol) preheated at 65^{o}C. After vortexing, 2 ml of 10% SDS was added to the mixture. The tube was vortexed again and incubated at 65^{o}C for 15 min. A 5 ml aliquot of 5 M potassium acetate was added, mixed thoroughly, and the homogenate incubated on ice for 20 min. The tube was centrifuged at 3,700 rpm for 20 min and the supernatant transferred to a new 50 ml Falcon tube. An equal volume of chloroform and isoamyl alcohol ($CHCl_3$:IAA) 24:1 was added and mixed thoroughly to form an emulsion. After centrifugation at 3,700 rpm for 10 min, the supernatant was transferred to a new 50 ml Falcon tube and 0.7 volume of isopropanol added. The tube was inverted several times and incubated for 20 min at -20^{o}C to precipitate the nucleic acids. DNA was pelleted by centrifugation for 15 min at 3,700 rpm and the pellet resuspended in 750 μl of TE buffer (10 mM Tris-HCl, pH 8.0, 1

mM EDTA pH 8.0). The DNA solution was further extracted with an equal volume of $CHCl_3$:IAA (24:1) and the DNA precipitated by addition of 0.1 volume of 3 M sodium acetate pH 7.0 and 0.7 volume of isopropanol. DNA was pelleted by centrifugation at 14,000 rpm for 5 min, then the pellet was washed twice with 70% ethanol and dried at room temperature for 20 min. The DNA was resuspended in TE buffer by heating at 65^{o}C for 5 min. The RNA was digested with RNase A (1 mg/ml, Sigma) and the DNA was stored at 4^{o}C.

3.2.2 RNA extraction.

RNA was extracted from 2 g of leaf or root tissues according to Schuler and Raymond (1989) with minor modifications. Frozen tissue was ground into a fine powder using a mortar and pestle, and immediately the powder was transferred to a 50 ml Falcon tube containing 15 ml of RNA extraction buffer (100 mM Tris-HCl pH 8.0, 50 mM EDTA pH 8.0, 500 mM NaCl, 10 mM mercaptoethanol) preheated at 65^{o}C. The tube was vortexed thoroughly and incubated at 65^{o}C for 15 min. The lysate was extracted with an equal volume of $CHCl_3$:IAA (24:1) and centrifuged at 3,700 rpm for 10 min. The supernatant was collected and nucleic acids precipitated by addition of an equal volume of isopropanol. Pelleted nucleic acids were resuspended in 900 μl of TE buffer, transferred to a 2 ml microcentrifuge tube and extracted with an equal volume of $CHCl_3$:IAA (24:1). The supernatant was transferred to a new 2 ml microfuge tube and 0.25 volume 8 M urea and 0.25 volume of 10 mM LiCl added. The solution was mixed by inverting the tube several times and incubated overnight on ice at 4^{o}C. Nucleic acids were pelleted by centrifugation at 14,000 rpm for 20 min and resuspended in 600 μl of TE buffer. The RNA was precipitated by addition of 0.1 volume of 3 M sodium acetate pH 7.0 and one volume of isopropanol. Pelleted RNA was washed twice with 70% ethanol and dried at room temperature for 20 min. Finally the pellet was resuspended in 40 μl of sterile distilled water (dH_2O) and stored at -80^{o}C.

3.3 Polymerase chain reaction (PCR)

PCR reactions were carried out according to Sambrook et al. (2000) with minor modifications using a Peltier Thermal Cycler-200 (MJ Research) and the Expand™ Long Template PCR system (Roche). Except where otherwise stated, reactions contained 300 μM dNTPs (Roche), 0.2 μM of each primer, 1x Expand™ buffer No. 1, 1 U of Expand™ DNA polymerase blend. All template nucleic acids were first denatured at 95°C for 5 min before cycling under the conditions described followed by a final extension at 68°C for 5 min.

3.4 Reverse transcription PCR (RT-PCR)

Prior to cDNA synthesis, total RNA was treated with RNase-free DNase (Promega) in a total volume of 20 μl following the manufacturer's instructions. Briefly, the reaction contained: ~20 μg of total RNA, 3 U of DNase (Promega), 1x DNase reaction buffer and 40 U of RNase inhibitor (Roche). The reaction was incubated at 37°C for 20 min and then purified. Subsequent cDNA synthesis involved addition of 1 μl of 50 μM oligo-dT primer to ~20 μg of treated total RNA in a total volume of 10 μl. This mixture was heated at 80°C for 5 min and then chilled on ice. After annealing, other reaction components were added in a total volume of 20 μl with the following final concentrations: 10 mM DTT, 1 mM dNTPs, 1 x RT buffer, 40 U RNase inhibitor (Roche) and 200 U of SUPERSCRIPT II polymerase (Invitrogen™). The mixture was incubated at 50°C for 90 min and then diluted to 100 μl with sterile dH_2O. The PCR reactions were performed in a total volume of 50 μl containing 300 μM of dNTPs, 0.2 μM of each primer, 5 μl of diluted cDNA, 1x PCR buffer and 1 U of Expand™ DNA polymerase blend (Roche). PCR conditions used were 95°C for 3 min, followed by 35 cycles of 95°C for 30 s, 50°C-55°C for 30 s and 72°C for 1 min; and additional 5 min extension at 72°C.

3.5 RNA Ligase Mediated-Rapid Amplification of cDNA ends (RLM-RACE)

RNA ligase-mediated rapid amplification of 5' and 3' cDNA ends (RLM-RACE) was carried out using the GeneRacer™ Kit (Invitrogen™) following

the directions of the manufacturer. Dephosphorylation of 15 μg of total RNA was carried in 1x CIP buffer, 40 U RNaseOut™, 10 U CIP and DEPC-treated water to a final volume of 10 μl, incubated at 50°C for 1 hr and then chilled on ice. After addition of 750 μl of TE, the reaction was extracted with $CHCl_3$:IAA (24:1) and the cDNA precipitated by addition of 0.1 volume of 3 M NaOAC pH 7.0 and 1 volume of isopropanol. DNA was pelleted by centrifugation and the pellet washed once with 70% ethanol. After air-drying, the pellet was resuspended in 7 μl of DEPC-treated water. The dephosphorylated RNA was decapped by addition of 1x TAP buffer, 40 U RNaseOut™ and 0.5 U TAP in a final volume of 10 μl and incubation at 37°C for 1 hr. After incubation, RNA was precipitated as described previously and resuspended in 7 μl of DEPC-treated water. The lyophilised GeneRacer™ RNA oligo (5'CGACUGGAGCACGAGGACACUGACAUGGACUGAAGGAG UAGAAA-3') (0.25 μg) was added to the decapped RNA, heated to 65°C for 5 min to relax RNA secondary structure and cooled on ice for 2 min. For ligation the reaction included 1x Ligase buffer, 1 mM ATP, 40 U RNaseOut™ and 5 U T4 RNA ligase in a final volume of 10 μl. Incubation was performed at 37°C for 1 hr. RNA was precipitated as described previously, resuspended in 10 μl DEPC-treated water and 1 μl of the GeneRacer™ Oligo dT primer (5'-GCTGTCAACGATACGCTACGTAACGGC ATGACAGTG$(T)_{18}$-3') (50 μM) plus 1 μl of dNTP mix (10 mM each) were added to the ligated RNA, incubated at 65°C for 5 min to remove RNA secondary structure and then cooled on ice. For the RT the reaction included, 1x First strand buffer, 10 mM DTT, 40 U RNaseOut™, 15 U Thermoscript™ RT and DEPC-treated water plus 12 μl ligated RNA/primer mixture in a final volume of 20 μl. Incubation was carried out at 50°C for 50 min and then cooled on ice. Two units of RNase H were added and the reaction mix incubated at 37°C for 20 min. After incubation, the reaction tube was stored at –20°C for later use. For the 3'RACE or 5'RACE steps, primary PCR reactions contained 1 μl of cDNA obtained from the previous step, 1x Expand™ buffer No. 1, 300 μM of dNTPs (Roche), 0.2 μM of the appropriate specific primer (see chapter 5), GeneRacer primers (Figure 3.5.1), 3.5 U of Expand™ Long Template polymerase (Roche) and dH_2O in a final volume of

50 μl. PCR conditions were 95°C for 3 min, followed by 35 cycles of 95°C for 30 s, 50-60°C for 30 s and 72°C for 1-3 min; and additional 5 min extension at 68°C was included. A secondary PCR reaction was conducted with 1 μl of the primary PCR using a nested gene-specific primer (see chapter 5) and the GeneRacer nested primer (Figure 3.5.1). The same reaction composition and cycle parameters were used as before.

GeneRacer 5' primer: 5'-CGACTGGAGCACGAGGACACTGA-3'
GeneRacer 5' nested primer: 5'-GGACACTGACATGGACTGAAGGAGTA-3'
GeneRacer 3' primer: 5'-GCTGTCAACGATACGCTACGTAACG-3'
GeneRacer 3' nested primer: 5'-CGCTACGTAACGGCATGACAGTG-3'

Figure 3.1 GeneRacer™ primers (Invitrogen™).

3.6 Genome walking by PCR

The protocol of Siebert et al. (1995) was used for walking along uncloned genomic DNA. Genomic DNA (5 μg) from *M. acuminata* ssp. *malaccensis* was digested in 100 μl reaction volumes with 80 U of PvuII or EcoR V restriction enzymes. The digested DNA was extracted once with phenol:$CHCl_3$:IAA (25:24:1) and once with chloroform, then precipitated by addition of 0.1 volume of 3 M sodium acetate pH 7.0 and 2 volumes of 95% ethanol. After mixing, the tubes were immediately centrifuged at 14,000 rpm for 10 min. The pellets were washed with 70% ethanol, centrifuged for 5 min, air dried and dissolved in 20 μl of TE buffer. Ten μl of DNA was then ligated to an excess of adaptor (Figure 3.6.1) overnight at 16°C under the following conditions: 1x T4 DNA ligase, 5 mM adaptor and 10 U T4 DNA ligase (New England Biolabs) in a total volume of 20 μl. The ligation reaction was terminated by incubation at 70°C for 5 min, then diluted 10-fold by addition of 180 μl of TE buffer and stored at –20°C. PCR amplifications were performed with the Expand™ polymerase (Roche). Primary PCR reactions were conducted in 50 μl volumes containing 1 μl of ligated, diluted DNA, 300 μM of dNTPs (Roche), 0.2 μM of specific primer (see chapter 5) and adaptor primer (AP1) (Fig. 3.6.1) respectively, 1x Expand™ buffer No. 1 and 3.5 U of

Expand^{TM} DNA polymerase blend (Roche). PCR conditions were 95^{o}C for 3 min, followed by 35 cycles of 95^{o}C for 30 s, 50-60^{o}C for 30 s and 72^{o}C for 1-3 min; and an additional 5 min extension at 72^{o}C. A secondary PCR reaction was conducted with 1 μl of a 50-fold dilution of the primary PCR using adaptor primer 2 (AP2) (Figure 3.6.1) and the nested gene-specific primer (see chapter 5). The same reaction composition and cycle parameters were used as before.

Adaptor

T7 promoter Not I

5'-CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT-3'

3'-H_2N-CCCGTCCA-PO_4-5'

Adaptor primer 1 (AP1)

5'-CCATCCTAATACGACTCACTATAGGGC-3'

Adaptor primer 2 (AP2)

5'-ACTCACTATAGGGCTCGAGCGGC-3'

Figure 3.2 Adaptor and primer sequences used in PCR genome walking (Siebert et al. 1995).

3.7 Gene cloning

3.7.1 Purification of PCR product.

The QIAquick gel extraction kit (QIAGEN) was used to extract and purify PCR products from agarose gels as per the manufacturer's instructions. In brief, the DNA fragment was excised from the agarose gel, weighed and three volumes of QG buffer was added per volume of gel (where 100 mg approximately equals 100 μl). Gel slices were incubated at 50^{o}C for 10 min with vortexing every 2-3 min. After the slice was completely dissolved, the sample was applied to a QIAquick spin column and centrifuged for 1 min. Flow-through was discarded and the column was washed with 0.75 ml of PE buffer. DNA was eluted upon addition of 30 μl of 10 mM Tris-HCl pH 7.5 and centrifugation for 3 min.

3.7.2 Ligation of PCR fragments in pGEM-T Easy.

For all ligation reactions PCR products were cloned into pGEMT-Easy (Promega) as per the manufacturer's instructions. In general, ligations were performed at 16^{o}C overnight in a 15 μl reaction volume containing 2 U of T4 DNA ligase (New England Biolabs), 1x ligation buffer, 10 μl aliquot of the purified PCR product and 50 ng of vector.

3.7.3 Preparation of competent cells.

The protocol of Inoue et al. (1990) was used to obtain XL1-blue *E. coli* ultra-competent cells for transformation. An aliquot of XL1-blue cells was streaked on an LB (appendix 1) plate. The plate was incubated at 37^{o}C overnight. One colony was picked next day and incubated in LB liquid medium at 37^{o}C overnight with shaking (225 rpm). An aliquot from the overnight culture was inoculated in 250 ml SOB media (see appendix 2) and shaken at 18^{o}C until OD_{600} = 0.6. After incubation, the culture was placed on ice for 10 min, then divided in four Falcon tubes of 50 ml volume and centrifuged at 2,500 g for 10 min at 4^{o}C. Pelleted cells were resuspended in 16 ml of ice-cold TB buffer (see appendix 2) and incubated on ice for 10 min. Cells were then centrifuged at 2,500 g for 10 min at 4^{o}C. Pellets were resuspended in a mix of 4 ml of ice cold TB plus 280 μl DMSO (Sigma), then incubated on ice for 10 min. After incubation, 100 μl and 200 μl aliquots were dispensed and frozen in liquid nitrogen, and immediately stored at –80 ^{o}C.

3.7.4 Transformation of competent cells.

100 μl of XL1-blue competent cells (cc) was transferred either to 2 ml sterile microcentrifuge tubes or Falcon tubes (No. 2059) previously chilled on ice. Ligation was added (5 μl), gently swirled to mix well and immediately incubated on ice for 30 min. After incubation, sample was given a heat pulse at 42^{o}C for 45 sec, then placed on ice for 2 min. Then, 600 μl of SOC media (appendix 1) was added and the tube was incubated at 37^{o}C for 1 hr with shaking at 225 rpm. After incubation, 100 μl of transformation sample (up to 200 μl) were plated on LB agar plates containing 100 μg/ml of ampicillin, 0.5

mM IPTG and 80 μg/ml of X-Gal. Plates were incubated overnight (16 hr) at 37^{o}C.

3.7.5 Plasmid purification.

For plasmid purification, a modified protocol of Sambrook et al. (2000) was used. Bacterial colonies were picked, transferred to 5 ml tubes containing LB liquid medium with 100 μg/ml ampicillin and incubated overnight (approximately 16 hours) with shaking at 225 rpm. After incubation, 1.5 ml of cell culture was centrifuged at 14,000 rpm for 1 min. Pelleted cells were resuspended in 500 μl of ice cold STE buffer (10 mM Tris HCl pH 8.0, 1 mM EDTA pH 8.0, 0.1 M NaCl) to remove culture media. Cells were pelleted by centrifugation, resuspended by vigorous vortexing in 100 μl of ice cold solution 1 (50 mM glucose, 25 mM Tris HCl pH 8.0, 10 mM EDTA pH 8.0) and incubated for 5 min on ice. Then the cells were lysed by addition of 200 μl of freshly prepared solution 2 (0.2 N NaOH, 1% SDS) and incubated for 5 min on ice. The solution was neutralised by addition of 150 μl of ice cold solution 3 (3 M sodium acetate pH 4.7) and incubated on ice for 5 min. After incubation, the lysate was centrifuged at 14,000 rpm for 5 min. The supernatant was transferred to 900 μl absolute ethanol, mixed and incubated at room temperature for 3 min. DNA was pelleted by centrifugation at 14,000 rpm for 5 min and the pellet washed with 70 % ethanol. The pellet was air dried and resuspended in sterile dH_2O (generally 30 μl) containing 1 μg/ml RNase A (Sigma).

3.8 Sequencing

The protocol of the Australian Genome Research Facility (AGRF, www.agrf.org.au) was used for sequencing. The sequencing reaction contained 2 μl of plasmid miniprep, 1 μl of 3.2 μM universal primer, forward (5'-CACGACGTTGTAAAACGAC-3' or reverse (5'-GAAACAGCTATGACCATG-3'), 1 μl of Big Dye terminator (BDT) version 3.1, 3.5 μl of BDT buffer, and sterile dH_2O to a final volume of 20 μl. The PCR amplification program consisted in one cycle of 95^{o}C for 3 min, then 35 cycles of 95^{o}C for 30 s, 50^{o}C for 30 s and 60^{o}C for 4 min. After PCR, DNA

samples were transferred to fresh 1.5 ml microfuge tubes and DNA precipitated by addition of 2 μl of 3 M sodium acetate pH 5.2 and 50 μl of 96% ethanol. DNA was pelleted by centrifugation at 14,000 rpm for 20 minutes, washed with 200 μl of 70 % ethanol and then air dried for approximately 10 minutes at room temperature. Samples were sent for gel separation to the AGRF at University of Queensland, Brisbane, Australia.

3.9 Southern blotting

3.9.1 Agarose gel electrophoresis

Unless stated otherwise, gels were prepared as 1% agarose in Tris-Acetate-EDTA (TAE) buffer containing 1 μg/mL ethidium bromide. Gels were run at 100V for approximately 1 h and visualised and photographed using a SYNGENE GelDock system.

3.9.2 Southern transfer of DNA

For Southern blotting genomic DNA samples were electrophoresed on 1.1% TAE gels at 55V for 3 h. DNA was depurinated by soaking in 0.25 M HCl for 10 min and denatured in 0.5 M NaOH, 1.5 M NaCl for 30 min. Then the gel was neutralised for 30 min in 0.5 M Tris-HCl, 3 M NaCl pH 7.5. DNA was transferred to pre-cut positively charged nylon membranes (Roche) overnight using the capillary method described in Sambrook et al. (2000). Following Southern transfer, DNA was fixed to the membranes by baking at 80^{o}C for 2 h.

3.9.3 Preparation of digoxigenin (DIG) labelled probes

Probes were PCR-labelled with DIG-dUTP (Roche) following the directions of the manufacturer. Labelling reactions contained 100 pg of plasmid DNA, 1 μl of 10 μM forward and reverse primer, 2 μl PCR DIG labelling mix (2 mM dATP, dCTP, dGTP, 1.5 mM dTTP and 0.5 mM DIG-11-dUTP labile), 2 μl of 10 X Expand^{TM} Buffer No. 1 (Roche), 1 U Expand^{TM} DNA polymerase blend (Roche) and sterile dH_2O in a final volume of 20 μl. PCR mix was denatured at 95^{o}C for 3 min followed by 30 cycles of 95^{o}C for 30 s, 50^{o}C for 30 s and

68°C for 1 min followed by 1 cycle of 68°C for 10 min. The PCR amplified product was electrophoresed in a 1% agarose gel and purified as described before. The purified product was resuspended in 50 µl of sterile dH_2O and stored at -20 °C.

3.9.4 DNA detection using digoxigenin

Membranes were prehybridised for 1 h at 42°C in DIG-Easy Hyb Solution (Roche). A DIG-labelled probe (25 µl) (see section 3.9.3) was denatured by boiling for 5 min and quenched on ice before adding to pre-warmed hybridisation solution and hybridising at 42°C overnight. The membranes were subjected to 2 x 5 min low stringency washes (2 x SSC, 0.1% SDS) at room temperature and 2 x 15 min high stringency washes (0.1 x SSC, 0.1% SDS) at 68°C, and then washed in wash solution (0.3% (v/v) Tween-20 in maleic acid buffer, pH 7.5). Membranes were blocked for 1 h in 1 x Blocking solution (Blocking reagent diluted in maleic acid buffer) (Roche) and then incubated for 30 min in an anti-DIG antibody/alkaline phosphatase conjugate (Roche) diluted to 1:10,000 in 1 x Bloking solution. Membranes were washed twice for 15 min in wash solution before equilibration in detection buffer (0.1 M Tris HCl, 0.1 M NaCl, pH 9.5). Detection with CDP-star (Roche) was carried out as per manufacturer´s instructions. Exposure times to X-ray film ranged from 30 s to 3 min.

3.10. Bioinformatic analysis

All sequences were assembled and edited using the Lasergene software package version 4.03 (DNASTAR, Madison, WI). Homology searches were performed using the the BLASTX and BLASTP algorithms (Altschul et al. 1997) through the National Center for Biotechnology Information (NCBI) GenBank database (www.ncbi.nlm.nih.gov). Percentages of identity and similarity between sequences were determined by the ALIGN program through the European Bioinformatic Institute server (www.ebi.ac.uk) and potential coiled-coil structures were predicted by the COILS program (Lupas 1996)(www.ch.embnet.org). Motif searches were performed by the PROSITE program (Hulo et al. 2006) and the hydropathy plot was performed according

to Kyte and Doolittle (1982) using the ExPASy proteomics server (www.us.expasy.org). Protein sequences were aligned with the CLUSTALX program version 1.81 (Thompson et al. 1997) with the default settings and the phylogenetic tree was constructed by the neighbour-joining (NJ) method (Saitou and Nei 1987) using the NJ algorithm implemented in the Molecular Evolutionary Genetics Analysis (MEGA) software package version 2.1 with the Poisson correction (Kumar et al. 2001). Bootstrapping was used to evaluate the degree of support for a particular grouping pattern in the phylogenetic tree. Putative cis-acting regulatory elements on the putative promoter region of RGC2 were predicted using the signal scan program (Higo et al, 1999) at the plant cis-acting regulatory DNA elements (PLACE) database (www.dna.affrc.go.jp/htdocs/PLACE).

3.11 References

Altschul, S. F., Madden, T.L., Schaffer, 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: 3389-3402.

Dellaporta, S.L., Wood, J. and Hicks, J. B. (1983). A plant DNA minipreparation: version II. *Plant Molecular Biology Reporters* 1: 19-21.

Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research* 27: 297-300.

Hulo, N., Bairoch A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P.S., Pagni, M., and Sigrist, C.J.A. (2006). The PROSITE database. *Nucleic Acids Research* 34: D227-D230.

Inoue, H., Nojima, H. and Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 96: 23-28.

Kumar, S., Tamura, K., Jakobsen, I.B., and Nei, M. (2001). MEGA2: Molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245.

Kyte, J. and R.F. Doolittle (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* 157: 105-132.

Lupas, A. (1996). Prediction and analysis of coiled-coil structures. *Methods in Enzymology* 266: 513-525.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425.

Sambrook, J. and Ruseell, D. (2001). Molecular cloning: A laboratory manual. 3rd edition. Cold Spring Harbor Laboratory. USA.

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A. and Lukyanov, S.A. (1995). An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Research 23*: 1087-1088.

Schuler, M. A. and Zielinski, R.E. (1989). RNA isolation from light- and dark-grown seedlings. In *Methods in Plant Molecular Biology*. Academic Press Inc. USA. pp. 89-96.

Smith, M.K., Hamill, S.D., Langdon, P.W. and Pegg, K.G. (1998). Selection of new banana varieties for the cool subtropics in Australia. Acta Horticulturae 490: 49-56.

Thompson, J.D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 24: 4876-4882.

3.12 Appendix 1

The composition of buffers and solutions for E. coli competent cell transformation

IPTG stock solution (0.1M)
1.2 g IPTG (Promega)
Add sterile dH_2O to 50 ml final volume. Sterilize by filtration and store at 4 ^oC.

X-Gal (2 ml)
100 mg 5-bromo-4-chloro-3-indolyl-β-D -galactoside (Promega)
Dissolve in 2 ml N,N'-dimethyl-formamide. Cover with aluminium foil and store at –20 ^oC.

LB media (1 L)
10 g Bacto-tryptone
5 g Bacto-yeast extract
5 g NaCl
Adjust pH to 7.0 with NaOH. Autoclave and store at 4 ^oC.

LB plates with ampicillin
Prepare 1 L of LB media and prior autoclaving add 15 g agar. Autoclave. Allow the media to cool to 50 ^oC before adding ampicillin to a final concentration of 100 μg/ml. Pour approximately 25 ml of media into 85 mm Petri dishes. Let the agar harden. Store at 4 ^oC for up to 1 month.

LB plates with ampicillin/IPTG/X-Gal
Make the LB plates with ampicillin as above; then supplement with 0.5 mM IPTG and 80 μg/ml X-Gal and pour the plates.

SOC media (100 ml)
2.0 g Bacto-tryptone
0.5 g Bacto-yeast extract

1 ml 1 M NaCl

0.25 ml 1M KCl

1 ml 2 M Mg^{2+} stock, filter sterilized (as prepared below)

1 ml 2 M glucose, filter sterilize

Add Bacto-tryptone, Bacto-yeast extract, NaCl and KCl to 97 ml dH_2O. Stir to dissolve. Autoclave and cool to room temperature. Add 2 M Mg^{2+} stock and 2 M glucose, each to a final concentration of 20 mM. Bring to 100 ml with sterile dH_2O. Filter the complete medium through a 0.2 μm filter unit. The final pH should be 7.0.

2M Mg^{2+} stock

20.33 g $MgCl_2 \cdot 6H_2O$

24.65 g MgSO4 · 7H2O

Add sterile dH_2O to 100 ml. Sterilize by filtration.

The Inoue method solutions (Inoue et al. 1990)

SOB media

2 % (w/v) bacto tryptone

0.5 % (w/v) yeast extract

10 mM NaCl

2.5 mM KCl

10 mM $MgCl_2$

10 mM $MgSO_4$

Adjust pH to 7.0. Sterilize by autoclaving.

TB solution

10 mM Pipes

55 mM $MnCl_2$

15 mM $CaCl_2$

250 mM KCl

Adjust pH to 6.7 with 5N KOH prior adding the $MnCl_2$. Sterilize by filtration.

Chapter 4

Structure, Phylogeny and Expression Analysis of Disease Resistance Gene Candidates of the Nucleotide Binding Site (NBS) Type from Banana (*Musa acuminata* ssp. *malaccensis*)

4.1 Abstract

The majority of known plant disease resistance (R) genes encode proteins with a nucleotide-binding site and leucine-rich repeat (NBS-LRR) structure. In this study, degenerate primers deduced from conserved motifs in the NBS domain of NBS-LRR resistance proteins were used to amplify genomic NBS-type sequences from the wild banana *Musa acuminata* ssp. *malaccensis*. At least five different classes of NBS sequences were identified and named as resistance gene candidates (RGCs). Deduced amino acid sequences of the banana RGCs showed the typical motifs (P-loop/kinase-1a, kinase-2, RNBS-B and GLPLA) present in the majority of known plant NBS-LRR resistance genes. Structural and phylogenetic analyses grouped the banana RGCs within the non-TIR (homology to Toll/interleukin-1 receptors) subclass of NBS sequences. The copy number of each class was estimated by Southern hybridisation and each RGC was found to be in low copy number. The expression of the RGCs was assessed by RT-PCR in leaf and root tissues of plants resistant or susceptible to *Fusarium oxysporum* f. sp. *cubense* (FOC) race 4. Four classes showed a constitutive expression profile whereas no expression was detected for one class in either roots or leaves. Interestingly, a transcriptional polymorphism was found for one of the RGCs, whose expression was associated with FOC race 4 resistance. The banana RGC sequences isolated in this study represent a valuable source of information that could be used to assist the cloning of functional R genes in this crop.

4.2 Introduction

Banana is one of the most important fruit crops in the world, with an annual production of more than 100 million metric tons (FAO 2005). This fruit represents both a staple food for millions of people in developing countries and an important export commodity for numerous agricultural-based economies. Several diseases threaten world banana production causing significant yield losses every year and, among these, fungal diseases are the major concern. The foliar fungal disease black Sigatoka caused by the air-borne *Mycosphaerella fijiensis* is considered the most economically important leaf disease of banana (Pasberg-Gauhl et al. 2000). For example, in Southern Mexico and throughout Central America has been estimated that the cost of fungicide sprays to control *M. fijiensis* accounts for 27-30% of productions costs (Stover and Simmonds 1987; Agrios 1997). The intensive fungicide applications are not only expensive for many farmers but also represent a serious health risk to plantation workers and threatens the environment. Another devastating fungal disease is Panama disease caused by the soil-borne *Fusarium oxysporum formae specialis* (f. sp.) *cubense* (FOC), which causes a lethal vascular wilt (Ploetz and Pegg 2000). FOC race 4 is a major concern to banana production since an effective chemical control does not exist. Therefore, the development of FOC race 4 resistance in banana cultivars is a matter of urgency. Sources of natural resistance for these two fungal diseases exist in wild banana species, however, resistance has been difficult to introgress into edible cultivars by conventional breeding techniques due to sterility problems (Roux et al. 2004). Genetic transformation could overcome this obstacle (Becker et al. 2000; Khanna et al. 2004) by introducing fungal resistance genes into susceptible cultivars. However, no resistance (R) gene capable of conferring resistance to these fungal diseases has been reported.

Most R genes identified to date are members of the cytoplasmic nucleotide-binding site and leucine-rich repeat (NBS-LRR) class. This class of R genes confers resistance to a wide variety of pathogens and pests including viruses, bacteria, fungi, nematodes and insects (Dangl and Jones 2001). The

NBS-LRR proteins are thought to recognize pathogens and respond by activating signal transduction pathways leading to disease resistance (Belkhadir et al. 2004). The C-terminal LRR has been considered the candidate pathogen recognition domain and the N-terminal region, including the NBS, the major signalling domain (Belkhadir et al. 2004). The LRR domain is the most variable region in closely related NBS-LRR proteins and is under diversifying selection (Michelmore and Meyers 1998; Richter and Ronald 2000). NBS-LRR genes are abundant in plant genomes with 149 found in *Arabidopsis* and 480 found in rice (Meyers et al. 2003; Zhou et al. 2004) that are often organized in clusters (Hulbert et al. 2001). The NBS-LRR genes present in a given cluster can confer resistance to different strains of the same pathogen or to diverse pathogens (Van der Vossen et al. 2000). This clustered genomic organization may provide a reservoir of genetic variation from which new specificities can evolve (Michelmore and Meyers 1998). The NBS-LRR class of R genes is divided into two distinct subclasses based on the presence or absence of an N-terminal domain with homology to the *Drosophila* Toll and human Interleukin-1 receptors (TIR) (Meyers et al. 1999; Pan et al. 2000), the TIR and non-TIR subclasses, respectively. The non-TIR subclass commonly has a predicted coiled-coil structure, sometimes in the form of a leucine zipper (Meyers et al. 1999; Pan et al. 2000; Hulbert et al. 2001). The non-TIR subclass is widely distributed in both monocotyledonous and dicotyledonous species, whereas the TIR subclass appears to be restricted to dicotyledonous species (Meyers et al. 1999; Pan et al. 2000; Cannon et al. 2002). The TIR and non-TIR subclasses of NBS-LRR genes can also be distinguished by the motifs found within the NBS domain or by a single amino acid residue in the final portion of the NBS kinase-2 motif, which in most cases is an aspartic acid for the TIR subclass and a tryptophan for the non-TIR subclass (Meyers et al. 1999).

The well conserved motifs of the NBS domain have been used to isolate NBS-type sequences using a PCR-based strategy with degenerate primers. Sequences obtained with this approach, called resistance gene candidates (RGCs), have been isolated from soybean (Kanazin et al. 1996; Yu et al. 1996; Graham et al. 2000), potato (Leister et al. 1996), lettuce (Shen et al.

1998), rice and barley (Leister et al. 1998), wheat (Seah et al. 2000), sunflower (Ayele-Gedil et al. 2001), common bean (Rivkin et al. 1999; López et al. 2003), strawberry (Martínez-Zamora et al. 2004), apple (Calenge et al. 2005) and other plant species. Significantly, numerous RGCs are arranged in clusters in the same way as the R genes and show close genetic linkage with known R gene loci. Consequently, degenerate PCR represents an attractive approach that may facilitate the isolation of resistance genes from banana.

The aim of this study was to isolate and characterize RGCs of the NBS-type from the wild banana *Musa acuminata* ssp. *malaccensis*, which is resistant to black Sigatoka and Panama disease.

4.3 Materials and methods

4.3.1 Plant material

Musa acuminata ssp*. malaccensis* (genotype AA) plantlets which were resistant (accessions 850 and 852) or susceptible (accessions 845 and 846) to FOC subtropical race 4 (Smith et al. 1998) were kindly provided by Dr. Mike Smith (Queensland Department of Primary Industries, Nambour, Australia). They were grown in pots in an incubator chamber and used as the source for the harvested tissues, leaves and roots of 4-month-old plants. Harvested tissues were frozen in liquid nitrogen and stored at –80^{o}C until DNA or RNA extraction.

4.3.2 Degenerate PCR and cloning

DNA was extracted from leaf tissue (line 850) as described by Dellaporta et al. (1983). A pair of degenerate primers designed by Kanazin et al. (1996) [forward 5'-GGiGGiGTiGGiAAiACiAC-3', reverse 5'-A(A/G)iGCTA(A/G)iGGiA(A/G)iCC] was used to amplify the NBS of R genes from the GGVGKTT kinase motif to the GLPLA motif. The degenerate PCR reaction was performed in a total volume of 50 μl containing 300 μM of dNTPs, 2 μM of each degenerate primer, approximately 200 ng of DNA, 1x PCR buffer and 1 U of Taq DNA polymerase (Roche). PCR conditions were

95°C for 3 min, followed by 35 cycles of 95°C for 30 s, 45°C for 30 s and 72°C for 1 min; and an additional 10 min extension at 72°C was included. Purified PCR products were cloned into pGEM-T easy vector (Promega) and transformed by heat-shock in *Escherichia coli* JM109 competent cells following Promega instructions. Plasmids were extracted using the High Pure Plasmid Isolation Kit (Roche).

4.3.3 Sequence analysis

All sequence reactions were carried out with the BigDye terminator sequencing kit version 3.1 and separated on an ABI 3730 automatic sequencer (Applied Biosystems) through the capillary separation service of the Australian Genome Research Facility (www.agrf.org.au). All sequences were assembled and edited using the Lasergene software package version 4.03 (DNASTAR, Madison, WI). Homology searches were performed with the BLASTX program (Altschul et al. 1997) through the National Center for Biotechnology Information GenBank database (www.ncbi.nlm.nih.gov). Percentages of identity between sequences were determined by the ALIGN program through the European Bioinformatic Institute server (www.ebi.ac.uk) and potential coiled-coil structures were predicted by the COILS program (Lupas 1996). Finally, the similarity plot was determined by PLOTSIMILARITY program using the EGCG extension of the Wisconsin package (ver. 8.1.0) through the Australian National Genomic Information Service (www.angis.org).

4.3.4 RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE)

RNA was extracted from leaf tissue (line 850), as described by Schuler and Zielinski (1989). Isolation of 5' cDNA ends of RGC1, RGC2, RGC3 and RGC5 was carried out using the GeneRacer™ Kit (Invitrogen™ Life Technologies) according to the manufacturer's protocol. Briefly, the RNA was treated with alkaline phosphatase and tobacco acid pyrophosphatase (TAP), then the GeneRacer™ RNA oligo was ligated to the 5' RNA end using RNA ligase followed by the synthesis of cDNA with reverse transcriptase. The 5' RACE was performed with the reverse specific primers GSP1 (5'-CAAGTCTTGTCGAATCGAAC-3') for RGC1, GSP2 (5'-

ACGCTTCCCCGTCCTCCGGCATC-3') for RGC2, GSP3 (5'-ACCCGCGATTACCATGTGG-3') for RGC3, and GSP5 (5-TCCACCTTGGTAGCAGACTC-3') for RGC5 in combination with the GeneRacer™ 5' primer, respectively. PCR products were subject to a second round of PCR, using the nested specific primers nGSP1 (5'-CTTCGCATCGAATGTTCGATTCG-3') for RGC1, nGSP2 (5'-ACACACACCCACATTCTCAATGG-3') for RGC2, nGSP3 (5'-CTGCATATCGACCACGTTGAGCG-3') for RGC3, and nGSP5 (5-TCTTTGGTCAGTCTCTTCAC-3') for RGC5 in combination with the GeneRacer™ nested 5' primer. The PCR products were cloned into pGEM-T Easy vector (Promega) and sequenced.

4.3.5 Genome walking by PCR

The sequence of the 5' genomic region of RGC4 and the sequence of the 3' genomic region (corresponding to the GLPLA motif) of all banana RGCs were determined by genome walking (Siebert et al. 1995). Briefly, genomic DNA from *M. acuminata* ssp. *malaccensis* was digested with PvuII and EcoR V blunt-cutting restriction enzymes. GenomeWalker adapters were ligated and PCR reactions were carried out on each genomewalker "library" using the adapter primer (AP1) and a reverse gene specific primer (5'-GTTGGACTTCATGGATGTG-3') for RGC4, and forward specific primers for RGC1 (5'-CAAGTCTTGTCGAATCGAAC-3'), RGC2 (5'-AGCCTGTTAGCCCCATTAGATGC-3'), RGC3 (5'-ACCCGCGATTACCATGTGG-3'), RGC4 (5'-GCCGTGTCACAATCTTACAAGG-3') and RGC5 (5'-CTGCTACCAAGGTGGAACAATC-3'). PCR products were subject to a second round of PCR, using the nested adapter primer (AP2) and nested reverse gene specific primer for RGC4 (5'-TACCTCTTGTCCTGGAGATGG-3'), and nested forward gene specific primers for RGC1 (5'-AGCTTCGACATCAGAAGAGAGGC-3'), RGC2 (5'-GTCACCGGCACGATGGAGCCATAC-3'), RGC3 (5'-CCTCGCGGTCGAACCAGGTGTGC-3'), RGC4 (5'-TCATAGGCCATCTCCAGGACAAGAG-3'), and RGC5 (5'-ACTCGAGACACAAAGATTGCCAGC-3'). The PCR products were cloned into pGEM-T Easy vector (Promega) and sequenced.

4.3.6 Protein sequence alignment and phylogenetic analysis

Protein sequences were aligned using the ClustalX program version 1.81 (Thompson et al. 1997) with the default settings. Phylogenetic tree was constructed by the neighbour-joining method (Saitou and Nei, 1987) implemented in the Molecular Evolutionary Genetics Analysis (MEGA) software version 2.1 (Kumar et al. 2001) with the Poisson correction. Bootstrapping (1000 replicates) was used to evaluate the degree of support for particular grouping patterns in the phylogenetic tree. The NBS sequence of the TIR-NBS-LRR R proteins N (A54810), L6 (T185460), RPP5 (AAF08790), P2 (AAK28806) and the non-TIR-NBS-LRR proteins I2

(AAD27815), Mi-1.2 (AAC67238), Pi-ta (AAK00132), Bs2 (AAF09256), Rx (CAB50786) were used in BLASTP searches to retrieve RGCs for the phylogenetic tree construction.

4.3.7 Southern hybridisation

Genomic DNA (5 μg) was digested independently with EcoR I, EcoR V, Hind III, Bgl II, and Sac I. Digests were electrophoresed on a 1.2% agarose gel, capillary-blotted onto a nylon membrane (Roche) and baked for 2 h at 80^{o}C. Prior to hybridisation, the membrane was blocked for 60 min at 42^{o}C with DIG Easy Hyb (Roche). DIG-labelled probes were PCR-amplified for each RGC using a mixture of DIG-labelled and standard dNTPs (1:3 ratio). The membrane was hybridised with DIG-labelled probes for at least 12 h at 42^{o}C followed by two washes at room temperature (10 min) in 2X SSC/0.1% SDS and two washes at 65^{o}C (15 min) in 0.1X SSC/0.1% SDS. Detection of the hybridised probe using CDP-STAR (Roche) was carried out according to the manufacturer's instructions.

4.3.8 RT-PCR

Total RNA of leaf or root tissues was extracted from line 850 and 845, respectively according to Schuler and Zielinski (1989). Prior to cDNA synthesis, total RNA was treated with RNase-free DNase (Promega) following the manufacturer's instructions. Synthesis of cDNA was carried out with SUPERSCRIPT II (Invitrogen) according to the manufacturer's protocol. Briefly, 1 μl of 50 μM oligo-dT primer was added to total RNA (20 μg) treated with RNase-free DNase (Promega) in a total volume of 10 μl. This mixture was heated at 80^{o}C for 5 min and then chilled on ice. After annealing, other reaction components were added in a total volume of 20 μl with the following final concentrations: 10 mM DTT, 1 mM dNTP, 1x RT buffer, 40 U RNase inhibitor (Roche) and 200 U of SUPERSCRIPT II (Invitrogen). The mixture was incubated at 42^{o}C for 90 min and then diluted to 1:20 with sterile water. PCR reactions were performed in a total volume of 50 μl containing 300 μM of dNTP, 0.2 μM of each primer forward and reverse, 5 μl of diluted cDNA, 1X PCR buffer and 1 U of Taq polymerase (Roche). The following forward

and reverse primers, respectively, were used: for RGC1, 5'-CAAGTCTTGTCGAATCGAAC-3' and 5'-TCGTCGGCATGCCAGAATAC-3'; for RGC2, 5'-CCATTGAGAATGTGGGTGTG-3' and 5'-ACTCCTCGAGAACGTATGG-3'; for RGC3, 5'-ACCCGCGATTACCATGTGG-3' and 5'-GCGCTTCTTCTCATGTCGC-3'; for RGC4, 5'-GCCGTGTCACAATCTTACAAGG-3' and 5'-GTTGGACTTCATGGATGTG-3'; for RGC5, 5'-CTGCTACCAAGGTGGAACAATC-3' and 5'-GCACAATTCTTGAACAGCTCC-3'. PCR conditions were 95^{o}C for 3 min, followed by 35 cycles of 95^{o}C for 30 s, 50^{o}C-55 ^{o}C for 30 s and 72^{o}C for 1 min; and additional 10 min extension at 72^{o}C was included. Twenty microliters of the PCR reaction were separated on a 1.2% TAE agarose gel stained with ethidium bromide. The PCR products were cloned into pGEM-T easy vector (Promega) and sequenced for verification. In RT-PCR experiments, primers specific for the banana *Actin 1* gene (Hermann et al. 2001), and spanning an intron (~100 bp) were included as a control to detect any genomic DNA contamination in the RNA samples. The *Actin 1* forward primer 5'-GATGCCCGGAGGTTCTCTTCC-3' was anchored in exon 3 while the reverse primer 5'AGTACAGGTACAACTCGAGC-3 was anchored in the 3' untranslated region (UTR).

4.4 Results

4.4.1 Amplification, cloning and sequence analysis of resistance gene candidates of the NBS-type in banana

Degenerate primers designed to amplify the NBS sequence (the region between the P-loop and GLPLA motifs) of the NBS-LRR class of R genes are predicted to amplify DNA fragments of around 530 bp due to the absence of introns in this region (Aarts et al. 1998). In agreement with this observation, a single PCR fragment of approximately 520 bp from banana genomic DNA was amplified, cloned and a total of 88 clones were sequenced. BLASTX searches revealed that all the clones were related to RGCs of the NBS-type. A 75% identity threshold value was used to determine those sequences that belong to the same class (Bai et al. 2002). A total of three distinct classes of RGCs of the NBS-type were identified. Each one of these classes contained redundant or highly similar clones (>97% nucleotide identity). A representative clone of each class, designated RGC1, RGC2 and RGC3, was chosen for further analysis. The class represented by RGC1 was the most abundant with 77 clones, whereas the classes represented to a lesser degree were RGC2 and RGC3 with four and seven clones, respectively. Two other classes of RGCs were previously isolated from *M. acuminata* ssp. *burmannicoides* in our laboratory also using the degenerate primers of Kanazin et al. (1996) (Taylor 2005). Based on the sequence of the *M. acuminata* ssp. *burmannicoides* RGCs, specific primers were used to isolate two more classes of RGCs from the genome of *M. acuminata* ssp. *malaccensis* (designated RGC4 and RGC5). In total, five different classes of NBS sequences that presented uninterrupted ORFs were identified in the genome of this *M. acuminata* subspecies. The 5' ends of RGC1, RGC2, RGC3 and RGC5 were isolated from leaf tissue by 5'RACE, while the 5' end of RGC4 and the sequences corresponding to the GLPLA motif of all banana RGCs were isolated using PCR genome walking. The nucleotide sequences and the conceptual translations of each class of RGC isolated from *M. acuminata* ssp. *malaccensis* are presented in Figures 4.1 to 4.5. Three of the banana RGCs (RGC2, RGC3 and RGC5) showed a

potential coiled-coil (CC) structure in the non-TIR domain, whereas RGC1 and RGC4 did not show this predicted structure.

```
   1 atggagtcttttctcatcctcgttgccgaaaagattgccgtggccatggccggcgaagct
   1 M  E  S  F  L  I  L  V  A  E  K  I  A  V  A  M  A  G  E  A

  61 atacaggcagctatgggcttcaatttaggagccgaagaatcgctgaagacggaagttaag
  21 I  Q  A  A  M  G  F  N  L  G  A  E  E  S  L  K  T  E  V  K

 121 gagacgatcagacggatcagaagcgagttcgagcacatgcaaatatttttaagctccgtg
  41 E  T  I  R  R  I  R  S  E  F  E  H  M  Q  I  F  L  S  S  V

 181 gacatgcagaagtataacaccaccattgagccatggctgaaacgagcgagggagatagca
  61 D  M  Q  K  Y  N  T  T  I  E  P  W  L  K  R  A  R  E  I  A

 241 gattccatggaagacgtgatcgacgagtacttgcatattaccgtagagcggtcacagggt
  81 D  S  M  E  D  V  I  D  E  Y  L  H  I  T  V  E  R  S  Q  G

 301 ggactcagatccttttttaatcaagctgtgagaagtcacaaaaagagtagcgcctggaat
 101 G  L  R  S  F  F  N  Q  A  V  R  S  H  K  K  S  S  A  W  N

 361 ctcatagctaatcggctgaaaagttatagaagctggcctatccatctcgaagccatgaag
 121 L  I  A  N  R  L  K  S  Y  R  S  W  P  I  H  L  E  A  M  K

 421 gatcgctatgacatcaggaagaatgagtccgaagtagatgatgatgacgccgaaggcgag
 141 D  R  Y  D  I  R  K  N  E  S  E  V  D  D  D  D  A  E  G  E

 481 aatgcaaacggccttgtcggaagagtgttcaattcgtcgagatcaaaccctgtcagggaa
 161 N  A  N  G  L  V  G  R  V  F  N  S  S  R  S  N  P  V  R  E

 541 gaagacgacaatatttacagagaacaaaggaaaattttgtttcagctgctaacagatgaa
 181 E  D  D  N  I  Y  R  E  Q  R  K  I  L  F  Q  L  L  T  D  E

 601 acgtctacacgcacggtgatatcggtttggggcatgggggggtgtaggtaagaccaccatg
 201 T  S  T  R  T  V  I  S  V  W  G  M  G  G  V  G  K  T  T  M

 661 gttgacaaagtttacgggaaccaggagatcgagaatcgcttcgactgcaaaatctgggtc
 221 V  D  K  V  Y  G  N  Q  E  I  E  N  R  F  D  C  K  I  W  V

 721 accgtttccaagtcttgtcgaatcgaacattcgatgcgaagaattctcaaggaactgctg
 241 T  V  S  K  S  C  R  I  E  H  S  M  R  R  I  L  K  E  L  L

 781 gacgcagatcaatcggatcatgatagtaatgggtcgtcggaccttaatcgtttacaggag
 261 D  A  D  Q  S  D  H  D  S  N  G  S  S  D  L  N  R  L  Q  E

 841 gacgtttgcagcattctacaggagaagaggtacttgctgattctcgatgatgtgtggagc
 281 D  V  C  S  I  L  Q  E  K  R  Y  L  L  I  L  D  D  V  W  S

 901 ggagagttgtcttcctatgtgcaacgtgctcttcccgataacaatcgtggaagcagaata
 301 G  E  L  S  S  Y  V  Q  R  A  L  P  D  N  N  R  G  S  R  I

 961 gtgatcacgacacggctaaacgaggtagcttcgacatcagaagagaggcaccggttgaag
 321 V  I  T  T  R  L  N  E  V  A  S  T  S  E  E  R  H  R  L  K

1021 cttcggaaaattgaagatgaaggccaagcgttcgatctgttctgtcgagaggtattctgg
 341 L  R  K  I  E  D  E  G  Q  A  F  D  L  F  C  R  E  V  F  W

1081 catgccgacgacaggcgttgccccaaacacttggagacggtggggagaaatattgtcagg
 361 H  A  D  D  R  R  C  P  K  H  L  E  T  V  G  R  N  I  V  R

1141 aagtgccaaggcctgccactggcc  1164
 381 K  C  Q  G  L  P  L  A    388
```

Figure 4.1 Nucleotide sequence and conceptual translation of the N-terminal region of banana RGC1. The NBS motifs defined by Meyers et al. (1999) are underlined.

```
   1 atggctggtgtcacatcacaggcagcggcggtgttctccctggtgaatgaaatctttaac
   1 M  A  G  V  T  S  Q  A  A  A  V  F  S  L  V  N  E  I  F  N

  61 cggtccatcaatttgatcgtcgcggaactccggttgcagttgaatgcgagagccgagctg
  21 R  S  I  N  L  I  V  A  E  L  R  L  Q  L  N  A  R  A  E  L

 121 aacaatctgcagagaacactattgaggactcactctctgctcgaggaggcaaaggcgagg
  41 N  N  L  Q  R  T  L  L  R  T  H  S  L  L  E  E  A  K  A  R

 181 cggatgactgacaagtctctcgtgctgtggctgatggagctcaaggaatgggcctacgac
  61 R  M  T  D  K  S  L  V  L  W  L  M  E  L  K  E  W  A  Y  D

 241 gccgacgacatcctcgacgagtacgaggccgcagcaatccgactgaaggtaacacgctcg
  81 A  D  D  I  L  D  E  Y  E  A  A  A  I  R  L  K  V  T  R  S

 301 accttcaaacgtcttatcgatcatgtgattataaatgttccattagcgcacaaagtagca
 101 T  F  K  R  L  I  D  H  V  I  I  N  V  P  L  A  H  K  V  A

 361 gacatcaggaaaaggttgaacggggtcactcttgagagggagctaaatctgggtgcgctg
 121 D  I  R  K  R  L  N  G  V  T  L  E  R  E  L  N  L  G  A  L

 421 gaagggtcgcagccgcttgattccacgaaaagaggtgtgaccacttctcttctgactgaa
 141 E  G  S  Q  P  L  D  S  T  K  R  G  V  T  T  S  L  L  T  E

 481 tcttgtattgtcgggcgagctcaagataaggagaatttgattcggttgctgttggagccc
 161 S  C  I  V  G  R  A  Q  D  K  E  N  L  I  R  L  L  L  E  P

 541 agcgatggggcggttcctgttgttcctatagttggattaggaggggcagggaagacgact
 181 S  D  G  A  V  P  V  V  P  I  V  G  L  G  G  A  G  K  T  T

 601 ctgtctcagcttatctttaatgacaagagagtggaggagcatttcccattgagaatgtgg
 201 L  S  Q  L  I  F  N  D  K  R  V  E  E  H  F  P  L  R  M  W

 661 gtgtgtgtgtctgacgattttgatgtgaagagaattactagagagatcacagagtacgcc
 221 V  C  V  S  D  D  F  D  V  K  R  I  T  R  E  I  T  E  Y  A

 721 accaacggaaggttcatggatctcaccaacttgaatatgcttcaagttaatctgaaagag
 241 T  N  G  R  F  M  D  L  T  N  L  N  M  L  Q  V  N  L  K  E

 781 gagataaggggacgacatttttgcttgtgctggatgatgtgtggaacgaagaccccgtg
 261 E  I  R  G  T  T  F  L  L  V  L  D  D  V  W  N  E  D  P  V

 841 aagtgggaaagcctgttagccccattagatgccggaggacggggaagcgtggtcattgtg
 281 K  W  E  S  L  L  A  P  L  D  A  G  G  R  G  S  V  V  I  V

 901 acgacacagagcaaaaaggtcgccgatgtcaccggcacgatggagccatacgttctcgag
 301 T  T  Q  S  K  K  V  A  D  V  T  G  T  M  E  P  Y  V  L  E

 961 gagttaacggaggatgacagttggtcactcatcgagagtcactccttcagggaggcgagc
 321 E  L  T  E  D  D  S  W  S  L  I  E  S  H  S  F  R  E  A  S

1021 tgctctagtacaaatcctagaatggaagagatcgggaggaagatagccaagaagatcagt
 341 C  S  S  T  N  P  R  M  E  E  I  G  R  K  I  A  K  K  I  S

1081 ggcctaccttacgga  1095
 361 G  L  P  Y  G     365
```

Figure 4.2 Nucleotide sequence and conceptual translation of the N-terminal region of banana RGC2. The NBS motifs defined by Meyers et al. (1999) are underlined.

```
   1 atgtgcgatctcgtctcccttgcatgccaagcctcacaacctttatgcacagcctgcctg
   1 M  C  D  L  V  S  L  A  C  Q  A  S  Q  P  L  C  T  A  C  L

  61 attcctgtacatgatgagattaaggaaactttgaccgcgtgctttcaactccgccggaac
  21 I  P  V  H  D  E  I  K  E  T  L  T  A  C  F  Q  L  R  R  N

 121 cggagctctctcacggaagcgctaagcgacctacgggccaccgcacagaaagtgaaggac
  41 R  S  S  L  T  E  A  L  S  D  L  R  A  T  A  Q  K  V  K  D

 181 aaggtcgaggaagaggaggctcaccagcggatctgcaatcctgatgtcagacggtggcag
  61 K  V  E  E  E  E  A  H  Q  R  I  C  N  P  D  V  R  R  W  Q

 241 aagaaggtcgaggagatactccgggaatgcgacgccgaccaggagcacgaggaaccaaag
  81 K  K  V  E  E  I  L  R  E  C  D  A  D  Q  E  H  E  E  P  K

 301 agatgcgcctgcctgtgtggctgcgacatggatctgctccaccgtcgccgagtcgccagg
 101 R  C  A  C  L  C  G  C  D  M  D  L  L  H  R  R  R  V  A  R

 361 aaagtcgtccagaatctgcaggacgtgaacaagctgaagtcagatggcgatgcattcact
 121 K  V  V  Q  N  L  Q  D  V  N  K  L  K  S  D  G  D  A  F  T

 421 ccccccttcacccacgagccgccaccggagccggtggaggaactgccgtttgaaacgcag
 141 P  P  F  T  H  E  P  P  P  E  P  V  E  E  L  P  F  E  T  Q

 481 accatcgggatggagtcggccctaagccagctcctatcccggtttgacgacgcggagaag
 161 T  I  G  M  E  S  A  L  S  Q  L  L  S  R  F  D  D  A  E  K

 541 agcatcatcggcgtccacgggctaggggcatgggcaagacgacgctcctcaaaacgctc
 181 S  I  I  G  V  H  G  L  G  G  M  G  K  T  T  L  L  K  T  L

 601 aacaacgagctcaaggagaatacccgcgattaccatgtggtgatcatgatcgaggttgcc
 201 N  N  E  L  K  E  N  T  R  D  Y  H  V  V  I  M  I  E  V  A

 661 aactccgagacgctcaacgtggtcgatatgcagaagatcatcgccaatcggctgggtctg
 221 N  S  E  T  L  N  V  V  D  M  Q  K  I  I  A  N  R  L  G  L

 721 ccgtggaacgagagcgagacggagagggagcgatccacatttctgcgcagggccctgagg
 241 P  W  N  E  S  E  T  E  R  E  R  S  T  F  L  R  R  A  L  R

 781 aggaagaagttcgttgtcctgctcgacgacgtctggaaaaagttccagttggcggacgtg
 261 R  K  K  F  V  V  L  L  D  D  V  W  K  K  F  Q  L  A  D  V

 841 ggaatccccacgccaagctccgacaacgggtggaagctgatcctcgcctcgcggtcgaac
 281 G  I  P  T  P  S  S  D  N  G  W  K  L  I  L  A  S  R  S  N

 901 caggtgtgcgtcgagatgggcgacaaggagcccatggagatgccctgcttgggcgacaat
 301 Q  V  C  V  E  M  G  D  K  E  P  M  E  M  P  C  L  G  D  N

 961 gaatcgctgaggttgttccggagcaacttgatggccgaggtcagtgccgccatcgaccat
 321 E  S  L  R  L  F  R  S  N  L  M  A  E  V  S  A  A  I  D  H

1021 gacagcgacatgagaagaagcgccatggatatcatacagagctgcggcggccttccacta
 341 D  S  D  M  R  R  S  A  M  D  I  I  Q  S  C  G  G  L  P  L

1081 gca   1083
 361 A     361
```

Figure 4.3 Nucleotide sequence and conceptual translation of the N-terminal region of banana RGC3. The NBS motifs defined by Meyers et al. (1999) are underlined.

```
   1 atgggcggggatgagctccctgggtggctaatggatgcgaagcagcctcagcttcaggtt
   1 M  G  G  D  E  L  P  G  W  L  M  D  A  K  Q  P  Q  L  Q  V

  61 agggtagttgccccgatgacattgcaggtgcgactcttgccagagaaatccaccacagc
  21 R  V  V  A  P  D  D  I  A  G  A  T  L  A  R  E  I  H  H  S

 121 ttggctattcccggcggccatttccgagcccacgctatggtgacggcgtcagagtcgcac
  41 L  A  I  P  G  G  H  F  R  A  H  A  M  V  T  A  S  E  S  H

 181 gacacggaggagcttctccgaaccatgattcgacagctgtccttcagtggtgagatgatt
  61 D  T  E  E  L  L  R  T  M  I  R  Q  L  S  F  S  G  E  M  I

 241 ccgcaggttctcggggatctggcgttgcattcggataaaacgggagtggagcaatgtctg
  81 P  Q  V  L  G  D  L  A  L  H  S  D  K  T  G  V  E  Q  C  L

 301 gataacctgggagaggtggatctagtgacgacgatcgtaaactatcttcaagataagagg
 101 D  N  L  G  E  V  D  L  V  T  T  I  V  N  Y  L  Q  D  K  R

 361 tatttagttgtcctagatgatataccgatcaattctgcctgggactgcctcaaagatgca
 121 Y  L  V  V  L  D  D  I  P  I  N  S  A  W  D  C  L  K  D  A

 421 ttacccgataaaaggaatgggagtaggatcataatgataaccgctgatgaggcggtggcc
 141 L  P  D  K  R  N  G  S  R  I  I  M  I  T  A  D  E  A  V  A

 481 ggcgcttggttttcccataactatcgttctgtgtcggaggaaggtgggctcgtcggtatc
 161 G  A  W  F  S  H  N  Y  R  S  V  S  E  E  G  G  L  V  G  I

 541 aagccgcagagggacgatctcatcaaacggattacgaaggagggccaggaccagtttggt
 181 K  P  Q  R  D  D  L  I  K  R  I  T  K  E  G  Q  D  Q  F  G

 601 gtaattgcaatcataggtttcggtggcctgggcaagacgactctagccatgcaagtcttc
 201 V  I  A  I  I  G  F  G  G  L  G  K  T  T  L  A  M  Q  V  F

 661 gagagcctgaaggtaaccggtagccactttcatgcctacgcttggattgccgtgtcacaa
 221 E  S  L  K  V  T  G  S  H  F  H  A  Y  A  W  I  A  V  S  Q

 721 tcttacaaggtggaggtgcttctgcgaagcatcattcgacaactctccatcagcgtgcag
 241 S  Y  K  V  E  V  L  L  R  S  I  I  R  Q  L  S  I  S  V  Q

 781 cagattcaacatgtcctacaactttctgcttcgaatcaagatatagaggtcgtggagcaa
 261 Q  I  Q  H  V  L  Q  L  S  A  S  N  Q  D  I  E  V  V  E  Q

 841 cttctagataagatgcgagaggaagatctgagaaggacgatcataggccatctccaggac
 281 L  L  D  K  M  R  E  E  D  L  R  R  T  I  I  G  H  L  Q  D

 901 aagaggtatttgattgttcttgatgatacatgggaaattagtgcctgggatagcttcaaa
 301 K  R  Y  L  I  V  L  D  D  T  W  E  I  S  A  W  D  S  F  K

 961 gctgcattaccttataatagaaatggtagtaggatcatagtcacaactcgaaatatgact
 321 A  A  L  P  Y  N  R  N  G  S  R  I  I  V  T  T  R  N  M  T

1021 gtggcacacacttgctgttctcataacagcttttgtaatcacatccatgaagtccaacct
 341 V  A  H  T  C  C  S  H  N  S  F  C  N  H  I  H  E  V  Q  P

1081 ctctccactcggcagtccatgaagctgttttgcaacagagtctttggcgaatctgcatgc
 361 L  S  T  R  Q  S  M  K  L  F  C  N  R  V  F  G  E  S  A  C

1141 cctggaaatttgataatgctcacggaagacatactgagaaaatgtgatggactaccactg
 381 P  G  N  L  I  M  L  T  E  D  I  L  R  K  C  D  G  L  P  L

1201 gcc   1203
 401 A     401
```

Figure 4.4 Nucleotide sequence and conceptual translation of the N-terminal region of banana RGC4. The NBS motifs defined by Meyers et al. (1999) are underlined.

```
   1 atgtcgacggcgctagtaatcggaggatggttcgcgcaaagcttcatccagacgttgctc
   1 M  S  T  A  L  V  I  G  G  W  F  A  Q  S  F  I  Q  T  L  L

  61 gacaaggccagcaactgcgcgatccaacaactcgcgcggtgccgcggccttcacgatgac
  21 D  K  A  S  N  C  A  I  Q  Q  L  A  R  C  R  G  L  H  D  D

 121 ctgaggcggctgcggacgtctctgctccggatccatgccatcctcgacaaggcagagacg
  41 L  R  R  L  R  T  S  L  L  R  I  H  A  I  L  D  K  A  E  T

 181 aggtggaaccataaaaacacgagcttggtggagctggtgaggcagctcaaggatgctgcc
  61 R  W  N  H  K  N  T  S  L  V  E  L  V  R  Q  L  K  D  A  A

 241 tatgacgccgaggacttactggaggagttggagtaccaagccgcgaagcaaaaggtcgag
  81 Y  D  A  E  D  L  L  E  E  L  E  Y  Q  A  A  K  Q  K  V  E

 301 caccggggagaccagataagcgacctcttttcttttttcccctagtactgcgagcgagtgg
 101 H  R  G  D  Q  I  S  D  L  F  S  F  S  P  S  T  A  S  E  W

 361 ttgggtgccgatggtgatgatgctgggactcgattgagggagatccaggagaagctgtgc
 121 L  G  A  D  G  D  D  A  G  T  R  L  R  E  I  Q  E  K  L  C

 421 aacattgctgccgatatgatggatgtcatgcagctattggcacccgatgatgggggggaga
 141 N  I  A  A  D  M  M  D  V  M  Q  L  L  A  P  D  D  G  G  R

 481 caattcgactggaaggtggtgggaagagaaacgagctctttcttgaccgaaaccgtcgtg
 161 Q  F  D  W  K  V  V  G  R  E  T  S  S  F  L  T  E  T  V  V

 541 tttggtcggggccaagaaagggagaaagtagtagaattgctgttggattcaggatctggt
 181 F  G  R  G  Q  E  R  E  K  V  V  E  L  L  L  D  S  G  S  G

 601 aacagtagcttctctgtcttacccctcgtcggaatcggaggggttgggaagacgactctg
 201 N  S  S  F  S  V  L  P  L  V  G  I  G  G  V  G  K  T  T  L

 661 gctcagctcgtgtacaacgacaatcgtgtcggcaactatttccacctcaaggtttgggtc
 221 A  Q  L  V  Y  N  D  N  R  V  G  N  Y  F  H  L  K  V  W  V

 721 tgtgtatccgacaatttcaatgtgaagagactgaccaaagagataatcgagtctgctacc
 241 C  V  S  D  N  F  N  V  K  R  L  T  K  E  I  I  E  S  A  T

 781 aaggtggaacaatctgacgaattgaacttggacaccctgcaacagatcctcaaggagaag
 261 K  V  E  Q  S  D  E  L  N  L  D  T  L  Q  Q  I  L  K  E  K

 841 attgcttcagagaggtttctgctagtcctcgatgatgtgtggagcgaaaacagggatgac
 281 I  A  S  E  R  F  L  L  V  L  D  D  V  W  S  E  N  R  D  D

 901 tgggaaaggctgtgcgcgccactaaggtttgcagcaagaggcagcaaggttatagtcaca
 301 W  E  R  L  C  A  P  L  R  F  A  A  R  G  S  K  V  I  V  T

 961 actcgagacacaaagattgccagcatcattggcacaatgaaggaaatttcgctcgatggt
 321 T  R  D  T  K  I  A  S  I  I  G  T  M  K  E  I  S  L  D  G

1021 ctccaggatgatgcttactgggagctgttcaagaaatgtgcatttggttctgtgaacccc
 341 L  Q  D  D  A  Y  W  E  L  F  K  K  C  A  F  G  S  V  N  P

1081 caggagcatctagagctcgaggttatcggtagaaagattgctggtaagttgaagggctca
 361 Q  E  H  L  E  L  E  V  I  G  R  K  I  A  G  K  L  K  G  S

1141 ccgctagca  1149
 381 P  L  A     383
```

Figure 4.5 Nucleotide sequence and conceptual translation of the N-terminal region of banana RGC5. The NBS motifs defined by Meyers et al. (1999) are underlined.

Sequence comparisons with other plant resistance gene candidates of the NBS-type and the characterised NBS-LRR resistance genes were performed using the region between the P-loop and GLPLA motifs (~170 aa) of the NBS domain. The highest level of identity among the deduced amino acid sequences of banana RGCs was found between RGC2 and RGC5 with 50% and the lowest level between RGC2 and RGC3 with 21% (Table 4.1). Overall, identities between the five banana RGCs were equivalent to those observed between other RGCs (Shen et al. 1998; Ayele-Gedil et al. 2001; López et al. 2003) or the NBS-LRR resistance genes of other plant species (Table 4.1). Homology searches of the GenBank database revealed that three of the isolated banana RGCs were identical or highly similar to NBS sequences of *Musa acuminata* already in GenBank. RGC1 was 100% identical to entry AAM97903, RGC3 was 96% identical to entry ABB96971 and finally RGC5 was 93% identical to entry AAM97908. Three partial *Musa* NBS sequences spanning the region between the P-loop and RNBS-B motifs (~100 amino acids) appear in GenBank (AAM97909, AAM97910 and AAM9711) and were considered as three novel classes of NBS-type sequences since they have no counterparts to any reported *Musa* sequences (they share <50% amino acid identity with *malaccensis* RGCs). Further research is required to isolate the GLPLA region for each one of them. Homology searches also revealed that each banana RGC showed a significant similarity to RGCs isolated from other monocots such as *Oryza sativa*, *Saccharum officinarum* and *Avena sativa* (Table 4.2); and also to known non-TIR-NBS-LRR resistance genes (Table 4.3).

Table 4.1 Percentage identity derived from pairwise comparisons between isolated banana RGCs and between the NBS domain of NBS-LRR resistance genes[a].

Class	RGC1	RGC2	RGC3	RGC4	RGC5	Rx	HERO	Fom-2
RGC1		27	24	34	33	33	34	28
RGC2	…..		21	26	50	26	34	36
RGC3	…..	…..		25	24	24	28	24
RGC4	…..	…..	…..		30	32	34	31
RGC5	…..	…..	…..	…..		35	37	36
Rx	…..	…..	…..	…..	…..		38	31
HERO	…..	…..	…..	…..	…..	…..		33
Fom-2	…..	…..	…..	…..	…..	…..	…..	

[a] The region between the P-loop and the GLPLA motifs (~170 aa) of the NBS domain was considered for the pairwise comparisons. Rx (CAB50786), confers resistance to *Potato virus X* (PVX) in *Solanum tuberosum*; HERO (CAD29729), confers resistance to *Globodera rostochiensis* in *Lycopersicon esculentum*; Fom-2 (AAS80152), confers resistance to *Fusarium oxysporum f.sp.melonis* in *Cucumis melo*.

Table 4.2 Best BLASTX hits of isolated banana RGCs with respect to RGCs from other plant species [a].

Class	Accession number	Plant	Identity %	Similarity %	Expect (E) value
RGC1	BAC79938	*Oryza sativa*	41	63	9e-35
RGC2	AAQ16581	*Saccharum officinarum*	49	72	1e-44
RGC3	AAC31552	*Avena sativa*	54	69	2e-36
RGC4	BAC15497	*Oryza sativa*	41	59	4e-34
RGC5	AAT47022	*Oryza sativa*	49	67	7e-44

[a] The region between the P-loop and the GLPLA motifs (~170 aa) of each banana RGC was used as a query in BLASTX searches.

Table 4.3 Two best BLASTX hits of isolated banana RGCs with respect to characterised resistance genes[a].

Class	R gene[b]	Accession number	Plant	Identity %	Similarity %	E-value
RGC1	*Rxo1*	AAX31149	*Zea mays*	39	63	6e-31
	RPM1	Q39214	*Arabidospsis thaliana*	37	55	1e-22
RGC2	*Rpg1-b*	AAR19097	*Glycine max*	49	65	9e-37
	I2	AAD27815	*Lycopersicon esculentum*	45	61	1e-32
RGC3	*RPS5*	O64973	*Arabidospsis thaliana*	38	56	4e-20
	RPS2	Q42484	*Arabidospsis thaliana*	36	57	5e-19
RGC4	*MLA1*	AAG37354	*Hordeum vulgare*	35	56	1e-22
	MLA10	AAQ55541	*Hordeum vulgare*	34	56	1e-22
RGC5	*I2*	AAD27815	*Lycopersicon esculentum*	47	65	4e-37
	Xa1	BAA25068	*Oryza sativa*	45	66	8e-38

[a] The region between the P-loop and the GLPLA motifs (~170 aa) of each banana RGC was used as a query in BLASTX searches.
[b] All the best hits found correspond to non-TIR-NBS-LRR resistance genes. *Rxo1*, confers resistance to *Xanthomonas oryzae* in *zea mays*; *RPM1*, *RPS5* and *RPS2*, confer resistance to *Pseudomonas syringae* in *Arabidopsis thaliana*; *Rpg1-b*, confers resistance to *Pseudomonas syringae* in *Glycine max*; *I2*, confers resistance to *Fusarium oxysporum* f.sp. *lycopersici* in *Lycopersicon esculentum*; *MLA1* and *MLA10*, confer resistance to *Blumeria graminis* in *Hordeum vulgare*; *Xa1*, confers resistance to *Xanthomonas oryzae* in *Oryza sativa*.

Alignment of the deduced amino acid sequences of the banana RGCs spanning the N-terminal region showed that the RGCs contained the typical consensus P-loop/kinase-1a, kinase-2, RNBS-B and GLPLA motifs of the NBS domain of R genes (Meyers et al. 1999) located at similar positions (Figure 4.6). Moreover, the NBS domain of all banana RGCs allow to predict the absence of the TIR domain by the presence of the motif RNBS-A-non-TIR near to the P-loop and also to the presence of a tryptophan residue (W) at end of the kinase-2 motif which are associated with non-TIR-NBS

sequences only (Meyers et al. 1999) (Figure 4.6). This prediction was correct since the N-terminus of each RGC lacks a TIR domain, instead the non-TIR (nT) motif determined by Bai et al. (2002) was found in this region for RGC1, RGC2, RGC3 and RGC5 (Figure 4.6). RGC4 did not present this motif, instead a duplicated NBS-type sequence was found at its N-terminus (RGC4-N-ter). Interestingly, the P-loop motif of this sequence was poorly conserved and the GLPLA motif was not present (Figure 4.6). In contrast to the NBS domain where several conserved motifs are found, the non-TIR domain shows a less degree of conservation.

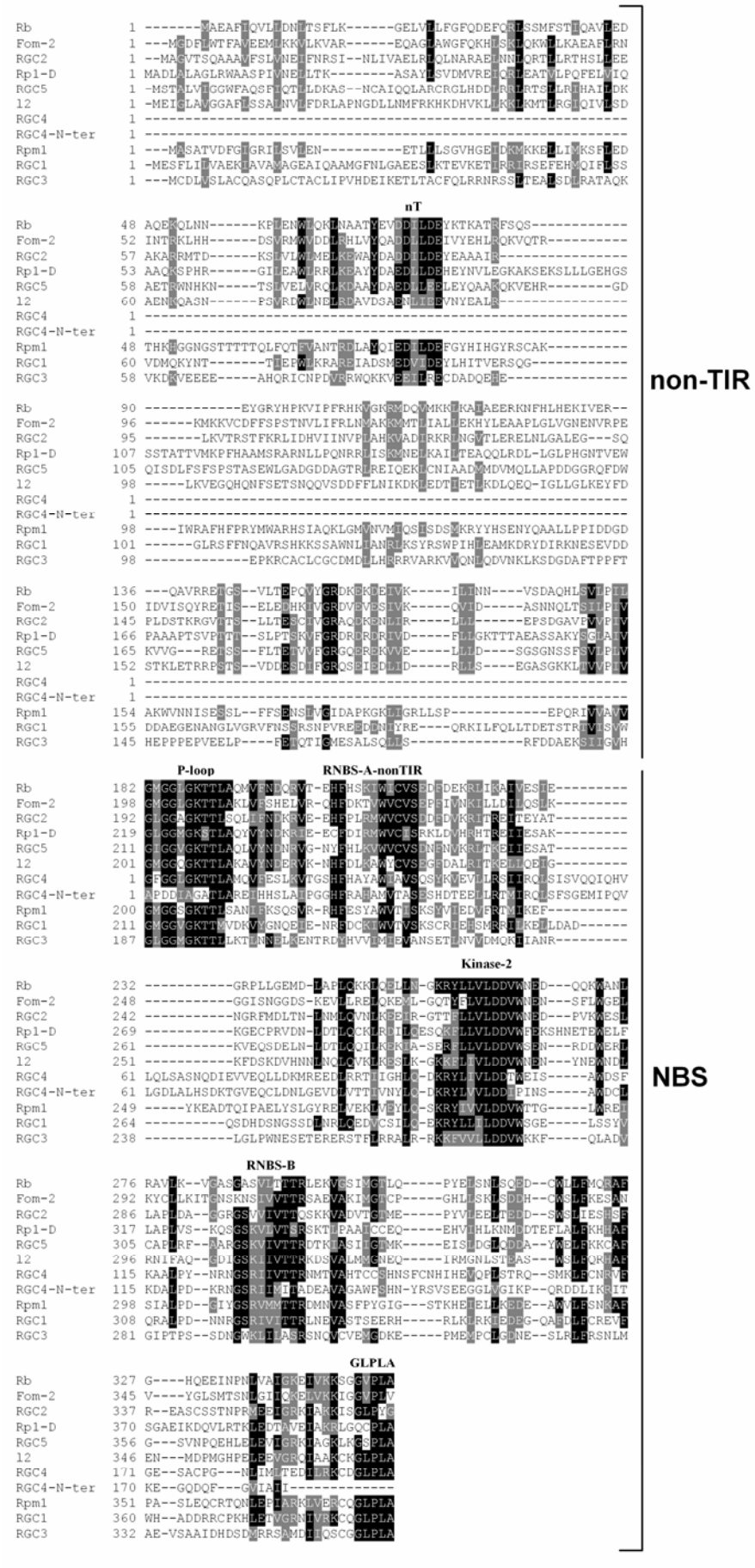

nT
non-TIR
P-loop
RNBS-A-nonTIR
Kinase-2
NBS
RNBS-B
GLPLA

Figure 4.6 ClustalX alignment of the deduced amino acid sequences of the N-terminal regions of banana RGC1 to RGC5 including the non-TIR domain. The corresponding region of known non-TIR-NBS-LRR R genes was included in the alignment for comparison. The non-TIR (nT) motif determined by Bai et al. 2002 and the conserved NBS motifs as determined by Meyers et al. (1999) are indicated. Identical amino acids are shaded in black and conservative substitutions are shaded in grey. Rp1-D (AAD47197), confers resistance to *Puccinia sorghi* in *Zea mays*; I2 (AAD27815) and Fom-2 (AAS80152), confer resistance to *Fusarium oxysporum* f.sp. *lycopersici* and *Fusarium oxysporum* f.sp. *melonis* in *Lycopersicon esculentum* and *Cucumis melo*, respectively; Rb (Q7XBQ9), confers resistance to *Phytophthora infestans* in *Solanum bulbocastanum*; RPM1 (A57072), confers resistance to *Pseudomonas syringae* in *Arabidopsis thaliana*.

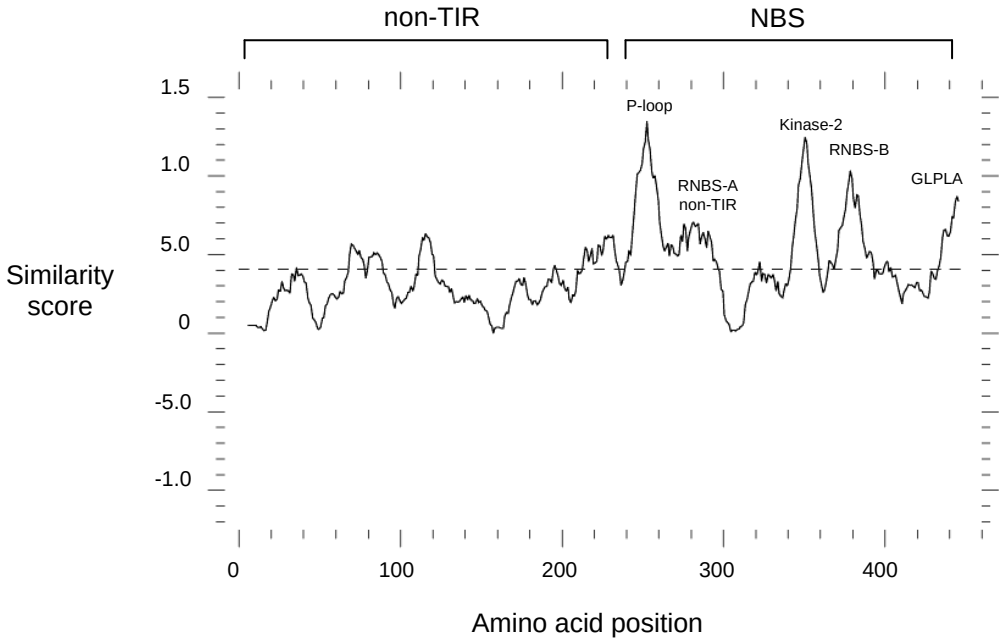


Figure 4.7 A similarity plot of the banana resistance gene candidates (RGC1 to RGC5). The conserved NBS motifs as determined by Meyers et al. (1999) are indicated. The dotted line indicates the average similarity.

4.4.2 Phylogenetic relationships of the banana RGCs

The phylogenetic analysis was performed using the region comprising only the NBS domain as it is present in both the TIR and non-TIR-NBS-LRR proteins and contains numerous conserved motifs that assist proper alignment (Meyers et al. 2003). The phylogenetic tree grouped the banana RGCs into the non-TIR subclass of NBS sequences described by Meyers et al. 1999 (Figure 4.8) which supports previous conclusions drawn from our sequence analysis indicating that the banana RGCs lack a TIR domain at the N-terminal region. The distribution of the banana RGCs in different branches of the non-TIR-NBS cluster reflects a high level of sequence divergence for

these sequences. In the case of RGC2 and RGC5, they are grouped in the same clade but separated by long branches, whereas the remaining banana RGCs are distributed in different clades. Overall, the banana RGCs were more closely related to NBS sequences from other species than each other. Interestingly, all banana RGCs were clustered with NBS sequences of known non-TIR-NBS-LRR resistance proteins suggesting they may encode resistance gene products of as yet unknown specificity.

4.4.3 Genomic copy number

Southern hybridisation analysis performed under conditions of high stringency revealed hybridisation patterns comprising one to three bands for each gene (Figure 4.9), suggesting that the banana RGCs are present as a single copy in the diploid genome of *M. acuminata ssp. malaccensis* or possibly as a few copies representing a small multigene family.

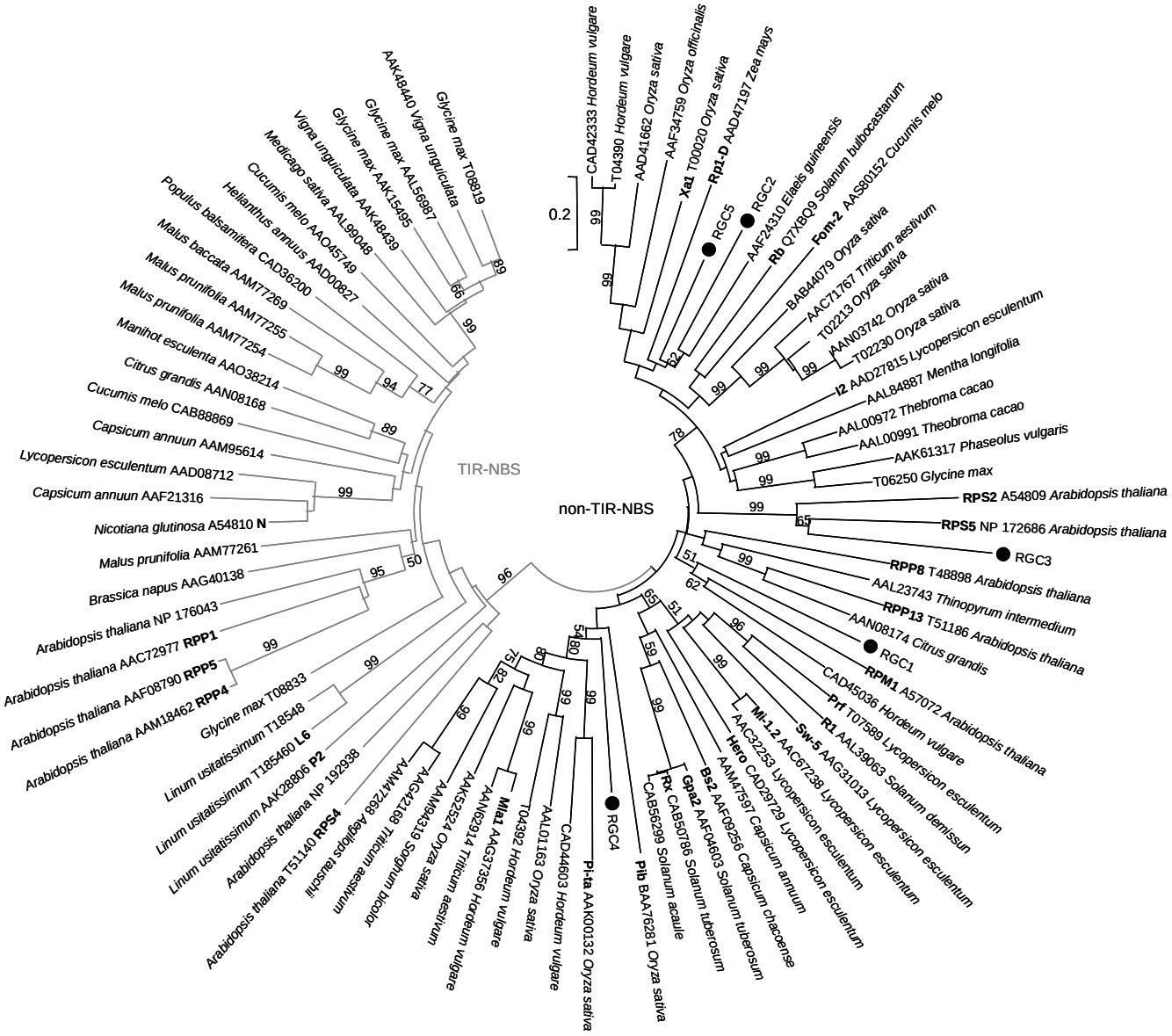


Figure 4.8 Neighbor-joining phylogenetic tree based on the ClustalX alignment of resistance gene candidate sequences from banana (black circles), other plants and the NBS of TIR- and non-TIR-NBS-LRR resistance proteins (in bold). Amino acid sequences from the P-loop to the GLPLA of the NBS domain were used for the analysis. The numbers below the branches indicate the percentage of 1000 bootstrap replications supporting the particular nodes, and only those with >50% support are shown. The tree was constructed using MEGA 2.1 with the Poisson correction.

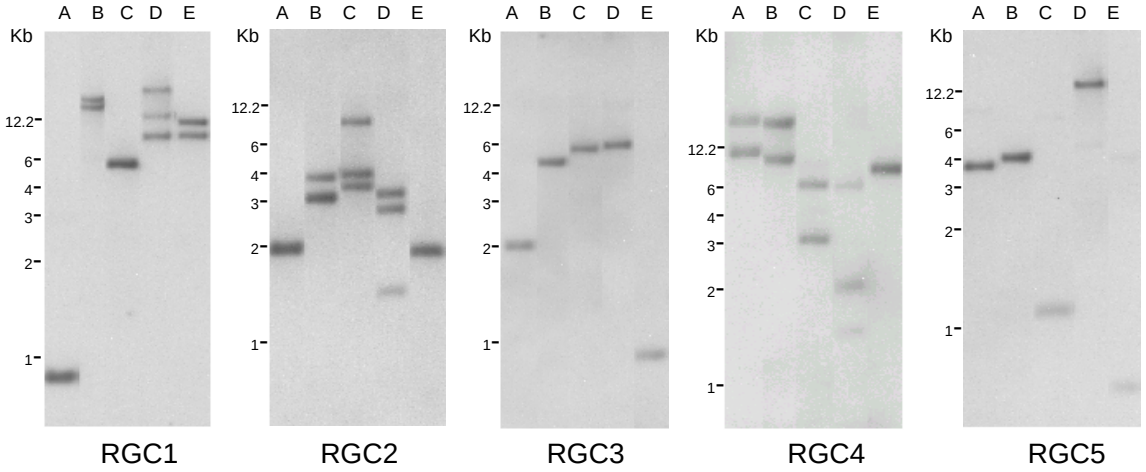


Figure 4.9 Southern blot analysis of each banana resistance gene candidate class (RGC1 to RGC5). Genomic DNA (5 μg) was digested with restriction enzymes, separated on a 1.1% agarose gel, and transferred to a nylon membrane by capillarity. Hybridization was done with DIG-labelled probes under high-stringency. The restriction enzymes used in lanes A-E are EcoR I, EcoR V, Hind III, Bgl II, and Sac I, respectively. Molecular weight markers are indicated on the left of each blot.

4.4.4 Expression profiles of the banana RGCs

To determine whether the banana RGCs were expressed, RT-PCR was carried out in unchallenged plants of *M. acuminata* ssp. *malaccensis* resistant or susceptible to FOC race 4 with specific primers for each of the RGCs. Amplification products were detected in both leaf and root tissue for RGC1, RGC2, RGC3 and RGC5 but not for RGC4 (Figure 4.10), suggesting that at least four of the five RGCs are expressed constitutively. Interestingly, an mRNA encoding RGC2 was detected in the resistant plant (line 850) but not in the susceptible plant (line 845) suggesting this transcriptional polymorphism correlates with resistance to the FOC race 4. Identical results were obtained when two other lines were tested, a transcript was detected in the resistant line 852 but not in susceptible line 846 (data not shown). As a control to ensure that the amplification products represented expressed mRNA and did not arise from contaminating DNA, PCR amplifications were carried out using primers targeting the gene encoding banana Actin 1 and spanning an intron of approximately 100 bp protein (Hermann et al. 2001). Only the expected cDNA fragment of ~ 480 bp was amplified indicating that cDNA preparations were not contaminated with genomic DNA. Therefore, these experiments provide qualitative evidence of gene expression for four of five classes of RGCs in banana.

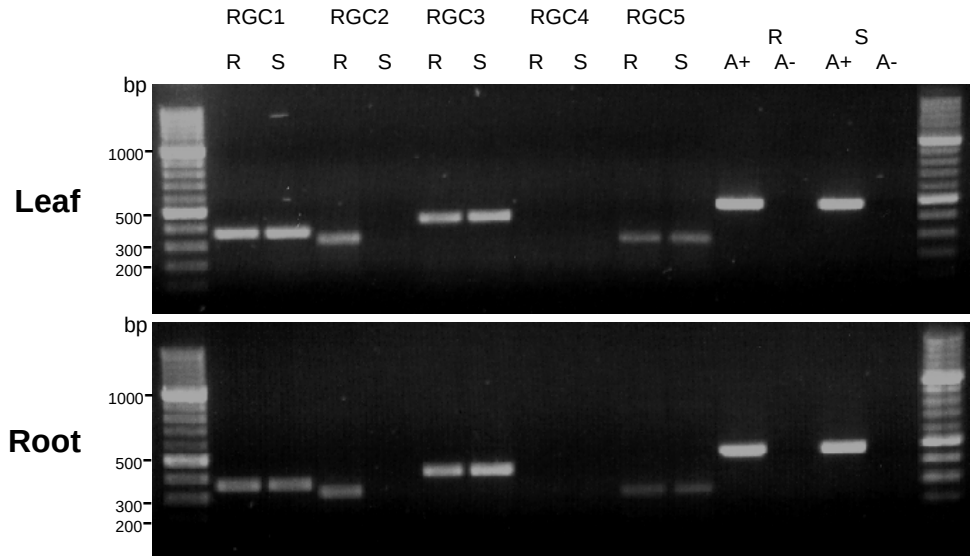


Figure 4.10 RT-PCR analysis of the five banana resistance gene candidates (RGC1 to RGC5) in FOC race 4 resistant (R) line 850 or susceptible (S) line 845 plants of *M. acuminata* ssp. *malaccensis*. Total RNA was extracted from leaf or root tissues and treated with DNase. A+ lanes (positive control), expected ~480 bp banana *Actin 1* cDNA fragment; A- lanes (negative control), no reverse transcriptase. 25 μl of PCR amplification was loaded for each banana RGC and 5 μl for the *Actin 1* gene fragment. Molecular weight markers are indicated on the left and right of each agarose gel.

4.5 Discussion

This study presents the isolation and characterisation of five different classes of RGCs of the NBS-type from the wild banana *Musa acuminata* ssp. *malaccensis*. Several features of the banana RGCs suggest they are related to the NBS of NBS-LRR disease resistance genes. For example, the characteristic motifs of the NBS domain of known resistance genes described by Meyers et al. (1999) and Pan et al. (2000) are present in each banana RGCs at similar positions (Figure 4.6). One of these motifs, the highly conserved P-loop, has been shown to bind ATP in the NBS-LRR resistance proteins I2 and Mi from tomato (Tameling et al. 2002) suggesting the banana RGC proteins may also bind ATP. The non-TIR (nT) motif (Bai et al. 2002), which is associated only with the non-TIR subclass of NBS sequences was found in the N-terminal region of four banana RGCs (Figure 4.6). Interestingly, RGC4 showed a duplicated NBS-type sequence in the N-terminal region. The presence of a duplicated NBS sequence at the N-terminus of NBS-LRR proteins has been reported in the rice genome as NBS-NBS-LRR genes (Zhou et al. 2004). Only four genes of the 480 rice NBS-LRR genes showed this structure (Zhou et al. 2004). None of the motifs associated with the TIR subclass were found in the corresponding region of the banana RGCs. This feature is consistent with the structure of R genes from monocot plants where the TIR domain appears to be absent (Meyers et al. 1999; Pan et al. 2000; Bai et al. 2002; Zhou et al. 2004). Moreover, the banana RGC1, RGC2 and RGC5 showed the presence of a putative coiled-coil (CC) structure in the non-TIR domain, which is another common feature of this region (Pan et al. 2000). For example, in the rice genome, 159 of the 480 NBS-LRR sequences found were identified as having a CC motif in the non-TIR domain (Zhou et al. 2004). The lengths of the non-TIR domain of the banana RGCs (Figures 4.1 to 4.6) were also similar to the lengths of the non-TIR domain of monocot and dicot R genes which range from 200-250 amino acids from the start of the coding region to the beginning of the NBS domain (P-loop) (Bai et al. 2002; Meyers et al. 2003). The majority of RGCs investigated so far, for example in soybean (Kanazin et al. 1996; Yu et al. 1996), potato (Leister et al. 1996), bean (Shen et al. 1998; López et al. 2003)

and others map to clusters of genetically defined R gene loci. Thus the banana RGC loci may encode functional R genes although an association with resistance to a pathogen remains to be demonstrated.

It is likely that the RGCs isolated in this study represent only a subset of the NBS sequences present in the banana genome as only one primer combination was used. Using PCR primers based on the conserved motifs of the NBS region, eight and 14 different classes of NBS sequences were isolated in *Arabidopsis* and rice, respectively (Aarts et al. 1998; Leister et al. 1998). This is much smaller than the numbers of NBS sequences known to be present in the genomes of *Arabidopsis* (Meyers et al. 2003) and rice (Zhou et al. 2004). These data show that the number of NBS sequences obtained using PCR primers based on the NBS region may represent only a small portion of the entire set present in the plant genome. The latter may be attributable to DNA polymorphisms in the motifs where the degenerate primers are designed which cause some sequences to be preferentially amplified. The use of other primer combinations might help in the identification of other NBS sequences in banana.

The phylogenetic analysis supports the classification of the banana RGCs into the non-TIR subclass since they all cluster with other NBS sequences of the non-TIR subclass. To date, the TIR domain has not been found in the structure of monocot NBS-LRR R genes even in the complete rice genome sequence (Goff et al. 2002; Bai et al. 2002; Cannon et al. 2002; Zhou et al. 2004). It has been hypothesised that the loss of the TIR domain from the NBS-LRR genes in monocot plants may have occurred during the divergence of the monocots and dicots early in the Cretaceous period about 100 million years ago (Pan et al. 2000). Because banana shares a common evolutionary origin with other monocot plants, it is likely that this domain is also absent from the structure of all banana R genes. Therefore, the fact that only NBS sequences of the non-TIR subclass were isolated in this study is unlikely to be an artefact of the PCR amplification process but instead the result from the absence of the TIR domain in the *Musaceae* family. The latter is supported by the study of Kanazin et al. (1996) who only found the

TIR subclass of NBS sequences in soybean using the same pair of degenerate primers. Indeed, the sequencing of the banana genome, which is currently in progress, will shed light on this matter (www.musagenomics.org).

All RGC sequences hybridised to a relatively small number of restriction fragments (one to three) (Figure 4.9), indicating that these RGCs are present within the banana genome as single copies or as members of a small gene family. Single copy RGCs also exist in other plants such as soybean (Kanazin et al. 1996), potato (Leister et al. 1996), *Arabidopsis* (The *Arabidopsis* Genome Initiative 2000; Meyers et al. 2003) and rice (Bai et al. 2002; Ramalingam et al. 2003; Zhou et al. 2004). However, most R genes so far isolated have been found as multicopy, clustered sequences (Hulbert et al. 2001). For example, the *Fusarium I2* resistance gene from tomato is within a 90 kb cluster of seven paralogues (Simons et al. 1998). The majority of RGCs isolated in different plant species show a clustered genomic organization (Aarts et al. 1998; Leister et al. 1998; Hulbert et al. 2001; Bai et al. 2002; Ramalingam et al. 2003). Although some of the isolated banana RGCs appear to be present in low copy number, it remains to be demonstrated whether they are organized in gene clusters.

Previous reports have shown that NBS-LRR resistance genes are not inducible but are expressed in a constitutive manner. Northern blot or RT-PCR analyses on different R genes such as *RPM1* (Grant et al. 1995), *Prf* (Salmeron et al. 1996), *RPP5* (Parker et al. 1997), *Mi* (Milligan et al. 1998), *I2* (Mes et al. 2000), and others have revealed the presence of low levels of transcripts in unchallenged plants. These findings are in agreement with the postulated role of NBS-LRR proteins acting as preformed receptors that recognize a pathogen (Van der Biezen and Jones 1998). The lack of a circulatory system in plants might be compensated for by a capacity to express each R protein constitutively in every cell that potentially could be attacked (Van der Biezen and Jones 1998). The finding of expression of RGC1, RGC2, RGC3, and RGC5 in both leaf and root tissue without any pathogen challenge suggests a constitutive expression for these sequences (Figure 4.10). The apparent absence of expression of RGC4 in both tissues

might suggest a non-functional promoter precedes this gene. Remarkably, RGC2 showed a transcriptional polymorphism that correlates to the FOC race 4 resistance phenotype of *M. acuminata* ssp. *malaccensis* plants. The RGC2 mRNA is present in two lines of resistant plants but absent in two lines of susceptible plants. This finding suggests that RGC2 might play a role in resistance to FOC race 4. A similar expression profile was shown by the *Fusarium I2* resistance gene from tomato where the expression of the *I2* gene was only present in both leaf and root tissue of plants resistant to *Fusarium oxysporum* f.sp *lycopersici* race 2 and absent in susceptible plants (Mes et al. 2000).

Apart from the transcriptional correlation of RGC2 in FOC race 4 resistance, this sequence as well as RGC5 showed a significant similarity to the *Fusarium I2* resistance gene from tomato. Since one goal of this project was to identify banana genes that may confer resistance to FOC race 4, RGC2 and RGC5 were chosen for further characterisation and the results of this work are presented in Chapter 5. Characterisation of the full open reading frame (ORF) of RGC2 and RGC5 revealed the presence of leucine reach repeats (LRR) at the C-terminal region of the predicted proteins (Chapter 5). Although, the C-terminal encoding region of the remaining banana RGCs was not isolated, it is highly probable that these sequences also contain a LRR domain since most plant NBS sequences contain this domain at the C-terminal region. For example, the *Arabidopsis* and rice genome sequences have shown that the majority of their NBS encoding sequences contain leucine reach repeats at the C-terminus. In *Arabidopsis*, 149 of the 178 NBS sequences identified contain a LRR domain (Meyers et al. 2003) and in the case of rice, 480 of the 535 NBS sequences identified contain this domain as well (Zhou et al. 2004).

This study has shown that the isolated banana sequences are NBS resistance gene candidates with a non-TIR domain at the N-terminus. The identification of RGCs in banana may assist in the identification of functional banana R genes based on sequence homology and expression analysis, since two banana RGCs (RGC2 and RGC5) showed a significant sequence

similarity to the *Fusarium I2* resistance gene from tomato and the expression of RGC2 correlated with FOC race 4 resistance. Cloning of the full ORF of these sequences (chapter 5) will allow testing their role in *Fusarium* resistance through genetic complementation. Alternatively, the identification of RGCs in banana may provide markers tightly linked to R gene loci that could be used in high-resolution genetic mapping as a tool for map-based cloning. New technologies of the post-genomic era, such as RNA interference (RNAi) (Waterhouse and Helliwell 2003) could facilitate testing the function of multiple RGCs in banana plants resistant to the most devastating pathogens. Those resistant plants that become susceptible after pathogen challenge would assist in the identification of a particular R gene. The RNAi technology has been recently used to determine the function of multiple genes involved in pathogen resistance in barley epidermal cells (Douchkov et al. 2005). Another recent technology that promises to facilitate the identification of multiple R genes in banana is the use of Binary Bacterial Artificial Chromosome (BIBAC) libraries, which can be used to transfer via *Agrobacterium tumefaciens* large DNA fragments (up to 120 kb) into the plant genome (He et al. 2003). Indeed, a BIBAC library has been constructed recently for the cultivar 'Tuu Gia' (*Musa acuminata*) which is resistant to the most serious diseases of banana such as black Sigatoka or Panama disease (Ortiz-Vázquez et al. 2005) and a highly efficient Agro-transformation method for banana is now available (Khanna et al. 2004). With these technologies, it would be possible to transform disease-susceptible banana cultivars with BIBAC clones harbouring RGCs organized either as singletons or clusters from the 'Tuu Gia' cultivar. This approach would lead to a RGC-BIBAC collection of banana lines ready to be used for pathogen-resistance screenings. The application of these technologies in banana foresees a promising future to unravel the function of RGCs in this crop and develop pathogen resistance in the field.

4.6 References

Aarts, M.G.M., Hekkert, B.L., Holub, E. B., Beynon, J.L., Stiekema, W.J. and Pereira, A. (1998). Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana*. *Molecular Plant Microbe Interactions* 11: 251-258.

Agrios, G.N. (1997). *Plant Pathology*. 4th ed. Academic Press. USA. p. 307.

Altschul, S. F., Madden, T.L., Schaffer, 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: 3389-3402.

Ayele-Gedil, M., Slabaugh, M.B., Berry, S., Johnson, R., Michelmore, R., Miller, J., Gulya, T. and Knapp, S. (2001). Candidate disease resistance genes in sunflower cloned using conserved nucleotide-binding site motifs: genetic mapping and linkage to the downy mildew resistance gene Pl1. *Genome* 44: 205-212.

Bai, J., Pennill, L.A., Ning, J., Lee, S.W., Ramalingam, J., Leach, J.E. and Hulbert, S.H. (2002). Diversity in nucleotide binding site-leucine rich repeat genes in cereals. *Genome Research* 12: 1871-1884.

Becker, D., Dugdale, B., Smith, M., Harding, R. and Dale, J. (2000). Genetic transformation of Cavendish banana (*Musa* spp. AAA group) cv 'Grand Nain' via microprojectile bombardment. *Plant Cell Reports* 19: 229-234.

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: 391-399.

Calenge, F., Van der Linden, C.G., Van de Weg E., Schouten, H.J., Van Arkel, G., Denance, C., and Durel, C.E. (2005). Resistance gene analogues identified through the NBS-profiling method map close to major genes and QTL for disease resistance in apple. *Theoretical and Applied Genetics* 110: 660-668.

Cannon, S. B., Zhu, H., Baumgarten, A. M., Spangler, R., May, G., Cook, D. R. and Young, N.D. (2000). Diversity, distribution and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. *Journal of Molecular Evolution* 54: 548-562.

Dangl, L. and Jones, J. (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411: 826-833.

Dellaporta, S.L., Wood, J. and Hicks, J. B. (1983). A plant DNA minipreparation: version II. *Plant Molecular Biology Reports* 1: 19-21.

Douchkov, D., Nowara, D., Zierold, U. and P. Schweizer (2005). A high-throughput gene-silencing system for the functional assessment of defense-related genes in barley epidermal cells. *Molecular Plant Microbe Interactions* 18:755-761.

FAO(2005).http://apps.fao.org/lim500/nphwrap.pi?production.crops.primary&Domain.

Goff, S., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). *Science* 296: 92-100.

Graham, M., Mareck, L., Lohnes, D., Cregan, P. and Shoemaker, R. (2000). Expression and genome organization of resistance gene analogues in soybean. *Genome* 43: 86-93.

Grant, M., Godard, L., Straube, E., Ashfield, T., Leward, J., Sattler, A., Innes, R. and Dangl, J. (1995). Structure of the *Arabidopsis* RPM1 enabling dual specifity disease resistance. *Science* 269: 843-846.

He, R.F., Wang, Y., Shi, Z., Ren, X., Zhu, L., Weng, Q. and He, G.C. (2003). Construction of a genomic library of wild rice and *Agrobacterium*-mediated transformation of large insert DNA linked to *BPH* resistance locus. *Gene* 321: 113-121.

Hermann, SR., Harding, RM. and Dale, J.L. (2001). The banana *Actin 1* promoter drives near-constitutive transgene expression in vegetative tissues of banana (*Musa* ssp). *Plant Cell Reports* 20: 525-530.

Hulbert, S.H., Webb, C.A., Smith, S.M. and Sun, Q. (2001). Resistance gene complexes: evolution and utilization. *Annual Review of Phytopathology* 39: 285-312.

Kanazin, V., Mareck, L. and Shoemaker, R. (1996). Resistance gene analogs are conserved and clustered in soybean. *Proceedings of the National Academy of Science (USA)* 93: 11746-11750.

Khanna, H., Becker, D., Kleidon, J. and Dale, J. (2004). Centrifugation assisted *Agrobacterium*-mediated transformation (CAAT) of embryogenic cell suspensions of banana (*Musa* spp. Cavendish AAA and Lady finger AAB). *Molecular Breeding* 14: 239-252.

Kumar, S., Tamura, K., Jakobsen, I.B. and Nei, M. (2001). MEGA2: Molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245.

Leister, D., Ballvora, A., Salamini, F. and Gebhardt, C. (1996). A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. *Nature Genetics* 14: 421-429.

Leister, D., Kurth, J., Laurie, D., Yano, M., Sasaki, T., Devos, K., Graner, A. and Schulze-Lefert, P. (1998). Rapid organization of resistance gene

homologues in cereal genomes. *Proceedings of the National Academy of Science (USA)* 95: 370-375.

López, C., Acosta, I., Jara, C., Pedraza, F., Gaitan-Solis, E., Gallego, G., Beebe, S. and Tohme, J. (2003). Identifying resistance gene analogs associated with resistance to different pathogens in common bean. *Phytopathology* 93: 88-95.

Lupas, A. (1996). Prediction and analysis of coiled-coil structures. *Methods in Enzymology* 266: 513-525.

Martínez-Zamora, M.G., Castagnaro, A.P. and Díaz-Ricci, J.C. (2004). Isolation and diversity analysis of resistance gene analogues (RGAs) from cultivated and wild strawberries. *Molecular Genetics and Genomics* 272: 480-487.

Mes, J., Van Doorn, A., Wijbrandi, J., Simons, G., Cornelissen, B. and Haring, M. (2000). Expression of the Fusarium resistance gene I-2 colocalizes with the site of fungal containment. *Plant Journal* 23: 183-193.

Meyers, B., Dickerman, A., Michelmore, R., Sivaramakrishnan, S., Sobral, B. and Young, N. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant Journal* 20: 317-332.

Meyers, B., Kozik, A., Griego, A., Kuang, H. and Michelmore, R.W. (2003). Genome-wide analysis of NBS-LRR encoding genes in *Arabidopsis*. *Plant Cell* 15: 809-834.

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: 1113-1130.

Milligan, S., Bodeau, J., Yaghoobi, J., Kaloshian, I., Zabel, P. and Williamson, V. (1998). The root knot resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding site, leucine-rich repeat family of plant genes. *Plant Cell* 10: 1307-1319.

Muskett, P., Kahn, K., Austin, M., Moisan, L., Sadanandom, A., Shirasu, K., Jones, J.D.G. and Parker, J.E. (2002). *Arabidopsis* RAR1 exerts rate-limiting control of R gene-mediated defenses against multiple pathogens. *Plant Cell* 14: 979-992.

Ortiz-Vázquez, E., Kaemmer, D., Zhang, H.B., Muth, J., Rodríguez-Mendiola, M., Arias-Castro, C. and James, A. (2005). Construction and characterization of a plant transformation-competent BIBAC library of the black Sigatoka resistant banana *Musa acuminata* cv. Tuu Gia (AA). *Theoretical and Applied Genetics* 110: 706-713.

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: 203-213.

Parker, J., Coleman, M., Szabo, V., Frost, V., Schmidt, R., Van der Biezen, E., Moores, T., Dean, C., Daniels, M. and Jones, J. (1997). The *Arabidopsis* downy mildew resistance gene RPP5 shares similarity to the toll/interleukin-1 receptors with N and L6. *Plant Cell* 9: 879-894.

Pasberg-Gauhl, C., Gauhl, F. and Jones D.R. (2000). Black Sigatoka: distribution and economic importance. pp. 37-44. *In Diseases of banana, abaca and enset*. Jones, D.R. (Ed.). CABI. Wallingford, UK.

Ploetz, R. and Pegg, K. (2000). Fungal disease of the root, corm and pseudostem. pp. 143-171. *In Diseases of banana, abaca and enset*. Jones, D.R. (Ed.). CABI. Wallingford, UK.

Ramalingam, J., Vera-Cruz, C.M., Kukreja, K., Chittoor, J.M., Wu, J.L., Lee, S.W., Baraoidan, M., George, M.L., Cohen, M.B., Hulbert, S.H., Leach, J.E., Leung H. (2003). Candidate defense genes from rice, barley and maize and their associations with qualitative and quantitative resistance in rice. *Molecular Plant Microbe Interactions* 16: 14-24.

Richter, T.E. and Ronald, P.C. (2000). The evolution of disease resistance genes. *Plant Molecular Biology* 42: 195-204.

Rivkin, M., Vallejos, C. and McClean, P. (1999). Disease-resistance related sequences in common bean. *Genome* 42: 41-47.

Roux, N.S., Toloza, A., Dolezel, J. and Panis, B. (2004). Usefulness of embryogenic cell suspensions cultures for the induction and selection of mutants in Musa ssp. pp. 33-43. In: Banana improvement: cellular, molecular biology and induced mutations. Mohan J. and R. Swennen. (Eds.) *Science Publishers Inc*. Enfield, NH, USA.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425.

Salmeron, J., Oldroyd, E., Rommens, C., Scofield, S., Kim, H., Lavelle, D., Dahlbeck, D. and Staskawicz, B. (1996). Tomato Prf is a member of a leucine-rich repeat class of plant disease resistance gene and lies embedded within the Pto kinase gene cluster. *Cell* 86: 123-133.

Seah, S., Spielmeyer, W., Jahier, J., Sivasithamparam, K. and Lagudah, S. (2000). Resistance gene analogs within and introgressed chromosomal segment derived from *Triticum ventricosum* that confers resistance to nematode and rust pathogens in wheat. *Molecular Plant Microbe Interactions* 13: 334-341

Shen, K., Meyers, B., Islam-Faridi, M., Chin, D., Stelly, D. and Michelmore, R. (1998). Resistance gene candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance genes in lettuce. *Molecular Plant Microbe Interactions* 11: 815-823.

Schuler, M. A. and Zielinski, R.E. (1989). RNA isolation from light- and dark-grown seedlings. In *Methods in Plant Molecular Biology*. Academic Press Inc. pp. 89-96.

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A. and Lukyanov, S.A. (1995) An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Research* 23: 1087-1088.

Simons, G., Groenendijk, J., Wijbrandi, J., Reijans, M., Groenen, J., Diergaarde, P., Van der Lee, T., Bleeker, M., Onstenk, J., Both, M., Haring, M., Mes, J., Cornelisse, B., Zabeau, M. and Vos, P. (1998). Dissection of the *Fusarium I2* gene cluster in tomato reveals six homologs and one active gene copy. *Plant Cell* 10: 1055-1068.

Smith, M.K., Hamill, S.D., Langdon, P.W. and Pegg, K.G. (1998). Selection of new banana varieties for the cool subtropics in Australia. *Acta Horticulturae* 490: 49-56.

Stover, R.H. and Simmonds, N.W. (1987). Bananas. 3rd edn. Longman Scientific and Technical. Harlow, Essex, UK. p. 468.

Tameling, W., Elzinga, S., Darmin, P., Vossen, J., Takken, F., Haring, M. and Cornelissen, B. (2002). The tomato R gene products I2 and Mi-1 are functional ATP binding proteins with ATPase activity. *Plant Cell* 14: 2929-2939.

Taylor, K. (2005). Characterization of potential fungal disease resistance genes in banana. PhD thesis. Queensland University of Technology, Brisbane, Australia. pp. 68-96.

The *Arabidopsis* genome initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 24: 4876-4882.

Van Der Biezen, E. and Jones, J.D.G. (1998). Plant disease-resistance proteins and the gene-for-gene concept. *Trends in Biochemical Science* 23: 454-456.

Van der Vossen, E.A., Van der Voort, J.N., Kanyuka, K., Bendahmane, A., Sandbrink, H., Baulcombe, D.C., Bakker, J., Stiekema, W.J., 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.

Waterhouse, P.M. and Helliwell, C.A. (2003). Exploring plant genomes by RNA-induced gene silencing. *Nature Reviews* 4: 29-38.

Yu, Y., Buss, G. and Maroof, M. (1996). Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. *Proceedings of the National Academy of Science (USA)* 93: 11751-11756.

Zhou, T., Wang, T., Chen, J. Q., Araki, H. and Jing, Z., Jiang, K., Shen, J. and Tian D. (2004). Genome-wide identification of NBS genes in *japonica* rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Molecular Genetics and Genomics* 271: 402-415.

Chapter 5

Structural and Phylogenetic Analysis of Two Potential *Fusarium* Resistance Genes from Banana (*Musa acuminata* ssp. *malaccensis*)

5.1 Abstract

The previous chapter (4) presented the isolation of five different classes of resistance gene candidates (RGCs) of the nucleotide binding site (NBS) type from the wild banana *Musa acuminata* ssp. *malaccensis*. Two RGCs (RGC2 and RGC5) showed significant sequence similarity to the *Fusarium* resistance gene *I2* from tomato. Furthermore, the expression of RGC2 correlated with resistance to *Fusarium oxysporum* formae specialis *cubense* (FOC) race 4 suggesting a possible role of RGC2 in resistance to this agronomically important pathogen. This chapter presents the isolation and characterisation of the full cDNA sequences of these banana RGCs. The open reading frames (ORFs) of RGC2 and RGC5 were predicted to encode proteins that showed the typical structure of the non-TIR (homology to Toll/interleukin-1 receptors) NBS (nucleotide binding site) LRR (leucine rich repeat) class of plant disease resistance proteins. Homology searches using the entire ORF of RGC2 and RGC5 revealed significant sequence similarity to the *Fusarium* resistance gene *I2* from tomato. Interestingly, the phylogenetic analysis showed that RGC2 and RGC5 were grouped within the same phylogenetic clade as the tomato *I2* gene and the recently cloned *Fusarium* resistance gene *Fom-2* from melon. These findings and the previous correlation of RGC2 in FOC race 4 resistance (Chapter 4) make RGC2 and RGC5 a pair of very interesting resistance gene candidates that could be associated to *Fusarium* resistance in banana. Finally, different expression cassettes with the RGC2 ORF were constructed. In one of the constructs, the expression is driven by a RGC2 putative promoter region isolated as part of this study. These constructs will be used to test the role of RGC2 in FOC race 4 resistance.

5. 2 Introduction

Banana is the most important fruit crop in the tropics. Over 100 million metric tons of fruit are produced annually and the amount that enters international commerce is worth ~$5 billion US dollars per year (Ploetz 2005). Locally consumed fruit are major staple foods in Africa and Latin America, and in countries such as Rwanda and Uganda, per capita consumption approaches 1kg/day (Ploetz 2005). Diseases caused by pathogens are major constraints in the production of this important crop and Panama disease is one of the most devastating. This disease is caused by the soil-borne fungus *Fusarium oxysporum* forma specialis (f. sp.) *cubense* (FOC), which systemically colonizes the xylem of susceptible banana plants through the roots causing a lethal vascular wilt (Ploetz and Pegg 2000). FOC is one of more than 120 formae speciales of *F. oxysporum* causing vascular wilts in flowering plants and severe economical losses every year in numerous crops (Di Pietro et al. 2003). In general, effective chemical control measures do not exist for this pathogen, therefore the use of resistant varieties is the only practical strategy for controlling *F. oxysporum* in the field (Di Pietro et al. 2003; Ploetz and Pegg 2000). Based on pathogenicity to host cultivars, four physiological races of FOC are recognised in banana (Ploetz and Pegg 2000). Race 1 destroyed the banana export industry in the 1950s which was based on the cultivar ‘Gros Michel’. Consequently, the banana export industry was saved with the introduction of Cavendish cultivars which were naturally resistant clones available at that time. Race 2 affects cooking bananas, such as ‘Bluggoe’ (ABB). Race 3 was reported to affect *Heliconia* spp. and was weakly pathogenic on ‘Gros Michel’. Finally, race 4 affects cultivars resistant to race 1 and race 2, making this race the most serious threat to the banana production worldwide. Sources of resistance to race 4 have been found in wild banana species (Ploetz and Pegg 2000), however it has been difficult to transfer this resistance into commercial cultivars by conventional breeding techniques because most banana cultivars are sterile polyploids. Genetic transformation in banana could overcome this obstacle (Becker et al. 2000; Khanna et al. 2004), however, a FOC resistance (R) gene has not been isolated.

R genes participate in gene-for-gene interactions, in which the R gene product appears to act as a receptor that recognises a product of the corresponding avirulence (Avr) gene from the pathogen, triggering signal transduction cascades that lead to defence responses to halt pathogen spread (Hammond-Kosack and Jones, 1997; Martin et al. 2003). In the absence of either R or Avr gene, no recognition occurs and the pathogen is able to colonize the host and cause disease (Hammond-Kosack and Jones, 1997; Hammond-Kosack and Parker, 2003). Most characterised R genes are predicted to encode intracellular proteins that contain nucleotide-binding sites and leucine-rich repeats domains (NBS-LRR) (Dangl and Jones, 2001; Belkhadir et al. 2004). The NBS-LRR class can be divided in two subclasses, the TIR and the non-TIR, depending on the presence of a domain at the N-terminus with homology to the *Drosophila* Toll and mammalian Interleukin-1 receptors (TIR) (Meyers et al. 1999).

A classical gene-for-gene relationship has been proposed to mediate the interaction between *F. oxysporum* races and host cultivars based on dominant monogenic resistance traits against known races (Di Pietro et al. 2003). This has been confirmed with the cloning of the *I2* gene from tomato (*Lycopersicon esculentum*), which is capable to confer resistance to *F. oxysporum* f.sp. *lycopersici* (FOL) race *2* but not to the other two races of FOL, race 1 and 3, whose resistance is located at different loci of the tomato genome (Sela-Buurlage et al. 2001). A map-based cloning approach was used to identify the *I2* gene, which is located within a cluster of seven paralogues on the long arm of chromosome 11 (Simons et al. 1998). The *I2* gene belongs to the non-TIR-NBS-LRR class of resistance genes. A second *Fusarium* resistance gene has been isolated recently, the *Fom-2* gene from melon (*Cucumis melo*), which confers resistance to *F. oxysporum* f. sp. *melonis* race 0 and 1 (Joobeur et al. 2004). This gene also belongs to the non-TIR-NBS-LRR class of R genes, but in contrast to the tomato *I2* gene, the *Fom-2* locus contains a single copy of a resistance gene and six retroelement-like sequences. The fact that these two *Fusarium* resistance genes belong to the non-TIR-NBS-LRR class of R genes suggests a similar

Fusarium resistance mechanism shared between the families Solanaceae and Cucurbitaceae. This observation opens the possibility to find *Fusarium* resistance genes in other plant families with a similar structure to the *I2* or *Fom-2* genes. Although the cloning of two *Fusarium* resistance genes from different plant families has been achieved, to our knowledge there are no published reports that explain their phylogenetic relationships or how their *Fusarium* recognition specificity may have arose. This information could be highly valuable in the quest for further *Fusarium* R genes in plants

The previous chapter presented the isolation and characterisation of five resistance gene candidates (RGCs) of the NBS type from the FOC resistant banana *Musa acuminata* ssp. *malaccensis*. Interestingly, two of these RGCs (RGC2 and RGC5) showed a significant sequence similarity to the tomato *I2* resistance gene. Moreover, the expression of RGC2 was found to correlate with FOC race 4 resistance. These data suggest a possible role of RGC2 and RGC5 in *Fusarium* resistance in banana. The present chapter reports the isolation and sequence characterisation of the full open reading frame (ORF) of RGC2 and RGC5.

5.3 Materials and methods

5.3.1 Plant material

A *Musa acuminata* ssp. *malaccensis* plant (accession 850) which shows resistance to FOC subtropical race 4 was provided by Dr. Mike Smith (Queensland Department of Primary Industries, Nambour, Australia). The plant was grown in an incubator chamber and used as the source for harvesting leaves and root tissues. Harvested tissues of a 4-month-old plant were immediately frozen in liquid nitrogen and stored at – 80^{o}C until nucleic acid extraction.

5.3.2 Isolation of the 3' cDNA ends of RGC2 and RGC5 using rapid amplification of cDNA ends (RACE)

The 3' ends of the cDNAs encoding RGC2 and RGC5 were isolated using the GeneRacer^{TM} Kit (Invitrogen^{TM} Life Technologies) according to the manufacturer's protocol. Briefly, the RNA was extracted from root tissue as described by Schuler and Zielinski (1989) and the cDNA synthesis was carried out according to the GeneRacer^{TM} protocol. The 3'RACE was performed with the forward specific primers 5'-AGCCTGTTAGCCCCATTAGATGC-3' for RGC2 and 5'-CTGCTACCAAGGTGGA

ACAATC-3' for RGC5 in combination with the GeneRacer^{TM} 3' primer, respectively. PCR products were subjected to a second round of PCR, using the nested specific primers 5'-GTCACCGGCACGATGGAGCCATAC-3' for RGC2, and 5'-ACTCGAGACACAAAGATTGCCAGC-3' for RGC5 in combination with the GeneRacer^{TM} nested 3'primer. PCR products were cloned and sequenced as described below.

5.3.3 Cloning and sequencing

PCR products were separated by electrophoresis in a 1.5% TAE agarose gel stained with ethidium bromide. The expected bands were eluted and purified with the High Pure PCR purification kit (Roche) according to the manufacturer's protocol. PCR-purified products were cloned into the pGEM-T Easy vector (Promega) and transformed by heat-shock into *Escherichia coli*

JM109 competent cells (Promega) following the manufacturer's instructions. Plasmids were extracted using the High Pure Plasmid Isolation Kit (Roche) and sequenced using the BigDye terminator sequencing kit version 3.1 (Applied Biosystems). Capillary separation was carried out using an ABI 3730 automatic sequencer (Applied Biosystems) at the Australian Genome Research Facility (University of Queensland). Selected clones were sequenced on both strands.

5.3.4 Amplification of the entire cDNA ORF of RGC2 and RGC5 by RT-PCR

The full intact ORFs of both RGC2 and RGC5 were isolated by RT-PCR from mRNA extracted from roots of *M. acuminata* ssp. *malaccensis* line 850 using specific forward and reverse primers that spanned the initiation and termination codons of each gene. For RGC2, the forward primer 5'-ATGGCTGGTGTCACATCACAGG-3' and the reverse primer 5'-TCAGGTGGTGCTACAGCGACATGG-3' were used. For RGC5 the forward primer 5'-ATGTCGACGGCGCTAGTAATCG-3' was used in combination with the reverse primer 5'-CTATGCATAAACTGGGTGAAATCG-3'. PCR reactions were performed in a total volume of 50 μl containing 300 μM of dNTPs, 0.2 μM of each specific forward and reverse primer, 1 μl of cDNA, 1x PCR buffer and 3.5 U of Expand^{TM} Long Template polymerase (Roche). PCR conditions were 95^{o}C for 3 min, followed by 35 cycles of 95^{o}C for 30 s, 55^{o}C for 30 s and 68^{o}C for 4 min; and included an additional 10 min extension at 68^{o}C. The PCR products were cloned and sequenced as described above.

5.3.5 Isolation of a putative promoter region of RGC2 by PCR genome walking

A putative promoter fragment of RGC2 was cloned using the genome walking protocol of Siebert et al. (1995). Briefly, genomic DNA from *M. acuminata* ssp. *malaccensis* (line 850) was digested with PvuII and EcoR V restriction enzymes. Genome walker adapters (see chapter 3) were ligated to the digested DNA and PCR reactions were carried out on each genome walker "library" using a gene-specific reverse primer (5'-

CACGAGAGACTTGTCAGTCATCCG-3) and the adapter primer (AP1, see chapter 3). PCR products were subjected to a second round of PCR, using the nested adapter primer (AP2, see chapter 3) and the nested, gene-specific reverse primer (5'-CTGCCTGTGATGTGACAC-3'). From the second PCR reaction, an approximately 2.3 kb fragment was isolated from the EcoR V library, cloned into pGEMT-Easy vector (designated pGEMT-RGC2PP) and sequenced as described above.

.

5.3.6 Bioinformatic analysis

RGC2 and RGC5 cDNA sequences were assembled and edited using the Lasergen software package version 4.03 (DNASTAR, Madison, WI). Homology searches were carried out using the BLASTX program (Altschul et al. 1997) at the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). Percentages of identity and similarity between sequences were determined by the ALIGN program through the European Bioinformatic Institute server (www.ebi.ac.uk). Potential coiled-coil structures were predicted by the COILS program (Lupas, 1996) using the EBI server. Motif searches were performed by the PROSITE program (Hulo et al. 2006) and the hydropathy plot was performed according to Kyte and Doolittle (1982) using the ExPASy proteomics server (www.us.expasy.org). Protein sequences were aligned by using the ClustalX program version 1.81 (Thompson et al. 1997) with the default settings. A phylogenetic tree was constructed by the neighbour-joining method (Saitou and Nei, 1987) using the NJ algorithm implemented in the Molecular Evolutionary Genetics Analysis (MEGA) software version 2.1 with the Poisson correction (Kumar et al. 2001). Bootstrapping (10,000 replicates) was used to evaluate the degree of support for a particular grouping pattern in the phylogenetic tree. Putative cis-acting regulatory elements on the putative promoter region of RGC2 were predicted using the signal scan program (Higo et al. 1999) at the plant cis-acting regulatory DNA elements (PLACE) database (www.dna.affrc.go.jp/htdocs/PLACE).

5.3.7 RGC2 constructs for banana transformation

In order to test the role of RGC2 in FOC resistance by genetic complementation, different expression cassettes for this banana RGC were constructed. The cDNA ORF of RGC2 was excised from pGEMT-Easy as a BamHI/Not I fragment and cloned into the pGEM-NOST plasmid containing the NOS 3' untranslated region (3'UTR) from pBI121 (Clontech). This construct was designated as pGEM-RGC2ORF-NOST. The RGC2ORF-NOST fusion was excised from the plasmid as a BamHI/Sal I fragment and cloned into the binary transformation vector pCAMBIA 2200 designated pCAMBIA 2200 (RGC2ORF-NOST). In order to generate different expression cassettes for RGC2, different promoters were cloned into the pCAMBIA 2200 (RGC2ORF-NOST) plasmid. First, the putative RGC2 promoter region was amplified by PCR from the pGEMT-RGC2PP plasmid using the forward primer 5'-ATACGAggtaccATCATGTATGCATATATATGAAACATTGG-3' and reverse primer 5'-TATTCGgatccTCTTCTTCTCTGATGGCAAATAGAG-3'. The PCR product was then cloned into the pCAMBIA 2200 (RGC2ORF-NOST) plasmid as a KpnI/BamHI fragment. This construct was designated as pCAMBIA 2200 (RGC2PP-RGC2ORF-NOST). Previous characterised promoters were also amplified by PCR as KpnI/BamHI fragments from different pGEM-derived plasmids. These promoters were: the polyubiquitin-1 (ubi-1) promoter from maize (Christensen et al. 1996) that drives a high level of gene expression in monocotyledonous plants, the nopaline synthase promoter (NOS) from *Agrobacterium* that drives a moderate level of gene expression in comparison to the strong *Cauliflower mosaic virus* 35S promoter (CaMV 35S) (Sanders et al. 1987), and finally the *Banana bunchy top virus* (BBTV) BT1 promoter that drives a low level of gene expression in banana vascular tissues (Dugdale et al. 2000). The following primer combinations were used to amplify each promoter:

1. for the Ubi-1 promoter (Christensen et al. 1996), the forward primer 5'-ATACGAggtaccCTGCAGTGCAGCGTGACC-3' and the reverse primer 5'-ATACGAggtaccGTGCAGAAGTAACACCAAAC-3 were used

2. for the nos promoter (Shaw et al. 1984), the forward primer 5'-ATACGAggtaccGGGTTTCTGGAGTTTAATGAGC-3' and the reverse primer 5'-TATTCGggatccTGCAGATTATTTGGATTGAGAGTG-3'
3. for the BT1 promoter (Dugdale et al. 2000), the forward primer 5'-ATACGAggtaccACAAGTAATGACTTTACAGC-3' and the reverse primer 5'- TATTCGggatccTCTGTCGTCGATGATGATCTTG-3' were used;

These constructs were designated as pCAMBIA 2200 (Ubi-1-RGC2ORF-NOST), pCAMBIA 2200 (BT1-RGC2ORF-NOST) and pCAMBIA 2200 (NOS-RGC2ORF-NOST), respectively. PCR reactions were performed in a total volume of 50 μl containing 300 μM of dNTP, 0.2 μM of each specific primer forward and reverse, 1 μl of cDNA, 1x PCR buffer and 3.5 U of Expand^{TM} Long Template polymerase (Roche). PCR conditions were 95^{o}C for 3 min, followed by 35 cycles of 95^{o}C for 30 s, 55^{o}C for 30 s and 68^{o}C for 4 min; and included an additional 10 min extension at 68^{o}C.

5.4 Results

5.4.1 Isolation of the banana RGC2 and RGC5 cDNAs and characterization of their predicted protein structures.

The expression of RGC2 and RGC5 was found to be constitutive in both leaf and root tissues of *M. acuminata* ssp. *Malaccensis* (Chapter 4). Because the roots are the primary contact site of *F. oxysporum* and the resistance reaction takes place in this tissue (Mes et al. 2000; Di Pietro et al. 2003), the cDNAs of RGC2 and RGC5 were isolated from roots. Primers specific for RGC2 and RGC5 were used to amplify the 3' ends of these banana RGCs by RACE. Three clones from the 3'RACE of both RGC2 and RGC5 were isolated and sequenced. This sequence information in combination with sequence information previously obtained by 5'RACE (chapter 4) facilitated isolation of the complete open reading frames (ORFs) of the cDNAs encoding RGC2 and RGC5.

In the case of RGC2, an expected ~3.7 kb PCR product was amplified. The band was cloned, sequenced and a large ORF of 3699 nucleotides (nt) was determined (Figure 5.1). A second sequence highly similar to RGC2 (97.8% identity at nucleotide level) was also identified but was not investigated further due to the presence of a premature stop codon that disrupts the ORF and therefore would result in a truncated polypeptide (Figure 5.1). The RGC2 ORF encoded a predicted polypeptide of 1232 amino acids (Figure 5.2) with an estimated molecular weight and isoelectric point of 139.61 kDa and 7.1, respectively. The putative RGC2 amino acid sequence shows the highest level of similarity to an uncharacterised NBS-LRR sequence (accession number NP915900) from *Oryza sativa* (37% identity, 59% similarity and E value = 0) in homology searches of the Genbank database using the BLASTX algorithm (Altschul et al. 1997). One of the most similar characterised R genes was the *Fusarium* resistance gene *I2* from *Lycopersicon esculentum* (31% identity, 48% similarity and E value = $1e^{-123}$). Figure 5.3 illustrates the predicted domains within RGC2, which belongs to the non-TIR-NBS-LRR class of R proteins (Chapter 4). The C-terminal region comprises a series of 27 leucine-rich repeats of irregular size that can be

aligned according to the consensus sequence LxxLxxLxxLxLxxC/N/Sx(x)LxxLPXX (where x is any residue) observed in other cytoplasmic R proteins (Jones and Jones, 1997). As noted in Chapter 4, the N-terminal region of the predicted protein contains a potential coiled-coil region (Figure 5.4) and six conserved motifs indicative of a nucleotide-binding site (NBS). The C-terminal region also contains a potential coiled-coil region between amino acids 600 and 800 (Figure 5.4). According to the hydropathy plot analysis, RGC2 deduced protein is predicted to be a cytoplasmic protein (Figure 5.5). PROSITE analysis identified 11 N-glycosylation, three tyrosine sulfation, five cAMP-and cGMP-dependent protein kinase phosphorylation, 14 protein kinase C phosphorylation, 19 casein kinase II phosphorylation, one tyrosine kinase phosphorylation, 11 N-myristoylation and one amidation putative sites in the predicted RGC2 amino acid sequence.

In the case of RGC5, an expected ~4.3 kb PCR product was amplified. This PCR product was cloned, sequenced and a large ORF of 4329 nt was determined (Figure 5.6). A second sequence highly similar to RGC5 (99.3% identity at nucleotide level) was also identified but was not investigated further due to the presence of a premature stop codon that disrupts the ORF and therefore would result in a truncated polypeptide (Figure 5.6). The RGC5 ORF encoded a predicted polypeptide of 1442 amino acids (Figure 5.7) with an estimated molecular weight and isoelectric point of 163.73 kDa and 7.2, respectively. The putative RGC5 protein showed the highest sequence similarity to an uncharacterised NBS-LRR sequence (accession number AAS49214) from *Glycine max* (33% identity, 49% similarity and E value = $6e^{-128}$). One of the most similar characterised R genes was the *Fusarium* resistance gene *I2* from *Lycopersicon esculentum* (31% identity, 48% similarity and E value = $7e^{-130}$). Figure 5.8 illustrates predicted domains in RGC5. Like RGC2, the C-terminal region of RGC5 comprises a series of leucine-rich repeats of irregular size; in this case 34 are predicted. As noted in Chapter 4, the N-terminal region of the predicted protein contains a potential coiled-coil region (Figure 5.9) and six conserved motifs indicative of

a nucleotide-binding site (NBS). Unlike RGC2, RGC5 did not show a potential coiled-coil region in the LRR region (Figure 5.9). According to hydropathy plot analysis, RGC5 deduced protein is predicted to be a cytoplasmic protein (Figure 5.10). PROSITE analysis identified seven N-glycosylation, two tyrosine sulfation, one cAMP- and cGMP-dependent protein kinase phosphorylation, 16 protein kinase C phosphorylation, 22 casein kinase II phosphorylation, three tyrosine kinase phosphorylation, 16 N-myristolylation, one amidation and one RGD cell attachment sequence putative sites in the predicted RGC5 amino acid sequence.

```
RGC2       1 ATGGCTGGTGTCACATCACAGGCAGCGGCGGTGTTCTCCCTGGTGAATGAAATCTTTAAC
RGC2fs     1 ATGGCTGGTGTCACATCACAGGCAGCGGCGGTGTTCTCCCTGGTGAATGAAATCTTTAAC
             ************************************************************

RGC2      61 CGGTCCATCAATTTGATCGTCGCGGAACTCCGGTTGCAGTTGAATGCGAGAGCCGAGCTG
RGC2fs    61 CGGTCCATCAATTTGATCGTCGCGGAACTCCGGTTGCAGTTGAATGCGAGAGCCGAGCTG
             ************************************************************

RGC2     121 AACAATCTGCAGAGAACACTATTGAGGACTCACTCTCTGCTCGAGGAGGCAAAGGCGAGG
RGC2fs   121 AACAATCTGCAGAGAACACTATTGAGGACTCACTCTCTGCTCGAGGAGGCAAAGGCGAGG
             ************************************************************

RGC2     181 CGGATGACTGACAAGTCTCTCGTGCTGTGGCTGATGGAGCTCAAGGAATGGGCCTACGAC
RGC2fs   181 TGGATGACTGACAAGTCTCTCGTGCTGTGGCTGATGGAGCTCAAGGAATGGGCCTACGAC
             ************************************************************

RGC2     241 GCCGACGACATCCTCGACGAGTACGAGGCCGCAGCAATCCGACTGAAGGTAACACGCTCG
RGC2fs   241 GCCGACGACATCCTCGACGAGTACGAGGCCGCAGCAATCCGACTGAAGGTAACTCGCTCG
             ************************************************** ******

RGC2     301 ACCTTCAAACGTCTTATCGATCATGTGATTATAAATGTTCCATTAGCGCACAAAGTAGCA
RGC2fs   301 ACCTTCAAACGTCTTATCGATCATGTGATTATAAATGTTCCATTAGCGCACAAAGTAGCA
             ************************************************************

RGC2     361 GACATCAGGAAAAGGTTGAACGGGGTCACTCTTGAGAGGGAGCTAAATCTGGGTGCGCTG
RGC2fs   361 GACGTCAGGAAAAGGTTGAACTGGGTCACTCTTGAGAGGGAGCTAAATCTGGGTGCGCTG
             *** ***************** ************************************

RGC2     421 GAAGGGTCGCAGCCGCTTGATTCCACGAAAAGAGGTGTGACCAC---TTCTCTTCTGACT
RGC2fs   421 GAAGGGTCGCAGCCGCTTGATGCCACAAAAAGAGGTGTGACCACCACTTCTCTTCTGACT
             ******************** **** *****************   *************

RGC2     478 GAATCTTGTATTGTCGGGCGAGCTCAAGATAAGGAGAATTTGATTCGGTTGCTGTTGGAG
RGC2fs   481 GAATCTTGTATTGTCGGGCGAGCTCAAGATAAGGAGAAATTGATTCGGTTGCTGTTGGAG
             ************************************ **********************

RGC2     538 CCCAGCGATGGGGCGGTTCCTGTTGTTCCTATAGTTGGATTAGGAGGGGCAGGGAAGACG
RGC2fs   541 CCCAGCGATGGGGCGGTTCCTGTTGTTCCTATAGTTGGATTAGGAGGGGCAGGGAAGACG
             ************************************************************

RGC2     598 ACTCTGTCTCAGCTTATCTTTAATGACAAGAGAGTGGAGGAGCATTTCCCATTGAGAATG
RGC2fs   601 ACTCTGTCTCAGCTTATCTTTAATGACAAGAGAGTGGAGGAGCATTTCCCATTGAGAATG
             ************************************************************

RGC2     658 TGGGTGTGTGTGTCTGACGATTTTGATGTGAAGAGAATTACTAGAGAGATCACAGAGTAC
RGC2fs   661 TGGGTGTGTGTGTCTGACGATTTTGATGTGAAGAGAATTACTAGAGAGATCACAGAGTAC
             ************************************************************

RGC2     718 GCCACCAACGGAAGGTTCATGGATCTCACCAACTTGAATATGCTTCAAGTTAATCTGAAA
RGC2fs   721 GCCACTAACGGAAGGTTGATGGATCTCACCAACTTGAATATGCTTCAAGTTAATCTGAAA
             ***** *********** ****************************************
```

```
RGC2    778 GAGGAGATAAGGGGGACGACATTTTTGCTTGTGCTGGATGATGTGTGGAACGAAGACCCC
RGC2fs  781 GAGGAGATAAGGGGGACGACATTTTTGCTTGTGCTAGACGACGTGTGGAACGAAGACCCC
            ********************************** ** ** ******************

RGC2    838 GTGAAGTGGGAAAGCCTGTTAGCCCCATTAGATGCCGGAGGACGGGGAAGCGTGGTCATT
RGC2fs  841 GTGAAGTGGGAAAGCCTGTTAGCCCCATTAGATGCCGGAGGACGGGGAAGCGTGGTCATT
            ************************************************************

RGC2    898 GTGACGACACAGAGCAAAAAGGTCGCCGATGTCACCGGCACGATGGAGCCATACGTTCTC
RGC2fs  901 GTGACGACACAGAGCAAAAAGGTCGCCGATATCACCGGCACGATGGAGCCATACGTTCTC
            ****************************** ****************************

RGC2    958 GAGGAGTTAACGGAGGATGACAGTTGGTCACTCATCGAGAGTCACTCCTTCAGGGAGGCG
RGC2fs  961 GAGGAGTTAACGGAGGATGACAGTTGGTCACTCATCGAGAGTCACTCCTTCAGGGAGGCG
            ************************************************************

RGC2   1018 AGCTGCTCTAGTACAAATCCTAGAATGGAAGAGATCGGGAGGAAGATAGCCAAGAAGATC
RGC2fs 1021 AGCTGCTCTAGTACAAATCCTAGAATGGAAGAGATCGGGAGGAAGATAGCCAAGAAGATC
            ************************************************************

RGC2   1078 AGTGGCCTACCTTACGGAGCAACAGCAATGGGGAGATATCTAAGATCTAAGCACGGAGAA
RGC2fs 1081 AGTGGCCTACCTTATGGAGCAACAGCAATGGGGAGATATCTAAGATCTAAGCACGGAGAA
            ************** ********************************************

RGC2   1138 AGCAGCTGGAGAGAAGTCTTGGAAACTGAGACTTGGGAGATGCCACCGGCTGCAAGTGAT
RGC2fs 1141 AGCAGCTGGAGAGAAGTCTTGGAAGCTGAGACTTGGGAGATGCCACCGGCTGCAAGTGAT
            ************************ **********************************

RGC2   1198 GTGTTATCCGCTCTAAGGAGAAGTTACGACAATCTACCCCCTCAGCTGAAGCTCTGTTTT
RGC2fs 1201 GTGTTATCCGCTCTAAAGAGAAGTTACGACAATCTAGCCCCTCAGCTGAAGCTCTGTTTT
            **************** ****************** ************************

RGC2   1258 GCCTTCTGTGCTCTGTTTACAAAGGGCTACAGGTTTCGAAAGGATACACTGATCCACATG
RGC2fs 1261 GCCTTCTGTGCTCTGTTTCCAAAGGGCTACAGGTTTCGAAAGGATACACTGATCCACATG
            ****************** ****************************************

RGC2   1318 TGGATAGCTCAAAATTTGATTCAATCAACAGAGTCGAAAAGATCGGAGGACATGGCAGAA
RGC2fs 1321 TGGATAGCTCAAAATTTGATTCAATCAATAGAGTCGAAAAGATCGGAGGACATGGCAGAA
            **************************** ******************************

RGC2   1378 GAATGCTTTGATGATTTGGTGTGCAGATTCTTCTTTCGGTACTCCTGGGGCAACTATGTG
RGC2fs 1381 GAATGCTTTGATGATTTGGTGTGCAGATTCTTCTTTCGGTACTCCTGGGGCAACTATGTG
            ************************************************************

RGC2   1438 ATGAATGACTCAGTCCATGACCTCGCTCGATGGGTTTCATTGGATGAATATTTTCGAGCA
RGC2fs 1441 ATGAATGACTCAGTCCATGACCTCGCTCGATGGGTTTCATTGGATGAATATTTTCGAGCA
            ************************************************************

RGC2   1498 GATGAAGACTCACCATTGCATATTTCAAAGCCAATTCGTCATTTGTCATGGTGCAGTGAA
RGC2fs 1501 GATGAAGACTCACCATTGCATATTTCAAAGCCAATTCGTCATTTGTCATGGTGCAGTGAA
            ************************************************************

RGC2   1558 AGAATAACCAAT------GTTCTTGAGGATAATAACACTGGTGGAGATGCTGTCAATCCG
RGC2fs 1561 AGTATAACCAATCTTCCTGTTCTTGAGGATAATAACACAGGTGGAGATGCTGTCAACCCG
            ** *********      ******************** ***************** ***

RGC2   1612 CTCAGCAGTTTGCGCACTCTCCTTTTCTTAGGCCAATCTGAGTTCCGGTCGTATCATCTT
RGC2fs 1621 CTCAGCAGTTTGCGCACTCTCCTTTTCTTAGGCCAATCTGAGTTCCAGTCGTATCATCTT
            ********************************************* **************

RGC2   1672 CTTGATAGAATGTTCAGGATGTTGAGCCGAATCCGTGTTTTGGATTTCAGCAACTGCCTTC
RGC2fs 1681 CTTGATACAATGTTCAGGATGTTGAGCCGAATCCGTGTTTTGGATTTCAGCAACTG--TC
            ******* ************************************************  **

RGC2   1732 ATAAGAAATTTGCCTTCTTCGGTTGGAAATCTGAAACATCTGCGTTACCTGGGCCTGTCT
RGC2fs 1739 ATAAGAAAGTTGCCTTCTTCGGTTGGAAATCTGAAACATCTGCGTTACCTGGGCCTGTCT
            ******** ****************************************************

RGC2   1792 AATACGAGAATTCAAAGGTTGCCGGAGTCTGTAACACGTCTTTGCCTCCTTCAGACATTG
RGC2fs 1799 AACACGAGAATTCAAAGGTTGCCGGAGTCTGTAACACGTCTTTGCCTCCTTCAGACATTG
            ** ************************************************************

RGC2   1852 CTACTAGAGGGCTGTGAACTGTGCAGGTTACCAAGAAGCATGAGCAGGCTCGTCAAACTG
RGC2fs 1859 CTACTAGAGGGCTGTGAACTGTGCAGGTTACCAAGAAGCATGAGCAGGCTTGTCAAACTG
            ************************************************* *********
```

```
RGC2   1912 AGGCAGCTCAAAGCAAATCCAGATGTAATTGCCGACATAGCCAAAGTCGGGAGATTGATC
RGC2fs 1919 AGGCAGCTCAAAGCAAATCCAGATGTAGTTGCCGACATAGCCAAAATCGGGACATTGATC
            ************************** ********** ****** *******

RGC2   1972 GAACTTCAAGAGCTGAAAGCCTATAATGTTGACAAGAAAAAAGGACATGGGATTGCAGAG
RGC2fs 1979 GAACTTCAAGAGCTGAAAGCCTATAATGTTGACAAGAAAAAAGGACATGGGATTGCAGAG
            ************************************************************

RGC2   2032 CTAAGTGCAATGAATCAGCTTCACGGTGATCTTTCCATTAGAAACCTTCAAAATGTAGAG
RGC2fs 2039 CTAAGTGCAATGAATCAGCTTCACGGTGATCTTTCCATTAGAAACCTTCAAAATGTAGAG
            ************************************************************

RGC2   2092 AAAACGCGAGAGTCTCGGAAGGCGAGGTTGGACGAGAAACAGAAGCTTAAGCTCTTGGAT
RGC2fs 2099 AAAACGCGAGAGTCTCGGAAGGCGAGGTTGGACGAGAAACAGAAGCTTAAGCTCTTGGAA
            ************************************************************

RGC2   2152 CTGCGATGGGCTGACGGTAGGGGTGCCGGAGAATGTGATCGTGACAGGAAAGTTCTTAAA
RGC2fs 2159 CTGCGATGGGCTGAGGGTAGGGGTGCCGGAGAATGTGATCGTGACAGCAAAGTTCTTGAA
            ************** ******************************** ********* **

RGC2   2212 GGCCTCCGACCACATCCAAACCTGAGAGAATTGAGTATCAAATACTACGGAGGCACTTCA
RGC2fs 2219 GGCCTCCGACCACATCCAAACCTGAGAGAATTGAGTATCAAATACTACGGAGGCACTTCA
            ************************************************************

RGC2   2272 TCTCCGAGTTGGATGACGGATCAGTATCTGCCCAACATGGAAACGATTCGCCTGCGTAGC
RGC2fs 2279 TCTCCGAGTTGGATGACGGATCAGTATCTGCCCAACATGGAAACGATTCGCCTGCGTAGC
            ************************************************************

RGC2   2332 TGCGCAAGGTTGACGGAACTCCCATGTCTCGGTCAGCTGCATATCCTTAGACATTTGCAC
RGC2fs 2339 TGCGCAAGGTTGACGGAACTCCCATGTCTCGGTCAGCTGCATCTCCTTAGACATTTGCAC
            ****************************************** ****************

RGC2   2392 ATCGATGGGATGTCCCAAGTGAGACAAATTAATCTGCAATTTTATGGCACCGGAGAAGTT
RGC2fs 2399 ATCGATGGGATGTCCCAAGTGAGACTAATCAATCTGCAATTTTATGGCACCGGAGAAGTT
            ************************ *** ******************************

RGC2   2452 TCAGGTTTTCCATTGCTGGAGCTCCTGAACATACGTCGCATGCCCAGTCTGGAGGAATGG
RGC2fs 2459 TCAGGTTTTCCATTGCTGGAGCTCCTGAACATATGTCGCATGCCCAGTCTGGAGGAATGG
            ********************************** ************************

RGC2   2512 TCGGAACCACGGAGAAACTGTTGCTACTTCCCTCGCCTCCATAAACTGCTGATCGAGGAT
RGC2fs 2519 TCGGAACCACGGAGAAACTGTTGCTACTTCCCTCGCCTCCATAAGCTGCTGATCGAGGAT
            ****************************************** ***************

RGC2   2572 TGTCCCAGGCTCAGGAATCTGCCCTCCCTCCCACCAACACTGGAAGAACTAAGGATATCA
RGC2fs 2579 TGTCCTAGCCTCAGGAATCTGCCCTCCCTCCCACCAACACTGGAAGAACTAAGGATATCA
            ***** ** ************************************************

RGC2   2632 AGAACAGGACTAGTTGATCTTCCAGGATTCCATGGAAACGGTGATGTGACGACGAATGTT
RGC2fs 2639 AGAACAGGACTAGTTGATCTTCCAGGATTCCATGGAAACGGTGATGTGATGACAAATGTT
            ********************************************** *** ******

RGC2   2692 TCCCTTTCTTCTTTGCATGTTTCGGAGTGTCGAGAACTGAGATCCCTAAGCGAAGGATTG
RGC2fs 2699 TCCCTTTCTTCTTTGCATGTTTCGGAGTGTCGAGAACTGAGATCCCTAAGCGAAGGATTG
            ************************************************************

RGC2   2752 TTGCAGCACAACCTCGTCGCCCTCAAGACAGCGGCATTTACCGATTGTGATTCTCTTGAG
RGC2fs 2759 TTGCAGCACCACCTCGTCGCGCTCAAGACAGCGGCATTTACCGATTGTGATTCTCTTGAA
            ********* ********** ************************************

RGC2   2812 TTTTTGCCGGCGGAAGGATTCAGAACAGCCATTTCACTTGAATCATTGATAATGACTAAT
RGC2fs 2819 TTTTTGCCCGCGGAAGGATTCAGAACAGCCATTTCACTTGAATCATTGATAATGACTAAT
            ******** ****************************************************

RGC2   2872 TGTCCACTGCCTTGCAGTTTTCTTTTGCCTTCCTCTCTCGAGCATCTAAAGTTGCAGCCA
RGC2fs 2879 TGTCCACTGCCTTGCAGTTTTCTTTTGCCTTCCTCTCTCGAGCATCTAAAGTTGCAACCA
            ************************************************************ ***

RGC2   2932 TGCCTCTATCCAAACAACAATGAGGATTCACTGTCAACATGCTTCGAGAACCTCACATCT
RGC2fs 2939 TGTCTCTATCCAAACAACAACGAGGATTCACTGTCAACATGCTTCCAGAACCTCACATCT
            ** ***************** ********************** ***************

RGC2   2992 CTTTCCTTCTTGGACATCAAAGATTGTCCAAATCTGTCATCATTTCCACCGGGTCCTCTA
RGC2fs 2999 CTTTCCTTCTTGGACATCAAAGATTGCCCAAATCTGTCATCATTTCTTCCTGGTCCTCTA
            ************************** ******************  ** *********
```

```
RGC2   3052 TGTCAGCTATCAGCACTCCAACATTTGTCCCTCGTCAATTGCCAGAGGCTACAATCTATT
RGC2fs 3059 TGTCAGCTATCAGCACTCCAACATTTGTCTCTCGTCAATTGCCAGAGGCTACAATCTATT
            *************************** ****************************

RGC2   3112 GGCTTCCAGGCACTCACCTCCCTCGAAAGCTTGACAATTCAGAACTGCCCTCGCCTCACC
RGC2fs 3119 GGCTTCCAGGCACTCACCTCCCTCGAAAGCTTGACAATTCAGAACTGCCCTCGCCTCACC
            ************************************************************

RGC2   3172 ATGTCACACAGTTTGGTTGAGGTGAATAACTCTTCCGATACAGGGCTCGCGTTTAATATC
RGC2fs 3179 ATGTCACACAGTTTGGTTGAGGTGAATAACTCTTCCGATACGAGGCTCGCGTTTAATATC
            ************************************  ******************

RGC2   3232 ACTCGATGGATGCGCAGACGAACAGGTGACGACGGCTTGATGCTCAGACACCGAGCACAA
RGC2fs 3239 ACTCGATGGATGCGCAGACGAACAGGTGACGACGACTTGATGCTCAGACACCGAGTACAA
            ********************************** ******************* ****

RGC2   3292 AATGATTCATTTTTCGGGGGACTTCTGCAACACCTCACCTTCCTCCAGTTTCTAAAGATC
RGC2fs 3299 AATGATTCATTTTTCGGGGGGCTTCTGCAACATCTCACCTTCCTCCAGTTTCTAAAGATC
            ******************* ********** ***************************

RGC2   3352 TGCCAGTGTCCACAACTCGTAACCTTCACCGGCGAAGAGGAAGAGAAGTGGAGAAACCTT
RGC2fs 3359 TGCCAGTGTCCACAACTTGTAACCTTCACCGGCGAAGAGGAAGAGAAGTGGAGAAACCTT
            ***************** ****************************************

RGC2   3412 ACTTCTCTTCAAATTCTGCACATCGTTGATTGTCCAAACCTGGAGGTACTGCCTGCAAAC
RGC2fs 3419 ACTTCTCTTCAAATTCTGCACATCGTTGATTGTCCAAACCTGGAGGTACTGCCTGCAAAC
            ************************************************************

RGC2   3472 TTGCAAAGCCTCTGCTCCCTCAGCACCTTGTACATCGTCAGATGCCCAAGAATCCATGCG
RGC2fs 3479 TTGCAATGCCTCTGCTCTCTCAGCACCTTGTACATCGTCAGATGCCCAAGAATCCATGCG
            ****** ********** ****************************************

RGC2   3532 TTTCCTCCCGGAGGTGTCAGCATGTCCCTGGCACATTTGGTCATCCATGAATGCCCTCAG
RGC2fs 3539 TTTCCTCCCCGAGGTGTCAGCATGTCCCTGGCACATTTGGTCATCCATGAATGCCCTCAG
            ********* ************************************************

RGC2   3592 CTGTGTCAGCGATGTGATCCACCGGGAGGTGATGATTGGCCCTTAATAGCTAATGTACCA
RGC2fs 3599 CTGTGTCAGCGATGTCATCCACGAGGAGGTAATGATTGGCCCTTAATAGCTAATGTACCA
            *************** ******  ****** ****************************

RGC2   3652 AGAATATGTCTTGGAAGGACTCATCCATGTCGCTGTAGCACCACCTGA  3699
RGC2fs 3659 AGAATATGTCTTGGAAGGACTCATCCATGACGCTGTAGCACCACCTGA  3706
            **************************** ******************
```

Figure 5.1 ClustalX alignment of the nucleotide sequences of RGC2 cDNA and a RGC2 cDNA homologue with a frameshift mutation (RGC2fs). The frameshift in RGC2fs occurred at position 1737 of the nucleotide sequence and it is indicated with an open box.

```
   1 atggctggtgtcacatcacaggcagcggcggtgttctccctggtgaatgaaatctttaac
  20 M  A  G  V  T  S  Q  A  A  A  V  F  S  L  V  N  E  I  F  N

  61 cggtccatcaatttgatcgtcgcggaactccggttgcagttgaatgcgagagccgagctg
  21 R  S  I  N  L  I  V  A  E  L  R  L  Q  L  N  A  R  A  E  L

 121 aacaatctgcagagaacactattgaggactcactctctgctcgaggaggcaaaggcgagg
  41 N  N  L  Q  R  T  L  L  R  T  H  S  L  L  E  E  A  K  A  R

 181 cggatgactgacaagtctctcgtgctgtggctgatggagctcaaggaatgggcctacgac
  61 R  M  T  D  K  S  L  V  L  W  L  M  E  L  K  E  W  A  Y  D

 241 gccgacgacatcctcgacgagtacgaggccgcagcaatccgactgaaggtaacacgctcg
  81 A  D  D  I  L  D  E  Y  E  A  A  A  I  R  L  K  V  T  R  S

 301 accttcaaacgtcttatcgatcatgtgattataaatgttccattagcgcacaaagtagca
 101 T  F  K  R  L  I  D  H  V  I  I  N  V  P  L  A  H  K  V  A

 361 gacatcaggaaaaggttgaacggggtcactcttgagagggagctaaatctgggtgcgctg
 121 D  I  R  K  R  L  N  G  V  T  L  E  R  E  L  N  L  G  A  L

 421 gaagggtcgcagccgcttgattccacgaaaagaggtgtgaccacttctcttctgactgaa
 141 E  G  S  Q  P  L  D  S  T  K  R  G  V  T  T  S  L  L  T  E

 481 tcttgtattgtcgggcgagctcaagataaggagaatttgattcggttgctgttggagccc
 161 S  C  I  V  G  R  A  Q  D  K  E  N  L  I  R  L  L  L  E  P

 541 agcgatggggcggttcctgttgttcctatagttggattaggaggggcagggaagacgact
 181 S  D  G  A  V  P  V  V  P  I  V  G  L  G  G  A  G  K  T  T

 601 ctgtctcagcttatctttaatgacaagagagtggaggagcatttcccattgagaatgtgg
 201 L  S  Q  L  I  F  N  D  K  R  V  E  E  H  F  P  L  R  M  W

 661 gtgtgtgtgtctgacgattttgatgtgaagagaattactagagagatcacagagtacgcc
 221 V  C  V  S  D  D  F  D  V  K  R  I  T  R  E  I  T  E  Y  A

 721 accaacggaaggttcatggatctcaccaacttgaatatgcttcaagttaatctgaaagag
 241 T  N  G  R  F  M  D  L  T  N  L  N  M  L  Q  V  N  L  K  E

 781 gagataaggggggacgacatttttgcttgtgctggatgatgtgtggaacgaagaccccgtg
 261 E  I  R  G  T  T  F  L  L  V  L  D  D  V  W  N  E  D  P  V

 841 aagtgggaaagcctgttagccccattagatgccggaggacggggaagcgtggtcattgtg
 281 K  W  E  S  L  L  A  P  L  D  A  G  G  R  G  S  V  V  I  V

 901 acgacacagagcaaaaaggtcgccgatgtcaccggcacgatggagccatacgttctcgag
 301 T  T  Q  S  K  K  V  A  D  V  T  G  T  M  E  P  Y  V  L  E

 961 gagttaacggaggatgacagttggtcactcatcgagagtcactccttcagggaggcgagc
 321 E  L  T  E  D  D  S  W  S  L  I  E  S  H  S  F  R  E  A  S

1021 tgctctagtacaaatcctagaatggaagagatcgggaggaagatagccaagaagatcagt
 341 C  S  S  T  N  P  R  M  E  E  I  G  R  K  I  A  K  K  I  S

1081 ggcctaccttacggagcaacagcaatggggagatatctaagatctaagcacggagaaagc
 361 G  L  P  Y  G  A  T  A  M  G  R  Y  L  R  S  K  H  G  E  S

1141 agctggagagaagtcttggaaactgagacttgggagatgccaccggctgcaagtgatgtg
 381 S  W  R  E  V  L  E  T  E  T  W  E  M  P  P  A  A  S  D  V

1201 ttatccgctctaaggagaagttacgacaatctacccccctcagctgaagctctgttttgcc
 401 L  S  A  L  R  R  S  Y  D  N  L  P  P  Q  L  K  L  C  F  A

1261 ttctgtgctctgtttacaaagggctacaggtttcgaaaggatacactgatccacatgtgg
 421 F  C  A  L  F  T  K  G  Y  R  F  R  K  D  T  L  I  H  M  W

1321 atagctcaaaatttgattcaatcaacagagtcgaaaagatcggaggacatggcagaagaa
 441 I  A  Q  N  L  I  Q  S  T  E  S  K  R  S  E  D  M  A  E  E
```

```
1381 tgctttgatgatttggtgtgcagattcttctttcggtactcctggggcaactatgtgatg
 461 C  F  D  D  L  V  C  R  F  F  F  R  Y  S  W  G  N  Y  V  M

1441 aatgactcagtccatgacctcgctcgatgggtttcattggatgaatattttcgagcagat
 481 N  D  S  V  H  D  L  A  R  W  V  S  L  D  E  Y  F  R  A  D

1501 gaagactcaccattgcatatttcaaagccaattcgtcatttgtcatggtgcagtgaaaga
 501 E  D  S  P  L  H  I  S  K  P  I  R  H  L  S  W  C  S  E  R

1561 ataaccaatgttcttgaggataataacactggtggagatgctgtcaatccgctcagcagt
 521 I  T  N  V  L  E  D  N  N  T  G  G  D  A  V  N  P  L  S  S

1621 ttgcgcactctccttttcttaggccaatctgagttccggtcgtatcatcttcttgataga
 541 L  R  T  L  L  F  L  G  Q  S  E  F  R  S  Y  H  L  L  D  R

1681 atgttcaggatgttgagccgaatccgtgttttggatttcagcaactgcgtcataagaaat
 561 M  F  R  M  L  S  R  I  R  V  L  D  F  S  N  C  V  I  R  N

1741 ttgccttcttcggttggaaatctgaaacatctgcgttacctgggcctgtctaatacgaga
 581 L  P  S  S  V  G  N  L  K  H  L  R  Y  L  G  L  S  N  T  R

1801 attcaaaggttgccggagtctgtaacacgtctttgcctccttcagacattgctactagag
 601 I  Q  R  L  P  E  S  V  T  R  L  C  L  L  Q  T  L  L  L  E

1861 ggctgtgaactgtgcaggttaccaagaagcatgagcaggctcgtcaaactgaggcagctc
 621 G  C  E  L  C  R  L  P  R  S  M  S  R  L  V  K  L  R  Q  L

1921 aaagcaaatccagatgtaattgccgacatagccaaagtcgggagattgatcgaacttcaa
 641 K  A  N  P  D  V  I  A  D  I  A  K  V  G  R  L  I  E  L  Q

1981 gagctgaaagcctataatgttgacaagaaaaaaggacatgggattgcagagctaagtgca
 661 E  L  K  A  Y  N  V  D  K  K  K  G  H  G  I  A  E  L  S  A

2041 atgaatcagcttcacggtgatctttccattagaaaccttcaaaatgtagagaaaacgcga
 681 M  N  Q  L  H  G  D  L  S  I  R  N  L  Q  N  V  E  K  T  R

2101 gagtctcggaaggcgaggttggacgagaaacagaagcttaagctcttggatctgcgatgg
 701 E  S  R  K  A  R  L  D  E  K  Q  K  L  K  L  L  D  L  R  W

2161 gctgacggtaggggtgccggagaatgtgatcgtgacaggaaagttcttaaaggcctccga
 721 A  D  G  R  G  A  G  E  C  D  R  D  R  K  V  L  K  G  L  R

2221 ccacatccaaacctgagagaattgagtatcaaatactacggaggcacttcatctccgagt
 741 P  H  P  N  L  R  E  L  S  I  K  Y  Y  G  G  T  S  S  P  S

2281 tggatgacggatcagtatctgcccaacatggaaacgattcgcctgcgtagctgcgcaagg
 761 W  M  T  D  Q  Y  L  P  N  M  E  T  I  R  L  R  S  C  A  R

2341 ttgacggaactcccatgtctcggtcagctgcatatccttagacatttgcacatcgatggg
 781 L  T  E  L  P  C  L  G  Q  L  H  I  L  R  H  L  H  I  D  G

2401 atgtcccaagtgagacaaattaatctgcaattttatggcaccggagaagtttcaggtttt
 801 M  S  Q  V  R  Q  I  N  L  Q  F  Y  G  T  G  E  V  S  G  F

2461 ccattgctggagctcctgaacatacgtcgcatgcccagtctggaggaatggtcggaacca
 821 P  L  L  E  L  L  N  I  R  R  M  P  S  L  E  E  W  S  E  P

2521 cggagaaactgttgctacttccctcgcctccataaactgctgatcgaggattgtcccagg
 841 R  R  N  C  C  Y  F  P  R  L  H  K  L  L  I  E  D  C  P  R

2581 ctcaggaatctgccctccctcccaccaacactggaagaactaaggatatcaagaacagga
 861 L  R  N  L  P  S  L  P  P  T  L  E  E  L  R  I  S  R  T  G

2641 ctagttgatcttccaggattccatggaaacggtgatgtgacgacgaatgtttccctttct
 881 L  V  D  L  P  G  F  H  G  N  G  D  V  T  T  N  V  S  L  S

2701 tctttgcatgtttcggagtgtcgagaactgagatccctaagcgaaggattgttgcagcac
 901 S  L  H  V  S  E  C  R  E  L  R  S  L  S  E  G  L  L  Q  H

2761 aacctcgtcgccctcaagacagcggcatttaccgattgtgattctcttgagttttgccg
 921 N  L  V  A  L  K  T  A  A  F  T  D  C  D  S  L  E  F  L  P

2821 gcggaaggattcagaacagccatttcacttgaatcattgataatgactaattgtccactg
 941 A  E  G  F  R  T  A  I  S  L  E  S  L  I  M  T  N  C  P  L
```

```
2881 ccttgcagttttcttttgccttcctctctcgagcatctaaagttgcagccatgcctctat
 961 P  C  S  F  L  L  P  S  S  L  E  H  L  K  L  Q  P  C  L  Y

2941 ccaaacaacaatgaggattcactgtcaacatgcttcgagaacctcacatctctttccttc
 981 P  N  N  N  E  D  S  L  S  T  C  F  E  N  L  T  S  L  S  F

3001 ttggacatcaaagattgtccaaatctgtcatcatttccaccgggtcctctatgtcagcta
1001 L  D  I  K  D  C  P  N  L  S  S  F  P  P  G  P  L  C  Q  L

3061 tcagcactccaacatttgtccctcgtcaattgccagaggctacaatctattggcttccag
1021 S  A  L  Q  H  L  S  L  V  N  C  Q  R  L  Q  S  I  G  F  Q

3121 gcactcacctccctcgaaagcttgacaattcagaactgccctcgcctcaccatgtcacac
1041 A  L  T  S  L  E  S  L  T  I  Q  N  C  P  R  L  T  M  S  H

3181 agtttggttgaggtgaataactcttccgatacagggctcgcgtttaatatcactcgatgg
1061 S  L  V  E  V  N  N  S  S  D  T  G  L  A  F  N  I  T  R  W

3241 atgcgcagacgaacaggtgacgacggcttgatgctcagacaccgagcacaaaatgattca
1081 M  R  R  R  T  G  D  D  G  L  M  L  R  H  R  A  Q  N  D  S

3301 tttttcgggggacttctgcaacacctcaccttcctccagtttctaaagatctgccagtgt
1101 F  F  G  G  L  L  Q  H  L  T  F  L  Q  F  L  K  I  C  Q  C

3361 ccacaactcgtaaccttcaccggcgaagaggaagagaagtggagaaaccttacttctctt
1121 P  Q  L  V  T  F  T  G  E  E  E  E  K  W  R  N  L  T  S  L

3421 caaattctgcacatcgttgattgtccaaacctggaggtactgcctgcaaacttgcaaagc
1141 Q  I  L  H  I  V  D  C  P  N  L  E  V  L  P  A  N  L  Q  S

3481 ctctgctccctcagcaccttgtacatcgtcagatgcccaagaatccatgcgtttcctccc
1161 L  C  S  L  S  T  L  Y  I  V  R  C  P  R  I  H  A  F  P  P

3541 ggaggtgtcagcatgtccctggcacatttggtcatccatgaatgccctcagctgtgtcag
1181 G  G  V  S  M  S  L  A  H  L  V  I  H  E  C  P  Q  L  C  Q

3601 cgatgtgatccaccgggaggtgatgattggcccttaatagctaatgtaccaagaatatgt
1201 R  C  D  P  P  G  G  D  D  W  P  L  I  A  N  V  P  R  I  C

3661 cttggaaggactcatccatgtcgctgtagcaccacctga
1221 L  G  R  T  H  P  C  R  C  S  T  T  *
```

Figure 5.2 Nucleotide sequence and conceptual translation of RGC2 cDNA. Conserved motifs in R proteins are underlined in the NBS domain. Two additional NBS motifs that were not covered in chapter 4 are shaded in gray.

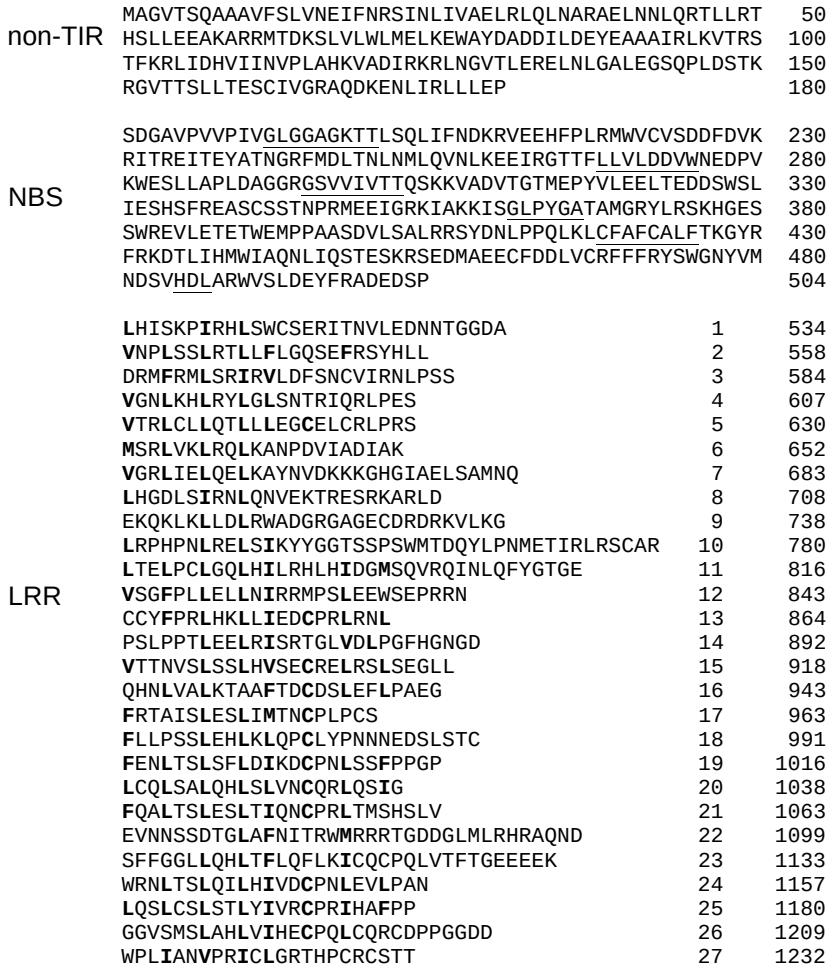


Figure 5.3. Predicted RGC2 protein structure. The amino acid sequence deduced from the RGC2 cDNA of roots is divided into three domains (non-TIR, NBS and LRR). Conserved motifs in R proteins are underlined. Amino acids matching the consensus of the cytoplasmic LRR are indicated in bold in the LRR domain.

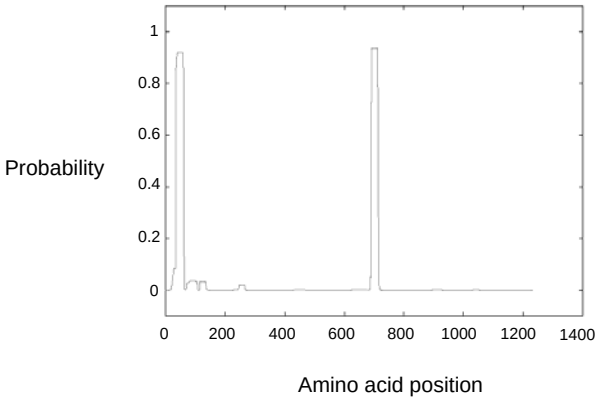


Figure 5.4 Predicted coiled-coil (CC) regions in the RGC2 deduced protein. The CC profile (window of 21 consecutive amino acids) was calculated according to Lupas (1996) and plotted against the amino acid number. Values shown on the y-axis of the graph indicates the probability to find potential CC regions in the amino acid sequence. One potential CC region is present in the non-TIR domain and another one in the first portion of the LRR domain.

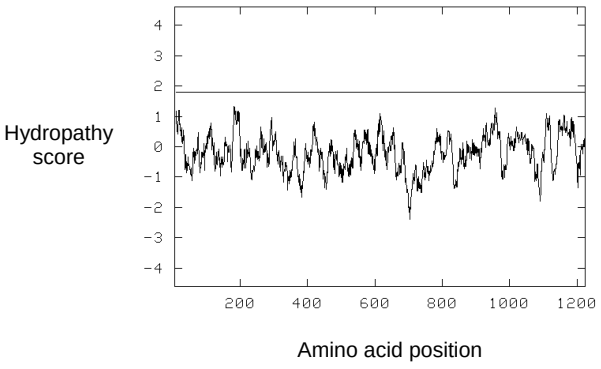


Figure 5.5 Hydropathy profile of the RGC2 deduced protein. The hydropathy profile (window of 19 consecutive amino acids) was calculated according to Kyte and Doolittle (1982) and plotted against the amino acid number. Peaks with hydropathy scores greater than 1.8 (horizontal line) indicate possible transmembrane regions.

```
RGC5       1 ATGTCGACGGCGCTAGTAATCGGAGGATGGTTCGCGCAAAGCTTCATCCAGACGTTGCTC
RGC5fs     1 ATGTCGACGGCGCTAGTAATCGGAGGATGGTTCGCGCAAAGCTTCATCCAGACGTTGCTC
             ************************************************************

RGC5      61 GACAAGGCCAGCAACTGCGCGATCCAACAACTCGCGCGGTGCCGCGGCCTTCACGATGAC
RGC5fs    61 GACAAGGCCAGCAACTGCGCGATCCAACAACTCGCGCGGTGCCGCGGCCTTCACGATGGC
             ************************************************************ *

RGC5     121 CTGAGGCGGCTGCGGACGTCTCTGCTCCGGATCCATGCCATCCTCGACAAGGCAGAGACG
RGC5fs   121 CTGAGGCGGCTGCGGACGTCTCTGCTCCGGATCCATGCCATCCTCGGCAAGGCAGAGGCG
             ********************************************** ********** **

RGC5     181 AGGTGGAACCATAAAAACACGAGCTTGGTGGAGCTGGTGAGGCAGCTCAAGGATGCTGCC
RGC5fs   181 AGGTGGAACCATAAAAACACGAGCTTGGTGGAGCTGGTGAGGCAGCTCAAGGATGCTGCC
             ************************************************************

RGC5     241 TATGACGCCGAGGACTTACTGGAGGAGTTGGAGTACCAAGCCGCGAAGCAAAAGGTCGAG
RGC5fs   241 TATGACGCCGAGGACTTACTGGAGGAGTTGGAGTACCAAGCCGCGAAGCAAAAGGTCGAG
             ************************************************************

RGC5     301 CACCGGGGAGACCAGATAAGCGACCTCTTTTCTTTTTCCCCTAGTACTGCGAGCGAGTGG
RGC5fs   301 CACCGGGGAGACCAGATAAGCGACCTCTTTTCTTTTTCCCTTAGTACTGCGAGCGAGTGG
             ************************************** ******************

RGC5     361 TTGGGTGCCGATGGTGATGATGCTGGGACTCGATTGAGGGAGATCCAGGAGAAGCTGTGC
RGC5fs   361 TTGGGTGCCGATGGTGATGATGCTGGGACTCGATTGAGGGAGATCCAGGGGAAGCTGTGC
             ************************************************ **********

RGC5     421 AACATTGCTGCCGATATGATGGATGTCATGCAGCTATTGGCACCCGATGATGGGGGGAGA
RGC5fs   421 AACATTGCTGCCGATATGATGGATGTCATGCAGCTATTGGCACCCGATGATGGGGGGAGA
             ************************************************************

RGC5     481 CAATTCGACTGGAAGGTGGTGGGAAGAGAAACGAGCTCTTTCTTGACCGAAACCGTCGTG
RGC5fs   481 CAATTCGACTGGAAGGTGGTGAGAAGAGAAACGAGCTCTTTCTTGACCGAAACCGTCGTG
             ******************** ************************************

RGC5     541 TTTGGTCGGGGCCAAGAAAGGGAGAAAGTAGTAGAATTGCTGTTGGATTCAGGATCTGGT
RGC5fs   541 TTTGGTCGGGACCAAGAAAGGGAGAAAGTAGTAGAATTGCTGCTGGATTCAGGATCTGGT
             ********** ******************************** *****************

RGC5     601 AACAGTAGCTTCTCTGTCTTACCCCTCGTCGGAATCGGAGGGGTTGGGAAGACGACTCTG
RGC5fs   601 AACAGCAGCTTCTCTGTCTTACCCCTCGTCGGAATCGGAGGG-TTGGGAAGACGACTCTG
             ***** ************************************ *****************

RGC5     661 GCTCAGCTCGTGTACAACGACAATCGTGTCGGCAACTATTTCCACCTCAAGGTTTGGGTC
RGC5fs   660 GCTCAGCTCGTGTACAACGACAATCGTGTCGGCAACTATTTCCACCTCAAGGTTTGGGTC
             ************************************************************

RGC5     721 TGTGTATCCGACAATTTCAATGTGAAGAGACTGACCAAAGAGATAATCGAGTCTGCTACC
RGC5fs   720 TGTGTATCCGACAATTTCAATGTGAAGAGACTGACCAAAGAGATAATCGAGTCTGCTACC
             ************************************************************

RGC5     781 AAGGTGGAACAATCTGACGAATTGAACTTGGACACCCTGCAACAGATCCTCAAGGAGAAG
RGC5fs   780 AAGGTGGAACGATCTGACAAATTGAACTTGGACACCCTGCAACAGATCCTCAAGGAGAAG
             ********** ******* ****************************************

RGC5     841 ATTGCTTCAGAGAGGTTTCTGCTAGTCCTCGATGATGTGTGGAGCGAAAACAGGGATGAC
RGC5fs   840 ATTGCTTCAGAGAGGTTTCTGCTAGTCCTCGATGATGTGTGGAGCGAAAACAGGGATGAC
             ************************************************************

RGC5     901 TGGGAAAGGCTGTGCGCGCCACTAAGGTTTGCAGCAAGAGGCAGCAAGGTTATAGTCACA
RGC5fs   900 TGGGAAAGGCTGTGCGCACCACTAAGGTTTGCAGCAAGAGGCAGCAAGGTTATAGTCACA
             ***************** ****************************************

RGC5     961 ACTCGAGACACAAAGATTGCCAGCATCATTGGCACAATGAAGGAAATTTCGCTCGATGGT
RGC5fs   960 ACTCGAGACACAAAGATTGCCAGCATCATTGGCACAATGAAGGAAATTTCGCTCGATGGT
             ************************************************************

RGC5    1021 CTCCAGGATGATGCTTACTGGGAGCTGTTCAAGAAATGTGCATTTGGTTCTGTGAACCCC
RGC5fs  1020 CTCCAGGATGATGCTTACTGGGAGCTGTTCAAGAAATGTGCATTTGGTTCTGTGAACCCC
             ************************************************************

RGC5    1081 CAGGAGCATCTAGAGCTCGAGGTTATCGGTAGAAAGATTGCTGGTAAGTTGAAGGGCTCA
RGC5fs  1080 CAGGAGCATCTAGAGCTCGAGGTTATCGGTAGAAAGATTGCTGGTAAGTTGAAGGGCTCA
             ************************************************************
```

```
RGC5   1141 CCGCTAGCAGCAAAAACACTAGGAAGCTTGTTGCGGTCGGATGTCAGCCAAGAACACTGG
RGC5fs 1140 CCGCTAGCAGCAAAAACACTAGGAAGCTTGTTGCGGTCGGATGTCAGCCAAGAACACTGG
            ************************************************************

RGC5   1201 AGAACTATAATGGAAAGTGAGGTATGGCAACTGCCACAAGCTGAAAATGAAATATTGCCT
RGC5fs 1200 AGAACTATAATGGGAAGTGAGGTATGGCAACTGCCACAAGCTGAAAATGAAATATTGCCT
            ************* ********************************************

RGC5   1261 GTTCTATGGCTGAGCTATCAACACCTTCCCGGACATCTTAGACAGTGTTTCGCTTTTTGC
RGC5fs 1260 GTTCTATGGCTGAGCTATCAACACCTTCCCGGACATCTTAGACAGTGTTTCGCTTTTTGC
            ************************************************************

RGC5   1321 GCTGTGTTTCACAAAGATTATTTATTCTATAAACATGAGTTGATCCAGACTTGGATGGCA
RGC5fs 1320 GCTGTGTTTCACAAAGATTATTTATTCTATAAACATGAGTTGATCCAGACTTGGATGGCA
            ************************************************************

RGC5   1381 GAAGGCTTCATTGCACCTCAAGGAAACAAGAGGGTGGAAGATGTCGGAAGCAGCTACTTC
RGC5fs 1380 GAAGGCTTCATTGCACCTCAAGGAAACAAGAGGGTGGAAGATGTCGGAAGCAGCTACTTC
            ************************************************************

RGC5   1441 CATGAGCTTGTTAATAGGTCTTTCTTTCAGGAATCTCAGTGGAGAGGGCGATACGTGATG
RGC5fs 1440 CATGAGCTTGTTAATAGGTCTTTCTTTCAGGAATCTCAGTGGAGAGGGCGATACGTGATG
            ************************************************************

RGC5   1501 CGTGACCTCATACACGATCTTGCCCAATTTATATCAGTGGGAGAGTGTCATAGGATAGAT
RGC5fs 1500 CATGACCTCATACACGATCTTGCCCAATTTATATCAGTGGGAGAGTGTCATAGGATAGAT
            * ************************************************************

RGC5   1561 GATGACAAGTCCAAAGAGACCCCTAGTACGACTCGTCATCTATCAGTAGCATTAACTGAG
RGC5fs 1560 GATGACAAGTCCAAAGAGACCCCTAGTACGACTCGTCATCTATCAGTAGCATTAACTGAG
            ************************************************************

RGC5   1621 CAAACGAAGTTGGTGGATTTTTCAGGTTACAATAAATTGCGGACCCTTGTGATCAACAAT
RGC5fs 1620 CAAACGAAGTTGGTGGATTTTTCAGGTTACAATAAATTGCGGACCCTTATGATCAACAAT
            ********************************************** ************

RGC5   1681 CAGAGAAATCAGTATCCATATATGACTAAAGTCAACAGCTGCTTATTGCCTCAGAGCTTG
RGC5fs 1680 CAGAGAAATCAGTATCCATATATGACTAAAGTCAACAGCTGCTTATTGCCTCAGAGCTTG
            ************************************************************

RGC5   1741 TTCAGAAGACTGAAAAGAATCCATGTTTTAGTTTTGCAGAAGTGTGGCATGAAAGAGTTG
RGC5fs 1740 TTCAGAAGACTGAAAAGAATCCATGTTTTAGTTTTGCAGAAGTGTGGCATGAAAGAGTTG
            ************************************************************

RGC5   1801 CCTGATATTATCGGTGACTTGATACAACTTCGGTACCTTGACATATCCTACAATGCTCGC
RGC5fs 1800 CCTGATATTATCGGTGACTTGATACAACTTCGGTACCTTGACATATCCTACAATGCTCGC
            ************************************************************

RGC5   1861 ATTCAGAGGTTGCCTGAGTCATTGTGCGACCTTTACAATCTGCAAGCACTGAGGCTATGG
RGC5fs 1860 ATTCAGAGGTTGCCTGAGTCATTGTGCGACCTTTACAATCTGCAAGCACTGAGGCTATGG
            ************************************************************

RGC5   1921 GGCTGTCAATTACAGAGTTTCCCACAAGGCATGAGCAAGCTGATCAACTTGAGGCAACTT
RGC5fs 1920 GGCTGTCAATTACAGAGTTTCCCACAAGGCATGAGCAAGCTGATCAACTTGAGGCAACTT
            ************************************************************

RGC5   1981 CATGTAGAAGATGAGATAATTTCCAAGATATACGAGGTTGGGAAGCTGATTTCTCTGCAA
RGC5fs 1980 CATGTAGAAGATGAGATAATTTCCAAGATATACGAGGTTGGGAAGCTGATTTCTCTGCAA
            ************************************************************

RGC5   2041 GAATTGTCTGCATTCAAAGTGCTAAAGAATCATGGAAACAAACTTGCAGAACTAAGTGGT
RGC5fs 2040 GAATTGTCTGCATTCAAAGTGCTAAAGAATCATGGAAACAAACTTGCAGAACTAAGTGGT
            ************************************************************

RGC5   2101 TTGACACAACTCCGCGGCACTCTACGAATTACAAATCTTGAAAATGTAGGGAGTAAAGAA
RGC5fs 2100 TTGACACAACTCCGCGGCACTCTACGAATTACAAATCTTGAAAATGTAGGGAGTAAAGAA
            ************************************************************

RGC5   2161 GAAGCAAGCAAGGCTAAACTGCACAGAAAACAGTATCTTGAAGCATTAGAGTTAGAGTGG
RGC5fs 2160 GAAGCAAGCGAGGCTAAACTGCACAGAAAACAGTATCTTGAAGCATTAGAGTTAGAGTGG
            ********* ************************************************

RGC5   2221 GCAGCTGGCCAGGTTTCCAGCTTGGAGCATGAGTTACTTGTCTCGGAGGAAGTATTTTTA
RGC5fs 2220 GCAGCTGGCCAGGTTTCCAGCTTGGAGCATGAGTTACTTGTCTCGGAGGAAGTATTTTTA
            ************************************************************
```

```
RGC5   2281 GGTCTCCAACCACATCACTTCCTCAAAAGTTCGACAATCAGAGGGTACAGTGGTGCTACA
RGC5fs 2280 GGTCTCCAACCACATCACTTCCTCAAAAGTTTGACAACCAGAGGGTACAGTGGTGCTACA
            ****************************** ***** **********************

RGC5   2341 GTACCCAGTTGGCTGGATGTGAAAATGCTACCGAACTTGGGAACTCTTAAACTAGAGAAC
RGC5fs 2340 GTACCCAGTTGGCTGGATGTGAAAATGCTACCGAACTTGGGAACTCTTAAACTAGAGAAC
            ************************************************************

RGC5   2401 TGTACAAGACTGGAGGGTCTTTCATATATTGGACAACTGCCACATCTCAAGGTCCTTCAT
RGC5fs 2400 TGTACAAGACTGGAGGGTCTTTCATATATTGGACAACTGCCACATCTCAAGGTCCTTCAT
            ************************************************************

RGC5   2461 ATAAAGAGAATGCCTGTGGTGAAACAAATGAGTCATGAATTATGTGGCTGTACGAAAAGC
RGC5fs 2460 ATAAAGAGAATGCCTGTGGTGAAACAAATGAGTCATGAATTATGTGGCTGTACGAAAAGC
            ************************************************************

RGC5   2521 AAGTTGTTCCCTAGGCTAGAAGAGTTGGTACTGGAGGATATGCCAACACTGAAAGAATTC
RGC5fs 2520 AAGTTGTTCCCTAGGCTAGAAGAGTTGGTACTGGAGGATATGCCAACACTGAAAGAATTC
            ************************************************************

RGC5   2581 CCGAATATTGCACAACTTCCTTGTCTCAAGATTATTCACATGAAGAACATGTTTTCAGTA
RGC5fs 2580 CCGAATATTGCACAACTTCCTTGTCTCAAGATTATTCACATGAAGAACATGTTTTCAGTA
            ************************************************************

RGC5   2641 AAACATATAGGTCGTGAATTATATGGTGATATAGAGAGCAATTGTTTTCCATCATTAGAA
RGC5fs 2640 AAACATATAGGTCGTGAATTATATGGTGATATAGAGAGCAATTGTTTTCCATCATTAGAA
            ************************************************************

RGC5   2701 GAGCTTGTGCTGCAGGACATGCTGACATTGGAGGAACTCCCAAATCTTGGACAACTTCCA
RGC5fs 2700 GAGCTTGTGCTGCAGGACATGCTGACATTGGAGGAACTCCCAAATCTTGGACAACTCCCA
            ********************************************************** ***

RGC5   2761 CATCTTAAGGTTATTCACATGAAGAACATGTCTGCACTGAAACTTATAGGTCGTGAATTA
RGC5fs 2760 CATCTTAAGGTTATTCACATGAAGAACATGTCTGCACTGAAACTTATAGGTCGTGAATTA
            ************************************************************

RGC5   2821 TGTGGTTCTAGAGAGAAAACTTGGTTTCCTAGGCTAGAAGTGCTAGTGCTGAAGAACATG
RGC5fs 2820 TGTGGTTCTAGGGAGAAAATTTGGTTTCCTAGGCTAGAAGTGCTAGTGCTGAAGAACATG
            *********** ******* ****************************************

RGC5   2881 CTGGCACTGGAGGAACTCCCAAGTCTTGGACAACTTCCATGTCTCAAGGTTCTTCGCATC
RGC5fs 2880 CTGGCACTGGAGGAACTCCCAAGTCTTGGACAACTTCCATGTCTCAAGGTTCTTCGCATC
            ************************************************************

RGC5   2941 CAGGTGTCGAAGGTAGGCCATGGACTCTTTAGTGCTACGAGGAGTAAATGGTTTCCAAGG
RGC5fs 2940 CAGGTGTCGAAGGTAGGCCATGGACTCTTTAGTGCTACGAGAAGTAAATGGTTTCCAAGG
            ****************************************** ******************

RGC5   3001 CTGGAAGAGCTAGAAATCAAGGGCATGCTGACATTTGAGGAACTCCATTCTCTTGAAAAA
RGC5fs 3000 CTGGAAGAGCTAGAAATCAAGGGCATGCTGACATTTGAGGAACTCCATTCTCTTGAAAAA
            ************************************************************

RGC5   3061 CTGCCGTGTCTCAAGGTTTTCCGCATCAAGGGATTGCCAGCAGTGAAAAAGATAGGCCAT
RGC5fs 3060 CTGCCATGTCTCAAGGTTTTCCGCATCAAGGGATTGCCAGCAGTGAAAAAGATAGGCCAT
            ***** ********************************************************

RGC5   3121 GGATTATTTGATTCTACCTGTCAGAGAGAGGGTTTTCCAAGGTTGGAAGAGCTTGTGTTA
RGC5fs 3120 GGATTATTTGATTCTACCTGTCAGAGAGAGGGTTTTCCAGGGTTGGAAGAGCTTGTGTTA
            ************************************** ********************

RGC5   3181 AGAGACATGCCAGCGTGGGAAGAGTGGCCTTGGGCTGAAAGGGAGGAGTTATTTTCCTGC
RGC5fs 3180 AGAGACATGCCAGCGTGGGAAGAGTGGTCTTGGGCTGAAAGGGAGGAGTTATTTTCCTGC
            *************************** ********************************

RGC5   3241 TTGTGTAGACTTAAAATTGAACAATGCCCCAAACTTAAATGCTTGCCTCCCGTCCCTTAT
RGC5fs 3240 TTGTGTAGACTTAAAATTGAACAATGCCCCAAACTTAAATGCTTGCCTCCCGTCCCTTAT
            ************************************************************

RGC5   3301 TCTCTCATAAAACTTGAATTATGGCAAGTTGGGCTGACAGGACTTCCAGGATTATGCAAA
RGC5fs 3300 TCTCTCATAAAACTTGAATTATGGCAAGTTGGGCTGACAGGACTTCCAGGATTATGCAAA
            ************************************************************

RGC5   3361 GGAATTGGTGGAGGTAGCAGCGCTAGAACTGCTTCTCTTTCACTCTTGCACATTATTAAA
RGC5fs 3360 GGAATTGGTGGAGGTAGCAGCGCTAGAACTGCTTCTCTTTCACTCTTGCACATTATTAAA
            ************************************************************
```

```
RGC5   3421 TGCCCAAATCTGAGAAATCTGGGAGAAGGGTTGCTGTCAAACCACCTGCCACATATCAAT
RGC5fs 3420 TGCCCAAATCTGAGAAATCTGGGAGAAGGGTTGCTGTCAAACCACCTGCCACATATCAAT
            ************************************************************

RGC5   3481 GCTATTCGGATATGGGAATGTGCTGAACTGTTGTGGCTGCCTGTCAAGAGGTTTAGAGAA
RGC5fs 3480 GCTATTCGGATATGGGAATGTGCTGAACTGTTGTGGCTGCCTGTCAAGAGGTTTAGAGAA
            ************************************************************

RGC5   3541 TTCACCACCCTTGAGAACTTGTCAATAAGGAACTGCCCCAAGCTCATGAGCATGACACAG
RGC5fs 3540 TTCACCACCCTTGAGAACTTGTCAATAAGGAACTGCCCCAAGCTCATGAGCATGACACAG
            ************************************************************

RGC5   3601 TGTGAGGAGAATGACCTCCTCCTCCCGCCGTCAATCAAGGCGCTAGAATTGGGTGACTGT
RGC5fs 3600 TGTGAGGAGAATGACCTCCTCCTCCCGCCGTCAATCAAGGCGCTAGAATTGGGTGACTGT
            ************************************************************

RGC5   3661 GGAAATCTTGGGAAATCGCTGCCTGGATGCCTACACAACCTCAGCTCACTAATTCAGTTG
RGC5fs 3660 GGAAATCTTGGGAAATCGCTGCCTGGATGCCTACACAACCTCAGCTCACTAATTCAGTTG
            ************************************************************

RGC5   3721 GCGATATCCAATTGTCCATACATGGTTTCCTTTCCAAGGGACGTAATGCTTCACTTGAAG
RGC5fs 3720 GCGATATCCAATTGTCCATACATGGTTTCCTTTCCAAGGGACGTAATGCTTCACTTGAAG
            ************************************************************

RGC5   3781 GAACTTGGAGCTGTAAGGATCATGAATTGTGATGGGCTGAGATCAATAGAGGGTTTACAA
RGC5fs 3780 GAACTTGGAGCTGTAAGGATCATGAATTGTGATGGGCTGAGATCAATAGAGGGTTTACAA
            ************************************************************

RGC5   3841 GTTCTCAAATCACTCAAGAGATTGGAAATCATAGGATGTCCCAGGCTTTTGCTAAATGAA
RGC5fs 3840 GTTCTCAAATCACTCAAGAGATTGGAAATCATAGGATGTCCCAGGCTTTTGCTAAATGAA
            ************************************************************

RGC5   3901 GGGGATGAGCAAGGGGAGGTCTTGTCACTGCTTGAATTATCAGTAGATAAAACAGCCCTA
RGC5fs 3900 GGGGATGAGCAAGGGGAGGTCTTGTCACTGCTTGAATTATCAGTAGATAAAACAGCCCTA
            ************************************************************

RGC5   3961 CTTAAACTCTCATTTATAAAAAATACACTGCCATTCATCCAGTCTCTCAGAATCATCTTG
RGC5fs 3960 CTTAAACTCTCATTTATAAAAAATACACTGCCATTCATCCAGTCTCTCAGAATCATCTTG
            ************************************************************

RGC5   4021 TCTCCTCAGAAAGTGTTGTTTGACTGGGAGGAGCAGGAATTGGTGCACAGCTTCACCGCT
RGC5fs 4020 TCTCCTCAGAAAGTGATGTTTGACTGGGGGGAGCAGGAATTGGTGCACAGCCTCACAGCT
            ************** ************ ********************** **** ***

RGC5   4081 CTCAGGCGCCTTGAATTCCTCAGTTGCAAGAATCTCCAGTCCTTGCCAACAGAGTTGCAT
RGC5fs 4080 CTCAGGCGCCTTGAATTCCTCAGTTGCAAGAATCTCCAGTCCTTGCCAACAGAGTTGCAT
            ************************************************************

RGC5   4141 ACCCTTCCTTCCCTCCATGCTTTGGTTGTAAGTGACTGTCCACAGATCCAATCACTGCCA
RGC5fs 4140 ACCCTTCCTTCCCTCCATGCTTTGGTTGTAAGTGACTGTCCACAGATCCAATCACTGCCA
            ************************************************************

RGC5   4201 TCGAAGGGACTCCCGACACTCCTCACAGATTTAGGATTTGACCATTGCCACCCAGTGCTG
RGC5fs 4200 TCGAAGGGACTCCCGACACTCCTCACAGATTTAGGATTTGACCATTGCCACCCAGTGCTG
            ************************************************************

RGC5   4261 ACTGCGCAACTGGAAAAGCACCTGGCAGAGATGAAGAGCTCAGGTCGATTTCACCCAGTT
RGC5fs 4260 ACTGCGCAACTGGAAAAGCACCTGGCAGAGATGAAGAGCTCAGGTCGATTTCACCCAGTT
            ************************************************************

RGC5   4321 TATGCATAG-4329
RGC5fs 4320 TATGCATAG-4328
            *********
```

Figure 5.6 ClustalX alignment of the nucleotide sequences of RGC5 cDNA and a RGC5 cDNA homologue with a frameshift mutation (RGC5fs). The frameshift in RGC5fs occurred at position 643 of the nucleotide sequence and it is indicated with an open box.

```
   1 atgtcgacggcgctagtaatcggaggatggttcgcgcaaagcttcatccagacgttgctc
   1 M  S  T  A  L  V  I  G  G  W  F  A  Q  S  F  I  Q  T  L  L

  61 gacaaggccagcaactgcgcgatccaacaactcgcgcggtgccgcggccttcacgatgac
  21 D  K  A  S  N  C  A  I  Q  Q  L  A  R  C  R  G  L  H  D  D

 121 ctgaggcggctgcggacgtctctgctccggatccatgccatcctcgacaaggcagagacg
  41 L  R  R  L  R  T  S  L  L  R  I  H  A  I  L  D  K  A  E  T

 181 aggtggaaccataaaaacacgagcttggtggagctggtgaggcagctcaaggatgctgcc
  61 R  W  N  H  K  N  T  S  L  V  E  L  V  R  Q  L  K  D  A  A

 241 tatgacgccgaggacttactggaggagttggagtaccaagccgcgaagcaaaaggtcgag
  81 Y  D  A  E  D  L  L  E  E  L  E  Y  Q  A  A  K  Q  K  V  E

 301 caccggggagaccagataagcgacctcttttcttttttcccctagtactgcgagcgagtgg
 101 H  R  G  D  Q  I  S  D  L  F  S  F  S  P  S  T  A  S  E  W

 361 ttgggtgccgatggtgatgatgctgggactcgattgagggagatccaggagaagctgtgc
 121 L  G  A  D  G  D  D  A  G  T  R  L  R  E  I  Q  E  K  L  C

 421 aacattgctgccgatatgatggatgtcatgcagctattggcacccgatgatgggggggaga
 141 N  I  A  A  D  M  M  D  V  M  Q  L  L  A  P  D  D  G  G  R

 481 caattcgactggaaggtggtgggaagagaaacgagctctttcttgaccgaaaccgtcgtg
 161 Q  F  D  W  K  V  V  G  R  E  T  S  S  F  L  T  E  T  V  V

 541 tttggtcggggccaagaaagggagaaagtagtagaattgctgttggattcaggatctggt
 181 F  G  R  G  Q  E  R  E  K  V  V  E  L  L  L  D  S  G  S  G

 601 aacagtagcttctctgtcttacccctcgtcggaatcggaggggttgggaagacgactctg
 201 N  S  S  F  S  V  L  P  L  V  G  I  G  G  V  G  K  T  T  L

 661 gctcagctcgtgtacaacgacaatcgtgtcggcaactatttccacctcaaggtttgggtc
 221 A  Q  L  V  Y  N  D  N  R  V  G  N  Y  F  H  L  K  V  W  V

 721 tgtgtatccgacaatttcaatgtgaagagactgaccaaagagataatcgagtctgctacc
 241 C  V  S  D  N  F  N  V  K  R  L  T  K  E  I  I  E  S  A  T

 781 aaggtggaacaatctgacgaattgaacttggacaccctgcaacagatcctcaaggagaag
 261 K  V  E  Q  S  D  E  L  N  L  D  T  L  Q  Q  I  L  K  E  K

 841 attgcttcagagaggtttctgctagtcctcgatgatgtgtggagcgaaaacagggatgac
 281 I  A  S  E  R  F  L  L  V  L  D  D  V  W  S  E  N  R  D  D

 901 tgggaaaggctgtgcgcgccactaaggtttgcagcaagaggcagcaaggttatagtcaca
 301 W  E  R  L  C  A  P  L  R  F  A  A  R  G  S  K  V  I  V  T

 961 actcgagacacaaagattgccagcatcattggcacaatgaaggaaatttcgctcgatggt
 321 T  R  D  T  K  I  A  S  I  I  G  T  M  K  E  I  S  L  D  G

1021 ctccaggatgatgcttactgggagctgttcaagaaatgtgcatttggttctgtgaacccc
 341 L  Q  D  D  A  Y  W  E  L  F  K  K  C  A  F  G  S  V  N  P

1081 caggagcatctagagctcgaggttatcggtagaaagattgctggtaagttgaagggctca
 361 Q  E  H  L  E  L  E  V  I  G  R  K  I  A  G  K  L  K  G  S

1141 ccgctagcagcaaaaacactaggaagcttgttgcggtcggatgtcagccaagaacactgg
 381 P  L  A  A  K  T  L  G  S  L  L  R  S  D  V  S  Q  E  H  W

1201 agaactataatggaaagtgaggtatggcaactgccacaagctgaaaatgaaatattgcct
 401 R  T  I  M  E  S  E  V  W  Q  L  P  Q  A  E  N  E  I  L  P

1261 gttctatggctgagctatcaacaccttcccggacatcttagacagtgtttcgctttttgc
 421 V  L  W  L  S  Y  Q  H  L  P  G  H  L  R  Q  C  F  A  F  C

1321 gctgtgtttcacaaagattatttattctataaacatgagttgatccagacttggatggca
 441 A  V  F  H  K  D  Y  L  F  Y  K  H  E  L  I  Q  T  W  M  A

1381 gaaggcttcattgcacctcaaggaaacaagagggtggaagatgtcggaagcagctacttc
 461 E  G  F  I  A  P  Q  G  N  K  R  V  E  D  V  G  S  S  Y  F

1441 catgagcttgttaataggtctttctttcaggaatctcagtggagagggcgatacgtgatg
 481 H  E  L  V  N  R  S  F  F  Q  E  S  Q  W  R  G  R  Y  V  M
```

```
1501 cgtgacctcatacacgatcttgcccaatttatatcagtgggagagtgtcataggatagat
 501 R  D  L  I  H  D  L  A  Q  F  I  S  V  G  E  C  H  R  I  D

1561 gatgacaagtccaaagagacccctagtacgactcgtcatctatcagtagcattaactgag
 521 D  D  K  S  K  E  T  P  S  T  T  R  H  L  S  V  A  L  T  E

1621 caaacgaagttggtggatttttcaggttacaataaattgcggacccttgtgatcaacaat
 541 Q  T  K  L  V  D  F  S  G  Y  N  K  L  R  T  L  V  I  N  N

1681 cagagaaatcagtatccatatatgactaaagtcaacagctgcttattgcctcagagcttg
 561 Q  R  N  Q  Y  P  Y  M  T  K  V  N  S  C  L  L  P  Q  S  L

1741 ttcagaagactgaaaagaatccatgttttagttttgcagaagtgtggcatgaaagagttg
 581 F  R  R  L  K  R  I  H  V  L  V  L  Q  K  C  G  M  K  E  L

1801 cctgatattatcggtgacttgatacaacttcggtaccttgacatatcctacaatgctcgc
 601 P  D  I  I  G  D  L  I  Q  L  R  Y  L  D  I  S  Y  N  A  R

1861 attcagaggttgcctgagtcattgtgcgacctttacaatctgcaagcactgaggctatgg
 621 I  Q  R  L  P  E  S  L  C  D  L  Y  N  L  Q  A  L  R  L  W

1921 ggctgtcaattacagagtttcccacaaggcatgagcaagctgatcaacttgaggcaactt
 641 G  C  Q  L  Q  S  F  P  Q  G  M  S  K  L  I  N  L  R  Q  L

1981 catgtagaagatgagataatttccaagatatacgaggttgggaagctgatttctctgcaa
 661 H  V  E  D  E  I  I  S  K  I  Y  E  V  G  K  L  I  S  L  Q

2041 gaattgtctgcattcaaagtgctaaagaatcatggaaacaaacttgcagaactaagtggt
 681 E  L  S  A  F  K  V  L  K  N  H  G  N  K  L  A  E  L  S  G

2101 ttgacacaactccgcggcactctacgaattacaaatcttgaaaatgtagggagtaaagaa
 701 L  T  Q  L  R  G  T  L  R  I  T  N  L  E  N  V  G  S  K  E

2161 gaagcaagcaaggctaaactgcacagaaaacagtatcttgaagcattagagttagagtgg
 721 E  A  S  K  A  K  L  H  R  K  Q  Y  L  E  A  L  E  L  E  W

2221 gcagctggccaggtttccagcttggagcatgagttacttgtctcggaggaagtattttta
 741 A  A  G  Q  V  S  S  L  E  H  E  L  L  V  S  E  E  V  F  L

2281 ggtctccaaccacatcacttcctcaaaagttcgacaatcagagggtacagtggtgctaca
 761 G  L  Q  P  H  H  F  L  K  S  S  T  I  R  G  Y  S  G  A  T

2341 gtacccagttggctggatgtgaaaatgctaccgaacttgggaactcttaaactagagaac
 781 V  P  S  W  L  D  V  K  M  L  P  N  L  G  T  L  K  L  E  N

2401 tgtacaagactggagggtctttcatatattggacaactgccacatctcaaggtccttcat
 801 C  T  R  L  E  G  L  S  Y  I  G  Q  L  P  H  L  K  V  L  H

2461 ataaagagaatgcctgtggtgaaacaaatgagtcatgaattatgtggctgtacgaaaagc
 821 I  K  R  M  P  V  V  K  Q  M  S  H  E  L  C  G  C  T  K  S

2521 aagttgttccctaggctagaagagttggtactggaggatatgccaacactgaaagaattc
 841 K  L  F  P  R  L  E  E  L  V  L  E  D  M  P  T  L  K  E  F

2581 ccgaatattgcacaacttccttgtctcaagattattcacatgaagaacatgttttcagta
 861 P  N  I  A  Q  L  P  C  L  K  I  I  H  M  K  N  M  F  S  V

2641 aaacatataggtcgtgaattatatggtgatatagagagcaattgttttccatcattagaa
 881 K  H  I  G  R  E  L  Y  G  D  I  E  S  N  C  F  P  S  L  E

2701 gagcttgtgctgcaggacatgctgacattggaggaactcccaaatcttggacaacttcca
 901 E  L  V  L  Q  D  M  L  T  L  E  E  L  P  N  L  G  Q  L  P

2761 catcttaaggttattcacatgaagaacatgtctgcactgaaacttataggtcgtgaatta
 921 H  L  K  V  I  H  M  K  N  M  S  A  L  K  L  I  G  R  E  L

2821 tgtggttctagagagaaaacttggtttcctaggctagaagtgctagtgctgaagaacatg
 941 C  G  S  R  E  K  T  W  F  P  R  L  E  V  L  V  L  K  N  M

2881 ctggcactggaggaactcccaagtcttggacaacttccatgtctcaaggttcttcgcatc
 961 L  A  L  E  E  L  P  S  L  G  Q  L  P  C  L  K  V  L  R  I

2941 caggtgtcgaaggtaggccatggactctttagtgctacgaggagtaaatggtttccaagg
 981 Q  V  S  K  V  G  H  G  L  F  S  A  T  R  S  K  W  F  P  R

3001 ctggaagagctagaaatcaagggcatgctgacatttgaggaactccattctcttgaaaaa
1001 L  E  E  L  E  I  K  G  M  L  T  F  E  E  L  H  S  L  E  K
```

```
3061 ctgccgtgtctcaaggttttccgcatcaagggattgccagcagtgaaaaagataggccat
1021 L  P  C  L  K  V  F  R  I  K  G  L  P  A  V  K  K  I  G  H

3121 ggattatttgattctacctgtcagagagagggttttccaaggttggaagagcttgtgtta
1041 G  L  F  D  S  T  C  Q  R  E  G  F  P  R  L  E  E  L  V  L

3181 agagacatgccagcgtgggaagagtggccttgggctgaaagggaggagttattttcctgc
1061 R  D  M  P  A  W  E  E  W  P  W  A  E  R  E  E  L  F  S  C

3241 ttgtgtagacttaaaattgaacaatgccccaaacttaaatgcttgcctcccgtcccttat
1081 L  C  R  L  K  I  E  Q  C  P  K  L  K  C  L  P  P  V  P  Y

3301 tctctcataaaacttgaattatggcaagttgggctgacaggacttccaggattatgcaaa
1101 S  L  I  K  L  E  L  W  Q  V  G  L  T  G  L  P  G  L  C  K

3361 ggaattggtggaggtagcagcgctagaactgcttctctttcactcttgcacattattaaa
1121 G  I  G  G  G  S  S  A  R  T  A  S  L  S  L  L  H  I  I  K

3421 tgcccaaatctgagaaatctgggagaagggttgctgtcaaaccacctgccacatatcaat
1141 C  P  N  L  R  N  L  G  E  G  L  L  S  N  H  L  P  H  I  N

3481 gctattcggatatgggaatgtgctgaactgttgtggctgcctgtcaagaggtttagagaa
1161 A  I  R  I  W  E  C  A  E  L  L  W  L  P  V  K  R  F  R  E

3541 ttcaccacccttgagaacttgtcaataaggaactgccccaagctcatgagcatgacacag
1181 F  T  T  L  E  N  L  S  I  R  N  C  P  K  L  M  S  M  T  Q

3601 tgtgaggagaatgacctcctcctcccgccgtcaatcaaggcgctagaattgggtgactgt
1201 C  E  E  N  D  L  L  L  P  P  S  I  K  A  L  E  L  G  D  C

3661 ggaaatcttgggaaatcgctgcctggatgcctacacaacctcagctcactaattcagttg
1221 G  N  L  G  K  S  L  P  G  C  L  H  N  L  S  S  L  I  Q  L

3721 gcgatatccaattgtccatacatggtttcctttccaagggacgtaatgcttcacttgaag
1241 A  I  S  N  C  P  Y  M  V  S  F  P  R  D  V  M  L  H  L  K

3781 gaacttggagctgtaaggatcatgaattgtgatgggctgagatcaatagagggtttacaa
1261 E  L  G  A  V  R  I  M  N  C  D  G  L  R  S  I  E  G  L  Q

3841 gttctcaaatcactcaagagattggaaatcataggatgtcccaggcttttgctaaatgaa
1281 V  L  K  S  L  K  R  L  E  I  I  G  C  P  R  L  L  L  N  E

3901 ggggatgagcaaggggaggtcttgtcactgcttgaattatcagtagataaaacagcccta
1301 G  D  E  Q  G  E  V  L  S  L  L  E  L  S  V  D  K  T  A  L

3961 cttaaactctcatttataaaaaatacactgccattcatccagtctctcagaatcatcttg
1321 L  K  L  S  F  I  K  N  T  L  P  F  I  Q  S  L  R  I  I  L

4021 tctcctcagaaagtgttgtttgactgggaggagcaggaattggtgcacagcttcaccgct
1341 S  P  Q  K  V  L  F  D  W  E  E  Q  E  L  V  H  S  F  T  A

4081 ctcaggcgccttgaattcctcagttgcaagaatctccagtccttgccaacagagttgcat
1361 L  R  R  L  E  F  L  S  C  K  N  L  Q  S  L  P  T  E  L  H

4141 acccttccttccctccatgctttggttgtaagtgactgtccacagatccaatcactgcca
1381 T  L  P  S  L  H  A  L  V  V  S  D  C  P  Q  I  Q  S  L  P

4201 tcgaagggactcccgacactcctcacagatttaggatttgaccattgccacccagtgctg
1401 S  K  G  L  P  T  L  L  T  D  L  G  F  D  H  C  H  P  V  L

4261 actgcgcaactggaaaagcacctggcagagatgaagagctcaggtcgatttcacccagtt
1421 T  A  Q  L  E  K  H  L  A  E  M  K  S  S  G  R  F  H  P  V

4321 tatgcatag
1441 Y  A  *
```

Figure 5.7 Nucleotide sequence and conceptual translation of RGC5 cDNA. Conserved motifs in R proteins are underlined in the NBS domain. Two additional NBS motifs that were not covered in chapter 4 are shaded in gray.

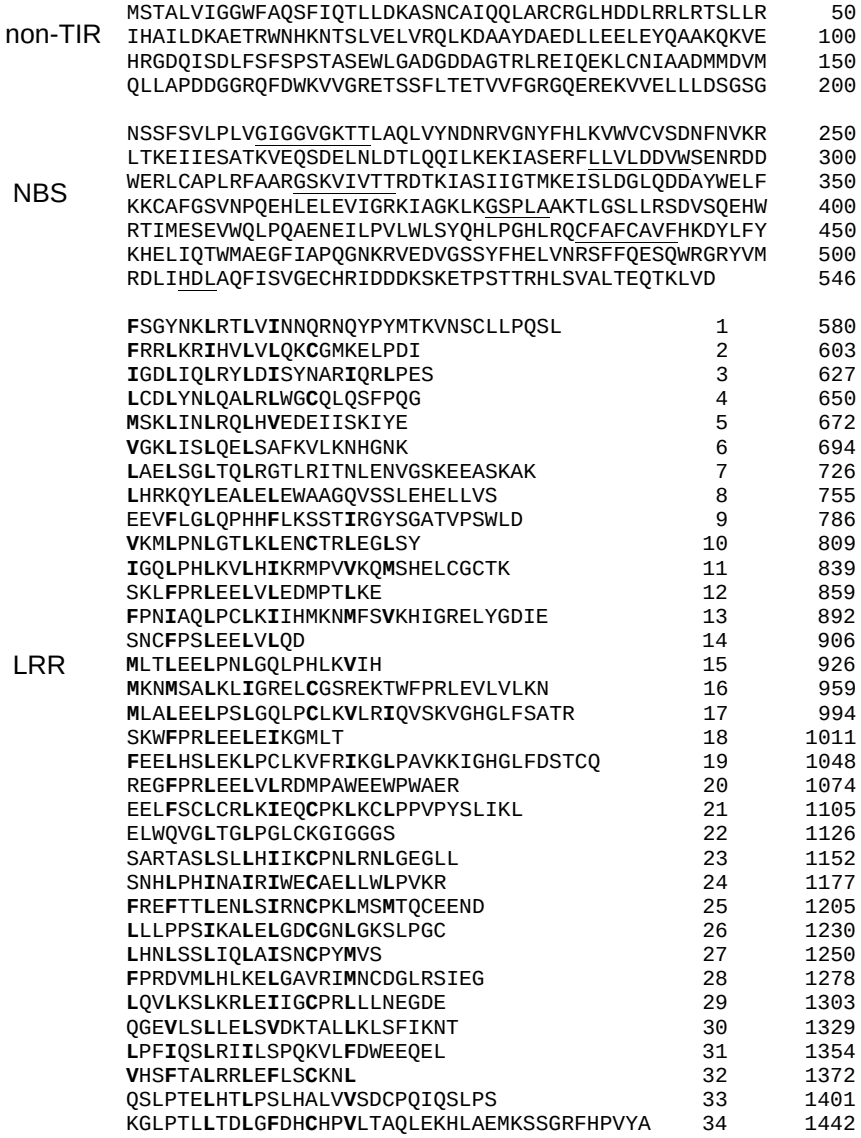


Figure 5.8. Predicted RGC5 protein structure. The amino acid sequence deduced from the RGC5 cDNA of roots is divided into three domains (non-TIR, NBS and LRR). Conserved motifs in R proteins are underlined. Amino acids matching the consensus of the cytoplasmic LRR are indicated in bold in the LRR domain.

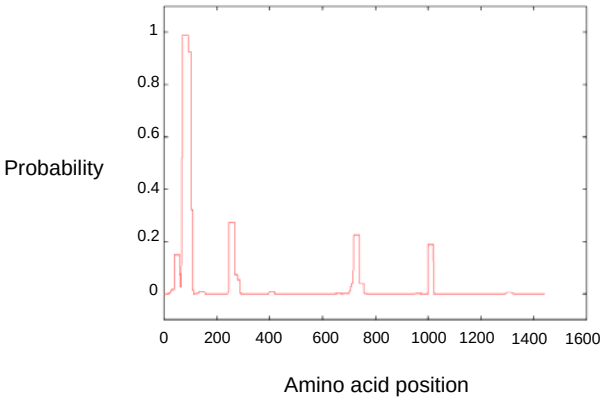


Figure 5.9 Predicted coiled-coil (CC) regions in the RGC5 deduced protein. The CC profile (window of 21 consecutive amino acids) was calculated according to Lupas (1996) and plotted against the amino acid number. Values shown on the y-axis of the graph indicates the probability to find potential CC regions in the amino acid sequence. One potential CC region is present in the non-TIR domain.

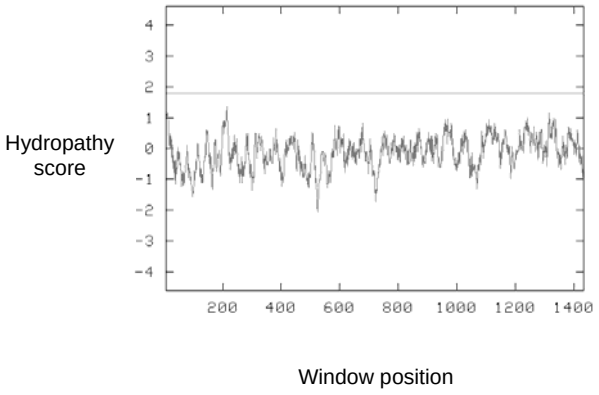


Figure 5.10 Hydropathy profile of the RGC5 deduced protein. The hydropathy profile (window of 19 consecutive amino acids) was calculated according to Kyte and Doolittle (1982) and plotted against the amino acid number. Peaks with hydropathy scores greater than 1.8 (horizontal line) indicate possible transmembrane regions.

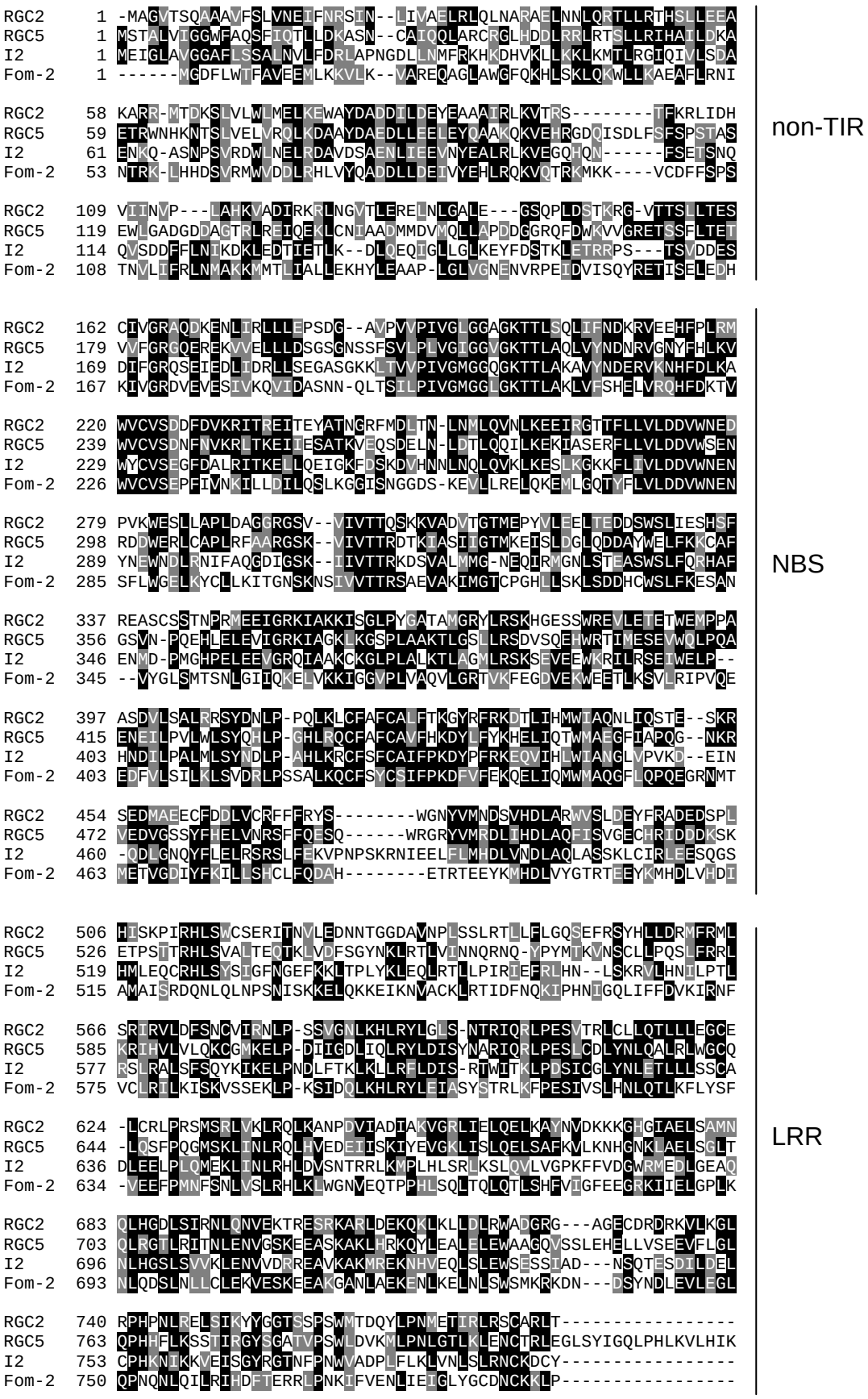
5.4.2 Sequence comparison of the banana RGC2 and RGC5 predicted proteins with the Fusarium resistance protein I2 from tomato*.*

Since the banana RGC2 and RGC5 showed significant sequence similarity to the *Fusarium* resistance gene *I2* from tomato, the level of similarity shared between the two proteins was examined at the level of the different domains. The *Fusarium* resistance gene *Fom-2* from melon was also included in the analysis as a reference to highlight the level of similarity between two known R genes that confer resistance to *F. oxysporum*. Overall, RGC2 and RGC5 share a relatively low level of amino acid sequence similarity to the I2 protein (Table 5.1). The highest levels of similarity shared by RGC2 vs I2 and RGC5 vs I2 were found in the NBS domain and the lowest levels were found in pairwise comparisons of the LRRs (Table 1). Pairwise comparison of the known *Fusarium* resistance proteins, tomato *I2* and melon *Fom-2* show similar patterns and levels of sequence similarity when the full proteins or individual domains are compared. Alignment of the four protein sequences revealed patterns of amino acid conservation along the different domains (Figure 5.11). A moderate level of conservation on the NBS domain and a relatively low level of conservation on the non-TIR and LRR domains were observed (Fig. 5.11). Notably, the difference in length of RGC5 with respect to RGC2, I2 and Fom-2 protein sequences was not restricted to short insertions spread at random over the sequence, but also two major insertions of similar size were found in the LRR domain; one of 106 amino acids in position 806-912 and the second one of 103 amino acids in position 943-1046 of the RGC5 deduced protein (Figure 5.11). A stretch of 27 amino acids shared by RGC2, RGC5, I2 and Fom-2 located between the two major insertions of RGC5 was detected in the LRR domain.

Table 5.1 Pairwise comparisons of the deduced amino acid sequences of RGC2, RGC5, *I2* and *Fom-2* genes using the ALIGN program (www.ebi.ac.uk). Percentages of amino acid sequence identity and similarity are indicated on the left and right of the slash, respectively.

		Domains[a]		
Comparison	Full protein	non-TIR	NBS	LRR
RGC2 vs RGC5	32.6/46.9	24.6/43.4	47.7/63.2	29.3/42
RGC2 vs I2	30.7/46.7	28.1/50	42/60.6	27.2/40.9
RGC5 vs I2	30.8/45.6	27.6/48.8	45.4/65.1	26.9/38.8
I2 vs Fom-2	28.8/48.5	23/38.3	34.9/55.3	27.2/48.1

[a]Domains analysed are defined in figures 5.3 and 5.8.


non-TIR
NBS
LRR

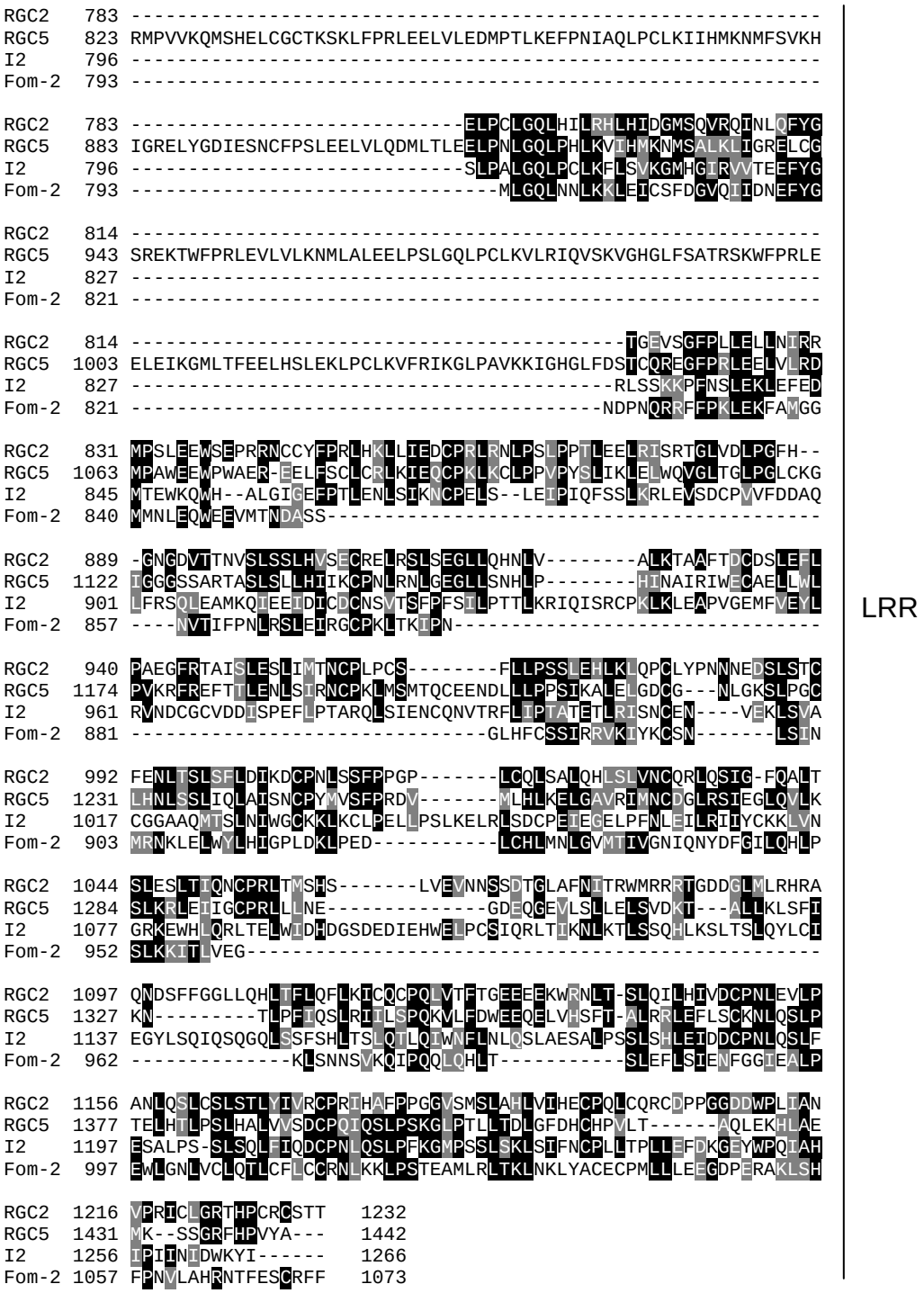


Figure 5.11 ClustalX alignment of the deduced amino acid sequences of RGC2, RGC5, I2 and Fom-2. Identical amino acids are shaded in black and conservative substitutions are shaded in grey. The non-TIR, NBS and LRR domains are indicated.

5.4.3 Phylogenetic relationships of the banana RGC2 and RGC5 sequences

RGC2 and RGC5 have been shown to belong to the non-TIR-NBS-LRR group of R genes (Chapter 4). In order to assess in more detail the phylogenetic relationships of RGC2 and RGC5 within the non-TIR group, a phylogenetic analysis using the deduced amino acid sequence of the NBS domain was undertaken. The NBS domain has been broadly used in phylogenetic studies of the NBS-LRR class of R genes because it contains numerous conserved motifs that assist proper alignment. These phylogenetic studies have shown that the NBS domain is very useful to predict the presence or absence of a TIR-like domain in the N-terminus of NBS-LRR proteins (Meyers et al. 1999; Pan et al. 2000) and also to define ancient phylogenetic clades (Cannon et al. 2000). NBS sequences from four previously defined ancient clades of non-TIR-NBS-LRR genes (N1, N2, N3 and N4) (Cannon et al. 2002) were included in the phylogenetic analysis along with NBS sequences from *Fusarium* resistance gene *I2* and other recently cloned R genes, such as the *Fusarium* resistance gene *Fom-2* from melon, and the other banana RGCs isolated in Chapter 4. The phylogenetic results (Figure 5.12) indicate that RGC2 and RGC5 are clustered in the N1 clade and this is well supported by a high bootstrap value (96%). Notably, the *Fusarium* resistance gene *I2* and *Fom-2* are also present in the N1 clade. The N1 clade also contains other characterized R genes such as *Xa1* that confers resistance to *Xanthomonas oryzae* in rice (Yoshimura et al. 1998), *Rp1-D* that confers resistance to *Puccinia sorghi* in *Zea mays* (Collins et al. 1999), *Rpg1-b* that confers resistance to *Pseudomonas syringae* in *Glycine max* (Ashfield et al. 2004), *RB* that confers resistance to *Phytophthora infestans* in *Solanum bulbocastanum* (Song et al. 2003) and *R3a* that confers resistance to *Phytophthora infestans* in *Solanum tuberosum* (Huang et al. 2005). Thus, in clade N1 we can find a set of highly divergent non-TIR-NBS-LRR genes that confer resistance to a diverse range of pathogens. It was interesting to observe that the tomato *I2* and melon *Fom-2* genes were not closely related in the N1 clade although they both confer resistance to *Fusarium oxysporum* (Figure 5.12). The R gene most closely related to tomato *I2* was the *Solanum R3a* gene, while the R gene most closely related

to melon *Fom-2* was the *Solanum RB* gene. These results suggest that *I2* and *Fom-2* are homologous in terms of structure, however due to the low level of sequence similarity and the distant phylogenetic relationship between these R genes it is difficult to determine whether these sequences are truly orthologous or their pathogen specificity arose by convergent evolution. Nevertheless, independent of their evolutionary origin, both sequences belong to the same phylogenetic clade, which opens the possibility that other divergent *Fusarium* resistance genes from different plant families may cluster in this phylogenetic clade. Thus, R genes that cluster in the N1 clade may serve as the first place to search for potential *Fusarium* resistance genes, such as the banana RGC2 and RGC5 sequences isolated in this study. In the case of the other banana RGCs, RGC1 and RGC4 clustered in clade N2, while RGC3 resolved in clade N3. None of the banana RGCs clustered in clade N4 (Figure 5.12).

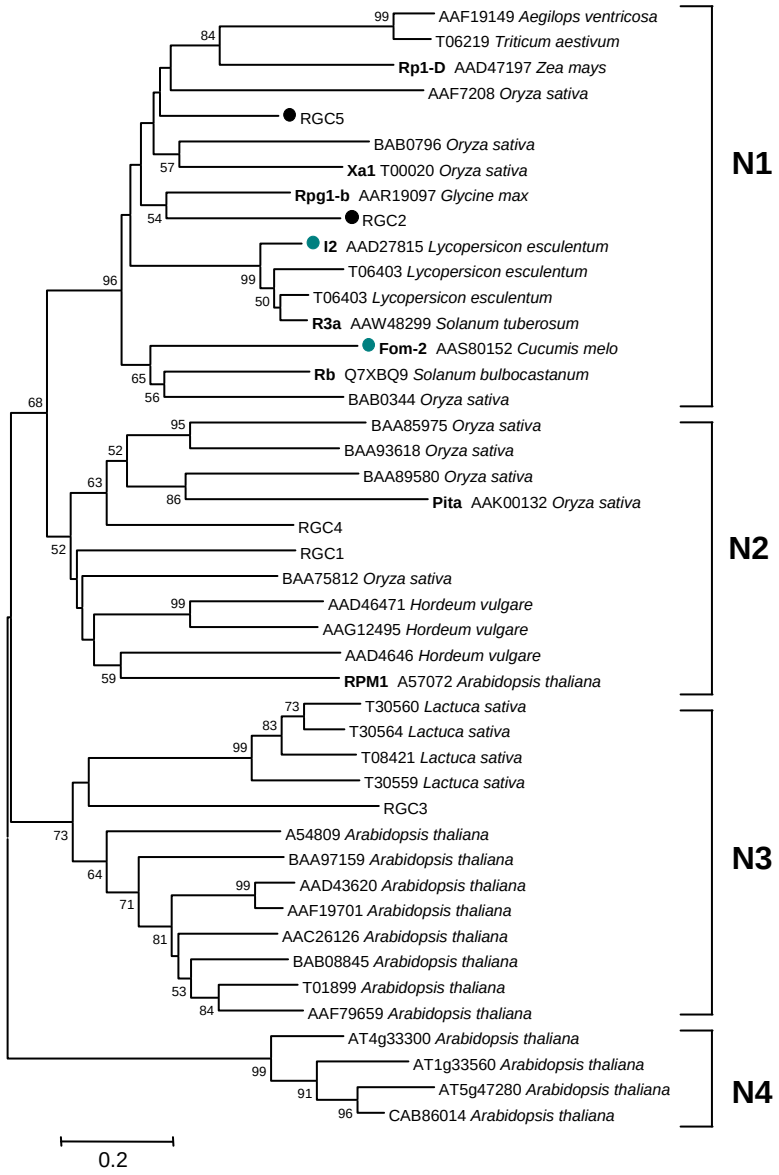


Figure 5.12 Phylogenetic analysis of banana RGC2 and RGC5 sequences (black circles). Sequences of four ancient non-TIR-NBS-LRR phylogenetic clades (N1, N2, N3 and N4) previously defined by Cannon et al. (2002) were used in the analysis. Three other banana RGCs (RGC1, RGC3 and RGC4) previously isolated (chapter 4) were also included in the phylogenetic tree. Characterised R genes are in bold and the *Fusarium* resistance genes *I2* and *Fom-2* from tomato and melon, respectively are highlighted with gray circles. Amino acid sequences from the P-loop to the GLPLA of the NBS domain were used for the analysis. The numbers below the branches indicate the percentage of 10,000 bootstrap replications supporting the particular nodes, and only those with >50% support are shown. The tree was constructed with the neighbor-joining method using the MEGA program version 2.1.

5.4.4 Isolation of a putative RGC2 promoter region and developing RGC2 constructs for genetic complementation tests.

Banana RGC2 mRNA expression was previously found to correlate with FOC race 4 resistance (Chapter 4). The absence of a transcript in susceptible plants suggests a mutation in the promoter region had abolished transcription. A genetic complementation approach in susceptible plants will be used to gain insights about the role of RGC2 in FOC race 4 resistance. In order to initiate genetic complementation experiments, four-expression cassettes harbouring different promoters, including a putative promoter region of RGC2, were developed.

Previous isolation of the 5' end of the cDNA of RGC2 using RNA Ligase Mediated Rapid Amplification of cDNA Ends (RLM-RACE) (Chapter 4), revealed a putative transcription start site based on the analysis of three RLM-RACE clones. In order to confirm this site, seven additional RLM-RACE clones were sequenced. The results indicated that all the 5'RACE sequences started at the same point (Figure 5.13). Thus the 5' untranslated region (5'UTR) of RGC2 has a length of 42 nt. The putative promoter region of RGC2 was isolated by PCR genome walking from an EcoR V genome walking library. A fragment of approximately 2.3 kb was amplified which included ~200 bp of the 5'UTR region and coding region. Sequencing revealed a putative promoter region of 2105 nt that did not show any significant sequence similarity hit to previously characterized R genes or any other sequence. The database of plant cis-acting regulatory DNA elements (PLACE) (www.dna.affrc.go.jp/htdocs/PLACE) was searched in order to find putative cis-acting regulatory elements in the promoter region of RGC2. Numerous putative cis-acting elements were found across the sequence (Appendix 1). A putative TATA-box-like motif was found at 86 nt upstream from the transcription start site (Figure 5.13). As eukaryotic genes transcribed by RNA polymerase II generally contain a TATA box located approximately 30 nt upstream from the transcription start site (Rombauts et al. 2003), the significance of this TATA-box-like motif, with respect to location and function, remains unclear. The putative promoter region of RGC2

including the 5'UTR was successfully fused to the RGC2 ORF in the binary vector pCAMBIA 2200 (RGC2ORF-NOST) as a Kpn I/BamH I fragment and the construct was named as pCAMBIA 2200 (RGC2PP-RGC2ORF-NOST)(Figure 5.14). This construct will be used to determine whether RGC2 is able to confer resistance to FOC race 4 in a susceptible banana genotype. Furthermore, the putative promoter region of RGC2 (RGC2PP) from the construct pCAMBIA 2200 (RGC2PP-RGC2ORF-NOST) was successfully replaced with three characterised promoters, the Ubi-1 promoter, the BT1 promoter and the NOS promoter, respectively (Figure 5.14). The use of different characterised promoters that drive different levels of gene expression may ensure the expression of the RGC2 in the transgenic banana plants if the uncharacterised promoter region of RGC2 fails to drive expression due to the absence of essential cis-acting regulatory elements in its sequence.

```
-2105 ATCATGTATGCATATATATGAAACATTGGGCCAAAAACTTTATTTAGCAAATTGATTGCCATGGATTAATCCGAGTAAAC 80
-2025 TTGACATGGGACAGCAAATTCATTTTGCAGCGACCATATCTAGGACTTGTGACAATGTGATGTTACAACAATAGAAAAAA 160
-1945 GGATTTTAGATTGGTGAACAATTTGATGTGTCACATCACAGGCAGACAACCCCACAATAAAGTGTGTTCCACCTAAACTT 240
-1865 ATTCTGCAACCACCCTTCCAGGACCAGCATGAAAGAAGGGTGTCATTAGTATGGCTTCTCTCTTGGTTCACACTAGACTA 320
-1785 GGTTCTCATCAGACACATCTTTTACATAGTTTTAATTCCTCTTTCCTGAAAAGATGTTTTTCTATGCAAGTCATGTTCAT 400
-1705 GTTACTGTTTGATCATAGTTTTCTGCATTAATGTCTCCAAGATAATTGGAAACATGACAATTGTGGCTTTTATAGCACAT 480
-1625 AGGATTATCATGTGTTTTGCAACTTGCATAGCTCTTAGTATTTGCTCCTTGGTCAGGAGAGTAGAAAGGCTCAGTAGTCC 560
-1545 TATTATTTTAGATCTTCAACAAGCCCTCAACTTGAGATGAAGATGTTCTTGTTAAGCACCCGTAGTTAATGAATTAAAGA 640
-1465 CTGTCCCTACTGGCCGTATAGGCAACAGGAGAGTCAGTCAACAGATAATTCAAGGTTCAAACAATCAATGTTGACTGGAC 720
-1385 TCAATCTTTTATTATGTTGGAGAATACTGTCATACCAAGTTAAATTATGTACCTTAAAAGAAGTAACACTTTGACCATGG 800
-1305 ACTAGAAATGAACGGAGCAGCCATTTTTTTTTATAGGACTTGAACCAAGAATTAGTACTCTGTTCCATGTCACAGGACGG 880
-1225 TCTGGGAGTAACCTGTCACTTATCTAATCCTCTTACCACACAATCCAAACTTAATTTCTAGCACAGTAAAGCAAAAAATA 960
-1145 AAGGGAGCAGAAGAGGATAAGATCGATGCAGTGTCACTGTAGGCTTGAAAAGGAGAGTGTCTTGACAAGGTGTGGAATTT 1040
-1065 GTCCAGGACAAATTGGAAGCCAATTGGCTCTTCTTATATCTTCTTCTCTTTTGTCAGCCTTGGGTATATTCACAGCATTT 1120
-985  TCTTGTGCTAAAAATAGGACCAATTATGTGATGTTGGTGTTGTTTCACCATCCACTGCTGTCATCCGATGTTTATATGCT 1200
-905  GCTAATTCGATAGTTGCGTCCGCCTACCGGCAACTCTTGTGTCAGCATATGAACTCTGCTTCTTTCTTTCTTTCCCTTTA 1280
-825  GCCTCAGCTTTTAGCCGAGCCATCATGTCTCAGTTAACATCAGCGTTTGGCCGCCAAGTCACTACAGGAGATAGAAAACA 1360
-745  AAGAAAGCACACAGTGGCCATTGAAAGGCCAAGCCACTTGGCCGCCAGGCAGCATCAAGACAAGTTGCTCTATGGACAGG 1440
-665  CAAAATTATTATTACTCTGTTGGCACATTGAAAGGGTAAACTTCCCTGCAAACGTAAATCCAACCCTCTATATCACAACT 1520
-585  GCTTTATGGTCACCCAACTGCCGGCACACAGTGGCCACCAACGCTGCCTGCCACACTTATCAACTTGATAGGGACCCCCT 1600
-505  TCCAGAAAAGCTCATGTCCATTTCAATGAATAAAGCAGTCTGGCAGGCAGTGTGGGCGTGCGCTGCAAGTCGTCTCTGTT 1680
-425  CACCCACTACTAGTGAGTTGCATAAGAACGAGCTATCTTCTTTAGGATTCACGAATTCACCAAAGTAGCTAATGGATAAA 1760
-345  CTTTGTCTTAAATTAATAAAGTATATTAACTTTTCGTCTTTGATCTAGAAAAGTATCTACGGTGTATGAACACCAAACTG 1840
-265  CTGTTTGTAATTTAGTGAAAGTTAACTGTAACTGTTCTTTATTTATCTGTTAGAATTCACTTCCCGCCCACTACGTTGCA 1920
-185  ATCGATGTTGTTGACAGTATTCACCGACAACATCGAGATAAACTGGTGTTTCCTGCATAATTTATGATTCCTGCCACCAA 2000
-105  CAATCACTTCGTCAAGTAATTATTTATCTTTTAGTTGACCGACTGGGTCATTGTTTATCCTCAAATTCATGGATCAAGAT 2080
-25   TTACAAGGGGAGCGACCTTGAATGGATCTTTGAGGTGATTCCCTCTATTTGCCATCAGAGAAGAAGAATG
                                 +1
```

Figure 5.13 Nucleotide sequence of the putative promoter region of banana RGC2. The transcription start site is designated +1 and the 5' untraslated region (5' UTR) is underlined. The ATG tranlation start codon is in bold. A putative TATA box near to the transcription start site is shaded in gray. This and other putative sites were predicted by the signal scan program of the plant cis-acting regulatory DNA elements (PLACE) database. Numbers on the left indicate the position of nucleotides upstream the transcription start site and numbers on the right are used as a reference to indicate the position of putative cis-acting regulatory elements predicted by the signal scan program (see table below).

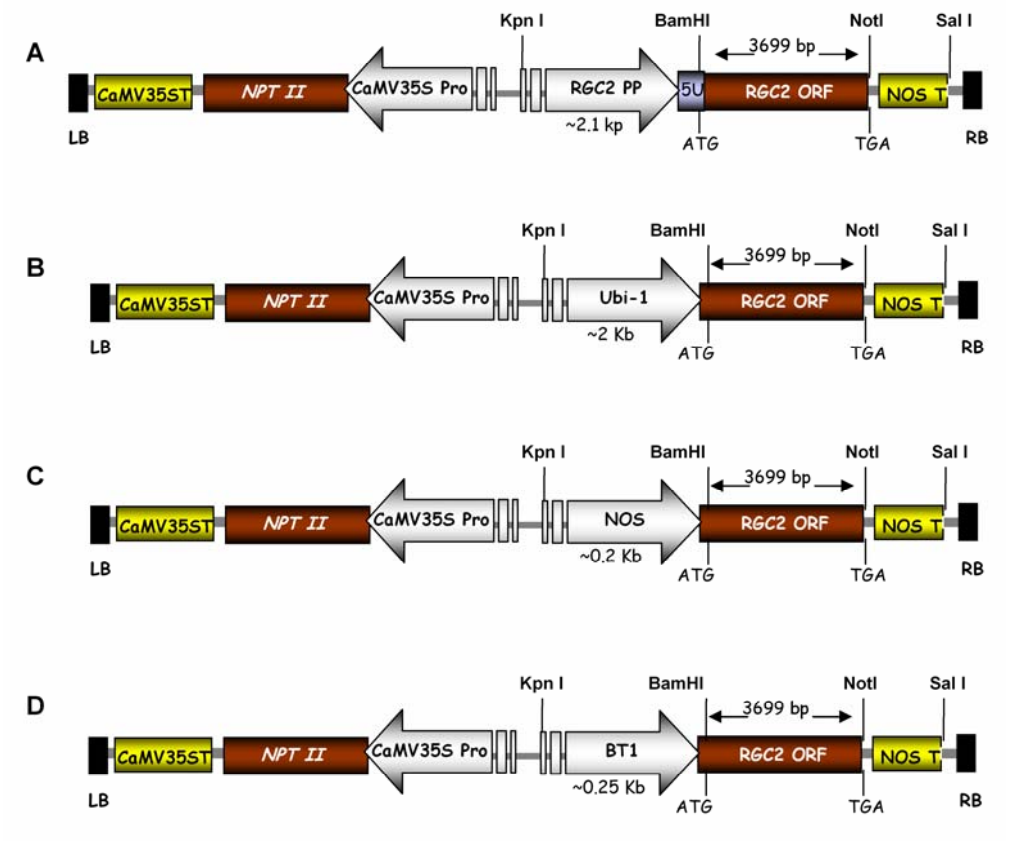


Figure 5.14 Schematic representation of four constructs to test the role of RGC2 in FOC race 4 resistance in a susceptible banana genotype. A, pCAMBIA 2200 (RGC2PP-RGC2ORF-NOST); B, pCAMBIA 2200 (Ubi-1-RGC2ORF-NOST); C, pCAMBIA 2200 (NOS-RGC2ORF-NOST); and D, pCAMBIA 2200 (BT1-RGC2ORF-NOST). The putative promoter (PP), the 5'UTR (5U) and ORF of RGC2 (RGC2ORF) are indicated in the expression cassette. The polyubiquitin-1 (Ubi-1), nopaline synthase (NOS), *Banana bunchy top virus* (BBTV) BT1 and *Cauliflower mosaic virus* 35S (CaMV35S pro) promoters are indicated. The neomycin phosphotransferase gene (*NPTII*) that confers resistance to kanamycin and the nopaline synthase terminator (NOS T), *Cauliflower mosaic virus* 35S terminator (CaMV35S T) are indicated. LB and RB represent the left and right T-DNA borders, respectively.

5.5 Discussion

This work provides structural and phylogenetic information of two potential *Fusarium* resistance genes from banana (RGC2 and RGC5). The deduced protein sequence of both RGC2 and RGC5 ORFs showed the typical motifs and domains of the non-TIR-NBS-LRR class of R genes (Meyers et al. 1999; Pan et al. 2000). This type of sequence is widespread in plant genomes, for example in the *Arabidopsis thaliana* and *Oryza sativa* genomes there are around 50 and 440 non-TIR-NBS-LRR related sequences, respectively (Meyers et al. 2003; Zhou et al. 2004). The genome of banana has yet to be determined and therefore the question remains as to the abundance of these types of sequences in this crop. By cloning the entire ORF of RGC2 and RGC5, the presence of a predicted coiled-coil (CC) structure at the non-TIR domain (Figures 5.4 and 5.9) and the typical NBS motifs present in the NBS-LRR class of R proteins (Figures 5.3 and 5.8) have been confirmed (Meyers et al. 1999; Pan et al. 2000). The full ORF of RGC2 and RGC5 also revealed two other NBS motifs (Figure 5.2 and 5.7) previously unobtainable as a consequence of the experimental approach adopted (Chapter 4). As expected, the C-terminus of both RGC2 and RGC5 contained LRR repeats of variable size, whose consensus sequences are related to cytoplasmic LRR proteins (Jones and Jones 1997) (Figure 5.3 and 5.8). It has been shown that the LRR domain of alleles of the flax rust resistance gene *L* determines recognition of specific races of the pathogen (Ellis et al. 1999). Therefore, the LRR of RGC2 and RGC5 may function in a similar way by recognizing an invading pathogen. The total number of LRRs found in RGC2 and RGC5 were similar to previously characterized R gene products, whose LRRs numbers vary from 14 to 40 (Jones and Jones 1997). The variation of LRR repeats may play a role in determining the recognition specificity of the RGC2 and RGC5 gene products. It has been demonstrated that expansion and contraction of LRR repeats are responsible for loss of function or recognition specificities of plant disease resistance genes. In flax, inactivation of the rust resistance gene *M* was associated with the loss of a single repeated unit within the LRR coding region (Anderson et al. 1997). Domain swapping and gene shuffling of tomato R proteins Cf-4 and Cf-9 also

demonstrated that variation in LRR copy number plays a major role in determining recognition specificity in these proteins against *Cladosporium fulvum* (Wulff et al. 2001). Scanning the LRR domain of RGC2 and RGC5 with the program COILS (Lupas et al. 1996) revealed a predicted CC in the LRR domain of RGC2 (Figure 5.4). The CC structure is quite common in the N-terminus (non-TIR domain) of R proteins (Meyers et al. 2003; Zhou et al. 2004) but rare in the C-terminus (LRR domain). There are just a few examples of R genes showing this predicted structure in the LRR, among these, we have the *Fusarium Fom-2* resistance gene from melon, which lacks a putative CC structure in the non-TIR domain but it appears to have one in the LRR domain (Joobeur et al. 2004). The function of this predicted CC structure in the LRR domain remains to be shown.

The *Fusarium* resistance gene *I2* from tomato was found to be one of the most similar R genes to RGC2 and RGC5 in homology searches using either the truncated N-terminal region (Chapter 4) or the entire ORF sequence. This prompts speculation that RGC2 as well as RGC5 may have a role in *Fusarium* resistance in bananas. Overall, the banana RGC2 and RGC5 gene products aligned with the *I2* gene predicted protein (Figure 5.11), although the similarity between them was relatively low (Table 1). This level of sequence similarity is quite common among NBS-LRR genes from different plant families (Bai et al. 2002) or even in NBS-LRR resistances genes whose pathogen recognition specificity is very similar (Joobeur et al. 2004; Ashfield at al. 2004; McDowell 2004). For example, homology searches using the non-TIR-NBS-LRR *Fom-2* resistance gene from melon as the query revealed that the most similar characterized R gene to the *Fom-2* was the *Fusarium* resistance gene *I2* from tomato (29% identity and 49% similarity; E value = $2e^{-88}$) (Joobeur et al. 2004). Another example is the non-TIR-NBS-LRR gene pair *RPM1* and *Rpg1-b* from *Arabidopsis thaliana* and *Glycine max*, respectively. Both gene products confer resistance to the bacterium *Pseudomonas syringae* by recognizing the same avirulence protein (AvrB) (Ashfield at al. 2004; McDowell 2004). Alignment of the predicted RPM1 and Rpg1 protein sequences revealed a relatively low level of amino acid sequence identity across the NBS domain (~34%) and they were not

phylogenetically closely related. Their respective LRR sequences were so divergent that they could not be aligned *in silico* beyond the first five LRR repeats (Ashfield at al. 2004).

In this study, it was found that the *I2* and *Fom-2* gene products share a limited sequence similarity. They are not closely related in the phylogenetic tree and they are more similar to other non-*Fusarium* R genes than between each other, however they both belong to the same phylogenetic clade N1 previously defined by Cannon et al. (2000). Thus, it is reasonable to hypothesize that other divergent R genes that confer resistance to *F. oxysporum* may cluster in the same phylogenetic clade. Based on this hypothesis, the N1 clade may be an interesting place to start the search for potential *Fusarium* resistance genes not only in the Musaceae familiy (RGC2 and RGC5) but in other plant species as well. The idea of having multiple *Fusarium* R genes from different plant families clustering in one single clade could have a practical implication in the isolation of novel *Fusarium* R genes, since the search for potential *Fusarium* R genes could be targeted to one particular phylogenetic clade. Taking into account that there are a large number of divergent NBS-LRR sequences in a plant genome (Meyers et al. 2003; Zhou et al. 2004), such an approach could facilitate the selection of a particular set of resistance gene candidates to be used in *Fusarium* resistance tests. A similar idea to narrow the search for potential functional resistance genes by using bioinformatic tools has been recently proposed for the Solanaceae family (Couch et al. 2006).

The banana RGC2 and RGC5 sequences did not only cluster within the N1 clade along with the *Fusarium I2*/*Fom-2* resistance genes but also they are non-TIR-NBS-LRR sequences that show significant sequence similarity to the *Fusarium* resistance gene *I2* from tomato. All these features make RGC2 and RGC5 interesting non-TIR-NBS-LRR sequences that could be associated with *Fusarium* resistance in banana. The particular case of RGC2 is even more notable since its expression was associated to *Fusarium oxysporum* f. sp. *cubense* race 4 (FOC race 4) resistance (Chapter 4). The function of RGC2 is currently being evaluated using a genetic

complementation approach (see below) in order to test whether this sequence does indeed confer FOC race 4 resistance. On the other hand, the role of RGC5 in FOC resistance will be tested against different FOC races. Taking into account that the *Fusarium* resistance in tomato and melon is controlled by more than a single resistance gene (Sela-Buurlage et al. 2001; Schreuder et al. 2000), it is probable that multiple R genes confer resistance to the four different and divergent races of FOC in banana (Koenig et al. 1997; O'Donnell et al. 1998). Furthermore, the isolation of more banana NBS sequences that potentially cluster within the clade N1 could provide a valuable resource of RGC genes that could be used in *Fusarium* resistance tests.

The function of most R genes has been validated in genetic complementation experiments where the R gene candidate with all its regulatory expression sequences, promoter, 5' and 3' UTR's are transferred via genetic transformation into a susceptible genotype in order to prove whether the R gene candidate is capable of confering resistance to a particular pathogen. Furthermore, the expression of most R genes is low and constitutive (Hulbert et al. 2001), thus in order to make constructs to validate the function of a R gene candidate the best choice of promoter is likely to be a native R gene promoter that drives the proper level of expression thus avoiding the activation of a constitutive hypersensitive response that could be deleterious to the plant or may not lead to pathogen resistance (Gurr and Rushton 2005). Taking into consideration the information above, in this study a putative native promoter region of RGC2 was isolated and an expression cassette was made with the RGC2 ORF (Figure 5.14). This construct will be used to test whether RGC2 is capable of confering FOC race 4 resistance in a susceptible genotype. Three other constructs with the RGC2 ORF were also made using heterologous and characterised promoters that drive different levels of gene expression in plants. One of the constructs contains the constitutive Ubi-1 promoter which is reported to drive high levels of gene expression in monocotyledonous plants (Christensen et al. 1996). Although a high level of expression of a resistance gene could be deleterious to the plant (Gurr and Rushton 2005), there are reports showing that this not

always necessarily the case. For example, the pepper *Bs2* gene under the control of the strong 35S promoter confers resistance to *Xanthomonas campestris* in both tomato and tobacco (Tai et al. 1999) which indicates that even the presence of a strong promoter controlling the expression of a R gene can lead to pathogen resistance without obvious detrimental effects on the plant. It remains to be seen the effect of overexpressing the RCG2 in FOC race 4 resistance in banana. Two other constructs were made using the promoters NOS and BT1, the relatively low expression that these promoters drive perhaps will mimic the expression of a R gene candidate such as RGC2 whose expression was detected to be low in comparison to the banana *Actin 1* gene (see Figure 4.10, Chapter 4). All these constructs with heterologous promoters may ensure the expression of RGC2 in the transgenic banana plants if the uncharacterised putative promoter region of this sequence fails to drive gene expression. The latter could occur if the isolated region containing the putative promoter region (~2.1 kb) does not contain all the necessary cis-acting regulatory elements that are essential to drive the proper expression.

In summary, this study reports the isolation of the first two banana non-TIR-NBS-LRR cDNA sequences (RGC2 and RGC5) from the roots of a FOC race 4 resistant banana *M. acuminata* ssp. *malaccensis*. Phylogenetic analysis grouped the *Fusarium* resistance genes *I2* and *Fom-2* into the same clade (N1), opening the possibility that other unknown *Fusarium* R genes from different plant families may share the same clade N1. This clade could be used as a platform to narrow the search for potential *Fusarium* R genes in banana and other crops. Both RGC2 and RGC5 cluster in the clade N1 making these banana sequences a pair of attractive resistance gene candidates that could be associated to FOC resistance. The interesting correlation of RGC2 expression with FOC race 4 resistance (Chapter 4) will allow to test its role against this particular FOC race using a genetic complementation approach, whereas RGC5 role in FOC resistance will be evaluated against different FOC races. These experiments are currently underway.

5.6 References

Altschul, S. F., Madden, T.L., Schaffer, 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: 3389-3402.

Anderson, P.A., Lawrence, G.J., Morrish, B.C., Ayliffe, M.A., Finnegan, E.J., and Ellis, J.G. (1997). Inactivation of the flax resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeat coding region. *Plant Cell* 9:641-651.

Ashfield, T., Ong, L., Nobuta, K., Schneider, C., and Innes, R. (2004). Convergent evolution of disease resistance gene specificity in two flowering plant families. *Plant Cell* 16: 309-318.

Bai, J., Pennill, L.A., Ning, J., Lee, S.W., Ramalingam, J., Leach, J.E. and Hulbert, S.H. (2002). Diversity in nucleotide binding site-leucine rich repeat genes in cereals. *Genome Research* 12: 1871-1884.

Becker, D., Dugdale, B., Smith, M., Harding, R., and Dale, J. (2000). Genetic transformation of Cavendish banana (Musa spp. AAA group) cv 'Grand Nain' via microprojectile bombardment. *Plant Cell Reports* 19: 229-234.

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: 391-399.

Cannon, S. B., Zhu, H., Baumgarten, A. M., Spangler, R., May, G., Cook, D. R., and Young, N.D. (2000). Diversity, distribution and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. *Journal of Molecular Evolution* 54: 548-562.

Christensen A.H. and Quail P.H. (1996). Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Research* 5: 213-218.

Collins, N., Drake, J., Ayliffe, M., Sun, Q., Ellis, J., Hulbert, S. and Pryor, T. (1999) Molecular characterization of the Maize *Rp1-D* rust resistance haplotype and its mutants. *Plant Cell* 11: 1365-1376.

Couch, B.C., Spangler, R., Ramos, C. and May, G. (2006). Pervasive purifying selection characterizes the evolution of *I2* homologues. *Molecular Plant Microbe Interactions* 19: 288-303.

Dangl, J., and Jones, J. (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411: 826-833.

Di Pietro, A., Madrid, M., Caracuel, Z., Delgado-Jarana, J., and Roncero, M.I. (2003). *Fusarium oxysporum*: exploring the molecular arsenal of vascular wilt fungus. *Molecular Plant Pathology* 4: 315-325.

Dugdale B., Becker D., Beetham P., Harding R.M., Dale J.L (2000). Promoters derived from banana bunchy top virus DNA-1 to 5 direct vascular-associated expression in transgenic banana (Musa spp.). *Plant Cell Reports* 19: 810-814.

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: 495-506.

Gurr, S.J. and Rushton, P.J. (2005). Engineering plants with increased disease resistance: how are we going to express it ? *Trends in Biotechnology* 23: 283-290.

Hammond-Kosack, K.E., and Jones, J.D.G. (1997). Plant disease resistance genes. *Annual Review of Plant Physiology and Plant Molecular Biology* 48: 575-607.

Hammond-Kosack, K., and Parker, J. (2003). Deciphering plant-pathogen communication: fresh perspectives for molecular resistance breeding. *Current Opinion in Biotechnology* 14: 177-193.

Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research* 27: 297-300.

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: 251-261.

Hulbert, S.H., Webb, C.A., Smith, S.M. and Sun, Q. (2001). Resistance gene complexes: evolution and utilization. *Annual Review of Phytopathology* 39: 285-312.

Hulo, N., Bairoch A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P.S., Pagni, M., and Sigrist, C.J.A. (2006). The PROSITE database. *Nucleic Acids Research* 34: D227-D230.

Joobeur, T., King, J., Nolin, S., Thomas, C., and Dean, R. (2004). The *Fusarium* wilt resistance locus *Fom-2* of melon contains a single resistance gene with complex features. *Plant Journal* 39: 283-297.

Jones, D.A., and Jones, J.D.G. (1997). The role of leucine-rich repeat proteins in plant defenses. *Advance in Botanical Research* 24: 89-167.

Khanna, H., Becker, D., Kleidon, J. and Dale, J. (2004). Centrifugation assisted *Agrobacterium*-mediated transformation (CAAT) of embryogenic cell suspensions of banana (*Musa* spp. Cavendish AAA and Lady finger AAB). *Molecular Breeding* 14: 239-252.

Koenig, R., Ploetz, R., and Kistler, H. (1997). *Fusarium oxysporum* f. sp. *cubense* consists of a small number of divergent and globally distributed clonal lineages. *Phytopathology* 87: 915-923.

Kumar, S., Tamura, K., Jakobsen, I.B., and Nei, M. (2001). MEGA2: Molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245.

Kyte, J. and R.F. Doolittle (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* 157: 105-132.

Lupas, A. (1996). Prediction and analysis of coiled-coil structures. *Methods in Enzymology* 266: 513-525.

Martin, G.B., Bogdanove, A.J. and G. Sessa (2003). Understanding the functions of plant disease resistance proteins. *Annual Review of Plant Biology* 54: 23-61.

Mes, J., Van Doorn, A., Wijbrandi, J., Simons, G., Cornelissen, B., and Haring, M. (2000). Expression of the Fusarium resistance gene *I-2* colocalizes with the site of fungal containment. *Plant Journal* 23: 183-193.

Meyers, B.C, Dickerman, A., Michelmore, R., Sivaramakrishnan, S., Sobral, B. and Young, N. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant Journal* 20: 317-332.

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: 809-834.

O'Donnell, K., Kistler, H., Cigelnik, E., and Ploetz, R. (1998). Multiple evolutionary origins of the fungus causing Panama disease of banana: Concordant evidence from nuclear and mitochondrial gene genealogies. *Proceedings of the National Academy of Science (USA)* 95: 2044-2049.

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: 203-213.

Ploetz, R.C., and Pegg, K.G. (2000). Fungal diseases of the root, corm and pseudostem. In Diseases of banana, abaca and enset. Ed. D.R. Jones. CAB international. pp. 143-172.

Ploetz, R.C. (2005). Panama disease, an old nemesis rears its ugly head: part 1, the beginnings of the banana export trades. Online. Plant Health Progress doi:10.1094/PHP-2005-1221-01-RV.

Rombauts, S., Florquin, K., Lescot, M., Marchal, K., Rouze, P., and Van de Peer, Y. (2003). Computational approaches to identify promoters and cis-regulatory elements in plant genomes. *Plant Physiology* 132: 1162-1176.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425.

Sanders, P.R., Winter, J.A., Barnason, A.R., Rogers, S.G. and R.T. Fraley (1987). Comparison of cauliflower mosaic virus 35S and nopaline synthase promoters in transgenic plants. *Nucleic Acids Research* 15: 1543-1558.

Schuler, M. A. and Zielinski, R.E. (1989). RNA isolation from light- and dark-grown seedlings. In *methods in plant molecular biology*. Academic Press Inc. pp. 89-96.

Sela-Buurlage, M., Budai-Hadrian, O., Pan, Q., Carmel-Goren, L., Vunsch, R., Zamir, D., and Fluhr, R. (2001). Genome-wide dissection of *Fusarium* resistance in tomato reveals multiple complex loci. *Molecular Genetics and Genomics* 265: 1104-1111.

Shaw C.H., Carter G.H., Watson M.D. and Shaw C.H. (1984). A functional map of the nopaline synthase promoter. *Nucleic Acids Research* 12: 7831-7846.

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A. and Lukyanov, S.A. (1995) An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Research* 23: 1087-1088.

Simons, G., Groenendijk, J., Wijbrandi, J., Reijans, M., Groenen, J., Diergaarde, P., Van der Lee, T., Bleeker, M., Onstenk, J., Both, M., Haring, M., Mes, J., Cornelisse, B., Zabeau, M., and Vos, P. (1998). Dissection of the Fusarium *I2* gene cluster in tomato reveals six homologs and one active gene copy. *Plant Cell* 10:1055-1068.

Tai, T., Dahlbeck, D., Clarck, E., Gajiwala, P., Pasion, R., Whalen, M., Stall, R. and Staskawicz, J. (1999). Expression of the *Bs2* pepper gene confers resistance to bacteria spot disease in tomato. *Proceedings of the National Academy of Science (USA)* 96: 14153-14158.

Thompson, J.D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 24: 4876-4882.

Van der Vossen, E., Sikkema, A., Lintel Hekkert, B., Gros, J., Stevens, P., Muskens, M., Wouters, D., Pereira, A., Stiekema, W., 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: 867-882.

Wulff, B.B., Thomas, C.M., Smoker, M., Grant, M. and Jones, J.D. (2001). Domain swapping and gene shuffling identify sequences required for induction of an Avr-dependent hypersensitive response by the tomato Cf-4 and Cf-9 proteins. *Plant Cell* 13: 255-272.

Yoshimura, S., Yamanouchi, U., Katayose, Y. and Toki, S. (1998) Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. *Proceedings of the National Academy of Science (USA)* 95: 1663-1668.

Zhou, T., Wang, T., Chen, J. Q., Araki, H. and Jing, Z., Jiang, K., Shen, J. and Tian D. (2004). Genome-wide identification of NBS genes in *japonica* rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Molecular Genetics and Genomics* 271: 402-415.

5.7 Appendix 1

Cis-acting regulatory elements of the RGC2 putative promoter region predicted by the signal scan program using the plant cis-acting regulatory DNA elements (PLACE) database.

Factor or Site Name		Strand	Signal Sequence	PLACE accession numer
CARGCW8GAT	site	11 (+)	CWWWWWWWWG	S000431
CARGCW8GAT	site	11 (-)	CWWWWWWWWG	S000431
LEAFYATAG	site	23 (-)	CCAATGT	S000432
CAATBOX1	site	25 (-)	CAAT	S000028
CCAATBOX1	site	25 (-)	CCAAT	S000030
SITEIIATCYTC	site	27 (+)	TGGGCY	S000474
SORLIP2AT	site	28 (+)	GGGCC	S000483
E2FCONSENSUS	site	28 (-)	WTTSSCSS	S000476
SEF4MOTIFGM7S	site	32 (-)	RTTTTTR	S000103
NTBBF1ARROLB	site	37 (+)	ACTTTA	S000273
CARGCW8GAT	site	38 (+)	CWWWWWWWWG	S000431
CARGCW8GAT	site	38 (-)	CWWWWWWWWG	S000431
DOFCOREZM	site	38 (-)	AAAG	S000265
TAAAGSTKST1	site	38 (-)	TAAAG	S000387
POLASIG1	site	39 (-)	AATAAA	S000080
TATABOX5	site	40 (+)	TTATTT	S000203
CAATBOX1	site	51 (-)	CAAT	S000028
ARR1AT	site	53 (+)	NGATT	S000454
CAATBOX1	site	55 (-)	CAAT	S000028
ARR1AT	site	63 (+)	NGATT	S000454
ARR1AT	site	68 (-)	NGATT	S000454
CACTFTPPCA1	site	74 (-)	YACT	S000449
WBOXATNPR1	site	81 (+)	TTGAC	S000390
WRKY71OS	site	82 (+)	TGAC	S000447
AUXREPSIAA4	site	86 (-)	KGTCCCAT	S000026
INRNTPSADB	site	99 (+)	YTCANTYY	S000395
-300ELEMENT	site	102 (-)	TGHAAARK	S000122
GATABOX	site	117 (-)	GATA	S000039
TGTCACACMCUCUMISIN	site	128 (-)	TGTCACA	S000422
GTGANTG10	site	129 (+)	GTGA	S000378
SEBFCONSSTPR10A	site	129 (-)	YTGTCWC	S000391
WRKY71OS	site	130 (+)	TGAC	S000447
CAATBOX1	site	133 (+)	CAAT	S000028
GTGANTG10	site	137 (+)	GTGA	S000378
RAV1AAT	site	146 (+)	CAACA	S000314
CAATBOX1	site	149 (+)	CAAT	S000028
SURE1STPAT21	site	150 (+)	AATAGAAAA	S000186
BOXIINTPATPB	site	151 (+)	ATAGAA	S000296
POLLEN1LELAT52	site	153 (+)	AGAAA	S000245
GT1CONSENSUS	site	154 (+)	GRWAAW	S000198
GT1GMSCAM4	site	154 (+)	GAAAAA	S000453
PYRIMIDINEBOXOSRAMY1A	site	157 (-)	CCTTTT	S000259
DOFCOREZM	site	158 (+)	AAAG	S000265
ARR1AT	site	161 (+)	NGATT	S000454
ARR1AT	site	168 (+)	NGATT	S000454
CAATBOX1	site	170 (-)	CAAT	S000028
CCAATBOX1	site	170 (-)	CCAAT	S000030
GTGANTG10	site	174 (+)	GTGA	S000378
CAATBOX1	site	179 (+)	CAAT	S000028
TGTCACACMCUCUMISIN	site	189 (+)	TGTCACA	S000422
WRKY71OS	site	190 (-)	TGAC	S000447
GTGANTG10	site	191 (-)	GTGA	S000378
GTGANTG10	site	196 (-)	GTGA	S000378
CAATBOX1	site	215 (+)	CAAT	S000028
POLASIG1	site	216 (+)	AATAAA	S000080
TAAAGSTKST1	site	218 (+)	TAAAG	S000387
NTBBF1ARROLB	site	218 (-)	ACTTTA	S000273
DOFCOREZM	site	219 (+)	AAAG	S000265
CACTFTPPCA1	site	221 (-)	YACT	S000449
-10PEHVPSBD	site	240 (+)	TATTCT	S000392
AMMORESIIUDCRNIA1	site	252 (-)	GGWAGGGT	S000374
DOFCOREZM	site	272 (+)	AAAG	S000265
WRKY71OS	site	282 (-)	TGAC	S000447
CACTFTPPCA1	site	288 (-)	YACT	S000449
NODCON2GM	site	300 (+)	CTCTT	S000462
OSE2ROOTNODULE	site	300 (+)	CTCTT	S000468
REALPHALGLHCB21	site	303 (-)	AACCAA	S000362
GTGANTG10	site	308 (-)	GTGA	S000378
DPBFCOREDCDC3	site	310 (+)	ACACNNG	S000292
CACTFTPPCA1	site	311 (+)	YACT	S000449
NODCON1GM	site	337 (-)	AAAGAT	S000461
OSE1ROOTNODULE	site	337 (-)	AAAGAT	S000467
-300ELEMENT	site	339 (-)	TGHAAARK	S000122

DOFCOREZM	site	339	(-)	AAAG	S000265
POLASIG2	site	351	(-)	AATTAAA	S000081
NODCON2GM	site	359	(+)	CTCTT	S000462
OSE2ROOTNODULE	site	359	(+)	CTCTT	S000468
DOFCOREZM	site	361	(-)	AAAG	S000265
DOFCOREZM	site	370	(+)	AAAG	S000265
NODCON1GM	site	370	(+)	AAAGAT	S000461
OSE1ROOTNODULE	site	370	(+)	AAAGAT	S000467
GT1CONSENSUS	site	377	(-)	GRWAAW	S000198
GT1GMSCAM4	site	377	(-)	GAAAAA	S000453
POLLEN1LELAT52	site	379	(-)	AGAAA	S000245
BOXIINTPATPB	site	380	(-)	ATAGAA	S000296
WBOXHVISO1	site	389	(-)	TGACT	S000442
WBOXNTERF3	site	389	(-)	TGACY	S000457
WRKY71OS	site	390	(-)	TGAC	S000447
CACTFTPPCA1	site	403	(+)	YACT	S000449
POLLEN1LELAT52	site	420	(-)	AGAAA	S000245
ARFAT	site	432	(+)	TGTCTC	S000270
GATABOX	site	441	(+)	GATA	S000039
GT1CONSENSUS	site	441	(+)	GRWAAW	S000198
IBOXCORE	site	441	(+)	GATAA	S000199
CAATBOX1	site	445	(-)	CAAT	S000028
CCAATBOX1	site	445	(-)	CCAAT	S000030
WRKY71OS	site	455	(+)	TGAC	S000447
CAATBOX1	site	458	(+)	CAAT	S000028
EBOXBNNAPA	site	458	(+)	CANNTG	S000144
MYCCONSENSUSAT	site	458	(+)	CANNTG	S000407
EBOXBNNAPA	site	458	(-)	CANNTG	S000144
MYCCONSENSUSAT	site	458	(-)	CANNTG	S000407
CAATBOX1	site	460	(-)	CAAT	S000028
SORLIP1AT	site	463	(-)	GCCAC	S000482
DOFCOREZM	site	467	(-)	AAAG	S000265
ARR1AT	site	482	(+)	NGATT	S000454
GT1CONSENSUS	site	484	(-)	GRWAAW	S000198
IBOXCORE	site	485	(-)	GATAA	S000199
GATABOX	site	486	(-)	GATA	S000039
EBOXBNNAPA	site	489	(+)	CANNTG	S000144
MYCATERD1	site	489	(+)	CATGTG	S000413
MYCCONSENSUSAT	site	489	(+)	CANNTG	S000407
DPBFCOREDCDC3	site	489	(-)	ACACNNG	S000292
EBOXBNNAPA	site	489	(-)	CANNTG	S000144
MYCATRD22	site	489	(-)	CACATG	S000174
MYCCONSENSUSAT	site	489	(-)	CANNTG	S000407
ANAERO4CONSENSUS	site	494	(+)	GTTTHGCAA	S000480
NODCON2GM	site	512	(+)	CTCTT	S000462
OSE2ROOTNODULE	site	512	(+)	CTCTT	S000468
CACTFTPPCA1	site	517	(-)	YACT	S000449
WBOXNTERF3	site	531	(-)	TGACY	S000457
WRKY71OS	site	532	(-)	TGAC	S000447
CACTFTPPCA1	site	540	(-)	YACT	S000449
POLLEN1LELAT52	site	543	(+)	AGAAA	S000245
DOFCOREZM	site	545	(+)	AAAG	S000265
CACTFTPPCA1	site	553	(-)	YACT	S000449
POLASIG3	site	562	(-)	AATAAT	S000088
TATABOX5	site	563	(+)	TTATTT	S000203
RAV1AAT	site	577	(+)	CAACA	S000314
CIACADIANLELHC	site	602	(-)	CAANNNNATC	S000252
GAREAT	site	608	(-)	TAACAAR	S000439
POLASIG2	site	632	(+)	AATTAAA	S000081
TAAAGSTKST1	site	635	(+)	TAAAG	S000387
DOFCOREZM	site	636	(+)	AAAG	S000265
CACTFTPPCA1	site	648	(+)	YACT	S000449
RAV1AAT	site	663	(+)	CAACA	S000314
MYBCORE	site	663	(-)	CNGTTR	S000176
WBOXHVISO1	site	672	(-)	TGACT	S000442
WBOXNTERF3	site	672	(-)	TGACY	S000457
WRKY71OS	site	673	(-)	TGAC	S000447
WBOXHVISO1	site	676	(-)	TGACT	S000442
WBOXNTERF3	site	676	(-)	TGACY	S000457
WBOXATNPR1	site	677	(-)	TTGAC	S000390
WRKY71OS	site	677	(-)	TGAC	S000447
RAV1AAT	site	679	(+)	CAACA	S000314
MYBCORE	site	679	(-)	CNGTTR	S000176
GATABOX	site	684	(+)	GATA	S000039
GT1CONSENSUS	site	684	(+)	GRWAAW	S000198
IBOXCORE	site	684	(+)	GATAA	S000199
CAATBOX1	site	702	(+)	CAAT	S000028
ARR1AT	site	703	(-)	NGATT	S000454
CAATBOX1	site	706	(+)	CAAT	S000028
RAV1AAT	site	709	(-)	CAACA	S000314
WBOXATNPR1	site	711	(+)	TTGAC	S000390
WBOXHVISO1	site	712	(+)	TGACT	S000442
WBOXNTERF3	site	712	(+)	TGACY	S000457
WRKY71OS	site	712	(+)	TGAC	S000447
INRNTPSADB	site	720	(+)	YTCANTYY	S000395
CAATBOX1	site	722	(+)	CAAT	S000028
ARR1AT	site	723	(-)	NGATT	S000454
NODCON1GM	site	724	(-)	AAAGAT	S000461

OSE1ROOTNODULE	site	724	(-)	AAAGAT	S000467
DOFCOREZM	site	726	(-)	AAAG	S000265
POLASIG1	site	728	(-)	AATAAA	S000080
RAV1AAT	site	735	(-)	CAACA	S000314
-10PEHVPSBD	site	741	(-)	TATTCT	S000392
CACTFTPPCA1	site	745	(+)	YACT	S000449
WRKY71OS	site	749	(-)	TGAC	S000447
DOFCOREZM	site	777	(+)	AAAG	S000265
CACTFTPPCA1	site	782	(-)	YACT	S000449
CACTFTPPCA1	site	787	(+)	YACT	S000449
TBOXATGAPB	site	788	(+)	ACTTTG	S000383
DOFCOREZM	site	789	(-)	AAAG	S000265
ELRECOREPCRP1	site	791	(+)	TTGACC	S000142
WBOXATNPR1	site	791	(+)	TTGAC	S000390
WBOXNTERF3	site	792	(+)	TGACY	S000457
WRKY71OS	site	792	(+)	TGAC	S000447
POLLEN1LELAT52	site	804	(+)	AGAAA	S000245
INRNTPSADB	site	805	(-)	YTCANTYY	S000395
ANAERO2CONSENSUS	site	816	(+)	AGCAGC	S000478
REALPHALGLHCB21	site	843	(+)	AACCAA	S000362
CACTFTPPCA1	site	854	(-)	YACT	S000449
CACTFTPPCA1	site	856	(+)	YACT	S000449
TGTCACACMCUCUMISIN	site	868	(+)	TGTCACA	S000422
WRKY71OS	site	869	(-)	TGAC	S000447
GTGANTG10	site	870	(-)	GTGA	S000378
PALBOXAPC	site	875	(-)	CCGTCC	S000137
CACTFTPPCA1	site	887	(-)	YACT	S000449
SEBFCONSSTPR10A	site	893	(+)	YTGTCWC	S000391
WRKY71OS	site	895	(-)	TGAC	S000447
GTGANTG10	site	896	(-)	GTGA	S000378
CACTFTPPCA1	site	897	(+)	YACT	S000449
IBOX	site	899	(-)	GATAAG	S000124
IBOXCORE	site	900	(-)	GATAA	S000199
GATABOX	site	901	(-)	GATA	S000039
ARR1AT	site	906	(-)	NGATT	S000454
NODCON2GM	site	910	(+)	CTCTT	S000462
OSE2ROOTNODULE	site	910	(+)	CTCTT	S000468
SV40COREENHAN	site	912	(-)	GTGGWWHG	S000123
CAATBOX1	site	921	(+)	CAAT	S000028
RBCSCONSENSUS	site	922	(+)	AATCCAA	S000127
ARR1AT	site	922	(-)	NGATT	S000454
POLLEN1LELAT52	site	935	(-)	AGAAA	S000245
CACTFTPPCA1	site	945	(-)	YACT	S000449
TAAAGSTKST1	site	947	(+)	TAAAG	S000387
DOFCOREZM	site	948	(+)	AAAG	S000265
TATABOX5	site	956	(-)	TTATTT	S000203
POLASIG1	site	957	(+)	AATAAA	S000080
TAAAGSTKST1	site	959	(+)	TAAAG	S000387
DOFCOREZM	site	960	(+)	AAAG	S000265
NODCON2GM	site	971	(-)	CTCTT	S000462
OSE2ROOTNODULE	site	971	(-)	CTCTT	S000468
MYBST1	site	975	(+)	GGATA	S000180
SREATMSD	site	975	(-)	TTATCC	S000470
GATABOX	site	976	(+)	GATA	S000039
IBOX	site	976	(+)	GATAAG	S000124
IBOXCORE	site	976	(+)	GATAA	S000199
IBOXCORENT	site	976	(+)	GATAAGR	S000424
ARE1	site	988	(-)	RGTGACNNNGC	S000022
CACTFTPPCA1	site	990	(-)	YACT	S000449
WRKY71OS	site	993	(-)	TGAC	S000447
GTGANTG10	site	994	(-)	GTGA	S000378
CACTFTPPCA1	site	995	(+)	YACT	S000449
-300ELEMENT	site	1006	(+)	TGHAAARK	S000122
PYRIMIDINEBOXOSRAMY1A	site	1008	(-)	CCTTTT	S000259
DOFCOREZM	site	1009	(+)	AAAG	S000265
CACTFTPPCA1	site	1016	(-)	YACT	S000449
WBOXATNPR1	site	1022	(+)	TTGAC	S000390
WRKY71OS	site	1023	(+)	TGAC	S000447
CAATBOX1	site	1052	(-)	CAAT	S000028
CCAATBOX1	site	1052	(-)	CCAAT	S000030
CCAATBOX1	site	1060	(+)	CCAAT	S000030
CAATBOX1	site	1061	(+)	CAAT	S000028
EBOXBNNAPA	site	1061	(+)	CANNTG	S000144
MYCCONSENSUSAT	site	1061	(+)	CANNTG	S000407
EBOXBNNAPA	site	1061	(-)	CANNTG	S000144
MYCCONSENSUSAT	site	1061	(-)	CANNTG	S000407
CAATBOX1	site	1063	(-)	CAAT	S000028
CCAATBOX1	site	1063	(-)	CCAAT	S000030
NODCON2GM	site	1068	(+)	CTCTT	S000462
OSE2ROOTNODULE	site	1068	(+)	CTCTT	S000468
GATABOX	site	1077	(-)	GATA	S000039
NODCON2GM	site	1086	(+)	CTCTT	S000462
OSE2ROOTNODULE	site	1086	(+)	CTCTT	S000468
DOFCOREZM	site	1088	(-)	AAAG	S000265
WRKY71OS	site	1093	(-)	TGAC	S000447
P1BS	site	1104	(+)	GNATATNC	S000459
P1BS	site	1104	(-)	GNATATNC	S000459
ROOTMOTIFTAPOX1	site	1106	(+)	ATATT	S000098

GTGANTG10	site	1110	(-)	GTGA	S000378
GT1CONSENSUS	site	1117	(-)	GRWAAW	S000198
POLLEN1LELAT52	site	1119	(-)	AGAAA	S000245
CARGCW8GAT	site	1128	(+)	CWWWWWWWWG	S000431
CARGCW8GAT	site	1128	(-)	CWWWWWWWWG	S000431
SEF4MOTIFGM7S	site	1129	(-)	RTTTTTR	S000103
CCAATBOX1	site	1140	(+)	CCAAT	S000030
CAATBOX1	site	1141	(+)	CAAT	S000028
GTGANTG10	site	1148	(+)	GTGA	S000378
RAV1AAT	site	1152	(-)	CAACA	S000314
RAV1AAT	site	1158	(-)	CAACA	S000314
GTGANTG10	site	1165	(-)	GTGA	S000378
BOXLCOREDCPAL	site	1167	(+)	ACCWWCC	S000492
CACTFTPPCA1	site	1173	(+)	YACT	S000449
WRKY71OS	site	1180	(-)	TGAC	S000447
ANAERO2CONSENSUS	site	1198	(-)	AGCAGC	S000478
GATABOX	site	1209	(+)	GATA	S000039
HBOXCONSENSUSPVCHS	site	1223	(+)	CCTACCNNNNNNNCT	S000200
MYBPZM	site	1223	(+)	CCWACC	S000179
CAREOSREP1	site	1231	(+)	CAACTC	S000421
NODCON2GM	site	1234	(+)	CTCTT	S000462
OSE2ROOTNODULE	site	1234	(+)	CTCTT	S000468
DPBFCOREDCDC3	site	1236	(-)	ACACNNG	S000292
WRKY71OS	site	1241	(-)	TGAC	S000447
CATATGGMSAUR	site	1246	(+)	CATATG	S000370
EBOXBNNAPA	site	1246	(+)	CANNTG	S000144
MYCCONSENSUSAT	site	1246	(+)	CANNTG	S000407
CATATGGMSAUR	site	1246	(-)	CATATG	S000370
EBOXBNNAPA	site	1246	(-)	CANNTG	S000144
MYCCONSENSUSAT	site	1246	(-)	CANNTG	S000407
DOFCOREZM	site	1262	(-)	AAAG	S000265
POLLEN1LELAT52	site	1263	(-)	AGAAA	S000245
DOFCOREZM	site	1266	(-)	AAAG	S000265
POLLEN1LELAT52	site	1267	(-)	AGAAA	S000245
DOFCOREZM	site	1270	(-)	AAAG	S000265
DOFCOREZM	site	1276	(-)	AAAG	S000265
TAAAGSTKST1	site	1276	(-)	TAAAG	S000387
DOFCOREZM	site	1288	(-)	AAAG	S000265
ARFAT	site	1306	(+)	TGTCTC	S000270
MYBCORE	site	1311	(+)	CNGTTR	S000176
MYB2AT	site	1311	(-)	TAACTG	S000177
MYB2CONSENSUSAT	site	1311	(-)	YAACKG	S000409
E2FCONSENSUS	site	1326	(+)	WTTSSCSS	S000476
GCCCORE	site	1330	(+)	GCCGCC	S000430
ARE1	site	1333	(-)	RGTGACNNNGC	S000022
WBOXHVISO1	site	1337	(-)	TGACT	S000442
WBOXNTERF3	site	1337	(-)	TGACY	S000457
WRKY71OS	site	1338	(-)	TGAC	S000447
GTGANTG10	site	1339	(-)	GTGA	S000378
CACTFTPPCA1	site	1340	(+)	YACT	S000449
GATABOX	site	1350	(+)	GATA	S000039
BOXIINTPATPB	site	1351	(+)	ATAGAA	S000296
POLLEN1LELAT52	site	1353	(+)	AGAAA	S000245
ANAERO1CONSENSUS	site	1356	(+)	AAACAAA	S000477
DOFCOREZM	site	1360	(+)	AAAG	S000265
POLLEN1LELAT52	site	1362	(+)	AGAAA	S000245
DOFCOREZM	site	1364	(+)	AAAG	S000265
CACTFTPPCA1	site	1373	(-)	YACT	S000449
SORLIP1AT	site	1374	(-)	GCCAC	S000482
CAATBOX1	site	1380	(-)	CAAT	S000028
DOFCOREZM	site	1384	(+)	AAAG	S000265
SORLIP1AT	site	1393	(+)	GCCAC	S000482
CACTFTPPCA1	site	1395	(+)	YACT	S000449
EBOXBNNAPA	site	1395	(+)	CANNTG	S000144
MYCCONSENSUSAT	site	1395	(+)	CANNTG	S000407
EBOXBNNAPA	site	1395	(-)	CANNTG	S000144
MYCCONSENSUSAT	site	1395	(-)	CANNTG	S000407
GCCCORE	site	1401	(+)	GCCGCC	S000430
POLASIG3	site	1445	(-)	AATAAT	S000088
POLASIG3	site	1448	(-)	AATAAT	S000088
CACTFTPPCA1	site	1453	(+)	YACT	S000449
MYBCORE	site	1457	(+)	CNGTTR	S000176
RAV1AAT	site	1458	(-)	CAACA	S000314
CAATBOX1	site	1467	(-)	CAAT	S000028
DOFCOREZM	site	1471	(+)	AAAG	S000265
GT1CONSENSUS	site	1475	(+)	GRWAAW	S000198
INTRONLOWER	site	1485	(-)	TGCAGG	S000086
ACGTATERD1	site	1492	(+)	ACGT	S000415
ACGTATERD1	site	1492	(-)	ACGT	S000415
RBCSCONSENSUS	site	1497	(+)	AATCCAA	S000127
ARR1AT	site	1497	(-)	NGATT	S000454
MYBPZM	site	1500	(+)	CCWACC	S000179
GATABOX	site	1511	(-)	GATA	S000039
GTGANTG10	site	1513	(-)	GTGA	S000378
EBOXBNNAPA	site	1516	(+)	CANNTG	S000144
MYB2CONSENSUSAT	site	1516	(+)	YAACKG	S000409
MYCCONSENSUSAT	site	1516	(+)	CANNTG	S000407
EBOXBNNAPA	site	1516	(-)	CANNTG	S000144

MYBCORE	site	1516	(-)	CNGTTR	S000176
MYCCONSENSUSAT	site	1516	(-)	CANNTG	S000407
DOFCOREZM	site	1522	(-)	AAAG	S000265
TAAAGSTKST1	site	1522	(-)	TAAAG	S000387
WBOXNTERF3	site	1528	(-)	TGACY	S000457
WRKY71OS	site	1529	(-)	TGAC	S000447
GTGANTG10	site	1530	(-)	GTGA	S000378
EBOXBNNAPA	site	1535	(+)	CANNTG	S000144
MYB2CONSENSUSAT	site	1535	(+)	YAACKG	S000409
MYCCONSENSUSAT	site	1535	(+)	CANNTG	S000407
EBOXBNNAPA	site	1535	(-)	CANNTG	S000144
MYBCORE	site	1535	(-)	CNGTTR	S000176
MYCCONSENSUSAT	site	1535	(-)	CANNTG	S000407
CACTFTPPCA1	site	1550	(-)	YACT	S000449
SORLIP1AT	site	1551	(-)	GCCAC	S000482
SORLIP1AT	site	1554	(+)	GCCAC	S000482
SORLIP1AT	site	1570	(+)	GCCAC	S000482
CACTFTPPCA1	site	1574	(+)	YACT	S000449
IBOX	site	1576	(-)	GATAAG	S000124
IBOXCORE	site	1577	(-)	GATAA	S000199
GATABOX	site	1578	(-)	GATA	S000039
GATABOX	site	1587	(+)	GATA	S000039
POLLEN1LELAT52	site	1604	(+)	AGAAA	S000245
DOFCOREZM	site	1607	(+)	AAAG	S000265
CAATBOX1	site	1624	(+)	CAAT	S000028
POLASIG1	site	1629	(+)	AATAAA	S000080
TAAAGSTKST1	site	1631	(+)	TAAAG	S000387
DOFCOREZM	site	1632	(+)	AAAG	S000265
CACTFTPPCA1	site	1649	(-)	YACT	S000449
GTGANTG10	site	1680	(-)	GTGA	S000378
CACTFTPPCA1	site	1685	(+)	YACT	S000449
CACTFTPPCA1	site	1688	(+)	YACT	S000449
CACTFTPPCA1	site	1692	(-)	YACT	S000449
GTGANTG10	site	1693	(+)	GTGA	S000378
CAREOSREP1	site	1695	(-)	CAACTC	S000421
GATABOX	site	1714	(-)	GATA	S000039
DOFCOREZM	site	1720	(-)	AAAG	S000265
TAAAGSTKST1	site	1720	(-)	TAAAG	S000387
ARR1AT	site	1725	(+)	NGATT	S000454
GTGANTG10	site	1729	(-)	GTGA	S000378
GTGANTG10	site	1737	(-)	GTGA	S000378
TBOXATGAPB	site	1741	(-)	ACTTTG	S000383
DOFCOREZM	site	1742	(+)	AAAG	S000265
CACTFTPPCA1	site	1744	(-)	YACT	S000449
AMYBOX2	site	1752	(-)	TATCCAT	S000021
TATCCAYMOTIFOSRAMY3D	site	1752	(-)	TATCCAY	S000256
TATCCAOSAMY	site	1753	(-)	TATCCA	S000403
MYBST1	site	1754	(+)	GGATA	S000180
SREATMSD	site	1754	(-)	TTATCC	S000470
GATABOX	site	1755	(+)	GATA	S000039
GT1CONSENSUS	site	1755	(+)	GRWAAW	S000198
IBOXCORE	site	1755	(+)	GATAA	S000199
TBOXATGAPB	site	1760	(+)	ACTTTG	S000383
DOFCOREZM	site	1761	(-)	AAAG	S000265
TATABOX3	site	1772	(-)	TATTAAT	S000110
POLASIG1	site	1775	(+)	AATAAA	S000080
TAAAGSTKST1	site	1777	(+)	TAAAG	S000387
NTBBF1ARROLB	site	1777	(-)	ACTTTA	S000273
DOFCOREZM	site	1778	(+)	AAAG	S000265
CACTFTPPCA1	site	1780	(-)	YACT	S000449
ROOTMOTIFTAPOX1	site	1783	(+)	ATATT	S000098
DOFCOREZM	site	1790	(-)	AAAG	S000265
DOFCOREZM	site	1798	(-)	AAAG	S000265
POLLEN1LELAT52	site	1807	(+)	AGAAA	S000245
DOFCOREZM	site	1810	(+)	AAAG	S000265
CACTFTPPCA1	site	1812	(-)	YACT	S000449
GATABOX	site	1814	(-)	GATA	S000039
MYBPLANT	site	1831	(+)	MACCWAMC	S000167
CACTFTPPCA1	site	1854	(-)	YACT	S000449
GTGANTG10	site	1855	(+)	GTGA	S000378
DOFCOREZM	site	1858	(+)	AAAG	S000265
MYB2AT	site	1863	(+)	TAACTG	S000177
MYB2CONSENSUSAT	site	1863	(+)	YAACKG	S000409
MYBCORE	site	1863	(-)	CNGTTR	S000176
MYB2AT	site	1869	(+)	TAACTG	S000177
MYB2CONSENSUSAT	site	1869	(+)	YAACKG	S000409
MYBCORE	site	1869	(-)	CNGTTR	S000176
DOFCOREZM	site	1877	(-)	AAAG	S000265
TAAAGSTKST1	site	1877	(-)	TAAAG	S000387
POLASIG1	site	1878	(-)	AATAAA	S000080
TATABOX5	site	1879	(+)	TTATTT	S000203
GT1CONSENSUS	site	1882	(-)	GRWAAW	S000198
IBOXCORE	site	1883	(-)	GATAA	S000199
GATABOX	site	1884	(-)	GATA	S000039
AMYBOX1	site	1886	(-)	TAACARA	S000020
GARE1OSREP1	site	1886	(-)	TAACAGA	S000419
MYBCORE	site	1887	(+)	CNGTTR	S000176
INRNTPSADB	site	1896	(+)	YTCANTYY	S000395

GTGANTG10	site	1897	(-)	GTGA	S000378
CACTFTPPCA1	site	1898	(+)	YACT	S000449
CACTFTPPCA1	site	1909	(+)	YACT	S000449
ACGTATERD1	site	1913	(+)	ACGT	S000415
ACGTATERD1	site	1913	(-)	ACGT	S000415
CAATBOX1	site	1919	(+)	CAAT	S000028
ARR1AT	site	1920	(-)	NGATT	S000454
CIACADIANLELHC	site	1924	(-)	CAANNNNATC	S000252
RAV1AAT	site	1926	(-)	CAACA	S000314
RAV1AAT	site	1929	(-)	CAACA	S000314
WBOXATNPR1	site	1931	(+)	TTGAC	S000390
WRKY71OS	site	1932	(+)	TGAC	S000447
CACTFTPPCA1	site	1936	(-)	YACT	S000449
GTGANTG10	site	1941	(-)	GTGA	S000378
DRE2COREZMRAB17	site	1943	(+)	ACCGAC	S000402
DRECRTCOREAT	site	1943	(+)	RCCGAC	S000418
LTREATLTI78	site	1943	(+)	ACCGACA	S000157
LTRECOREATCOR15	site	1944	(+)	CCGAC	S000153
RAV1AAT	site	1948	(+)	CAACA	S000314
GATABOX	site	1957	(+)	GATA	S000039
GT1CONSENSUS	site	1957	(+)	GRWAAW	S000198
IBOXCORE	site	1957	(+)	GATAA	S000199
DPBFCOREDCDC3	site	1963	(-)	ACACNNG	S000292
INTRONLOWER	site	1972	(-)	TGCAGG	S000086
ARR1AT	site	1985	(+)	NGATT	S000454
SORLIP1AT	site	1993	(+)	GCCAC	S000482
RAV1AAT	site	1998	(+)	CAACA	S000314
CAATBOX1	site	2001	(+)	CAAT	S000028
ARR1AT	site	2002	(-)	NGATT	S000454
GTGANTG10	site	2004	(-)	GTGA	S000378
CACTFTPPCA1	site	2005	(+)	YACT	S000449
ASF1MOTIFCAMV	site	2010	(-)	TGACG	S000024
WBOXATNPR1	site	2011	(-)	TTGAC	S000390
WRKY71OS	site	2011	(-)	TGAC	S000447
CACTFTPPCA1	site	2015	(-)	YACT	S000449
POLASIG3	site	2019	(-)	AATAAT	S000088
TATABOX5	site	2020	(+)	TTATTT	S000203
GT1CONSENSUS	site	2023	(-)	GRWAAW	S000198
IBOXCORE	site	2024	(-)	GATAA	S000199
GATABOX	site	2025	(-)	GATA	S000039
NODCON1GM	site	2026	(-)	AAAGAT	S000461
OSE1ROOTNODULE	site	2026	(-)	AAAGAT	S000467
DOFCOREZM	site	2028	(-)	AAAG	S000265
ELRECOREPCRP1	site	2035	(+)	TTGACC	S000142
WBOXATNPR1	site	2035	(+)	TTGAC	S000390
WBOXNTERF3	site	2036	(+)	TGACY	S000457
WRKY71OS	site	2036	(+)	TGAC	S000447
DRE2COREZMRAB17	site	2038	(+)	ACCGAC	S000402
DRECRTCOREAT	site	2038	(+)	RCCGAC	S000418
LTRECOREATCOR15	site	2039	(+)	CCGAC	S000153
WBOXNTERF3	site	2046	(-)	TGACY	S000457
WRKY71OS	site	2047	(-)	TGAC	S000447
CAATBOX1	site	2050	(-)	CAAT	S000028
GT1CONSENSUS	site	2054	(-)	GRWAAW	S000198
SREATMSD	site	2055	(+)	TTATCC	S000470
IBOXCORE	site	2055	(-)	GATAA	S000199
GATABOX	site	2056	(-)	GATA	S000039
MYBST1	site	2056	(-)	GGATA	S000180
ARR1AT	site	2077	(+)	NGATT	S000454

Chapter 6

General Discussion and Conclusions

6.1 Introduction

As the world population continues to increase, food supplies must also grow to meet nutritional requirements. One means of ensuring the availability and stability of the food supply is to mitigate crop losses caused by plant pathogens (Campbell et al. 2002). The goal of producing crops with increased and durable resistance to a spectrum of diseases is therefore a major focus of current research in plant biotechnology. In nature, plants are continually challenged by fungi, bacteria, viruses and nematodes, but comparatively few of these pathogens are successful in gaining entry into a prospective host. That is, disease is rare in nature because plants carry different 'layers' of defence from structural barriers and pre-formed antimicrobials, to adaptive defence mechanisms that encompass non-host, race-specific and race non-specific resistance (Thordal-Christensen 2003; Mysore and Ryu 2004). However, with cultivation of vast areas of genetically identical crops the situation can be quite different. In this case, protection relies on a small number of in-bred disease resistance genes per crop species. Unfortunately, control is transient because pathogens can overcome disease resistance genes and/or become resistant to pesticides (Hulbert et al. 2001; Stuiver and Clusters 2001). Genetic engineering has the potential to solve these problems by inserting carefully selected and possibly multiple disease resistance genes into the plant in an efficient and systematic manner (Hulbert et al. 2001; Stuiver and Clusters 2001). The resistance genes can come from resistant genotypes of the same species or from other species. Overall, genetic improvement of plants via genetic engineering has several advantages over traditional breeding approaches such as: crossing of the species barrier, the ability to eliminate unwanted genetically linked traits, and also the ability to rapidly transfer genes into commercially elite cultivars (Campbell et al. 2002).

There have been some notable reports of success in developing pathogen resistance in plants using genetic engineering. For example, in the Solanaceae family several R genes of the NBS-LRR class have been isolated and transferred from one species to another with successful results (Hulbert et al. 2001). For example, the tobacco *N* gene encoding resistance to *Tomato mosaic virus* (TMV) has been transferred to tomato thereby conferring resistance to TMV resistance in the latter (Whitham et al. 1996). Moreover, the *Bs2* gene, which confers resistance to *Xanthomonas campestris* in pepper has been introduced into tomato making the latter plant species resistant to *X. campestris* (Tai et al. 1999). Another example is the *Solanum bulbocastanum* resistance gene *Rpi-blb1* that confers resistance to *Phytophtora infestans* which has been transferred to potato and tomato (Van der Vossen et al. 2003). Attempts to demonstrate function in species outside of the family from which the gene was isolated have, however, been unsuccessful. For example, the *Arabidopsis RPS2* gene that confers resistance to *Pseudomonas syringae* is non-functional in transgenic tomato and this phenomenon has been referred to as “restricted taxonomic functionality” (Tai et al. 1999). This phenomenon is probably an indication that other components of the resistance signal transduction pathway are not present in a form that can interact with the resistance gene in the recipient species. Nevertheless, the transfer of resistance genes even between related species will be a great step forward for plant breeders (Rommens and Kishore 2000).

Amongst the fruit crops, banana is the most important. It represents a staple food to at least 400 million people in developing countries and also income and employment for farming communities (Sagi et al. 1995). Unfortunately, banana production is seriously compromised by a range of diseases. Some of these are temporarily under control as a result of an intensive and extensive use of chemicals. However, there are other diseases that cannot be controlled successfully with chemical methods such as Panama disease, caused by the fungus *F. oxysporum* f.sp. *cubense* (FOC)(Ploetz and Pegg 2000). Currently, FOC race 4 represents a serious threat to the banana production worldwide since the majority of the most important banana

cultivars are highly susceptible to this pathogen. As a result, the destruction that this pathogen could potentially cause in banana has been considered as a ‘time bomb’ (Frison 2003). Conventional breeding schemes attempting to enhance *Fusarium* resistance of banana cultivars have been largely unsuccessful due to long generation time and to the virtually sterile nature of banana cultivars. Therefore, genetic improvement of banana is urgently needed in order to protect the banana production for present and future generations.

Banana transformation protocols either by biolistics or *Agrobacterium* are now a reality (Becker et al. 2000; Khanna et al. 2004), however the identification of a *Fusarium* resistance gene in banana has not been accomplished yet. Cloning of genes capable to confer resistance to *Fusarium oxysporum* has been achieved successfully in the family Solanaceae (*I2* gene) and Cucurbitaceae (*Fom-2* gene). Extending these achievements in the family Musaceae will surely allow the development of new *Fusarium* resistant varieties in banana through genetic engineering. Therefore, research directed towards the cloning of potential *Fusarium* resistance genes in this crop is required. Taking into consideration that the proteins encoded by the *Fusarium* resistance genes *I2* and *Fom-2* share the same basic non-TIR-NBS-LRR structure open the possibility to find *Fusarium* R genes with this structure in other plant families such as the Musaceae. This study reports the cloning, bioinformatic and expression analysis of five disease resistance gene candidates (RGCs) of the NBS-type from the *Fusarium*-resistant banana *Musa acuminata* ssp. *malaccensis*. Moreover, as two of these banana RGCs (RGC2 and RGC5) appear to have potential to confer *Fusarium* resistance, this study has undertaken the cloning and structural characterisation of their full cDNA sequences and the development of expression cassettes to test the function of RGC2 in FOC race 4 resistance using a genetic complementation approach.

6.2 Molecular cloning and characterisation of disease resistance gene candidates of the NBS-type in banana

The isolation, sequencing and bioinformatic analysis of resistance gene candidates in banana should provide a platform for further functional analysis and determination of pathogen specificities. In chapter 4, the isolation and characterisation of five different classes of NBS sequences (RGC1 to RGC5) from the wild banana *Musa acuminata* ssp. *malaccensis* is described. These RGC genes show significant sequence similarity to the corresponding domain of the NBS-LRR family of R genes. These five classes are also present in the genome of other bananas, such as the commercial cultivar 'Grand Nain' and the wild banana *Musa acuminata* ssp. *burmannicoides* 'Calcutta 4' (Taylor 2005). Structural and phylogenetic analysis of the five banana NBS classes showed that they belong to the non-TIR subclass of NBS sequences (Meyers et al. 1999; Pan et al. 2000). Further analysis of the N-terminus confirmed the absence of a TIR-like structure in the five banana NBS classes. To date, the TIR domain has not been found in the structure of monocot NBS-LRR R genes, even in the complete rice genome sequence (Goff et al. 2002; Bai et al. 2002; Zhou et al. 2004; Monosi et al. 2004). It has been hypothesised that the loss of the TIR domain from the NBS-LRR genes in monocot plants may have occurred subsequent to the divergence of the monocot and dicot lineages in the early Cretaceous about 100 million years ago (Pan et al. 2000). Because the genus *Musa* shares a common evolutionary origin with other monocot plants, it is likely that this domain is also absent in the structure of banana NBS-LRR genes. Thus, the lack of the TIR-like domain in the N-terminus of the five RGCs that we isolated in this study may be an indication of a broader absence of this domain in banana NBS-LRR genes. Indeed, the full genomic sequence of banana will shed light on this matter.

In this study, one combination of degenerate primers to isolate NBS related sequences was used, so the number of NBS classes that we found in the genome of *M. acuminata* ssp. *malaccensis* could be underestimated. To expand the search for other possible NBS classes in the banana genome, other combinations of degenerate primers need to be utilised. There are

other highly conserved motifs in the NBS of R genes that might lead to the isolation of other banana NBS sequences by degenerate PCR, such as the motifs LLVLDDVW/D, GSRIIITTRD and CFLYCALFP (Meyers et al. 1999; Pan et al. 2000). Another genetic resource which would significantly assist the progression of research into banana R genes is the complete sequencing of the banana genome. Efforts to accomplish this remarkable task are currently in progress (www.musagenomics.org).

Apart from providing information about the evolution of NBS-LRR genes, the NBS domain may be useful to isolate potential disease resistance genes based on homology to previously isolated R genes. Four of the five classes (RGC1, RGC2, RGC3 and RGC5) showed homology to R genes that confer resistance to related pathogens in banana (Table 4.3). RGC1 and RGC3 show significant sequence similarity to the *RPM1* (Grant et al. 1996) and *RPS5* (Warren et al. 1998) resistance genes from *Arabidopsis thaliana*, which encode two non-TIR-NBS-LRR proteins conferring resistance to the bacterium *Pseudomonas syringae.* This leads to the speculation that either or both of these two resistance gene candidates (RGC1 or RGC3) could be involved in conferring resistance to bacterial-related pathogens in banana such as *Ralstonia solanacearum*, causal agent of Moko disease (Thwaites et al. 2000). Further research involving the cloning of the entire genes and genetic complementation experiments is required to address this issue. Interestingly, RGC2 and RGC5 showed a significant sequence similarity to the resistance gene *I2* from tomato, which confers resistance to *F. oxysporum* f.sp. *licopersici* (FOL) race 2 (Simons et al. 1998) (Table 4.3). Remarkably, the expression of RGC2 was found to be present only in FOC race 4 resistant plants of *M. acuminata ssp. malaccensis* (Figure 4.10) suggesting that RGC2 might play a role in FOC race 4 resistance. A similar expression profile was shown by the tomato *I2* resistance gene whose expression was only present in both leaf and root tissues of tomato plants with resistance to FOL race 2 (Mes et al. 2000). Further research is required to determine the function (pathogen recognition specificity) of each one of the banana resistance gene candidates isolated in this study. New technologies such as RNA interference (Waterhouse and Helliwell 2003)

could facilitate testing the function of RGCs by silencing multiple RGCs in banana plants resistant to the most devastating pathogens. Those resistant plants that become susceptible would assist in the identification of a particular R gene. Another recent technology that promises to facilitate the identification of R genes in banana is the use of Binary Bacterial Artificial Chromosome (BIBAC) libraries. A BIBAC vector can potentially contain inserts of up to 120 Kb ready to be transferred via *Agrobacterium*-mediated transformation into a plant host genome (He et al. 2003). Recently, BIBAC transformation libraries for *Arabidopsis* and rice have been made (Tao et al., 2002; Chang et al., 2003). One of the typical features of R genes is that they are organised in gene clusters in the plant genome. For example, the *Xa21* gene of rice, an LRR-kinase R gene, is within a cluster of 8 homologues spanning about 230 Kb while *RPP5* of *Arabidopsis*, an NBS-LRR *R* gene, is within a cluster of 7 homologues spanning about 50 Kb. Another example, is the *Fusarium* resistance gene *I2* from tomato which is within a 90 kb cluster of 7 paralogues (Michelmore and Meyers, 1998; Simons et al. 1998). Thus, a single BIBAC vector could contain a cluster of several homologue R gene candidates including their regulatory sequences ready to be tested in one single step of transformation. Although RGC5 appears to be a single gene (Figure 4.9), the RGC2 seems to belong to a small gene family that could be organised in a cluster, if this is the case it will be relatively easy to test the function of all the RGC2 homologue sequences using a BIBAC vector since the *Agrobacterium*-mediated transformation of banana is well standardised in the Plant Biotechnology laboratory at QUT (Khanna et al. 2004). A BIBAC library from the cultivar 'Tuu Gia' (*Musa acuminata*), which is resistant to the most serious diseases of banana has been made recently (Ortiz-Vázquez et al. 2005). In this manner, it would be possible to transform disease-susceptible banana cultivars with BIBAC clones harbouring RGCs organized either as clusters or singletons from the 'Tuu Gia' resistant cultivar. This approach would lead to a RGC-BIBAC collection of banana lines ready to be used for pathogen-resistance screenings. Indeed, when combined with appropriate screening strategies these high-throughput technologies will greatly improve knowledge on disease resistance in banana and lead to the

development of pathogen resistance in this crop through genetic engineering in the near future.

6.3 Molecular cloning and characterisation of two potential *Fusarium* resistance genes in banana

The major goal of this study was to identify potential FOC race 4 resistance genes using a combination of bioinformatics and gene expression analysis. The previous discovery that two *Fusarium* R genes from different plant families belong to the non-TIR subclass of NBS-LRR genes (Simons et al. 1998; Joobeur et al. 2004) opened the possibility to use a rational approach to search for similar sequences in the banana genome with the aim to find functional NBS-LRR genes that confer resistance to FOC. Given that there are eight major classes of known disease resistance genes in plants (Hammond-Kosack and Parker 2003), such rational approach could facilitate the search for potential *Fusarium* R genes. Chapter 5 examined the structural and phylogenetic information of two banana NBS-LRR RGCs (RGC2 and RGC5) with significant sequence similarity to the *Fusarium* resistance gene *I2* from tomato. The deduced protein sequence of both RGC2 and RGC5 full ORFs showed the typical motifs and domains of the non-TIR subclass of NBS-LRR disease resistance genes (Meyers et al. 1999; Pan et al. 2000). Sequence similarity searches using the entire ORF sequence of RGC2 and RGC5 reveal that the *I2* gene ranks as one of the most similar R gene to RGC2 and RGC5. Previous correlation of RGC2 in FOC race 4 resistance (Chapter 4) and the finding that this sequence along with the RGC5 sequence are both significantly similar to the tomato *I2* resistance gene prompted comparison the two banana RGCs with the tomato *I2* gene in more detail. Overall, the banana RGC2 and RGC5 gene products showed a relatively low level of sequence similarity to the I2 protein. This level of sequence similarity is quite common among NBS-LRR genes from different plant families (Bai et al. 2002), or even in NBS-LRR resistance genes whose pathogen recognition specificity is very similar (Joobeur et al. 2004; Ashfield at al. 2004; McDowell 2004). For example, similarity searches using the non-TIR-NBS-LRR *Fom-2* resistance gene revealed that the most similar characterized R gene to the *Fom-2* was the *Fusarium* resistance gene

I2 from tomato (29% identity and 49% similarity; E value = $2e^{-88}$) (Joobeur et al. 2004). Another example is the non-TIR-NBS-LRR gene pair *RPM1* and *Rpg1-b* from *Arabidopsis thaliana* and *Glycine max*, respectively. Both gene products confer resistance to the bacterium *Pseudomonas syringae* by recognizing the same avirulence protein (AvrB) (Ashfield at al. 2004; McDowell 2004). Alignment of the predicted RPM1 and Rpg1 protein sequences revealed a relatively low level of amino acid sequence identity across the NBS domain (~34%) and they were not phylogenetically closely related.

Although the cloning of two *Fusarium* resistance genes from different plant families has been achieved in the past eight years using a map-based cloning approach, to our knowledge there are not previous published reports that explain their phylogenetic relationships. This information could be highly valuable in the quest for further *Fusarium* R genes in plants. In chapter 5, it was found that the *I2* and *Fom-2* gene products share a limited sequence similarity. They are not phylogenetically closely related and they are more similar to other non-*Fusarium* R genes than between each other, however they share the same phylogenetic clade N1 previously defined by Cannon et al. (2000). These results suggest that probably other divergent *Fusarium* R genes may cluster in the same phylogenetic clade. Based on this hypothesis, the N1 clade may be an interesting place to search for potential *Fusarium* resistance genes not only in the Musaceae familiy (RGC2 and RGC5) but in other plant species as well. Given the large number of very divergent NBS-LRR genes in a plant genome (Meyers et al. 2003; Zhou et al. 2004), this phylogenetic information could narrow even more the search for potential *Fusarium* R genes and lead to further research focus on how the *Fusarium* recognition specificity evolved in two different plant families. Taking into consideration that the *Fusarium* resistance in tomato and melon is controlled by more than a single resistance gene (Sela-Buurlage et al. 2001; Schreuder et al. 2000), it is probable that multiple R genes confer resistance to the four divergent races of FOC in banana (Koenig et al. 1997; O'Donnell et al. 1998). The isolation by degenerate PCR of more banana RGCs of the NBS-type that potentially cluster within the clade N1 could provide a valuable

resource of RGC genes that could be used in *Fusarium* resistance tests. Likewise, the future bioinformatic analysis of the full genome sequence of banana which is currently in progress (www.musagenomics.org) will permit determining the total number of RGCs that cluster in clade N1. Indeed, this information will permit assessment in a systematic way of the role of this set of RGCs in *Fusarium* resistance. The remarkable finding of RGC2 correlation with FOC race 4 resistance (Chapter 4) will permit testing of the role of this sequence in FOC race 4 resistance using a genetic complementation approach. In order to achieve this task, different expression cassettes with the RGC2 ORF were constructed in the binary vector pCAMBIA 2200 (Figure 5.14). One of the constructs contains a putative promoter region of RGC2 isolated in this work and three other constructs contain characterised heterologous promoters that drive from high to low levels of gene expression in transgenic plants (Sanders et al. 1987; Christensen et al. 1996; Dugdale et al. 2000). The use of different heterologous promoters may ensure the expression of the RGC2 in the transgenic banana plants if the putative promoter region which has not been characterized fails to drive expression. The latter could occur if the isolated region containing the putative promoter region (~2.1 kb) does not contain all the necessary cis-acting regulatory elements that are essential for the proper expression of RGC2. Because of time constraints associated with this project, the construction of RGC5 constructs for banana transformation was not possible, however these constructs were recently made by the Plant Biotechnology staff at QUT and they will be used to test the potential of RGC5 to confer *Fusarium* resistance.

The results of Chapter 5 provide interesting insights into the structure and phylogeny of two potential *Fusarium* resistance genes from banana and provide a rational starting point for their functional characterization. The availability of the full ORF of these sequences will make possible testing their role in *Fusarium* resistance using a genetic complementation approach.

6.4 Conclusions

This thesis reports the characterisation of disease resistance gene candidates of the NBS-type and the discovery of two potential *Fusarium* resistance genes from the wild banana *Musa acuminata* ssp. *malaccensis*. This research is of particular importance in the development of banana genotypes with resistance to the devastating pathogen *F. oxysporum* race 4. The major research outcomes have been **(i)** cloning of disease resistance gene candidates of the NBS-type from the wild banana *M. acuminata ssp. malaccensis* and demonstration that they are associated to the non-TIR subclass of NBS sequences, **(ii)** identification of a resistance gene candidate (RGC2) whose expression is associated to FOC race 4 resistance, **(iii)** cloning of the full ORF of two potential *Fusarium* resistance genes (RGC2 and RGC5) that show significant sequence similarity to the tomato gene *I2* that confers resistance to *Fusarium oxysporum*, **(iv)** finding that the banana RGC2 and RGC5 are grouped within an ancient phylogenetic clade along with the *Fusarium* resistance genes *I2* and *Fom-2*, and finally **(v)** development of different expression cassettes containing the RGC2 ORF sequence with the aim of testing its role in FOC race 4 resistance using a genetic complementation approach. One of the expression cassettes contains a putative promoter region of RGC2 successfully isolated in this study.

The information generated in this thesis may lead to the identification of a FOC race 4 resistance gene in banana in further studies and may also assist the cloning of *Fusarium* resistance genes in other plant species. The future identification of a *Fusarium* resistance gene in banana will certainly have a tremendous implication for millions of people who depend on this crop as a staple food and also in the protection of the banana export industry. Furthermore, the identification of a *Fusarium* R gene in banana will provide interesting insights about the evolution of *Fusarium* R genes in plants and will facilitate the dissection of the signal transduction cascades leading to *Fusarium* resistance, which may give rise to novel and durable resistance strategies to control Panama disease.

6.5 References

Agrios, G.N. (1997). *Plant Pathology*. 4th ed. Academic Press, San Diego, California, USA p. 307.

Ashfield, T., Ong, L., Nobuta, K., Schneider, C., and Innes, R. (2004). Convergent evolution of disease resistance gene specificity in two flowering plant families. *Plant Cell* 16: 309-318.

Bai, J., Pennill, L.A., Ning, J., Lee, S.W., Ramalingam, J., Leach, J.E. and Hulbert, S.H. (2002). Diversity in nucleotide binding site-leucine rich repeat genes in cereals. *Genome Research* 12: 1871-1884.

Becker, D., Dugdale, B., Smith, M., Harding, R. and Dale, J. (2000). Genetic transformation of Cavendish banana (*Musa* spp. AAA group) cv 'Grand Nain' via microprojectile bombardment. *Plant Cell Reports* 19: 229-234.

Campbell, M.A., Fitzgerald, H.A. and Ronald, P.C. (2002). Engineering pathogen resistance in crop plants. *Transgenic Research* 11: 599-613.

Cannon, S. B., Zhu, H., Baumgarten, A.M., Spangler, R., May, G., Cook, D.R., and Young, N.D. (2000). Diversity, distribution and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. *Journal of Molecular Evolution* 54: 548-562.

Chang, Y.L., Henriquez, X., Preuss, D., Copenhaver, G.P., Zhang, H-B. (2003). A plant-transformation competent BIBAC library from the *Arabidopsis thaliana* Landsberg ecotype for functional and comparative genomics. *Theoretical and Applied Genetics* 106: 269-276.

Christensen A.H. and Quail P.H. (1996). Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Research* 5: 213-218.

Di Pietro, A., Madrid, M., Caracuel, Z., Delgado-Jarana, J. and Roncero, M.I. (2003). *Fusarium oxysporum*: exploring the molecular arsenal of vascular wilt fungus. *Molecular Plant Pathology*: 4: 315-325.

Dugdale B., Becker D., Beetham P., Harding R.M., Dale J.L (2000). Promoters derived from banana bunchy top virus DNA-1 to 5 direct vascular-associated expression in transgenic banana (*Musa* spp.). *Plant Cell Reports* 19: 810-814.

Frison E. (2003). Going bananas. *New Scientist*. 18 January. pp. 26-29

Goff, S. *et al*. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). *Science* 296: 92-100.

Grant, M., Godard, L., Straube, E., Ashfield, T., Leward, J., Sattler, A., Innes, R. and Dangl, J. (1995). Structure of the *Arabidopsis* RPM1 enabling dual specifity disease resistance. *Science* 269: 843-846.

Gurr, S.J. and Rushton, P.J. (2005). Engineering plants with increased disease resistance: how are we going to express it ? *Trends in biotechnology* 23: 283-290.

Hammond-Kosack, K., and Parker, J. (2003). Deciphering plant-pathogen communication: fresh perspectives for molecular resistance breeding. *Current Opinion in Biotechnology* 14: 177-193.

He, R.F., Wang, Y., Shi, Z., Ren, X., Zhu, L., Weng, Q. and He, G.C. (2003). Construction of a genomic library of wild rice and *Agrobacterium*-mediated transformation of large insert DNA linked to *BPH* resistance locus. *Gene* 321: 113-121.

Hua-Van, A., Langin, T. and Daboussi, M. (2001). Evolutionary history of the impala transposon in *Fusarium oxysporum*. *Molecular Biology and Evolution* 18: 1959-1969.

Hulbert, S., Webb, C., Smith, S. and Sun, Q. (2001). Resistance gene complexes: evolution and utilization. *Annual Review of Phytopathology* 39: 285-312.

Joobeur, T., King, J., Nolin, S., Thomas, C., and Dean, R. (2004). The *Fusarium* wilt resistance locus *Fom-2* of melon contains a single resistance gene with complex features. *Plant Journal* 39: 283-297.

Khanna, H., Becker, D., Kleidon, J. and Dale, J. (2004). Centrifugation assisted *Agrobacterium*-mediated transformation (CAAT) of embryogenic cell suspensions of banana (*Musa* spp. Cavendish AAA and Lady finger AAB). *Molecular Breeding* 14: 239-252.

Koenig, R., Ploetz, R. and Kistler, H. (1997). *Fusarium oxysporum* f. sp. *cubense* consists of a small number of divergent and globally distributed clonal lineages. *Phytopathology* 87: 915-923.

McDowell, J. (2004). Convergent evolution of disease resistance genes. *Trends in Plant Science* 9: 315-317

Mes, J., Van Doorn, A., Wijbrandi, J., Simons, G., Cornelissen, B. and Haring, M. (2000). Expression of the *Fusarium* resistance gene *I-2* colocalizes with the site of fungal containment. *Plant Journal* 23: 183-193.

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: 1113-1130.

Meyers, B.C., Dickerman, A., Michelmore, R., Sivaramakrishnan, S., Sobral, B. and Young, N. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant Journal* 20: 317-332.

Meyers, B.C., Kaushik, S. and Nandety, R.C. (2005). Evolving disease resistance genes. *Current Opinion in Plant Biology* 8: 129-134.

Monosi, B., Wisser, R.J., Pennill, L. and Hulbert, S.H. (2004). Full-genome analysis of resistance gene homologues in rice. *Theoretical and Applied Genetics* 109: 1434-1447.

Mysore, K.S. and Choong-Min, R. (2004). Non-host resistance: how much do we know? *Trends in Plant Science* 9: 97-104.

O'Donnell, K., Kistler, H., Cigelnik, E. and Ploetz, R. (1998). Multiple evolutionary origins of the fungus causing Panama disease of banana: Concordant evidence from nuclear and mitochondrial gene genealogies. *Proceedings of the National Academy of Science (USA)* 95: 2044-2049.

Ortiz-Vázquez, E., Kaemmer, D., Zhang, H.B., Muth, J., Rodríguez-Mendiola, M., Arias-Castro, C. and James, A. (2005). Construction and characterization of a plant transformation-competent BIBAC library of the black Sigatoka resistant banana *Musa acuminata* cv. Tuu Gia (AA). *Theoretical and Applied Genetics* 110: 706-713.

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: 203-213.

Ploetz, R.C. and Pegg, K.G. (2000). Fungal diseases of the root, corm and pseudostem. pp. 143-172. In *Diseases of Banana, Abacá and Enset*. (Jones, D.R., ed). CAB International, Wallingford, UK.

Rommens, C. and Kishore, G. (2000). Exploiting the full potential of disease-resistance genes for agricultural use. *Current Opinion in Biotechnology* 11: 120-125.

Sagi, L., Panis, B., Remy, S., Schoofs, H., De Smet, K., Swennen, R. and Cammue, P. (1995). Genetic transformation of banana and plantain (*Musa* spp) via particle bombardment. *Bio/Techniques* 13: 481-485.

Sanders, P.R., Winter, J.A., Barnason, A.R., Rogers, S.G. and R.T. Fraley (1987). Comparison of cauliflower mosaic virus 35S and nopaline synthase promoters in transgenic plants. *Nucleic Acids Research* 15: 1543-1558.

Simons, G., Groenendijk, J., Wijbrandi, J., Reijans, M., Groenen, J., Diergaarde, P., Van der Lee, T., Bleeker, M., Onstenk, J., Both, M., Haring, M., Mes, J., Cornelisse, B., Zabeau, M. and Vos, P. (1998).

Dissection of the *Fusarium I2* gene cluster in tomato reveals six homologs and one active gene copy. *Plant Cell* 10: 1055-1068.

Stuiver, M. and Custers, J. (2002). Engineering disease resistance in plants. *Nature* 411: 865-868.

Tai, T., Dahlbeck, D., Clarck, E., Gajiwala, P., Pasion, R., Whalen, M., Stall, R. and Staskawicz, J. (1999). Expression of the *Bs2* pepper gene confers resistance to bacteria spot disease in tomato. *Proceedings of the National Academy of Science (USA)* 96: 14153-14158.

Tao, Q., Wang, A. and Zhang, H-B. (2002).. One large-insert plant-transformation-competent BIBAC library and three BAC libraries of *japonica* rice for genome research in rice and other grasses. *Theoretical and Applied Genetics* 105: 1058-1066.

Taylor, K. (2005). Characterization of potential fungal disease resistance genes in banana. PhD thesis. Queensland University of Technology, Brisbane, Australia. pp. 68-96.

Thordal-Christensen, H. (2003). Fresh insights into processes of non-host resistance. *Current Opinion in Plant Biology* 6: 351-357.

Thwaites, R., Eden-Green, S.J and Black, R. (2000). Diseases caused by bacteria. pp. 213-239. In *Diseases of Banana, Abacá and Enset*. (Jones, D.R., ed). CAB International, Wallingford, UK.

Warren, R., Henk, A., Mowery, P., Holub, E. and Innes, R. (1998). A mutation within the leucine-rich repeat domain of the Arabidopsis disease resistance gene *RPS5* partially suppresses multiple bacterial and downy mildew resistance genes. *Plant Cell* 10: 1439-1452.

Waterhouse, P.M. and Helliwell, C.A. (2003). Exploring plant genomes by RNA-induced gene silencing. *Nature Reviews* 4: 29-38.

Whitham, S. (1996). The *N* gene of tobacco confers resistance to tobacco mosaic virus in transgenic tomato. *Proceedings of the National Academy of Science (USA)* 93: 8776-8781.

Van der Vossen, E., Sikkema, A., Lintel Hekkert, B., Gros, J., Stevens, P., Muskens, M., Wouters, D., Pereira, A., Stiekema, W., 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* 36:867-882.

Zhou, T., Wang, Y., Chen, J.Q., Araki, H., Jing, Z., Jiang, K., Shen, J. and Tian, D. (2004). Genome-wide identification of NBS genes in *japonica* rice reveals significant expansion of divergent non-TIR-NBS-LRR genes. *Molecular Genetics and Genomics* 271: 402-415.