TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Pflanzenzüchtung

Characterization of the *Pl*$_{ARG}$ locus mediating resistance against *Plasmopara halstedii* in sunflower

Silke Wieckhorst

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Agrarwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. C. Schwechheimer

Prüfer der Dissertation:

1. Univ.-Prof. Dr. C.-C. Schön
2. Univ.-Prof. Dr. R. Hückelhoven

Die Dissertation wurde am 06.10.2011 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 14.11.2011 angenommen.

Abbreviations

AFLP	amplified fragment length polymorphism
Avr	avirulence
BAC	bacterial artificial chromosome
BIBAC	binary bacterial artificial chromosome
BES	BAC end sequence
bph	bases per hash
BR	resistant bulk
BS	susceptible bulk
BSA	bulked segregant analysis
BSTA	bulked segregant transcriptome analysis
CAPS	cleaved amplified polymorphic sequence
CC	coiled-coil domain
CLI	cotyledon limited infection
cM	centi Morgan
CMS	cytoplasmic male sterility
CUGI	Clemson University Genomics Institute
DNA	deoxyribonucleic acid
E	expect value
EST	expressed sequence tag
FLI	Fritz Lipmann Institute
HICF	high information content fingerprinting
HR	hypersensitive cell death response
hss	hash saving steps
Indel	insertion / deletion events
LRR	leucine-rich repeat
MAMPs	microbial-associated molecular patterns
MAS	marker assisted selection
NBS	nucleotide binding site
NCBI	National Centre of Biotechnology Information
NGS	next generation sequencing
PAMPs	pathogen-associated molecular patterns
PCR	polymerase chain reaction
PK	serine/threonine kinase

Pl_{ARG}	*Pl* locus in sunflower introgressed from the wild species *Helianthus argophyllus* which mediates resistance to downy mildew caused by *Plasmopara halstedii*
QTL	quantitative trait locus
RAPD	random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
RGC	resistance gene candidate
R genes	resistance genes
RNA	ribonucleic acid
SNP	single nucleotide polymorphism
SSCP	single strand conformation polymorphism
SSR	simple sequence repeat
TDF	transcript-derived fragment
TE	transposable element
TIR	*Drosophila* toll and mammalian interleukin-1 receptor
TM	transmembrane domain

1 Introduction

Plant breeding is a perpetual task to develop high yielding varieties which are well adapted to the target environment. Modern crops are mainly grown on large scale as monoculture and therefore are vulnerable to pathogen attacks. To protect plants effectively against this threat resistance genes are introduced into elite breeding material. Molecular tools help to elucidate the genetic basis and genomic organization of the resistance mechanism and thus facilitate the use of resistance genes in an effective and durable way. Application of molecular tools in the breeding process can speed up the development of resistant varieties and allows the combination of multiple resistance genes. The aim of this study is the molecular characterization of the downy mildew resistance gene Pl_{ARG} originating from the wild species *Helianthus argophyllus*.

1.1 Origin and distribution of cultivated sunflower

Sunflower (*Helianthus annuus* L.) is a New World crop which originates from North America. It is nowadays the second largest hybrid crop after maize and one of the major oilseed crops after soybean, rapeseed, cottonseed, and groundnut cultivated worldwide on 24 million ha (Seiler and Jan 2010). The largest producers of sunflower seed are the Russian Federation and the Ukraine. In 2009, Germany had an area of 23,600 ha under cultivation (Table 1-1).

The domesticated sunflower was introduced to Europe in the early 16th century, and initially cultivated as an ornamental it spread quickly throughout Europe (Seiler and Jan 2010). Sunflower gained importance as an oil seed crop with the introduction to Russia in 1697. First breeding success was achieved in the first sixty years of the 20th century, where achene oil content of Soviet cultivars increased from less than 300 g/kg to 500 g/kg.

Till 1946 only open-pollinated cultivars were grown which were heterogeneous with regard to height, flowering, maturity, dry-down rate after maturity and were susceptible to many prevalent diseases. The introgression of the cytoplasmic male sterility (CMS) plasma PET1 originating from *H. petiolaris* (Leclercq 1969) as well as the identification of nuclear fertility restorer genes (Kinman 1970) allowed the development of hybrids. The first hybrids produced using CMS were introduced 1972 in the USA and within five years they were grown on about 90% of the production area (Seiler and Jan 2010). Another important improvement in sunflower breeding was the creation of a high-oleic acid sunflower mutant (Soldatov 1976) which had a modified amount of unsaturated

fatty acids and thus increased the utilization of sunflower oil for both food and industrial applications (Vick and Hu 2010). Availability of hybrids and the success of sunflower as a source of food and edible oil resulted in worldwide increase of area cultivated with sunflower and was followed by the occurrence of several diseases (Sackston 1981a). Thus, resistance breeding is of pivotal importance to provide farmers with healthy sunflower hybrids. This study aims to discover the structural and functional basis of the downy mildew resistance Pl_{ARG} and to develop molecular tools to facilitate the selection of resistant plants which might replace expensive phenotypic tests.

Table 1-1: Sunflower seed production area in the year 2009 in ha presented for the five continents, for ten countries which have the highest area harvested worldwide, and area under cultivation in Germany (FAOSTAT 2009).

Continent / Country	Area harvested [ha]
Europe	14,109,736
Asia	5,171,979
Americas	3,141,362
Africa	1,384,102
Oceania	51,757
Russian Federation	5,597,900
Ukraine	4,193,000
Argentina	1,820,030
India	1,800,000
China	970,000
Spain	851,600
United States of America	790,562
Romania	761,093
France	724,800
Bulgaria	683,711
Germany	23,600

1.2 Disease resistance in plants

Plants are attacked by fungi, oomycetes, bacteria, nematodes, insects which feed on the photosynthetic products or by viruses which use plants as a replication machinery. No antibody-based defense system like in mammals exists in plants, but they have evolved sophisticated mechanisms to perceive attacks of pathogens (Dangl and Jones 2001; Toyoda et al. 2002). Flor (1971) propagated the gene-for-gene resistance concept after analyzing the flax/flax-rust pathosystem and observing specificity between single host resistance *(R*) genes and single pathogen avirulence (*Avr*) genes. Meanwhile, studies also report about cases where two *R* genes are required for disease resistance or a single *R* gene like *Mi-1* functions against multiple species of

root-knot nematode and aphids (Bent and Mackey 2007; Eitas and Dangl 2010).

The plant immune system was described by Jones and Dangl (2006) as a “zigzag” model with four phases of interaction between pathogens and the plant. A similar model was suggested by Bent and Mackey (2007) who called it the new “Central Dogma” of plant pathology. In phase one, the basal immune system of plants recognizes and responds to the so called microbial- or pathogen-associated molecular patterns (MAMPs or PAMPs) like bacterial flagellins, lipopolysaccharides, or fungal chitins. However, pathogens can overcome the MAMP/PAMP triggered immunity in phase two by evolving virulence factors that actively suppress parts of the general defense response in the hosts resulting in a susceptible phenotype. In phase three, adapted pathogens can be repelled by the host when the host species evolves specific *R* genes which recognize a given effector, the *Avr* protein, resulting in effector-triggered immunity. The recognition of the effector initiates disease resistance and usually a hypersensitive cell death response (HR). Under selection pathogens evolve further and escape detection by the *R* gene products in phase four resulting again in effector-triggered susceptibility. Thus, the spectrum of pathogens and resistance genes is dynamic and changes over time. One challenge for plant breeders is the identification, introduction and maintenance of resistance genes in elite germplasm to protect the plants against pathogens.

Many plant *R* genes have been discovered and cloned in recent years. Five major classes of *R* genes are known (Fig. 1-1) which show common structural motifs and conserved domains. Most of the cloned *R* genes belong to the nucleotide binding site - leucine-rich repeat (NBS-LRR) class, consisting of the TIR-NBS-LRR and CC-NBS-LRR subclasses (Dangl and Jones 2001). The TIR domain has homology to the *Drosophila* Toll and mammalian interleukin-1 receptor (TIR), whereas CC stands for a putative coiled-coil domain. The resistance genes *L* in flax against *Melampsora lini* (flax rust) and *N* in tobacco against the tobacco mosaic virus (Lawrence et al. 1995; Whitham et al. 1994) belong to the TIR-NBS-LRR subclass. One member of the CC-NBS-LRR subclass is the resistance gene *Mi* in tomato against the nematode *Meloidogyne incognita* and the potato aphid *Macrosiphum euphorbiae* (Milligan et al. 1998). The second class harbors the LRR-transmembrane (LRR-TM) domains to which the resistance genes *Cf-2* and *Cf-9* belong. These genes protect tomato against *Cladosporium fulvum* (Dixon et al. 1996; Jones et al. 1994). The third class, the LRR-TM-protein kinase (LRR-TM-PK) class, comprises the resistance gene *Xa21* which causes resistance in rice against *Xanthomonas oryzae* (Song et al. 1995). Resistance genes that have a serine/threonine protein kinase (PK) belong to the fourth class. One

member is *Pto* in tomato against the bacterium *Pseudomonas syringae* (Martin et al. 1993). A few *R* proteins do not fit into the other classes and therefore build the fifth class like *RPW8* which confers non-race specific resistance in *Arabidopsis* against powdery mildew (Xiao et al. 2001). Molecular cloning of the resistance gene PI_{ARG} will provide sequence information of the target region. Thus, it will be possible to assign PI_{ARG} to its resistance gene class.

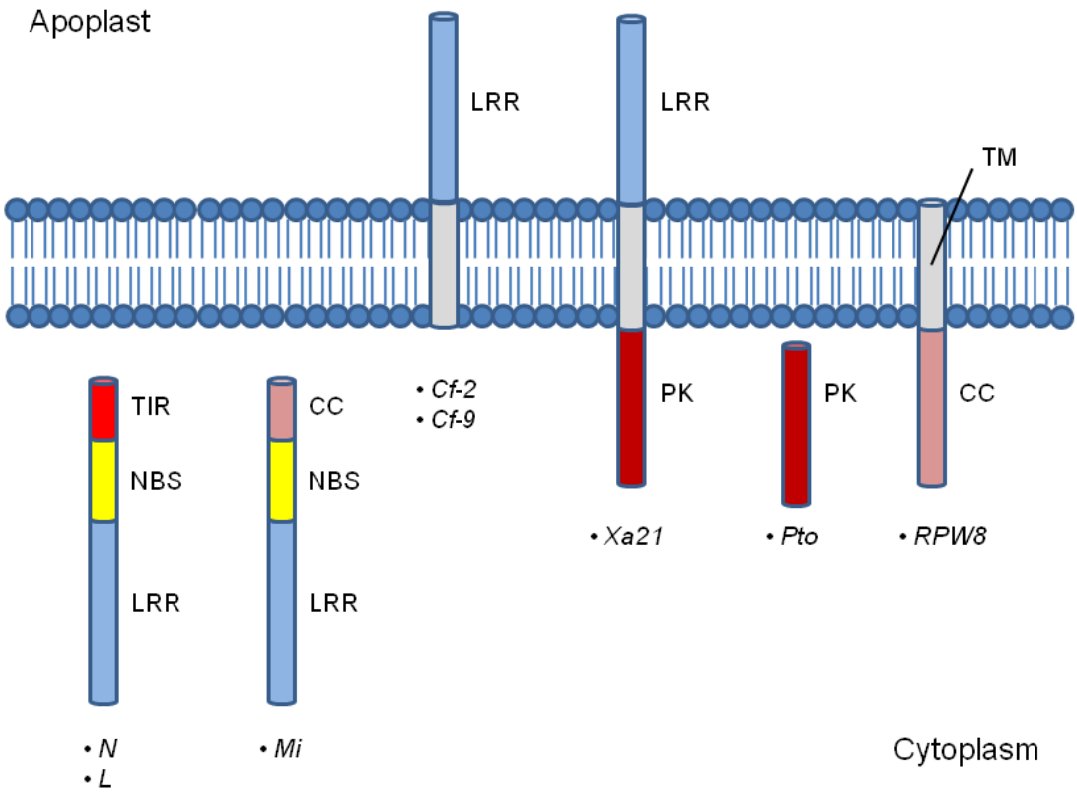


Figure 1-1: Five classes of plant disease resistance proteins and examples of cloned resistance genes modified after Dangl and Jones (2001).

New *R* specificities evolve by single base mutations or small insertions/deletions and also by intragenic recombination with unequal exchange to make longer/shorter LRRs and/or by equal or unequal extragenic recombination to generate more/fewer *R* genes at one locus (Bent and Mackey 2007). Therefore, different arrangements of *R* genes are known (Hammond-Kosack and Jones 1997): 1) The *R* gene may exist as a single copy that is present in resistant lines but absent from susceptible lines (*RPM1* locus, *Arabidopsis*). 2) *R* genes can consist of a single gene with an array of distinct alleles, providing different recognition specificity (*L* locus, flax). *3*) For many genes, the *R* locus

is comprised of tandem arrays of closely linked *R* gene homologues with differing specificities (*M* locus, flax). 4) In particular genomic regions, *R* genes to viral, bacterial, and fungal pathogens are clustered like *Rpg1* against *Pseudomonas syringae* and *Rps3* against *Phythophthora soja* in soybean (Ashfield et al. 1998). The goal of this study is the structural and functional characterization of the downy mildew resistance locus Pl_{ARG}.

Disease resistance in sunflower

Sunflowers are attacked by many diseases and therefore one of the major breeding goals besides yield is the development of resistant genotypes. The spectrum of pathogens is dynamic and changed over the years. In the last decade (1992 – 2004) white rot, caused by *Sclerotinia sclerotiorum*, was the most important sunflower disease worldwide, followed by Phomopsis (*Diaporthe helianthi*), charcoal rot (*Macrophomina phaseolina*), downy mildew (*Plasmopara halstedii*) and broomrape (*Orobanche cumana*) (Vear 2004). Since 2004 the importance of downy mildew has grown and it is nowadays one of the most prevalent diseases besides broomrape (Virányi 2008). Plant breeders have to provide commercial hybrids which carry resistance genes against several diseases and resistance against *P. halstedii* is a "must have" trait for sunflower.

Plasmopara halstedii

Plasmopara halstedii (Farl.) Berl. & de Toni is a soil-, seed- and wind-born pathogen causing downy mildew in sunflower (Fig. 1-2). The pathogen belongs to the family of Oomycetes. Oospores of *P. halstedii* can survive up to ten years in the soil (EPPO/CABI 1997). The pathogen was first described in the USA on *Asteraceae* in 1882 and was named *Peronospora halstedii*. In the year 1888 it was re-named to *Plasmopara halstedii.* In Eastern Europe *P. halstedii* was first found on sunflower in 1941 (Sackston 1981b).

Downy mildew causes significant yield losses. Infection of up to 90% of sunflower plants with typical symptoms of systemic infection has been reported in the fields and yield losses were estimated to be up to 50% (Sackston 1981b). The pathogen can be controlled by fungicides and by cultivation of resistant hybrids. Several major *Pl* loci (Pl_1-Pl_{15}, Pl_{ARG}, Pl_v-Pl_z) have been described, which confer resistance to one or more races of *P. halstedii* (Table 1-2). Two types of resistance are known: In seedlings with type I resistance *P. halstedii* growth is limited to the roots, whereas in seedlings with type II resistance *P. halstedii* grows through the hypocotyl and a slight sporulation may

be observed on the cotyledons (Mouzeyar et al. 1994). Sackston (1990) called the phenomenon observed in type II resistance “cotyledon limited infection” (CLI). Races of *P. halstedii* are determined according to the international nomenclature based on a series of differential lines, which has been proposed by Tourvieille de Labrouhe et al. (2000a).

Figure 1-2: Sunflower seedling with spores of *P. halstedii* 15 days after inoculation.

Initially, only the race 100 of *P. halstedii* was known in Europe. This race has a specific virulence profile different from that of American races (Delmotte et al. 2008). However, in the last decades, new races of *P. halstedii* were discovered worldwide in the cultivation areas of sunflower (Gulya et al. 1991; Molinero-Ruiz et al. 2002; Tourvieille de Labrouhe et al. 2000b). Delmotte et al. (2008) investigated 24 isolates collected in France between 1966 and 2006 and concluded that the biology of *P. halstedii*, the multiple introductions of the pathogen into France, and the selection pressure caused by the use of host resistance genes may have caused the spread of new races of *P. halstedii*. Gulya (2007) reported the existence of at least 35 pathotypes worldwide. Several races developed tolerance to metalaxyl and mefenoxam, the only effective fungicides available (Albourie et al. 1998; Molinero-Ruiz et al. 2005; Molinero-Ruiz and Melero-Vara 2003; Spring et al. 2006). Therefore, investigation of the structure and functionality of *Pl* loci is necessary in order to exploit them effectively and durably in plant breeding.

Plasmopara (*Pl*) resistance loci

The genus *Helianthus* contains 51 species, 14 annual and 37 perennial, and belongs to the family *Compositae* (*Asteraceae*) (Schilling 2006). To enlarge the genetic variability

in cultivated sunflower interspecific hybridization between *Helianthus* species is an important tool. Plant breeding benefits from the introduction of disease resistance genes from wild species into cultivated sunflower by broadening the genetic basis of disease resistance in breeding material (Seiler 2010; Seiler 1992). Pl_{ARG} was introgressed from the wild species *H. argophyllus* and mediates resistance to all known races of *P. halstedii* (Seiler et al. 1991). Till now no race of *P. halstedii* is known which overcomes the resistance of Pl_{ARG} (Gerald Seiler, personal communication). Pl_{ARG} is a new source for broad-spectrum resistance against *P. halstedii* and thus is of high interest for improvement of elite breeding materials.

Table 1-2: Overview of known *Pl* loci, source of origin, linkage group (LG) and references.

Loci	Source	LG	Reference
Pl_1	*H. annuus*	LG8	Vranceanu and Stoenescu 1970; Mouzeyar et al. 1995; Gedil et al. 2001
Pl_2	*H. annuus*	LG8	Zimmer and Kinman 1972; Vear et al. 1997; Brahm et al. 2000
Pl_3 (Pl_1)	*H. annuus*		Vear and Leclercq 1971; Zimmer and Kinman 1972
Pl_4 (Pl_2)	*H. tuberosus*		Vear 1974
Pl_5	*H. tuberosus*	LG13	Vranceanu et al. 1981; Bert et al. 2001; Radwan et al 2003
Pl_6	*H. annuus*	LG8	Miller and Gulya 1991; Roeckel-Drevet et al. 1996 ; Vear et al. 1997; Bouzidi et al. 2002
Pl_7	*H. praecox*	LG8	Miller and Gulya 1991
Pl_8	*H. argophyllus*	LG13	Miller and Gulya 1991; Radwan et al. 2003
Pl_9	*H. annuus*		Gulya et al. 1991
Pl_{10}	*H. annuus*		Gulya et al. 1991
Pl_{11}	*H. annuus*		Rahim et al. 2002
Pl_{12}	*H. annuus*		Rahim et al. 2002
Pl_{13}	*H. annuus*	LG1	Gulya et al. 1985; Mulpuri et al. 2009
Pl_{14}	*H. annuus*	LG1	Gulya et al. 1985; Bachlava et al. 2010a
Pl_{15}	*H. annuus*	LG8	Bertero de Romano et al. 2010
Pl_{ARG}	*H. argophyllus*	LG1	Seiler et al. 1991; Dußle et al. 2004; Wieckhorst et al. 2010
Pl_v-Pl_z	*H. annuus*		Molinero-Ruiz et al. 2002b; Molinero-Ruiz et al. 2003

Two categories of *P. halstedii* resistance exist: Qualitative resistance caused by single major *Pl* loci (Table 1-2) and quantitative resistance which is controlled by several genes with minor effects (Tourvieille de Labrouhe et al. 2008; Vear et al. 2008a). Pl_{ARG} is a major *Pl* locus which is effective against *P. halstedii*. Classical genetic analysis by phenotyping segregating populations elucidated that Pl_{ARG} is unlinked to the previous known major resistance loci Pl_1, Pl_2, Pl_5, Pl_6, Pl_7 and Pl_8 which are mainly used in breeding material (Röcher 1999; Vear et al. 2003).

The major *Pl* loci were initially considered as single independent genes (Miller and Gulya 1991), but genetic mapping and segregation studies revealed that some loci consist of clusters of resistance genes with different specificities. Mouzeyar et al. (1995) mapped the first *Pl* locus (Pl_1) on group 1 of the CARTISOL map (Gentzbittel et al. 1995), which corresponds to linkage group (LG) 8 of the public SSR map constructed by Tang et al. (2002). Roeckel-Drevet et al. (1996) and Vear et al. (1997) mapped Pl_2 and Pl_6, respectively, and found that both loci are located in the same genomic region as Pl_1. Vear et al. (1997) showed for Pl_6 that it is a cluster of at least two resistance genes conferring resistance to the races 100 and 300 on one hand and 700, 703 and 710 on the other hand. Another cluster of *Pl* loci was identified on LG 13 which carries Pl_5 and Pl_8 (Bert et al. 2001; Radwan et al. 2003). Mulpuri et al. (2009) and Bachlava et al. (2010a) mapped Pl_{13} and Pl_{14} at the bottom of LG 1. As centromere positions of sunflower chromosomes are unknown the terms “at the bottom” and “at the top” of LGs are used in this thesis to describe the position of genes or markers. Orientations of LGs are based on the reference map of Tang et al. (2002). For the clusters on LG1 (Pl_{13}, Pl_{14}), LG8 (Pl_1, Pl_2, Pl_6, Pl_7) and LG13 (Pl_5, Pl_8) it was shown that several resistance gene candidates (RGC) of the TIR- and CC-NBS-LRR classes are closely linked with the downy mildew resistance loci (Bachlava et al. 2010a; Bouzidi et al. 2002; Gedil et al. 2001; Gentzbittel et al. 1998; Radwan et al. 2004; Radwan et al. 2003; Radwan et al. 2008; Slabaugh et al. 2003). This work aims at testing the hypothesis that Pl_{ARG} is a cluster of TIR- or CC-NBS-LRR resistance gene candidates like Pl_{14} on LG1, Pl_1, Pl_2, Pl_6, P_7 on LG8, and Pl_5, Pl_8 on LG13.

1.3 Molecular resources in sunflower

The availability of molecular tools and resources is a prerequisite for identification and cloning of candidate genes for the Pl_{ARG} locus. A large segregating population and the availability of molecular markers allow the fine mapping of the target region and the genetic localization of Pl_{ARG} which is the basis for cloning the gene.

Sunflower (2n=2x=34) has a relatively large genome size of 3,000 - 3,500 Mb

(Arumuganathan and Earle 1991; Baack et al. 2005) and is not yet sequenced. However, over the past years, a large resource of molecular markers and expressed sequence tag (EST) sequences has been developed. The National Centre of Biotechnology Information (NCBI) dbEST now contains 133,695 cultivated sunflower EST sequences (30.04.2011). Moreover, EST sequences of wild *Helianthus* species like *H. argophyllus* (35,720 ESTs), *H. ciliaris* (21,590 ESTs), *H. exilis* (33,961 ESTs), *H. paradoxus* (30,517 ESTs), *H. petiolaris* (27,484 ESTs), and *H. tuberosus* (40,362 ESTs) are available.

Molecular marker development is advanced in sunflower and over the years different marker types have been developed and used for marker assisted selection (MAS). Pérez-Vich and Berry (2010) described three different generations of markers in sunflower research: Firstly, anonymous deoxyribonucleic acid (DNA) markers like restriction fragment length polymorphism (RFLP), random amplification of polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and genomic SSR markers were developed (Al-Chaarani et al. 2002; Berry et al. 1995; Gentzbittel et al. 1995; Lawson et al. 1998; Lu et al. 2000; Quagliaro et al. 2001; Tang et al. 2002). The second generation of markers were gene-targeted markers based on sequenced sunflower cDNA-RFLP probes, single nucleotide polymorphisms (SNP), insertion / deletion events (Indel), and SSR markers developed from sunflower expressed sequence tags (ESTs) (Kolkman et al. 2007; Lai et al. 2005a; Pashley et al. 2006; Yu et al. 2003). Functional markers belong to the third generation of DNA markers, which detect causal DNA sequence differences between alleles of a gene underlying a given phenotype. This type of markers has been developed for traits determining oil quality and herbicide resistance (Kolkman et al. 2004; Tang et al. 2006).

Identification of markers closely flanking the gene of interest can be obtained with bulked segregant analysis (BSA) (Michelmore et al. 1991). The concept is that two bulk samples are generated from a population segregating for the gene of interest. The individuals of each bulk are genetically identical for a particular trait or genomic region, but arbitrary at all unlinked regions. Thus, the bulks are genetically dissimilar in the selected region but heterozygous at all other regions. The large resource of molecular markers in combination with BSA allows the fine mapping of the Pl_{ARG} locus in a segregating population and closely linked markers can be the starting point for physical mapping.

For molecular cloning of Pl_{ARG} bacterial artificial chromosome (BAC) libraries are required. Several *H. annuus* BAC libraries have been developed. The first published

BAC library was derived from the inbred line HA821 with a coverage of four to five genome equivalents and an insert size of 80 kb (Gentzbittel et al. 2002). A second sunflower BAC library with 1.9 x coverage was developed for RHA325 which carries the restorer gene *Rf1* and the Pl_2 locus (Özdemir et al. 2004). Bouzidi et al. (2006) developed a BAC library suitable for screening with hybridization probes as well as with polymerase chain reaction (PCR). The BAC library has an insert size of around 80 kb, five-fold genome coverage and was developed from sunflower line YDQ. Feng et al. (2006) developed complementary BAC and binary bacterial artificial chromosome (BIBAC) libraries from sunflower cultivar HA89 using two different restriction enzymes to provide a greater probability of obtaining a clone of interest. The BAC library has an average insert size of 140 kb and the BIBAC library of 137 kb. Together, the two libraries are equivalent to approximately 8.9 haploid genomes of sunflower. A further BAC library is publicly available at the Clemson University (http://www.genome.clemson.edu). It has an insert size of 125 kb and is spotted on filters with coverage of 8.5 x. This library was used for molecular characterization of the target region of Pl_{ARG}, because of the large average insert size, high genome coverage, and public availability.

1.4 Forward genetic approaches

Cloning disease resistance genes with a known phenotype is a typical forward genetic approach. Forward genetics aims to identify the sequence information that underlies a specific phenotype. Towards the characterization of the resistance locus Pl_{ARG} different options were at hand: The classical map based cloning approach was combined with a RGC approach. As a complementary approach two different genome wide transcriptome based strategies were followed: the non-targeted cDNA-AFLP method to identify differentially expressed fragments associated with the resistance reaction and the targeted bulk segregant transcriptome analysis (BSTA), aiming at the identification of expressed resistance gene candidates from the target region.

Map based cloning

A map based cloning approach requires the identification of markers closely linked to the gene of interest, the establishment of a high resolution mapping population, and the availability of a large-insert library for physical mapping. In this approach, a large-insert BAC library is screened with probes of closely linked markers delimiting the target interval and BAC end-sequences are generated from the identified BAC clones. After re-mapping of BAC end-sequences a further screening of the BAC library is necessary until a minimum-tiling path is identified. Sequencing of BAC clones and annotation

provide candidates for the gene of interest. Resistance genes like *Lr10* in wheat and *Mi* in tomato (Feuillet et al. 2003; Milligan et al. 1998; Stein et al. 2000) were identified through map based cloning.

Resistance gene candidates

After several resistance genes had been cloned, it became obvious that many of them shared conserved sequence motifs (Gebhardt 1997). Leister et al. (1996) developed the RGC approach. They designed degenerate primers for the conserved domains such as the P-loop and GLPL-motif of the NBS domain from resistance genes of tobacco and *Arabidopsis thaliana*. The application of these markers in potato resulted in the identification of sequences which showed homology to known resistance genes.

The approach of amplifying and mapping resistance gene candidates was widely and successfully used in different species, for different pathogens, and for different resistance gene classes. Hattendorf et al. (2007) isolated RGCs in roses and identified RGC clusters which are associated with quantitative trait loci (QTL) against powdery mildew. Hunger et al. (2003) amplified RGC sequences from *Beta vulgaris* L. and found linkage between RGCs and loci for rhizomania and *Cercospora* resistance. In sunflower, resistance gene candidates with TIR-NBS-LRR motifs were identified in the vicinity of Pl_1, Pl_2, Pl_6, Pl_7 (Bouzidi et al. 2002; Gedil et al. 2001; Gentzbittel et al. 1998; Slabaugh et al. 2003), and with CC-NBS-LRR motifs in the Pl_5 / Pl_8 and Pl_{14} region (Bachlava et al. 2010a; Radwan et al. 2004; Radwan et al. 2003). In a genome wide approach, Radwan et al. (2008) identified 630 NBS-LRR homologs in sunflower by database analysis and sequencing of DNA fragments harboring conserved NBS sequences. They derived amplicons from common and wild sunflower species and developed DNA markers from 196 unique NBS-LRR sequences from which 167 NBS-LRR loci were mapped. NBS-LRR loci were distributed in clusters or singletons throughout the sunflower genome. On LG 8, in the region of the downy mildew resistance loci Pl_1, Pl_2, Pl_6, the largest and most complex cluster with 38 RGCs within 9.3 cM was identified. On LG 13 (Pl_5, Pl_8) the second largest NBS-LRR cluster with 27 TIR-NBS-LRR and CC-NBS-LRR loci within 31.6 cM was found. Thus, Radwan et al. (2008) provided a source for identification and cloning of *R* genes in sunflower. This resource was employed to identify RGCs in the target region of Pl_{ARG}.

Non-targeted transcriptome analysis

In addition to the structural analysis of forward genetics candidate genes can also be identified at the transcriptome level in case of differential expression after inoculation

with the pathogen. It enables the discovery of genes involved in plant-pathogen interaction without prior sequence knowledge. The cDNA-AFLP approach was used in several experiments to identify genes involved in the incompatible reaction between plants and pathogens (Baldo et al. 2010; Eckey et al. 2004; Wang et al. 2010). For example, Chapman et al. (2009) identified candidate genes associated with *Pch2* eyespot resistance in wheat and detected two resistance gene candidates which showed homology to a callose synthase protein and to a NBS-LRR disease resistance protein. Chacón et al. (2010) used the suppression subtractive hybridization approach and identified the *N. tabacum protein kinase gene* (*NtPK*) as a candidate for disease resistance against *Rhizoctonia solani*. Functional studies confirmed the involvement of the *NtPK* gene in defense against *R. solani*.

To identify candidate genes for Pl_{ARG} in this study samples were taken at different time points from *P. halstedii* inoculated and non-inoculated seedlings of the susceptible line HA342 and the resistant line ARG1575-2. These samples were analyzed with the cDNA-AFLP method to identify differentially expressed genes.

Targeted bulk segregant transcriptome analysis

Next-generation high-throughput DNA sequencing techniques open new opportunities in genome analysis. Different platforms exist which are reviewed by Ansorge et al. (2009). Next generation sequencing (NGS) technologies have been used for a broad spectrum of applications such as *de novo* sequencing of the cucumber genome in combination with Sanger sequencing (Huang et al. 2009), for *de novo* sequencing in barley (Steuernagel et al. 2009), for resequencing of *Arabidopsis* (Turner et al. 2010), and for transcriptome sequencing in American (*Castanea dentate*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection (Barakat et al. 2009). The decreased costs of NGS enable application of this technology for new approaches to identify causal mutations in selected phenotypic bulks which has been successfully shown by Schneeberger et al. (2009) in *Arabidopsis thaliana*. The approach can be adapted to the transcriptome level by sequencing the transcriptome of one susceptible and one resistant phenotypic bulk using NGS, thereby targeting the genomic region of the resistance locus. Each bulk contains individuals that are homozygous susceptible and homozygous resistance to Pl_{ARG} respectively. The bulks are therefore genetically dissimilar in the target region but heterozygous at all unlinked regions. Transcripts from the target region can be identified by searching for SNPs which show distinct polymorphism between the two bulks. This approach is further on referred to as bulk segregant transcriptome analysis (BSTA).

1.5 Objectives

Sunflower production has increased after release of high oil content cultivars and the discovery of cytoplasmic male sterility and restorer genes (Sackston 1981b). One challenge in plant breeding is the development of healthy varieties, because sunflowers are attacked by many diseases and one of the most widespread diseases is downy mildew caused by *Plasmopara halstedii*. Pl_{ARG} was introduced from the wild species *H. argophyllus* and mediates resistance against all known races of *P. halstedii*. This work aims at analyzing the structural basis of the resistance locus Pl_{ARG} and at predicting candidate genes which can be used purposefully and durably in plant breeding. In particular the objectives were i) the fine mapping of Pl_{ARG} to identify closely linked molecular markers which can be used as starting point for map-based cloning, ii) the characterization of the introgression derived from the wild species *H. argophyllus*, iii) the identification of candidate genes for Pl_{ARG} using different approaches as outlined above, and iv) the evaluation of the different forward genetic approaches to identify candidate genes for Pl_{ARG}.

2 Material and methods

2.1 Plant material

2.1.1 Sunflower lines

In the present study, five resistant sunflower inbred lines were analyzed: ARG1575-2, RHA419, RHA420, RHA443, and 79ARGMTP. ARG1575-2 is a homozygous inbred line that carries the PI_{ARG} locus and mediates resistance to all known races of *P. halstedii* (Seiler et al. 1991). ARG1575-2 was derived by crossing *H. argophyllus* (accession 1575, PI 468651) with cmsHA89 followed by two generations of backcrossing with cmsHA89 and five selfing generations (Seiler et al. 1991). ARG1575-2 was crossed with the susceptible line HA342 to establish a segregating population for mapping of PI_{ARG}.

The homozygous inbred lines RHA419, RHA420, and RHA443 are all derived from ARG1575-2 and were chosen for comparison of marker alleles on LG 1 with the recurrent parents used in backcrossing like cmsHA89, and RHA373 for estimating the introgression size in each resistant line which originates from the wild species *H. argophyllus*. RHA419 and RHA420 were derived from the cross RHA373 × ARG1575-2 (Miller et al. 2002), whereas RHA443 is a $F_{6:7}$ restorer line selected from the cross RHA426/RHA419//RHA377/AS4379 (Miller et al. 2006). 79ARGMTP carries a downy mildew resistance gene on LG 1, which could also be PI_{ARG}, and was developed from the cross of *H. argophyllus* MPHE-92 with *H. annuus* FS20-6-2 at INRA, Montpellier (Vear et al. 2008b).

Three susceptible inbred lines, HA342, $NDBLOS_{sel}$ and KWS04 were analyzed in this study. HA342 was used as susceptible parent to establish a segregating population for mapping PI_{ARG}. The susceptible inbred line HA342 was available in the normal and in the PET1 CMS cytoplasm. HA342 is derived from a single BC_1F_4 plant from the cross HA89*2/Pervenets and has high oleic acid content (Miller et al. 1987). $NDBLOS_{sel}$ is an inbred line selected from the germplasm pool ND-BLOS (Roath et al. 1987) and was used for QTL mapping of *Sclerotinia* midstalk rot resistance (Micic et al. 2005). KWS04 is a proprietary line of KWS SAAT AG (Einbeck, Germany) and is used for the development of hybrid varieties. $NDBLOS_{sel}$ and KWS04 are genetically divergent and thus suitable to establish a segregating population.

2.1.2 Mapping populations

To fine map the target *PI*$_{ARG}$ region on LG 1, a segregating population from the cross of (cms)HA342 and ARG1575-2 was developed, which comprised two subpopulations totaling 2,145 F_2 individuals. The first subpopulation cmsHA342 × ARG1575-2 consisted of 1,065 F_2 individuals and was provided by Prof. W. Friedt (University of Giessen), whereas the second subpopulation HA342 × ARG1575-2 consisted of 1,080 F_2 individuals and was established by Dr. V. Hahn (University of Hohenheim). For convenience (cms)HA342 × ARG1575-2 refers to the whole population of 2,145 F_2 individuals. Suppressed recombination was observed for LG 1 of (cms)HA342 × ARG1575-2 and therefore markers from the target region could not be mapped with high resolution. Thus, the population NDBLOS$_{sel}$ × KWS04 comprising 2,780 F_2 individuals was developed to increase the mapping resolution of the targeted genomic region. NDBLOS$_{sel}$ and KWS04 are highly polymorphic but susceptible to *P. halstedii*; therefore NDBLOS$_{sel}$ × KWS04 did not segregate for *PI*$_{ARG}$.

2.2 Resistance tests

Phenotypic analysis of the (cms)HA342 × ARG1575-2 population is the basis for genetic mapping of the resistance locus *PI*$_{ARG}$. The resistance of F_2 plants was investigated by testing 16-40 F_3 seedlings per F_2 individual. The resistance tests were performed in the greenhouse using artificial inoculation after Gulya (1996). In each experiment the susceptible line HA342 and the resistant line ARG1575-2 were analyzed to control the success of the experiment. To characterize the resistance reaction of recombinants four races (100, 330, 710, and 730) were available which were provided by Prof. O. Spring (University of Hohenheim). The inoculum was prepared from infected plants which showed inoculation on cotyledons and leaves. The spores were washed with water from leaves and inoculum was adjusted to 40,000 spores / ml. Three day old seedlings were inoculated for three hours at 18°C. Afterwards the inoculum was decanted and seedlings were stored over night at 18°C. On the next day seedlings were planted into a mixture of sand and perlite (1:1) and were grown in a climate chamber for 13 days at 18°C, 16 h light and 8 h night. To initiate sporulation the plants were covered with a box for 24 h to reach humidity of 100%.

Phenotypic evaluation of the progenies was complicated by CLI observed in the resistant lines (Fig. 2-1). However, genotypes could be clearly distinguished, because a hypersensitive reaction (HR) could be observed on the cotyledons of resistant

genotypes which generally limits the spread of *P. halstedii*, whereas the inoculated susceptible genotypes showed a dense sporulation of *P. halstedii* on the cotyledons. Seedlings were considered either susceptible when high sporulation was evident on the cotyledons or resistant when no or only spurious sporulation accompanied by HR was observed on the cotyledons. Because of the occurrence of CLI in plants which carry the resistance gene PI_{ARG}, progenies with ambiguous phenotype were re-tested using 2-10 $F_{3:4}$ families. F_2 plants were classified as homozygous susceptible, homozygous resistant or heterozygous according to the phenotypes of F_3 or F_4 families.

Figure 2-1: Cotyledons of a susceptible seedling (left) and the resistant line ARG1575-2 (right), 14 days after inoculation with *P. halstedii*. On susceptible genotypes a dense sporulation of *P. halstedii* was observed on the cotyledons without showing HR reactions. Cotyledon limited infection (CLI) was observed on the resistant line, but sporulation was restricted by a hypersensitive reaction (HR).

Race 730 was used for phenotyping to determine the position of PI_{ARG} on LG1 and to identify informative F_2 individuals. Initially, resistance to downy mildew was evaluated in a subset of unselected 183 $F_{2:3}$ families of the subpopulation cmsHA342 × ARG1575-2. Informative lines with a recombination between the closely flanking markers were additionally analyzed with races 100, 330, and 710. Twenty-five homozygous F_4 seedlings for each F_2 individual with sufficient seeds available were evaluated with each of the *P. halstedii* races (Fig. 2-3).

For cDNA-AFLP and expression analyses seeds of HA342 and ARG1575-2 were germinated on filter paper for three days at 24°C. Half of the seedlings were inoculated for 3 h at 18°C in tap water and half of them were inoculated in a suspension of *P. halstedii*, race 730. The next day the seedlings were planted in a mixture of sand

and perlite (1:1). Cotyledons and hypocotyls of each seedling were harvested separately 1.5, 3, 6, 9, 12, and 15 days after inoculation for cDNA-AFLP analysis and 0, 0.5, 1.5, 3, 6, 9, 12, and 15 days after inoculation for RT-PCR. After harvest, the plant material was immediately frozen in liquid nitrogen and stored at -80°C.

2.3 Molecular analyses

2.3.1 Nucleic acid isolation

Genomic DNA

For genetic analyses DNA was extracted from all lines and from 2,145 F_2 plants of (cms)HA342 × ARG1575-2 and 2,780 F_2 individuals of $NDBLOS_{sel}$ × KWS04. DNA was isolated from dried or fresh leaf material of each F_2 plant or inbred lines using the CTAB extraction method (Doyle and Doyle 1990). Dried leaf material was ground to a fine powder with the TissueLyser II (Qiagen, Hilden, Germany) and fresh leaf material was pestled in liquid nitrogen. DNA pellets were resolved in TE buffer and adjusted to a concentration of 10 ng/µl. Quantification and quality control of DNA was performed on 0.8 % agarose gels compared to lambda DNA (Fermentas, St. Leon-Rot, Germany) or lambda DNA/*Hind*III (Fermentas, St. Leon-Rot, Germany).

Plasmid DNA

Plasmid DNA was isolated following manufacturer's instructions with the High Pure Isolation Kit (Kit number: 11754785001, Roche, Mannheim, Germany). Quality and quantity of the plasmid DNA were checked on 1.0 % agarose gels compared to lambda DNA (Fermentas, St. Leon-Rot, Germany) or lambda DNA/*Hind*III (Fermentas, St. Leon-Rot, Germany).

BAC plasmid extraction and insert size determination

Positive BAC clones were identified by screening the HA_HBa sunflower BAC library with specific overgo probes. The BAC library, HA_HBa, was ordered by the group of S.J. Knapp (University of Georgia) and has been publicly available since October 2004 at the Clemson University Genomics Institute (CUGI, www.genome.clemson.edu). The BAC library was constructed from the *H. annuus* line HA383. DNA fragments resulting from a partial *Hind*III digestion were ligated into the pIndigoBAC536 vector (Luo et al. 2001). The BAC library comprised 202,752 clones with an average insert size of 125.5 kb, providing a 7.3 - 8.5x genome coverage.

To determine the insert size of the positive BAC clones each single BAC clone obtained from CUGI was inoculated in 10 ml LB medium (12.5 µg/ml chloramphenicol) and cultivated for 18 h at 37°C on a shaker. The bacterial culture was transferred to a 10 ml Falcon tube and centrifuged at 9,000 x g for 5 min at room temperature. The supernatant was discarded and the pellet was resuspended in 300 µl suspension buffer (Table 8-4) with 3 µl of 10 mg/ml RNase A (Qiagen, Hilden, Germany). After resuspending the pellet, the solution was transferred to a 1.5 ml tube and incubated for 5-10 min on ice. Afterwards 400 µl lysis buffer (Table 8-4) was added and incubated for 5 min. For neutralization 300 µl neutralizing buffer (Table 8-4) was used and incubated for 15 min. The tube was centrifuged for 5 min at 15,000 x g, 900 µl of the supernatant was transferred to a new tube and 900 µl isopropanol was added. After inverting, incubation at room temperature for 10 min and centrifugation for 10 min at 15,000 x g the supernatant was discarded. The pellet was washed in 1 ml 70 % ethanol for 30 min and after centrifugation for 5 min at 15,000 x g, the pellet was dried under the fume hood and dissolved in 60 µl *A. bidest.* H_2O.

The insert size of the BAC plasmids was determined by performing a *Not*I digestion (New England BioLabs, Ipswich, MA, USA). The BAC plasmid (12 µl) was digested with 5 U *Not*I for 4 h at 37°C. After adding 8 µl of 6x loading dye buffer the samples and the MidRangeII PFG Marker (New England BioLabs, Ipswich, MA, USA) were loaded on a 1% pulsed field-agarose gel (peqLab, Erlangen, Germany) and separated in 0.5x TBE by pulsed field gel electrophoresis in a CHEF-DR II Pulsed Field Electrophoresis System (BioRad, Munich, Germany) for 17 h with a cooler setting of 14°C, pumping setting 80, initial pulse time 5 sec, final pulse time 15 sec and 6 V/cm. The gel was stained in a 0.5 µg/ml ethidium bromide bath (Roth, Karlsruhe, Germany) and watered for 30 min before a picture was taken with the Herolab Geldocumentation (Herolab, Wiesloch, Germany).

RNA isolation

RNA isolation was performed for the cDNA-AFLP approach, for RT-PCR and to establish cDNA-bulks for sequencing. Sampling for cDNA-AFLPs and RT-PCR was already described under 2.2. For BSTA a susceptible bulk (BS) and a resistant bulk (BR) were established each containing 16 F_2 genotypes of population (cms)HA342 x ARG1575-2 which are homozygous in the target region, but carrying a recombination event above or below the target region (Tables 8-13, 8-14). Seeds were sterilized with DanKlorix (1:1) (Colgate Palmolive GmbH, Hamburg, Germany) for three minutes and were germinated on filter paper for three days at 24°C. The seedlings

were planted and grown in the greenhouse at 18°C (night 16°C) and 14 h light for eight days. Hypocotyls, cotyledons, and leaves were harvested separately.

Plant material for RNA extraction was immediately frozen in liquid nitrogen after harvesting and stored at -80°C. The material was ground in a pre-cooled mortar with liquid nitrogen under RNase-free conditions. RNA for cDNA-AFLP analysis was extracted with the RNeasy® Plant Mini Kit (Kit number: 74904, Qiagen, Hilden, Germany) following manufacturer's instructions. DNA was digested at 37°C for 30 min with 5U DNaseI (Fermentas, St. Leon-Rot, Germany). DNaseI was inactivated by adding 0.1 Vol. of 50 mM EDTA at 65°C for 10 min.

RNA extraction and DNA digestion for BSTA 454 pyrosequencing and RT-PCR were performed with the NucleoSpin® RNA Plant Kit (Kit number: 740949.50, Macherey-Nagel, Düren, Germany) according to the manufacturer's instruction.

For RNA quantification and quality control the absorption was measured at 260 nm and 280 nm in a 100 µl quartz cuvette with the SPECTRONIC GENESYSTM 10 BIO spectrometer (Thermo ELECTRON CORPORATION, Madison, WI, USA). RNA was separated on 1.2% 1x TAE agarose gels to check for genomic DNA contamination and RNA degradation.

2.3.2 cDNA synthesis for RT-PCR and 454 sequencing

cDNA for semi-quantitative RT-PCR

For expression analyses with RT-PCR first strand cDNA was synthesized from 0.6 µg total RNA using a oligo-dT primer and the RevertAid™ H Minus First Strand cDNA Synthesis Kit (Kit number: K1631, Fermentas, St. Leon-Rot, Germany) following manufacturer's instructions. The cDNA was diluted 1:50 with *A. bidest.* H_2O and used as a template in PCR reactions (see 2.3.3).

cDNA for 454 pyrosequencing

Total RNA (50 µg of each sample) of a resistant bulk (BR) and a susceptible bulk (BS) was sent to vertis Biotechnologie AG (Freising, Germany) for cDNA synthesis, normalization, size fractionating and adapter ligation.

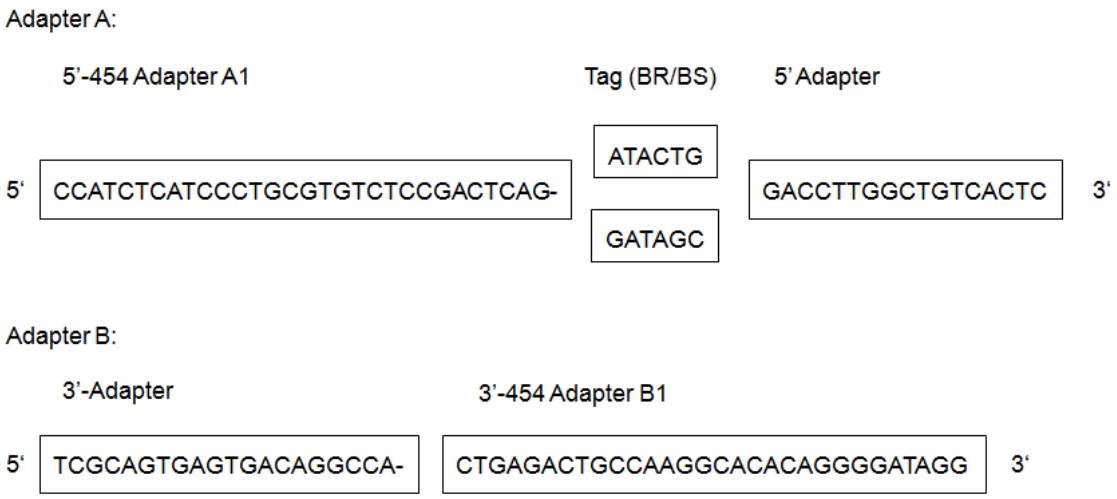


Figure 2-2: Sequences of adapters A and B containing the 5'- and 3'-454 adapter, a specific tag for the resistant bulk (BR) and the susceptible bulk (BS), and the cDNA synthesis primers' 5'-adapter and 3'-adapter. The first four bases of adapter primers A1 and B1 represent phosphorothioate-modified bases as specified by Roche. In addition, primer B1 is 5' biotinylated.

From total RNA poly(A)$^+$ RNA was isolated and first-strand cDNAs were synthesized with $(N)_6$ random primers. After ligation of adapter A which contains bulk specific tags and adapter B (see Fig. 2-2) to the 5' and 3' ends, cDNAs were amplified with 18 PCR cycles using a proof reading enzyme. The quality of the product was checked on a 1.5 % agarose gel. For normalization the amplified cDNAs were denaturated and reassociated. The reassociated ds-cDNAs were separated from the ss-cDNAs by passing the mixture over a hydroxylapatite column. The ss-cDNAs were amplified with 13 PCR cycles for the BR sample and 14 PCR cycles for the BS sample. The products were separated on a 1.5 % agarose gel and cDNAs in the size range of 600-800 bp were eluted from the agarose gel. The double stranded cDNA of BR and BS were pooled and 200 ng were sent for 454 sequencing to the Fritz Lipmann Institute (Jena, Germany).

2.3.3 Analyses of molecular markers

Polymerase chain reaction

PCR was performed with primers for SSR (Tang et al. 2003; Tang et al. 2002), SNP (Lai et al. 2005a), and RGC (Radwan et al. 2008) markers. For BAC contig mapping, primers were designed using the software Primer3 v0.4.0 (Rozen and Skaletsky 2000) or BatchPrimer3 (You et al. 2008). PCR was carried out in a final volume of 10 µl using

275 nM of each primer. In the presence of 1x Taq polymerase buffer, 2 - 2.5 mM $MgCl_2$, 0.2 mM of each dNTP, and 0.4 U Taq polymerase (Q-Biogene, Illkirch, Germany) the initial denaturation at 95°C for 5 min was followed by 35 cycles of 95 °C for 30 sec, 52-60°C (primer specific annealing temperature) for 30 s and 72 °C for 30-90 sec depending on the length of the fragment to amplify. For touch-down PCR, annealing temperature was reduced in the first 10 cycles in steps of 0.5 °C per cycle. PCR was completed with an extension step at 72 °C for 10 min.

Separation of PCR products

SNP markers were analyzed with the single strand conformation polymorphism (SSCP) method (Slabaugh et al. 1997) using a 0.5x SERDOGEL SSCP (SERVA, Heidelberg, Germany) and the electrophoresis chamber S2 (GibcoBRL, Karlsruhe, Germany). The gel was run with 2 W for 16 h and stained with a 0.25 % silver nitrate solution (Table 8-4) (Bassam et al. 1991). SNP markers which were converted into cleaved amplified polymorphic sequence (CAPS) markers were resolved with 3.0% agarose gels after restriction enzyme digestion. PCR products amplified with primers labeled with IRD700/IRD800 were analyzed on a LI-COR DNA-Analyzer 4300 (LI-COR Biosciences, Lincoln, NE, USA). PCR products labeled with the fluorescent dyes FAM, TET and HEX were separated on an ABI3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) together with a size standard (DeWoody 2004).

Cloning and sequencing of PCR products

In case of multiple PCR fragments isolation of single fragments was achieved via cloning. PCR products were cloned into the pCR®2.1-TOPO vector and transformed into chemically competent One Shot *E. coli* cells using the TOPO TA Cloning® Kit (Kit number: K4500-40, Invitrogen, Karlsbad, Germany) according to manufacturer's instructions. 20 µl, 40 µl and 100 µl of each sample were plated on LB agar plates containing 50 µg/ml ampicillin and 60 µg/ml X-gal. After overnight incubation at 37 °C white colonies were picked and grown overnight in 3 ml LB medium containing 50 µg/ml ampicillin on a shaker at 37 °C. Plasmid isolation was performed as described in 2.3.1. For each sample a long term culture storable at -80 °C was generated by adding 800 µl of 70% glycerin to 400 µl of the liquid culture.

PCR products or recombinant plasmids were sequenced with the ABI Prism BigDye® v1.1 Cycle Sequence Kit (Kit number: 4337450, Applied Biosystems, Foster City, CA, USA) following manufacturer's instructions but using only one quarter (2 µl) of the recommended volume (8 µl). Either PCR product specific forward and reverse primers

or M13 forward (5´ GTA AAA CGA CGG CCA G 3´) and M13 reverse (5´ CAG GAA ACA GCT ATG AC 3´) primers were used for the sequence PCR reaction. The PCR product was purified by isopropanol precipitation and diluted with 12 µl Hi-Di (Kit number: 4337450, Applied Biosystems, Foster City, CA, USA). Data analysis of samples was performed using Sequencing Analysis v5.3.1 (Applied Biosystems, Foster City, CA, USA) and Sequencher v4.8 (Gene Codes Corporation, Ann Arbor, MI, USA).

2.3.4 BAC clone and cDNA sequencing

BAC sequencing was performed at the Fritz Lipmann Institute (FLI), Jena, Germany. 29 BAC clones derived from the sunflower BAC library HA383 were prepared and barcoded as previously described by Steuernagel et al. (2009). Of the barcoded and pooled BACs a GS FLX Titanium shotgun library was prepared and sequenced on a half 70 x 75 picotiterplate of a Roche GS FLX sequencer.

Emulsion PCR and sequencing were performed for normalized random primed sunflower cDNA libraries (see 2.3.2) following the protocol from the GS FLX manual (Roche Diagnostics, Penzberg, Germany). The amplified mixture of cDNA libraries was sequenced on a full 70 x 75 picotiterplate using Titanium chemistry.

2.3.5 Copy number analysis

RGC151SW1 was identified as a candidate for Pl_{ARG} based on RGC analysis and confirmed by BAC analysis. To determine copy number variation of RGC151 between the resistant line ARG1575-2 and the recurrent parent cmsHA89 Southern blot analysis was performed at CUGI. DNA samples of ARG1575-2 and cmsHA89 were digested with the restriction enzymes *Eco*RI, *Hind*III, *Xba*I, *Hind*III/*Xba*I and separated by gel electrophoresis. The gel was blotted onto a Hybond N^+ filter. Afterwards the membrane was hybridized with a P^{32} labeled PCR product amplified with the primers RGC151 forward and RGC151 reverse using ARG1575-2 DNA as template.

2.4 Genetic mapping of the Pl_{ARG} locus

2.4.1 Linkage mapping analysis

To determine the correct marker order an anchor map of LG 1 for 475 F_2 individuals of cmsHA342 × ARG1575-2 was calculated using seven polymorphic codominant SSR markers of the public sunflower linkage map, which were identified by Dußle et al. (2004). To increase the density of markers in the target Pl_{ARG} region, 22 publicly

available markers from LG 1 were screened for polymorphisms between (cms)HA342, ARG1575-2, $NDBLOS_{sel}$, and KWS04 (Table 8-5). Thirteen SSR markers (Tang et al. 2003; Tang et al. 2002; Yu et al. 2003), six SNP markers (Lai et al. 2005a) and the three resistance gene candidate makers RGC151, RGC52a and RGC52b (Radwan et al. 2008) were analyzed using a LI-COR DNA-Analyzer 4300 (LI-COR Biosciences, Bad Homburg, Germany) or the SSCP method (Slabaugh et al. 1997). HT211 and RGC151 were converted into CAPS markers and resolved on 3.0% agarose gels after digesting the PCR fragments with *Taq*I and *Rsa*I, respectively.

Overall, 19 SSR, SNP, CAPS and RGC markers were used for linkage analysis of (cms)HA342 × ARG1575-2, and ten SSR, SNP and CAPS markers were used to construct the linkage map of $NDBLOS_{sel}$ × KWS04.

To investigate the portion of *H. argophyllus* genome in the resistant inbred line ARG1575-2 marker scores of ARG1575-2 and cmsHA89 were compared with 14 codominant SSR and two CAPS markers of LG 1, and with 94 randomly distributed SSR markers covering the remainder of the genome (LG 2 – LG 17). The resistant lines RHA419, RHA420, RHA443 and 79ARGMTP were included in the comparison to estimate the introgression originating from *H. argophyllus* on LG 1.

A chi-square goodness-of-fit test was performed for the null hypothesis to observe a segregation ratio of 1:2:1 for the polymorphic markers under study in the mapping populations. Maps for the populations (cms)HA342 × ARG1575-2 and $NDBLOS_{sel}$ × KWS04 were constructed with JOINMAP 4.0 (Van Ooijen 2006) using a LOD threshold of >3.0 and the Kosambi mapping function (Kosambi 1944).

2.4.2 Selection of recombinant lines

The strategy outlined in Fig. 2-3 was used to identify recombination events in the region of PI_{ARG}. Only recombinant lines were phenotypically tested for fine mapping. In total, 2,145 F_2 individuals of (cms)HA342 × ARG1575-2 were genotyped with ORS610 and ORS371, which flank PI_{ARG}, and with ORS662, which cosegregated with PI_{ARG}. Additional 16 polymorphic markers were analyzed in 188 recombinant F_2 individuals to increase the marker density in the target region. $F_{2:3}$ or $F_{2:4}$ progenies of 108 recombinant F_2 individuals produced sufficient seeds and were subsequently phenotyped to confirm the genomic location of PI_{ARG}. Next, ORS509, HT244 and HT446 were used to narrow down the interval harboring PI_{ARG}. Recombinant lines were selfed and homozygous progenies were phenotyped with the *P. halstedii* races 100, 330, 710, and 730.

A map was constructed with JOINMAP 4.0 (Van Ooijen 2006) using a LOD threshold of >3.0 and the Kosambi mapping function (Kosambi 1944). Only recombinant lines of (cms)HA342 × ARG1575-2 were analyzed as described above. For all other F_2 individuals values were imputed according to the marker results of ORS610, ORS662 and ORS371.

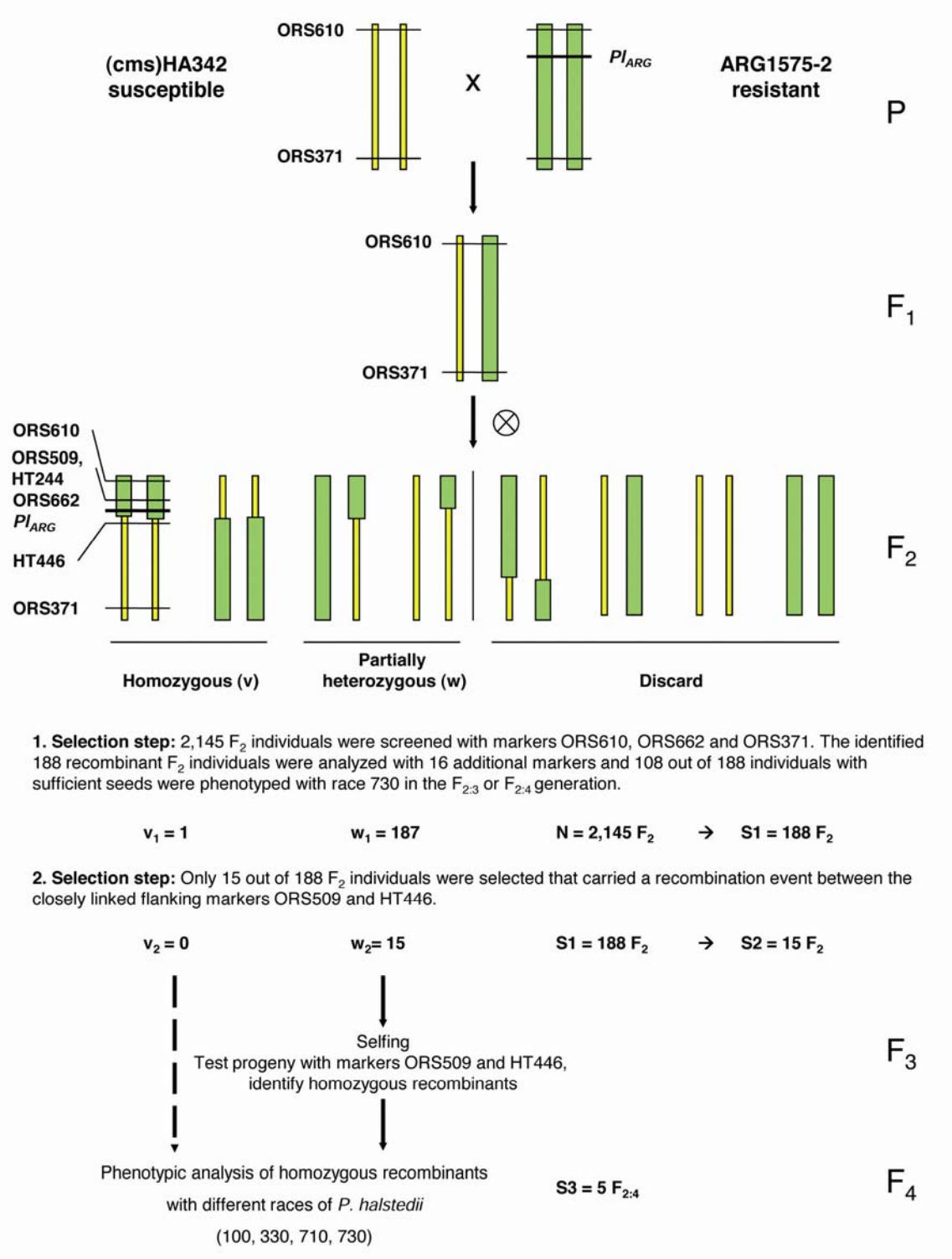


Figure 2-3: High resolution mapping strategy for the identification of informative recombinant lines in cross (cms)HA342 × ARG1575-2. The F_2 population consisted of 2,145 lines and was screened in two steps. The whole population (N) was screened with markers ORS610, ORS662 and ORS371 and in selection step one (S1) 188 recombinant lines were selected and, subsequently, genotyped and phenotyped. Lines with recombination events between the closely linked flanking markers ORS509, HT244 and HT446 were identified in the second selection step (S2) and were selfed and homozygous recombinant lines with sufficient seeds (S3) were tested with the four *P. halstedii* races 730, 100, 330, and 710. Numbers of homozygous recombinant plants are indicated under v and numbers of heterozygous plants are indicated under w.

2.4.3 BAC library screening and mapping of BAC end sequences

Overgo probes (Han et al. 2000; Zheng et al. 2006) were designed from the sequences cosegregating with PI_{ARG} (RGC151, ORS662, and HT211) (Table 2-1) to screen the HA_HBa BAC library. Positive BAC clones were picked, propagated and confirmed by colony PCR. Plasmid DNA was used for High Information Content Fingerprinting (HICF) (Ding et al. 1999; Luo et al. 2003) and BAC end sequencing. BACs were assembled into contigs using the fingerprinted contigs (FPC) software v8.5.3 (Soderlund et al. 1997). Inserts of positive BAC clones were sequenced from both ends using the T7 (5' TTG TAA TAC GAC TCA CTA TAG G 3') and M13 (5' TCA CAC AGG AAA CAG CTA TGA C 3') sequencing primers to generate BAC end sequences (BES). Positive BAC clones of RGC52 identified in a BAC screening by Wenxiang Gao at University of Georgia were delivered by Clemson University. BES of these RGC52 clones were publicly available at NCBI. Primers were designed on BES with Primer3 (Rozen and Skaletsky 2000) (Table 8-6). To map the assembled BAC contigs in both populations (cms)HA342 × ARG1575-2 and $NDBLOS_{sel}$ × KWS04, polymorphic markers between the parents were analyzed with the SSCP method or on 1.5% agarose gels, respectively.

Table 2-1: Markers and sequences of overgo probes selected for the BAC library screen. Overlap of forward and reverse primer sequences is highlighted in grey.

Marker	Overgo sequence
HT211-F (5'-3')	GGGAAGTAAAGCCTGATGTACTCC
HT211-R (3'-5')	ACATGAGGAACCCAACAGACGTCA
RGC151-F (5'-3')	TTTGGTTGGGTTTCGGGAATGT
RGC151-R (3'-5')	CCCTTACACAAGGAGTAGAGCA
ORS662-F (5'-3')	AAAAACCGGGTATGGCAATAACCG
ORS662-R (3'-5')	TTATTGGCCCAACCTATACCTCAG

2.5 Sequence analysis of BAC clones

2.5.1 Assembly of BAC 454 sequences

Vector sequences and reads derived from *E. coli* DNA were removed from the sequence data set. The remaining sequences were assembled using MIRA [http://chevreux.org/projects_mira.html]. Assemblies were performed separately for each of the 29 sequenced BACs as well as contig-wise using the reads of all BAC clones that belong to the same HICF contig. Assembly parameters (bases per hash: bph, hash saving steps: hss) were iteratively changed as described by Steuernagel et

al. (2009). For further analyses, the assembly producing the longest HICF subcontig was used.

2.5.2 Gene annotation of BAC 454 sequences

The contigs obtained from the HICF contig-wise assembly were screened for coding sequences using a custom made analysis pipeline (KFX Mayer, personal communication) which combines several public tools: FGENESH (http://softberry.com/), GeneID (http://genome.crg.es/geneid.html), PASA (http://pasa.sourceforge.net), and GenomeThreader (http://www.genomethreader.org/). The following data sets and program settings were used to fit the pipeline requirements using a parameter set for dicotyl plants:

- 33,410 protein sequences from the *Arabidopsis* information resource (TAIR9, http://www.arabidopsis.org/) were used for annotation
- EST sequences of *Helianthus* species including *H. annuus* (133,682), *H. argophyllus* (35,720), *H. celiaris* (21,590), *H. exilis* (33,961), *H. paradoxus* (30,517), *H. petiolaris* (27,484), and *H. tuberosus* (40,387) sequences (NCBI, ENTREZ, http://www.ncbi.nlm.nih.gov) were applied to improve the gene models.
- ESTs of dicotyledonous flowering plants including sequences of *Lactuca sativa* (33,115), *Arabidopsis thaliana* (148,368), *Glycine max* (116,965), *Trifolium pratense* (13,644), *Medicago truncatula* (55,182), *Malus domestica* (77,604), *Solanum lycopersicum* (53,791), and *Populus trichocarpa* (31,082) (the TIGR Plant Transcript Assemblies, http://plantta.jcvi.org/cgi-bin/plantta_release.pl)

Two different analyses were performed, one where the gene was predicted with a minimum of interruptions by gaps (shortest path) and one trying to cover the whole gene model with the highest possible similarity (longest path).

Repeats and transposable elements were not masked before gene prediction was performed, because the repeat database might contain pseudogenes which could cause masking of functional genes. To identify predicted genes with a high similarity to transposable elements and to adjust the function of “true” genes the protein sequences of the predicted gene models were used as query sequences to search the NCBI nonredundant protein database (http://blast.ncbi.nlm.nih.gov) by blastp. Based on the blast results, all hits with a maximum expect value (e) equal or smaller than 1.00 e^{-10} were manually evaluated to exclude transposable elements from gene models.

The sequences carrying resistance gene candidates based on the annotation pipeline results were also analyzed with the gene prediction software AUGUSTUS (http://augustus.gobics.de) and GeneSeqer (http://bioservices.usd.edu/geneseqer/) to identify alternative gene models. The multiple-sequence alignment of protein sequences of predicted gene models along with 93 amino acid sequences of known resistance genes (reference data set, Table 8-12) was built using ClustalW (Thompson et al. 1994) to identify the gene models of the candidate genes which fit best to known resistance gene sequences. Based on the results of tblastn analysis against the reference data set manual correction of predicted gene models was performed using GENEDOC (multiple-sequence alignment editor, http://www.nrbsc.org/gfx/genedoc/). Blastp analysis against the Domain DB of NCBI was performed to identify the different domains of the candidate genes. The alignment was also visually inspected for protein motifs (Meyers et al. 1999) and highlighted in GeneDoc.

To analyze the sequence relationship of newly identified candidate genes to known sunflower RGC genes (Radwan et al. 2008) and cloned resistance genes from different dicotyledonous plant species the corresponding NBS domain regions (P-loop to GLPL motif) were aligned with ClustalW and a neighbor-joining tree was generated by bootstrap resampling with 1,000 replicates using MEGA4 (Kumar et al. 2008).

2.6 Transcriptome analysis

2.6.1 cDNA-AFLP

The cDNA-AFLP protocol of Vuylsteke et al. (2007) was followed to identify differentially expressed sequences from inoculated and non-inoculated plants of the susceptible line HA342 and the resistant line ARG1575-2. Double-stranded cDNA was synthesized with biotinylated oligo-$(dT)_{25}$ primer (Biomers, Ulm, Germany) out of 2 µg total RNA and was purified applying the Nucleospin Extraction II Kit (Kit number: 740609.50, Macherey-Nagel, Düren, Germany). Firstly, cDNA was digested with 10 U *Bst*YI (New England Biolabs, Ipswich, MA, USA) and after immobilizing 3'-terminal cDNA fragments on streptavidin-coated Dynabeads (Invitrogen, Karlsbad, Germany) and five washing steps with 1x STEX, a second digestion was performed with 10 U *Mse*I (New England Biolabs, Ipswich, MA, USA). Dynabeads were discarded, adapters were ligated to the fragments of the supernatant, and pre-amplification was carried out using a PCR program with 25 cycles according to the cDNA-AFLP protocol and the restriction site specific primers. The products were separated on a 1 % agarose gel and diluted depending on the concentration 200–900 fold with TE buffer. A selective

amplification was performed according to the protocol using fluorescently labeled primers (IRD700/IRD800) with one selective base. 0.75 µl of the PCR product and 1 µl of formamide loading dye (DeWoody 2004) were mixed, denatured at 90 °C for 3 min, and separated on a 6 % denaturing polyacrylamide gel (SequaGel®XR, National Diagnostics, Atlanta, USA) using the LI-COR DNA-Analyzer 4300 (LI-COR Biosciences, Lincoln, NE, USA). After identifying differentially expressed fragments a selective PCR with a second selective base at the 3' end of the unlabeled primer was carried out to reduce the number of fragments and simplify the isolation of the desired fragment.

To clone the differentially expressed fragments, products of the second selective PCR were separated using the electrophoresis chamber S2 (GibcoBRL, Karlsruhe, Germany) and a 6 % denaturing polyacrylamide gel (SequaGel®XR, National Diagnostics, Atlanta, USA). The gel was run with 62 W for 60-90 min depending on the size of the fragment which had been selected for isolation. Afterwards the gel was stained with silver nitrate according to the following protocol for denaturing polyacrylamide gels: 20 min fixation with the Fix/Stop solution (Table 8-4), 3x washing with *A. bidest.* H_2O for 5 min, 30 min staining with the silver nitrate solution (1 mg/ml) (Table 8-4), 10 sec washing with *A. bidest.* H_2O. Developer solution (Table 8-4) at 6 °C was used to stain the gel to the desired intensity. Afterwards the process was stopped by adding the Fix/Stop solution for several minutes and the gel was washed again two times with *A. bidest.* H_2O. The differentially expressed fragments were cut from the gel with a scalpel and were incubated for one hour in 100 µl TE buffer. With 5 µl of this mixture re-amplification was performed with unlabeled selective primers. The size and the quality of the PCR products were checked on a 1.5 % agarose gel and the PCR products were cleaned with the Nucleo®Spin Extract II Kit (Kit number: 740609.50, Macherey-Nagel, Düren), cloned, and sequenced as described in 2.3.3. For selected sequences primers were designed with Primer3 v0.4.0 (Rozen and Skaletsky 2000) and these were tested by BSA to analyze if the isolated differentially expressed fragments map to the target region.

2.6.2 Assembly of bulked segregant transcriptome analysis 454 sequences

After 454 sequencing of cDNAs from the phenotypic bulks, barcode and cDNA synthesis primers were trimmed from the raw sequences. Three assemblies were performed with the Roche 454 program Newbler v2.0.01.12-64 choosing the default parameters for the sequences of the resistant bulk (BR), the susceptible bulk (BS), and

for all reads of bulk R and bulk S (BRBS) combined.

2.6.3 Identification of resistance gene candidates in the cDNA assemblies

Automated SNP detection without a reference genome is not reliable. Therefore, two filter strategies were performed to identify transcripts involved in disease resistance and to analyze them afterwards for SNPs differentiating for the susceptible and the resistant bulk. First, with EMBOS getorf (http://www.molgen.mpg.de/~beck/EMBOSS/getorf.html) open reading frames were identified with a minimum length of 90 nt within sequences of the BRBS assembly. The translated protein sequences were used for domain search with PFAM (http://pfam.sanger.ac.uk/) to identify resistance gene candidates. Second, a reference data set of 93 plant resistance genes obtained from the Plant Resistance Genes database (http://prgdb.cbm.fvg.it/index.php, Table 8-12) was used for tblastn analysis to identify the best hits from the BRBS assembly (e-value cutoff 1.00 e^{-10}).

Contigs showing homology to PFAM domains typical for resistance genes were compared to the set of contigs identified by tblastn analysis against the reference data set. Only contigs which had been identified with both methods were evaluated in more detail. The presence of putative SNPs was checked by visual inspection of raw assembly data using the Tablet software (Milne et al. 2010).

The sequenced BAC clones from the PI_{ARG} interval were used for proof of concept of the BSTA approach. cDNA sequences matching to the predicted genes on the HICF BAC contigs were identified by blastn analysis (e-value cutoff 1.00 e^{-100}). The BRBS assembly data of these sequences were visually inspected for putative SNPs using the Tablet software (Milne et al. 2010).

Validation of candidate genes was performed by re-amplification, re-sequencing and genetic mapping. Primer pairs were designed on the BRBS assembly using BatchPrimer3 (You et al. 2008) (Table 8-7). Amplicons were generated and sequenced with standard protocols (2.3.3). Only recombinant lines of (cms)HA342 × ARG1575-2 were analyzed. For all other F_2 individuals values were imputed. A map was constructed with JOINMAP 4.0 (Van Ooijen 2006) using a LOD threshold of >3.0 and the Kosambi mapping function (Kosambi 1944).

3 Results

3.1 Localization of *PI_{ARG}*

3.1.1 *PI_{ARG}* on linkage group 1

In order to determine the correct marker order on LG 1 seven polymorphic SSR markers were mapped in 475 F_2 individuals of cmsHA342 × ARG1575-2. ORS1182, ORS610, ORS1128, and ORS543 cosegregated at the upper end of LG 1, whereas ORS662, ORS959, and ORS371 mapped 0.5 cM, 3.6 cM, and 5.2 cM from the upper end, respectively. $F_{2:3}$ or $F_{2:4}$ families of 183 F_2 individuals that had produced sufficient seeds were evaluated for resistance to downy mildew. A segregation of 25 homozygous resistant, 114 segregating, and 44 homozygous susceptible progenies was observed, differing significantly from the expected 1:2:1 Mendelian segregation ratio of *PI_{ARG}* (X^2 = 15.01, DF = 2, $p \leq 0.0006$). *PI_{ARG}* cosegregated with ORS662.

3.1.2 Fine mapping of the *PI_{ARG}* target region

Population (cms)HA342 × ARG1575-2

To increase the genetic resolution around *PI_{ARG}*, 2,145 F_2 individuals of (cms)HA342 × ARG1575-2 were genotyped as described in Fig. 2-3. ORS610 and ORS371 mapped at the upper and lower end of LG 1, respectively and flank *PI_{ARG}*, which cosegregates with ORS662. The three markers were used to screen 2,145 F_2 individuals. All markers showed significant segregation distortion towards the susceptible line in the subpopulation cmsHA342 × ARG1575-2, but did not deviate from the expected 1:2:1 segregation ratio in HA342 × ARG1575-2 (Table 3-1).

In total, 188 F_2 individuals were identified that carried a recombination event between the two flanking markers ORS610 and ORS371. The F_2 individuals were phenotyped and genotyped with 16 polymorphic markers between ORS610 and ORS371. To confirm the map position of *PI_{ARG}* $F_{2:3}$ or $F_{2:4}$ progenies of 108 out of the 188 recombinant F_2 individuals were tested (Fig. 2-3). The remaining 80 recombinant F_2 individuals could not be tested due to the lack of seeds for resistance tests especially for the subpopulation cmsHA342 × ARG1575-2.

Table 3-1: Segregation ratios and X^2-values of the codominant SSR markers ORS610, ORS662, ORS053, and ORS371 analyzed in three sunflower populations.

	Population		
	cmsHA342×ARG1575-2 (N=1,065)	HA342×ARG1575-2 (N=1,080)	$NDBLOS_{sel}$×KWS04 (N=528)
ORS610	272 : 587 : 206[a]	272 : 561 : 247	143 : 253 : 131
$X^2_{DF=2}$	19.34 (*p*=0.0001)	2.79 (*p*=0.2478)	1.38 (*p*=0.5016)
ORS662	271 : 585 : 209	270 : 561 : 249	131 : 269 : 128
$X^2_{DF=2}$	17.57 (*p*=0.0002)	2.45 (*p*=0.2938)	0.22 (*p*=0.8958)
ORS053	rec. F_2.[b]	263 : 569 : 248	130 : 266 : 132
$X^2_{DF=2}$		3.53 (*p*=0.1712)	0.05 (*p*=0.9753)
ORS371	272 : 582 : 204	268 : 559 : 244	n.a.
$X^2_{DF=2}$	19.36 (*p*=0.0001)	3.13 (*p*=0.2091)	

[a] Numbers are given for the marker genotypes (homozygous parent 1 allele : heterozygous : homozygous parent 2 allele).
[b] rec. F_2: only recombinant F_2 individuals were analyzed
n.a.: not analyzed

The genetic map of LG 1 based on 2,145 F_2 individuals of (cms)HA342 × ARG1575-2 spanned 4.2 cM (Fig. 3-1a). ORS610, ORS543, ORS1128, CRT272, and ORS1182 cosegregated at the upper end of LG 1, whereas PI_{ARG} cosegregated with ORS662, HT211, RGC52a, RGC52b, RGC151, HT722, and ORS716. ORS509, and HT244 on the one side and HT446 on the other side flanked PI_{ARG} at a distance of 0.2 cM and 0.1 cM, respectively. More recombination events were observed below PI_{ARG}. ORS053, ORS959, and ORS371 mapped 1.9 cM, 2.3 cM, and 3.9 cM from PI_{ARG}, respectively. Thus, closely linked markers on both sides of PI_{ARG} were identified for fine mapping of the PI_{ARG} genomic region.

Of the 188 F_2 individuals selected initially (Fig. 2-3), only 15 F_2 individuals were selected in the second step and revealed a recombination event between the closely linked flanking markers ORS509, HT244, and HT446. Five of the 15 recombinant lines produced sufficient seed to develop homozygous recombinant progenies (Fig. 3-2). Four homozygous lines were resistant to *P. halstedii* race 730. One homozygous line was susceptible to *P. halstedii* race 730. These lines were tested with three additional races of *P. halstedii* (100, 330, 710) and showed the same resistance pattern for all four races.

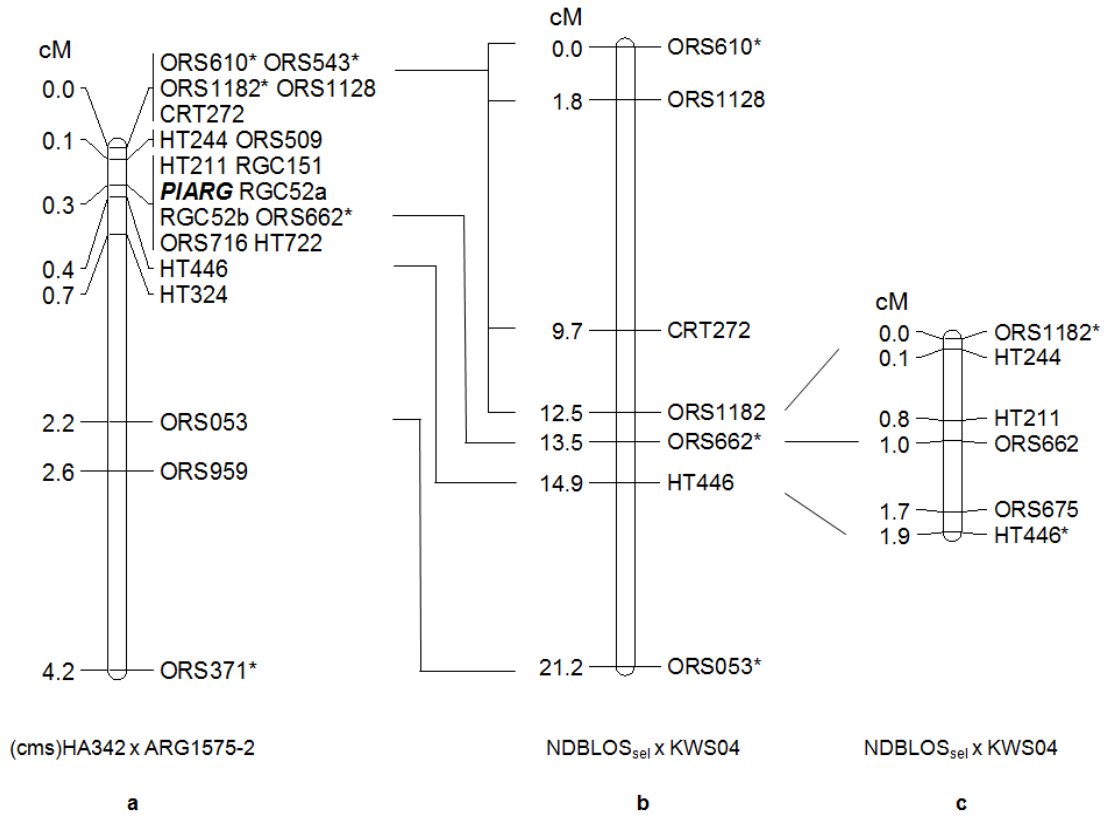


Figure 3-1: (a) Linkage group (LG) 1 map of (cms)HA342 × ARG1575-2 constructed with 2,145 F_2 individuals. (b) Partial LG 1 map of NDBLOS$_{sel}$ × KWS04 constructed with 528 F_2 individuals. (c) Partial LG 1 map of NDBLOS$_{sel}$ × KWS04 constructed with 2,780 F_2 individuals. Markers with an asterisk were screened in all individuals, while all other markers were screened in recombinant lines only. *PI_{ARG}* is shown in bold. Maps are not drawn to scale and distances on the left side of the map correspond to centiMorgan. Lines connect common markers.

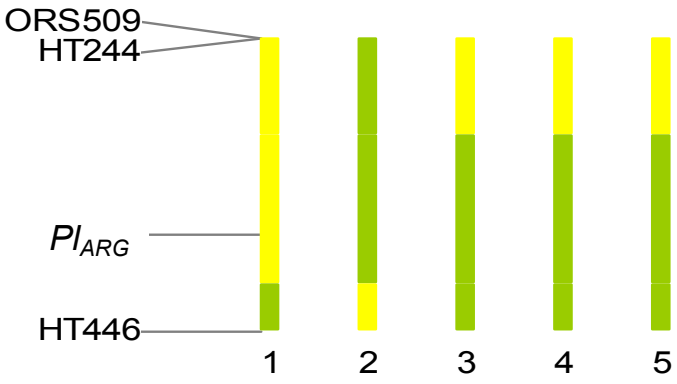


Figure 3-2: Graphical genotype of the five recombinant lines (1-5) selected in (cms)HA342 x ARG1575-2 which were tested with the four *P. halstedii* races 100, 330, 710, and 730. HT509, HT244 and HT446 are SSR markers closely linked to *PI_{ARG}*. Alleles in yellow originate from HA342 and in green from ARG1575-2.

Population NDBLOS$_{sel}$ × KWS04

Suppressed recombination was evident in (cms)HA342 × ARG1575-2 compared with the linkage map of Tang et al. (2002, 2003), which is likely due to the introgression of parts of the *H. argophyllus* genome. To obtain a higher mapping resolution in the target region and to estimate the degree of suppression, the intraspecific cross NDBLOS$_{sel}$ × KWS04 of 2,780 F_2 individuals was developed. Initially, 528 F_2 individuals were screened with ORS610, ORS662, and ORS053 from LG 1 (Table 3-1). For 198 F_2 individuals a recombination event was observed between ORS610 and ORS053. These individuals were screened with four additional polymorphic SSR markers resulting in a map based on 528 F_2 individuals (Fig. 3-1b) that spanned 21.2 cM between ORS610 and ORS053. ORS610, ORS1128, CRT272, and ORS1182 which cosegregate on LG 1 of (cms)HA342 x ARG1575-2 were mapped within 12.5 cM on LG 1 of NDBLOS$_{sel}$ × KWS04. Based on this map ORS1182 is the closest flanking SSR marker above the PI_{ARG} region in the population NDBLOS$_{sel}$ × KWS04.

Additional 2,252 F_2 individuals of NDBLOS$_{sel}$ × KWS04 were screened with ORS1182 and HT446 to identify recombinants in this high-resolution mapping population. Altogether, 99 recombinant F_2 individuals were identified in the target region and were screened with HT244, HT211, ORS662, and ORS675. HT211 and ORS662, which cosegregated with PI_{ARG} in (cms)HA342 × ARG1575-2, were mapped with higher resolution in NDBLOS$_{sel}$ × KWS04 at a distance of 0.2 cM (Fig. 3-1c). HT244 and HT446 are the flanking markers of PI_{ARG} in (cms)HA342 x ARG1575-2. In NDBLOS$_{sel}$ × KWS04 a genetic distance of 1.8 cM was observed.

3.1.3 Linkage group 1 originates from *H. argophyllus*

To investigate the reason for suppressed recombination of (cms)HA342 × ARG1575-2, the alleles of ARG1575-2 and of the recurrent parent cmsHA89 were compared at 14 SSR and two CAPS marker loci of LG 1. The resistant line ARG1575-2 is expected to carry only 12.5% of the *H. argophyllus* genome after two generations of backcrossing with cmsHA89. It was found that ARG1575-2 and cmsHA89 carried different alleles at all markers tested on LG 1 with the exception of HT324 (Table 3-2). It could not be investigated whether the HT324 allele of ARG1575-2 originated from *H. argophyllus* or from cmsHA89 because the original *H. argophyllus* accession 1575 is no longer available. It can be assumed that accession 1575 and cmsHA89 are monomorphic at this locus. Analysis of 94 SSR markers randomly distributed on LG 2 to 17 revealed that 12.7 % of them had common alleles between ARG1575-2 and

cmsHA89 (Fig. 3-3). To estimate the size of the introgressed segment on LG 1 in other resistant inbred lines carrying PI_{ARG}, the marker alleles of ARG1575-2 and six inbred lines were compared. The haplotypes of RHA419, RHA443, and RHA420 originate mostly from ARG1575-2 (Table 3-2). A recombination event occurred in RHA419 and RHA443 below the PI_{ARG} locus in the interval ORS053 - ORS959. Comparison with 79ARGMTP which probably also carries PI_{ARG}, and the susceptible recurrent parent FS20-6-2 which was used for backcrossing showed also no evidence of recombination on LG 1 (Table 3-2).

Table 3-2 Shared haplotypes between the recurrent parent cmsHA89 and ARG1575-2, RHA419, RHA420, RHA443 and between the recurrent parent FS20-6-2 and 79ARGMTP for markers on LG 1. Marker position in cM corresponds to the map from (cms)HA342 × ARG1575-2 (2,145 F_2 individuals). The symbols (+) were used for markers that have the same allele as ARG1575-2 (dark grey), (x) for markers that have the same allele as 79ARGMTP (light grey) and (-, %) for marker alleles that differ from line ARG1575-2 or 79ARGMTP (white).

		Sunflower line						
Marker	cM	cmsHA89	ARG1575-2	RHA419	RHA420	RHA443	FS20-6-2	79ARGMTP
ORS1128	0.0	-	+	+	+	+	%	x
ORS543	0.0	-	+	+	+	+	%	x
ORS610	0.0	-	+	+	+	+	%	x
ORS1182	0.0	-	+	+	+	+	x	x
CRT272	0.0	-	+	+	+	+	%	x
ORS509	0.1	-	+	+	+	+	%	x
ORS716	0.3	-	+	+	+	+	%	x
ORS662	0.3	-	+	+	+	+	%	x
HT211	0.3	-	+	+	+	n.d.	%	x
RGC151	0.3	-	+	+	+	n.d.	%	x
HT722	0.3	-	+	+	+	+	%	x
HT446	0.4	-	+	+	+	+	%	x
HT324	0.7	+	+	+	+	+	%	x
ORS053	2.2	-	+	+	+	+	%	x
ORS959	2.6	-	+	-	+	-	%	x
ORS371	4.2	-	+	-	+	-	%	x

n.d.: not determined

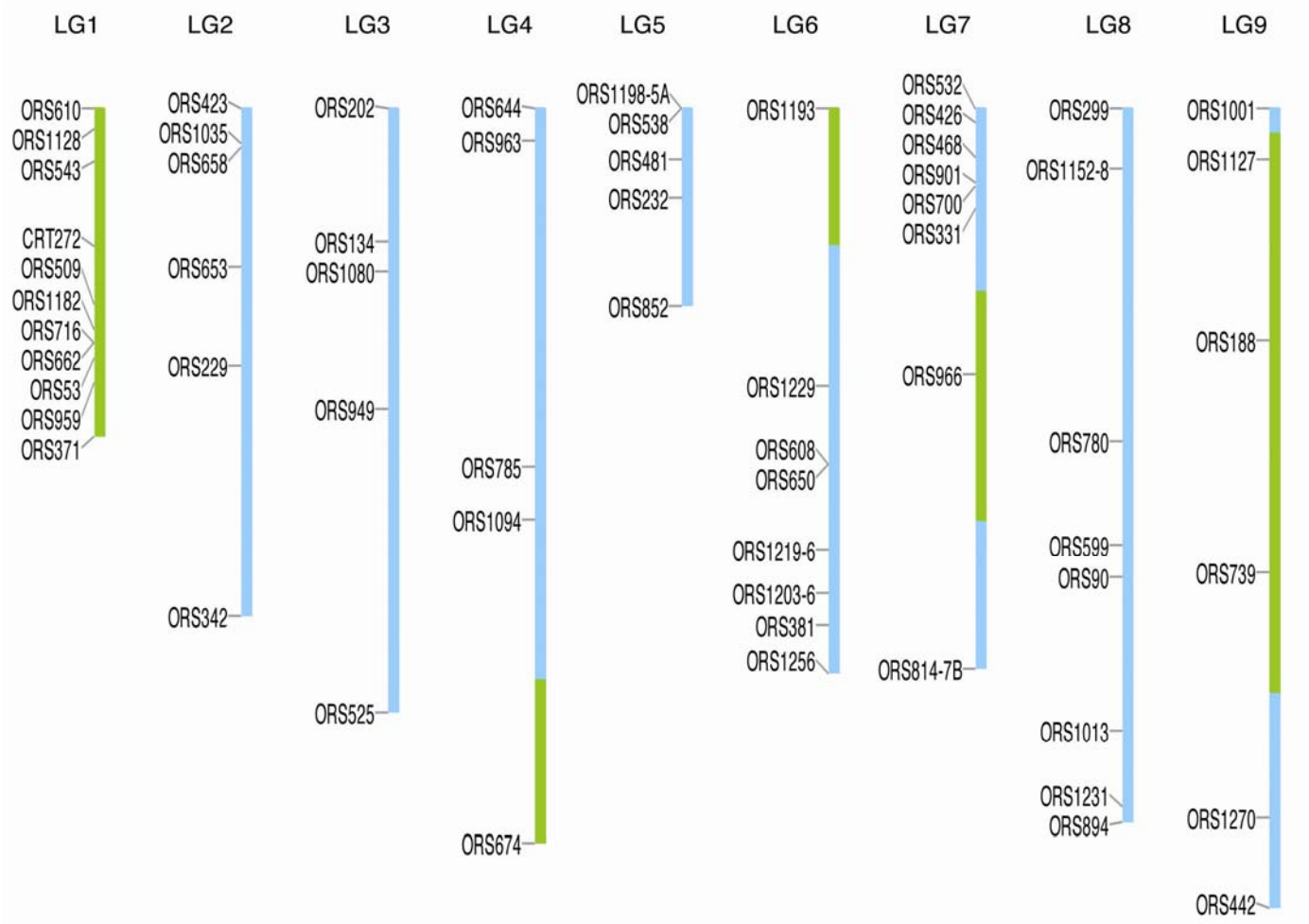


Figure 3-3: Graphical genotype of ARG1575-2 showing distribution of alleles originating from *H. argophyllus* and cmsHA89. Marker distances are drawn according to the map of Tang et al. (2003). Markers with alleles different from HA89 are labeled in green and originate from *H. argophyllus*. Markers with alleles like HA89 are labeled in blue. Numbers on the top of the maps indicate the linkage groups.

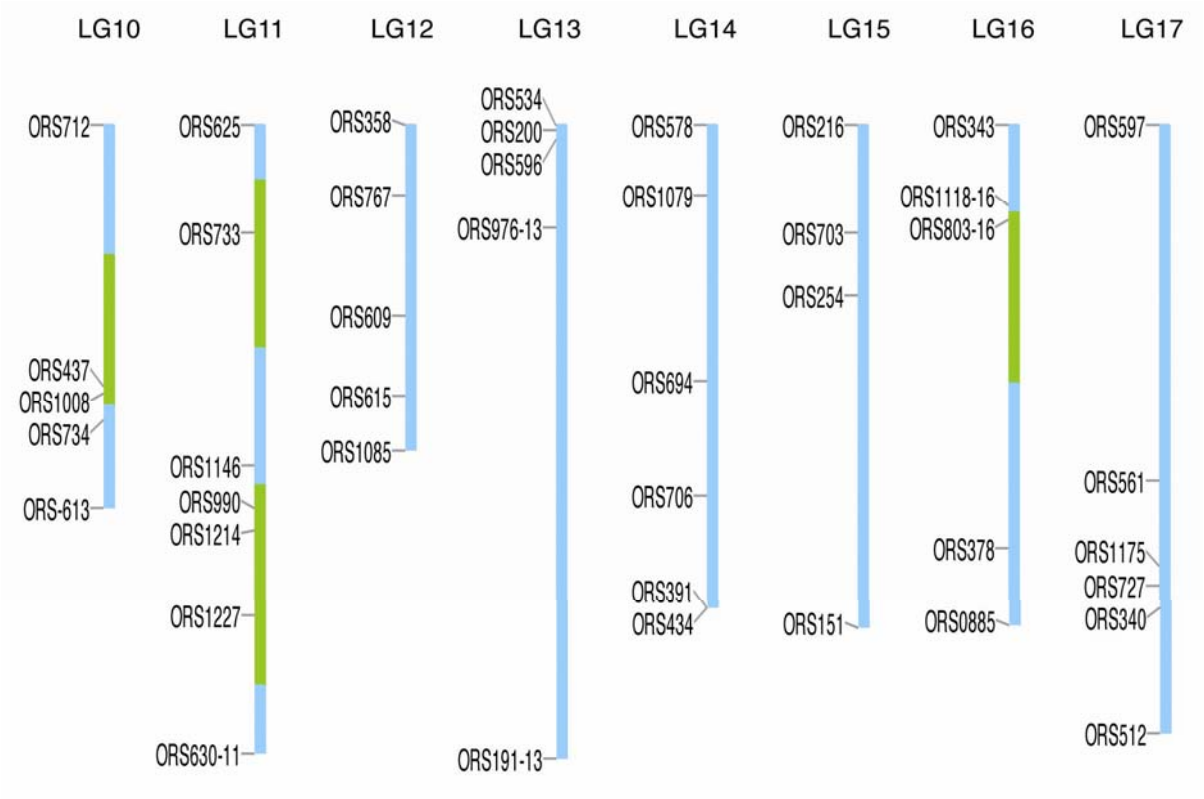


Figure 3-3 (continued): Graphical genotype of ARG1575-2 showing distribution of alleles originating from *H. argophyllus* and cmsHA89. Marker distances are drawn according to the map of Tang et al. (2003). Markers with alleles different from HA89 are labeled in green and originate from *H. argophyllus*. Markers with alleles like HA89 are labeled in blue. Numbers on the top of the maps indicate the linkage groups.

3.2 Characterization of the target region

3.2.1 Identification of BAC clones

To establish sequence information for the PI_{ARG} region the large-insert sunflower BAC library HA383 was screened with three overgo probes designed from markers from the target region (RGC151, HT211, ORS662). In total, 30 BACs were identified and for all of them HICF was performed to identify overlap of BAC clones. Out of the 30 BAC clones 20 were confirmed by colony PCR (Table 3-3). All BAC clones identified with probe RGC151 and confirmed by colony PCR fell together in HICF contig03. For time and cost efficiency, overgo probes HT211 and ORS662 were pooled before screening the BAC library. Six of the 18 positive BAC clones gave a signal with marker HT211 and HICF assembled them into HICF contig01. BAC clones confirmed by marker ORS662 were assembled by HICF into two contigs containing three (HICF contig05) and two (HICF contig06) clones, respectively. No confirmation with HT211 or ORS662 was achieved for five clones which build HICF contig04 and for the singleton BACs which assembled into no HICF contig. HICF contig04 and the singleton BAC clones were excluded from further analysis.

Positive BAC clones of RGC52 were delivered by Clemson University. These BACs assembled into contig 2 and contig 7 (Table 3-3).

Table 3-3: Number of BAC clones identified by screening the BAC library HA383 with overgo probes designed on sequences of cosegregating markers. Positive BAC clones were confirmed with colony PCR and assembled into six high information content fingerprinting (HICF) contigs.

Marker	Identified BACs	Colony PCR	HICF contig (number of clones)
RGC151	12	9	Contig03 (9)
HT211	18	6	Contig01 (6)
ORS662		5	Contig05 (3), Contig06 (2)
RGC52	18	18	Contig02 (11), Contig07 (7)

Sequence information from both ends of the sunflower DNA insert of each BAC clone which was confirmed by colony PCR and assembled into one of the HICF contigs was generated using the T7 and M13 primers. In total, 39 BES were generated from the 20 BACs of contig01, 03, 05, and 06 (Table 8-8). BES of contig02 and contig07 were available at NCBI (Table 8-8). The BES lengths ranged from 149 to 834 bp with an average of 600 bp. The GC content of the BES had an average of 38 % (19.7-51.6 %). Blastn and blastx analyses were performed. Sequences which showed similarity to

repetitive sequences and retrotransposons with an e-value cutoff of 1.00 e^{-10} were not considered for primer design. BES HBA323D15_M13 from a BAC clone identified with overgo probe RGC151 showed similarity to a NBS-LRR gene of *Lactuca sativa*.

3.2.2 Anchoring of BAC contigs

To investigate if the HICF contigs map to the target region, primers were designed for BES of BAC clones (Table 8-6) and amplicons were tested for polymorphism between (cms)HA342 and ARG1575-2 and between $NDBLOS_{sel}$ and KWS04. HICF contig02 was assigned to LG 16 in population RHA280 x RHA801 (Wenxiang Gao, personal communication). Therefore, only primer pairs designed on the BES of HICF contig01, 03, 05, 06, and 07 were tested. In total, PCR for 52 primer pairs based on BES of HICF contig 1 (8), contig 3 (12), contig 5 (4), contig 6 (4), and contig 7 (14) were optimized and the products were checked for polymorphism on agarose gels or on non-denaturing polyacrylamid gels. For convenience, marker names (e.g. Co1-4_M13) in the following include the HICF contig (**Co1**-4_M13), the numbered BES (Co1-**4**_M13), and the primer of the BES which was used to generate the BES (M13 or T7) (Co1-4_**M13**).

HICF contig01 was anchored to the target region using the marker Co1-4_M13 which was polymorphic between HA342 and ARG1575-2. For HA342 a PCR product of 555 bp was observed whereas no product was obtained for ARG1575-2.

HICF Contig03 was mapped on LG 1 in (cms)HA342 × ARG1575-2 using the dominant marker Co3-4_T7. Homozygous susceptible genotypes and heterozygous genotypes amplified a 617 bp PCR fragment, but for ARG1575-2 no fragment could be amplified. Marker Co3-2_T7 was polymorphic between (cms)HA342 and ARG1575-2 as well as between $NDBLOS_{sel}$ and KWS04.

HICF contig06 was anchored to the target region of PI_{ARG} using the marker Co6-1_M13. Marker Co7-11_M13 showed a polymorphism between HA342 / ARG1575-2 and $NDBLOS_{sel}$ / KWS04 and was used to anchor the HICF contig07. For HA342 and KWS04 a PCR product of 502 bp was amplified, but with ARG1575-2 as well as with $NDBLOS_{sel}$ no product was obtained.

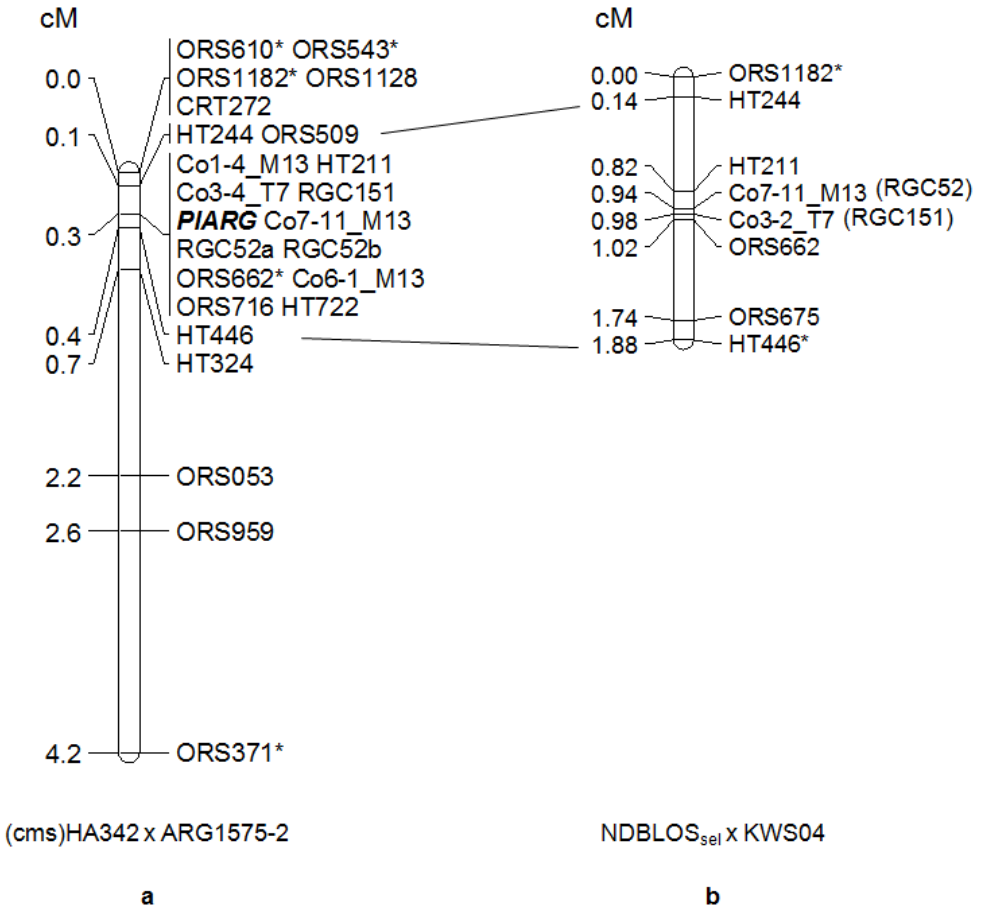


Figure 3-4: Anchored HICF contigs Co1, Co3, Co6 and, Co7 on linkage group (LG) 1. (a) LG 1 map of (cms)HA342 × ARG1575-2 constructed with 2,145 F_2 individuals. (b) Partial LG 1 map of $NDBLOS_{sel}$ × KWS04 constructed with 2,780 F_2 individuals. Markers with an asterisk were screened in all individuals, while all other markers were screened in recombinant lines only. PI_{ARG} is shown in bold. Maps are not drawn to scale and distances on the left side are given in centiMorgan.

Recombinant F_2 individuals of (cms)HA342 × ARG1575-2 (ORS610/ORS371) and $NDBLOS_{sel}$ × KWS04 (ORS1182/HT446) were genotyped with the polymorphic markers. Map calculation of the final maps was based on the whole population (cms)HA342 × ARG1575-2 with 2,145 F_2 individuals and $NDBLOS_{sel}$ × KWS04 with 2,780 F_2 individuals by imputing marker alleles of the non-recombinant lines. Markers Co1-4_M13, Co3-4_T7, Co3-2_T7, Co6-1_M13, and Co7-11_M13 cosegregated with PI_{ARG} and consequently HICF contig01, 03, 06, and 07 mapped in the target region (Fig. 3-4a). In $NDBLOS_{sel}$ × KWS04 HICF contigs 03 and 07 were mapped between HT211 and ORS662 and one recombination event was observed between the two contigs (Fig. 3-4b).

BAC clones of HICF contig05 were identified with an overgo probe designed on the sequence of marker ORS662. No polymorphic marker could be developed for BES

from contig05.

In conclusion, HICF contig01, contig03, and contig06 mapped on LG 1 into the PI_{ARG} target region whereas HICF contig02 mapped to LG 16.

3.2.3 BAC clone sequencing and assembly

Due to large overlaps of the BAC clones in the HICF contigs, 29 clones out of the 38 identified BACs were selected for sequencing on the GS FLX Titanium platform (Roche Diagnostics).

An overview of the sequenced BAC clones, insert sizes, barcodes and sequencing data is given in Table 8-9. The insert size of each clone was determined by pulsed field gel electrophoresis to calculate the expected sequence length and sequence depth of each BAC (Fig. 3-5). BAC clones of the HA383 library have an average insert size of 125.5 kb and the observed inserts of the selected BACs ranged from 60 kb to 252 kb.

Each of the 29 BAC clones was barcoded to allow separation of reads after sequencing. For 98% of the reads the correct barcoding adapter was recognized allowing the unambiguous assignment of sequences to the corresponding BACs. In total, 376,543 reads were assigned to the corresponding BACs. This resulted in 12,984 reads per clone in average with a range from 4,347 to 28,793 reads per single BAC clone.

On average 1% of the reads were derived from *E. coli* and 6% from the BAC vector resulting in 357,159 clean reads with a range of 3,956 – 27,309 reads per clone. Fig. 3-6 shows the read distribution of the clean reads. The sequencing coverage by 454 reads ranged from 7–84 x per individual BAC clone with an average of 33 x.

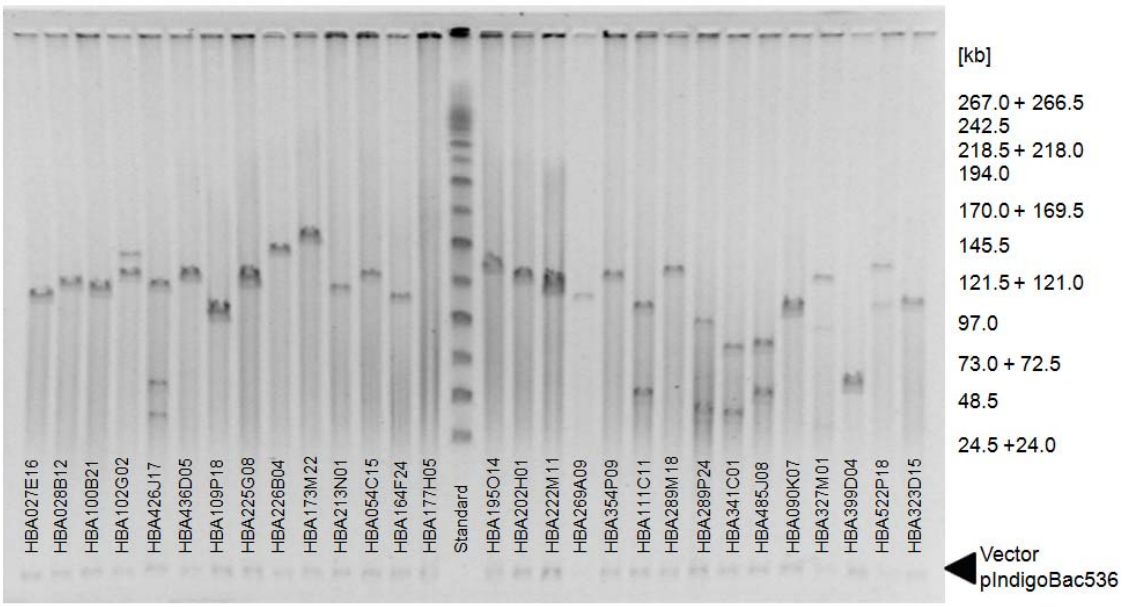


Figure 3-5: Insert size determination of 29 BAC clones by pulsed field gel electrophoresis compared to the MidRange II PFG marker (standard). Fragment size of the standard are indicated on the right side.

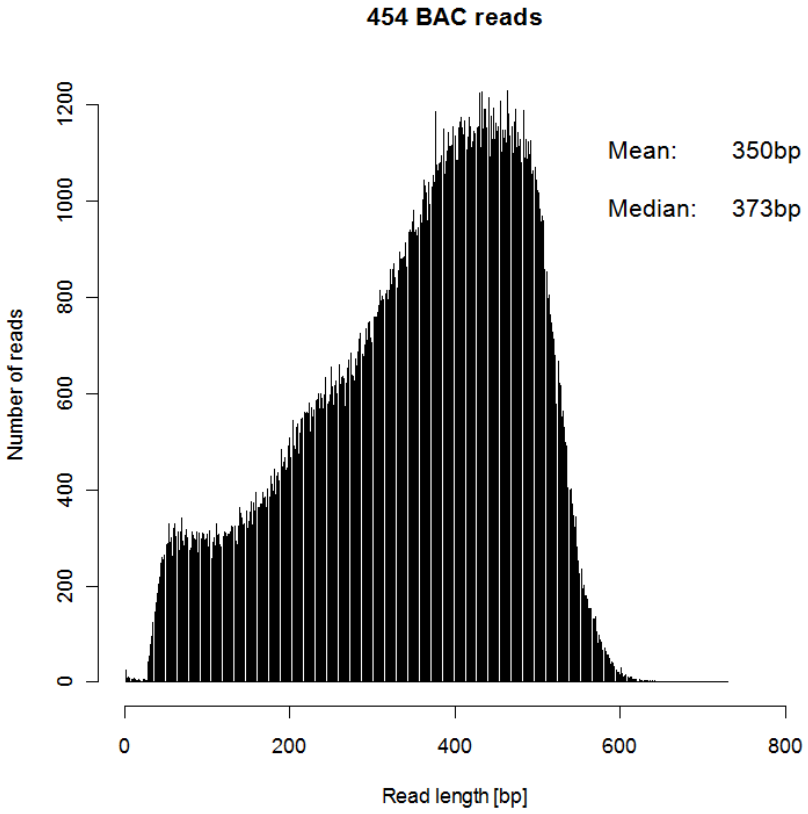


Figure 3-6: 454 read distribution of 29 sequenced BAC clones after quality control. On the abscissa the read length in bp and on the ordinate the number of observed reads is shown.

The assembly strategy for *de novo* sequenced BAC clones suggested by Steuernagel et al. (2009) was followed. They showed that optimal *ab inito* assembly parameters could not be defined, thus the clipped sequences were assembled with MIRA running several iterative changes of the hss/bph parameter set for each of the 29 BAC data sets. For each clone the assembly with the largest contig was defined as the 'best assembly'. Table 8-10 gives an overview about the assembly statistic for the 'best' assembly of each BAC clone. Presuming that the combined length of all contigs (>1kb) per assembly represent the overall BAC insert length, the observed insert size and the assembled insert size were compared and were found to be in agreement for 25 of the BAC clones. The predicted insert size of BAC clones HBA102G02, HBA426J17, and HBA164F24 differed from the observed insert size more than 10 % (Table 8-10). Plasmid DNA of BAC clone HBA177H05 was degraded and insert size was therefore not determined.

A second assembly was performed with reads of BAC clones belonging to the same HICF contig to obtain one assembly with a deep coverage for each of the HICF contigs which can be used for gene annotation. Again, the assembly with the largest contig was defined as the 'best assembly'. The largest contig in each assembly ranged between 61,836 bp for HICF contig05 and 123,158 bp for HICF contig07. Numbers of contigs larger than 100 bp per HICF contig varied from 80 in HICF contig05 to 686 in HICF contig03. Average coverage of subcontigs between 100 – 2,000 bp, 2,000 – 10,000 bp and greater than 10,000 bp length were calculated and subcontigs smaller than 2,000 bp were excluded from further analyses due to low coverage (Table 3-4). The contig-wise assembly was used for all further analysis. Gene prediction was performed on the contig-wise assembly for contigs larger then 2,000 bp (see 3.3.1).

A blastn analysis of the BES against the assembled BAC sequences showed that 65 of the 72 BES were found in the assembly with 99-100% identity and 100% coverage. Five BES identified from BAC clones of HICF contig05 and contig06 were found in the assembled sequences of HICF contig05 as well as contig06. Alignments of the appropriate sequences showed 100% sequence similarity of the overlaps. High information content fingerprinting demonstrated that no physical overlap exists between the BAC clones of HICF contig05 and contig06. Thus, sequence similarity between clones may be the result of duplications in the genome.

No overlap of the HICF contigs could be demonstrated on the basis of BES for HICF contig01, 03, and 07 (RGC52, RGC151, HT211) indicating a gap of unknown size between them.

Table 3-4: Summary of best MIRA assemblies for six BAC High Information Content (HICF) contigs.

HICF contig	Contig01	Contig02	Contig03	Contig05	Contig06	Contig07
Linkage group	**LG 1**	**LG 16**	**LG 1**	**unknown**	**LG 1**	**LG 1**
No. of BAC clones	6	5	9	3	2	4
No. of reads	57,653	65,442	151,402	26,660	24,784	31,218
No. of subcontigs [bp] (n>100 bp)	362	185	686	80	108	232
100 - 2,000	338	177	648	76	103	172
2,000 – 10,000	14	5	27	0	3	57
>10,000	10	3	11	4	2	3
Length [bp]						
Min	103	176	102	194	134	121
Mean	1,331	1,676	1,221	2,343	1,943	2,144
N50	6,602	26,210	1,904	40,067	106,444	4,872
Max	73,419	92,740	116,032	61,836	106,444	123,158
Average coverage in subcontigs						
100 - 2,000 bp	4	3	11	5	5	3
2,000 – 10,000 bp	17	74	58	0	54	13
>10,000 bp	79	118	139	62	54	48

3.2.4 BAC clone characterization

Employing the programs FgenesH, GeneID, PASA and GenomeThreader for gene prediction, 180 genes and gene fragments were identified on the six HICF contigs. For annotation only assembled subcontigs longer than 2,000 bp were used, because the read coverage was 13-139 x for these sequences (Table 3-4). Table 8-11 shows the predicted genes and their designated function identified by blast analysis. For convenience, genes predicted on subcontigs of contig-wise assembled sequences were denoted like contig03_c2_00001 (HICF contig _ subcontig _ gene number). Of the 180 predicted gene models, 48 were with known function, 15 showed homology to hypothetical proteins and predicted proteins with unknown function, 105 were transposable elements (TEs) and for 12 predicted genes no blast hits were found (Table 3-5). This leads to the conclusion that the majority of the predicted gene models belong to transposable elements (58%). The gene density in sunflower was calculated

for the longest subcontig of each HICF contig including only genes with known function. Among the six HICF contigs, an unequal distribution of gene density was detected. Density varied from 1 gene per 15 kb (HICF contig02) to a relatively low density of 1 gene in 106 kb (HICF contig06).

Table 3-5: Predicted gene models on the high information content fingerprinting (HICF) contigs. Predicted gene models were differentiated after homology search against NCBI in genes with known function, hypothetical proteins, transposable elements (TEs) and gene models with no homology to known genes (no hit).

	HICF Contig01 LG1	HICF Contig02 LG16	HICF Contig03 LG1	HICF Contig05 unknown	HICF Contig06 LG1	HICF Contig07 LG1
No. of gene models	34	26	50	8	14	48
No. of genes	9	7	16	3	3	11
of hypothetical proteins	0	0	5	1	3	5
of TEs	23	16	25	4	7	30
of no hit	2	3	4	0	1	2

3.2.5 Candidate genes belong to the TIR-NBS-LRR type

In total, six predicted genes or gene fragments which showed homology to resistance gene candidates were identified on the sequence level of BAC clones (Table 8-11): contig02_c2_000011, contig02_c2_000012, contig03_c2_00009, contig07_c1_00003, contig07_c1_00007, and contig07_c1_000015. HICF contig02 was assigned to LG 16. Therefore, the resistance gene candidates cannot be candidate genes for PI_{ARG}, but they were included in further analyses to investigate whether the genes belong to the same gene family. HICF contig03 and 07 were mapped in the target region of PI_{ARG} and therefore the resistance gene candidates were classified as potential candidates for PI_{ARG}.

The predicted genes from the annotation pipeline which showed homology to *R* genes were analyzed in more detail. The sequences were analyzed with the gene prediction programs AUGUSTUS and GeneSeqer. Several gene models for each gene were predicted. The predicted amino acid sequences were aligned to the reference sequences (Table 8-12) to identify the best fitting gene model. The predicted genes contig02_c2_000011 and contig02_c2_000012 were combined to one gene which was named RGC52SW1. Contig07_c1_00003 and contig07_c1_00007 were combined to RGC52SW23 and contig07_c1_000015 was named RGC52SW4. The predicted candidate gene including contig03_c2_00009 was named RGC151SW1. The protein sequences of RGC151SW1, RGC52SW1, RGC52SW23 and RGC52SW4 which

showed best homology to known resistance genes are presented in Chapter 8.2.2. The exon–intron assignments of the resistance gene candidates are shown in Fig. 3-7 to 3-10 and Table 3-6 to 3-9.

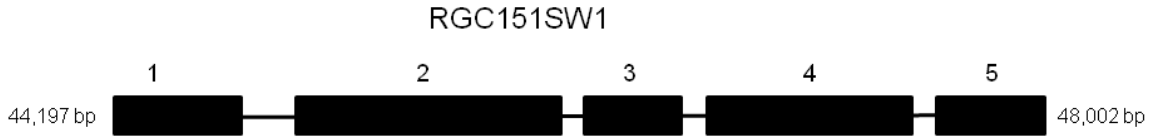


Figure 3-7: Predicted exon (filled boxes) – intron (line) assignment of resistance gene candidate e RGC151SW1 (BAC contig03_c2)

Table 3-6: Genomic coordinates of candidate gene RGC151SW1 exons based on contig03_c2, exon length, identity between BAC sequence and modified coding sequence to predict an complete open reading frame.

	Genomic coordinates	Length [bp]	Identity [%]	Mismatches	Gaps
Exon 1	44197-44729	533	100.0	0	0
Exon 2	44934-46056	1123	100.0	0	0
Exon 3	46140-46544	405	100.0	0	0
Exon 4	46636-47481	846	100.0	0	0
Exon 5	47564-48002	438	100.0	0	0

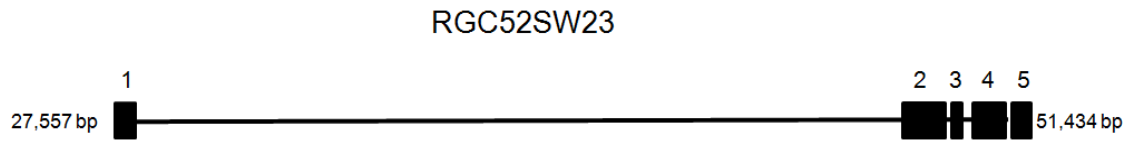


Figure 3-8: Predicted exon (filled boxes) – intron (line) assignment of resistance gene candidate RGC52SW23 (BAC contig07_c1)

Table 3-7: Genomic coordinates of candidate gene RGC52SW23 exons based on contig07_c1, exon length, identity between BAC sequence and modified coding sequence to predict an complete open reading frame.

	Genomic coordinates	Length [bp]	Identity [%]	Mismatches	Gaps
Exon 1	27557-28100	545	99.8	1	1
Exon 2	47995-49131	1138	99.7	3	3
Exon 3	49244-49567	324	100.0	0	0
Exon 4	49877-50773	897	99.9	1	0
Exon 5	50814-51434	621	100.0	0	0

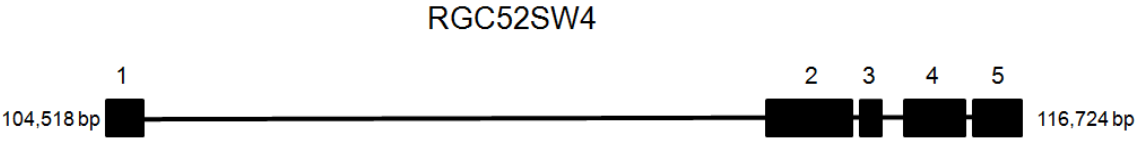


Figure 3-9: Predicted exon (filled boxes) – intron (line) assignment of resistance gene candidate RGC52SW4 (BAC contig07_c1)

Table 3-8: Genomic coordinates of candidate gene RGC52SW4 exons based on contig07_c1, exon length, identity between BAC sequence and modified coding sequence to predict an complete open reading frame.

	Genomic coordinates	Length [bp]	Identity [%]	Mismatches	Gaps
Exon 1	104518-105016	500	99.8	1	1
Exon 2	113245-114381	1138	99.9	1	1
Exon 3	114492-114815	324	100.0	0	0
Exon 4	115095-115955	861	100.0	0	0
Exon 5	116035-116724	690	100.0	0	0

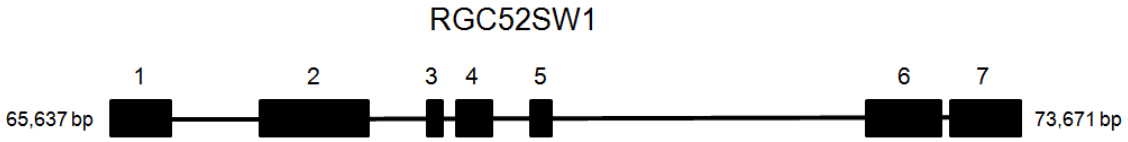


Figure 3-10: Predicted exon (filled boxes) – intron (line) assignment of resistance gene candidate RGC52SW1 (BAC contig02_c2)

Table 3-9: Genomic coordinates of candidate gene RGC52SW1 exons based on contig07_c1, exon length, identity between BAC sequence and modified coding sequence to predict an complete open reading frame.

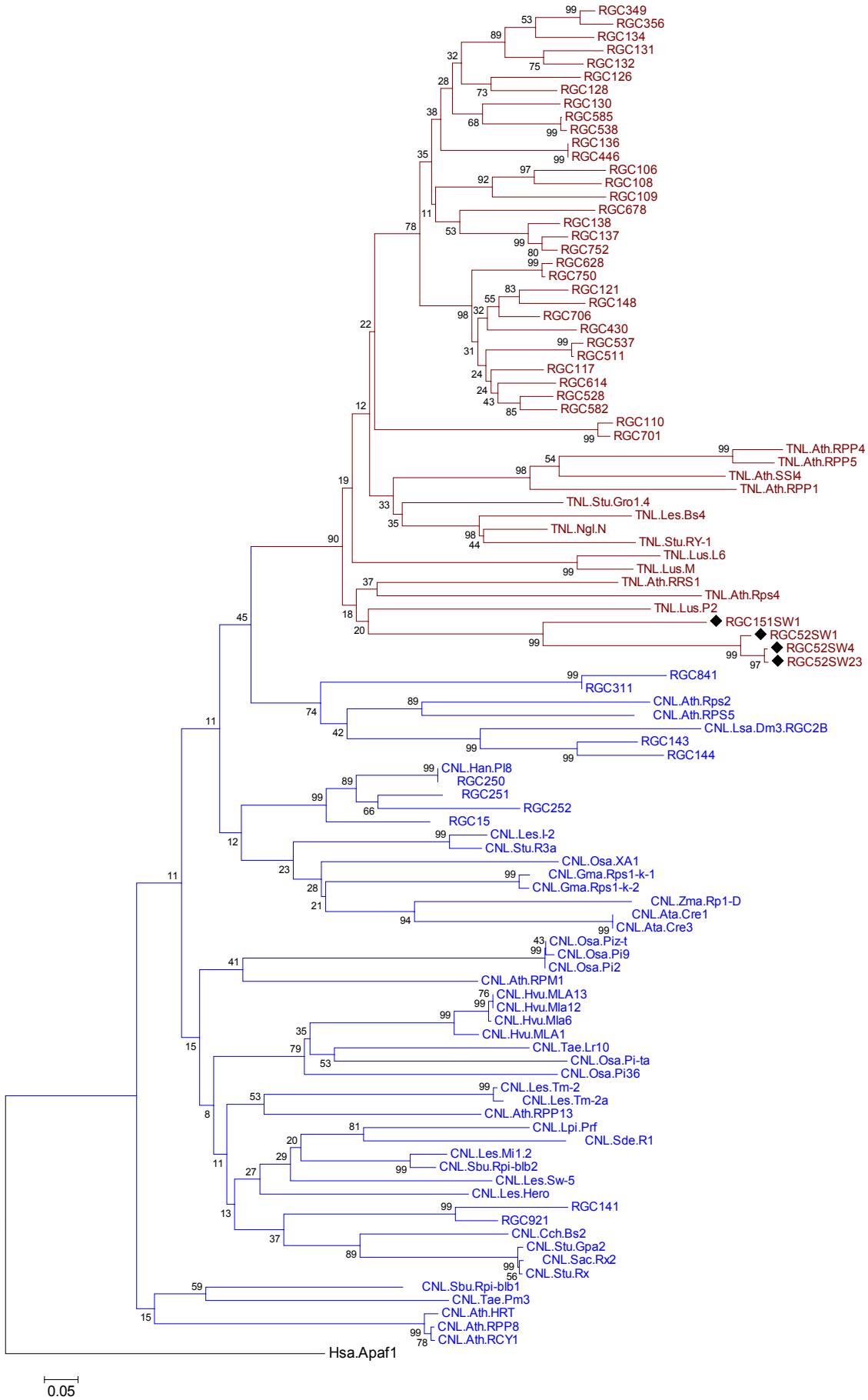
	Genomic coordinates	Length [bp]	Identity [%]	Mismatches	Gaps
Exon 1	65637-66181	545	100.0	0	0
Exon 2	66948-67914	967	99.6	4	4
Exon 3	68407-68577	171	98.8	2	0
Exon 4	68690-69018	330	99.7	1	1
Exon 5	69351-69558	200	96.0	8	8
Exon 6	72305-72968	664	100.0	0	0
Exon 7	73021-73671	651	100.0	0	0

The NBS motif, *i.e.* the region between P-loop and GLPL motifs, was used for pair wise comparison and phylogenetic analysis of the four RGCs. Pair wise comparisons of the NBS amino acid sequences showed 93 – 96% sequence identity between RGC52SW1, RGCS52W23, and RGC52SW4, but showed higher differences to the NBS sequence of candidate gene RGC151SW1. The amino acid sequence identity between RGC151SW1 and the three RGC52 candidates was only 54%. The phylogenetic relationship of the newly identified RGCs was determined by calculating a neighbor-joining tree with MEGA 4 (Kumar et al. 2008). For the analysis, 41 resistance genes of the CC-NBS-LRR subclass (dicotyledonous plants) and 13 reference sequences belonging to the TIR-NBS-LRR subclass were used (Table 8-12). In addition, 43 sunflower RGC sequences of the TIR-NBS-LRR and CC-NBS-LRR subclasses identified by Radwan et al. (2008) were selected for the analysis.

Two major clades representing CC and TIR sequences were observed (Fig. 3-11). The known resistance genes from the reference data set (http://prgdb.cbm.fvg.it/index.php, Table 8-12) clearly separate between these two clades. The newly identified resistance gene candidates for PI_{ARG} belong to the clade of TIR sequences, but were clustered in

a different subclass than the other known sunflower TIR RGCs. Sequence alignments of the NBS domain between the newly identified sunflower RGCs and known TIR-NBS-LRR sequences (reference data set, Table 8-12) showed six major motifs which had been described by Meyers et al. (1999) (Fig. 3-12).

▼Figure 3-11: Neighbor-joining tree of an amino acid alignment of the NBS-encoding region of 102 sequences. The tree was calculated with 13 reference sequences from the TIR-NBS-LRR subclass, 41 reference sequences from the CC-NBS-LRR subclass, 43 sunflower RGCs (Radwan et al. 2008) and four newly identified (♦) resistance gene candidates. The human Apaf1 sequence was used as out-group. The TIR (red) and the CC (blue) subclass split into two clades. Sequence names of the reference sequences contain the species (first three letters) and the name of the cloned gene. Numbers shown on branches are percentages of bootstrap replications supporting nodes.


RGC349
RGC356
RGC134
RGC131
RGC132
RGC126
RGC128
RGC130
RGC585
RGC538
RGC136
RGC446
RGC106
RGC108
RGC109
RGC678
RGC138
RGC137
RGC752
RGC628
RGC750
RGC121
RGC148
RGC706
RGC430
RGC537
RGC511
RGC117
RGC614
RGC528
RGC582
RGC110
RGC701
TNL.Ath.RPP4
TNL.Ath.RPP5
TNL.Ath.SSI4
TNL.Ath.RPP1
TNL.Stu.Gro1.4
TNL.Les.Bs4
TNL.Ngl.N
TNL.Stu.RY-1
TNL.Lus.L6
TNL.Lus.M
TNL.Ath.RRS1
TNL.Ath.Rps4
TNL.Lus.P2
RGC151SW1
RGC52SW1
RGC52SW4
RGC52SW23
RGC841
RGC311
CNL.Ath.Rps2
CNL.Ath.RPS5
CNL.Lsa.Dm3.RGC2B
RGC143
RGC144
CNL.Han.Pl8
RGC250
RGC251
RGC252
RGC15
CNL.Les.I-2
CNL.Stu.R3a
CNL.Osa.XA1
CNL.Gma.Rps1-k-1
CNL.Gma.Rps1-k-2
CNL.Zma.Rp1-D
CNL.Ata.Cre1
CNL.Ata.Cre3
CNL.Osa.Piz-t
CNL.Osa.Pi9
CNL.Osa.Pi2
CNL.Ath.RPM1
CNL.Hvu.MLA13
CNL.Hvu.Mla12
CNL.Hvu.Mla6
CNL.Hvu.MLA1
CNL.Tae.Lr10
CNL.Osa.Pi-ta
CNL.Osa.Pi36
CNL.Les.Tm-2
CNL.Les.Tm-2a
CNL.Ath.RPP13
CNL.Lpi.Prf
CNL.Sde.R1
CNL.Les.Mi1.2
CNL.Sbu.Rpi-blb2
CNL.Les.Sw-5
CNL.Les.Hero
RGC141
RGC921
CNL.Cch.Bs2
CNL.Stu.Gpa2
CNL.Sac.Rx2
CNL.Stu.Rx
CNL.Sbu.Rpi-blb1
CNL.Tae.Pm3
CNL.Ath.HRT
CNL.Ath.RPP8
CNL.Ath.RCY1
Hsa.Apaf1
0.05

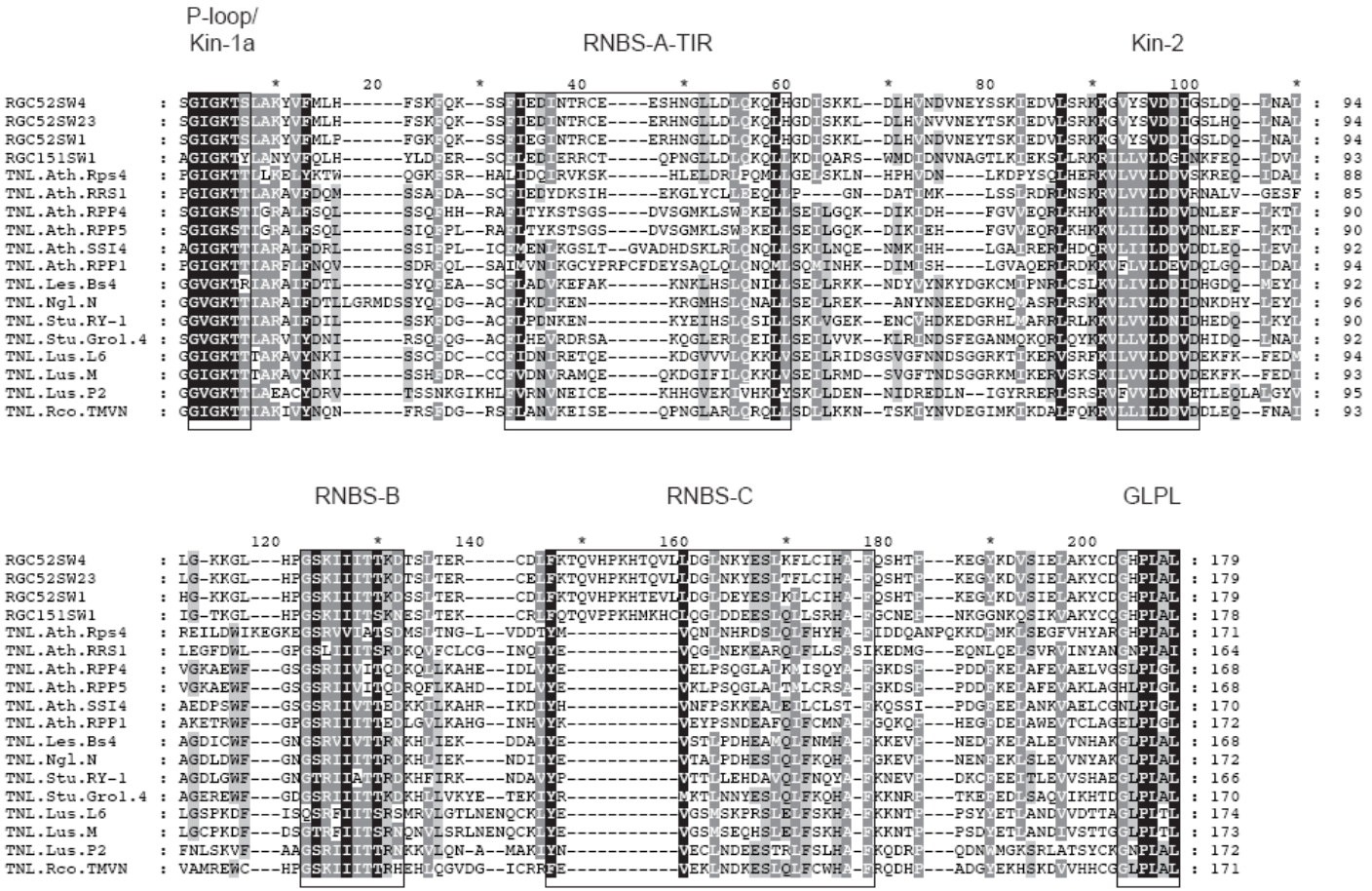


Figure 3-12: Partial alignment of deduced amino acid sequences of the NBS-encoding region for the four newly identified resistance gene candidates from *Helianthus annuus* and 13 cloned TIR-NBS-LRR (TNL) resistance genes from the reference data set identified in different species. ClustalW was used for the sequence alignment. Homologous regions were shaded using the Genedoc software. The P-loop/Kin-1a, RNBS-A-TIR, Kin-2, RNBS-B, RNBS-C and GLPL motifs described by Meyers et al. (1999) are boxed.

3.2.6 Copy number variation of RGC151

Several fragments were observed in cmsHA89 as well as in ARG1575-2 after digestion with different restriction enzymes (Fig. 3-13). This observation suggests that both lines have several copies of RGC151 homologous genes. The fragments have different intensities which may result from sequence variation between the DNA on the filter and the hybridization probe. All fragments, except one (*EcoR*I, 1.6 kb), which were observed for cmsHA89 could also be found for ARG1575-2. Two to three fragments were observed exclusively for ARG1575-2 (Table 3-10). These fragments must originate from the wild species *H. argophyllus*. The chromosomal location of these fragments is unknown. It cannot be excluded that one or more of the copies are located on other linkage groups than LG 1.

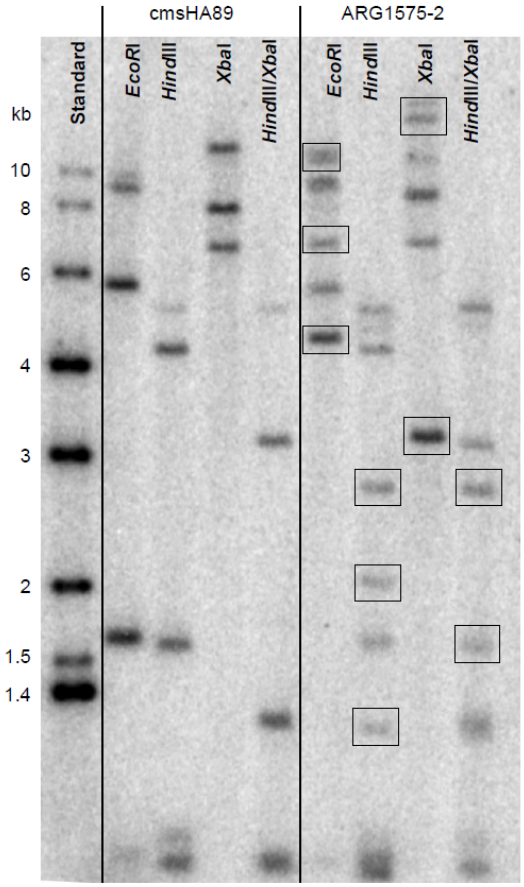


Figure 3-13: Genomic Southern blot analysis of cmsHA89 and the resistant line ARG1575-2. Genomic DNA was digested with *EcoR*I, *Hind*III, *Xba*I or *Hind*III/*Xba*I and hybridized to a 550 bp fragment of RGC151. The size marker is indicated on the left. Fragments which were unique for ARG1575-2 are highlighted by boxes.

Table 3-10: Number of observed fragments identified in the Southern blot for cmsHA89 and ARG1575-2 digested with different restriction enzymes.

Line	*Eco*RI	*Hind*III	*Xba*I	*Xba*I/*Hind*III
cmsHA89	3	4	3	4
ARG1575-2	5	7	5	6
Common fragments	2	4	3	4

Resistance gene candidate RGC151SW1 was partially sequenced in line ARG1575-2 (Fig. 3-14). It was attempted to sequence the whole gene for ARG1575-2, but not for all tested primer combinations an amplicon could be generated. All tested primer combinations amplified fragments of the expected length for HA342. For ARG1575-2 only fragments CG1_a, Co3-2_M13, RGC151, and CG1_e were generated. No product of the expected length or only a weak product was amplified with the primer pairs CG1_b, CG1_c, and CG1_d for ARG1575-2. PCR products of CG1_a, Co3-2_M13, RGC151 and CG1_e were cloned into a pCR®2.1-TOPO vector and eight plasmids per fragment were sequenced and analyzed. Sequence comparison by clustalW analysis suggests the existence of multiple copies of RGC151 in the sunflower genome (Alignments 8-1 to 8-3). Pair wise comparison of the CG1_a, CG1_e, and RGC151 nucleotide sequences demonstrated diversity within HA342 and ARG1575-2 and a high level of diversity between the two lines. The nucleotide sequence identity of CG1_a ranged from 99 – 100% within HA342 and ARG1575-2, respectively. Between the two lines 93 – 94% similarity was found. For RGC151 a similarity of 96 – 100% was found within HA342 and 99 – 100% within ARG1575-2. The two lines had a similarity of 87% for RGC151. Fragment CG1_e had a similarity of 100% within HA342 and ARG1575-2, respectively, and a similarity of 88 % was found between the two lines.

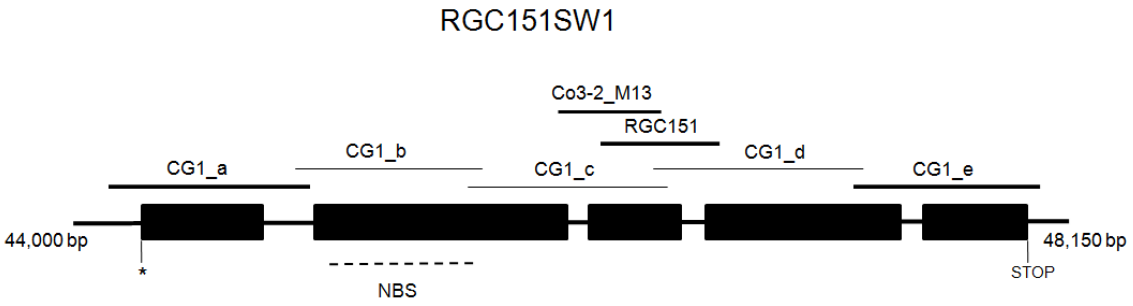


Figure 3-14: Predicted exon (filled boxes) – intron (line) assignment of resistance gene candidate RGC151SW1 which was identified on HICF contig03_c2 (44,000 – 48,200 bp). Primers were designed on the BAC sequence of HA383 and PCR was performed to amplify the fragments CG1_a, CG1_b, CG1_c, CG1_d, CG1_e, Co3-2_M13 and RGC151. Bold labeled fragments could be amplified for ARG1575-2 and were sequenced. For fragments CG1_b, CG1_c, and CG1_d no amplicon with the expected length could be amplified for ARG1575-2. The NBS domain is highlighted with a dashed line.

Recently, Radwan et al. (2011) published the full length cDNA clone of the resistance-like protein RGC151 (GU085221) derived from RHA419 and the appropriate protein sequence of 1021 aa (ACY69610). Alignment of the genomic DNA sequence of HA383 and the full length cDNA clone of RHA419 confirmed the predicted five exons of RGC151SW1. One deletion in exon 5 of RGC151SW1 (HA383) causes a frame shift. A similarity of 79% was observed between the predicted protein sequence of RGC151SW1 (HA383) and the protein sequence of RGC151 (RHA419) until the frame shift occurred (1 – 991 aa). Blastn analysis of the 550 bp PCR product of RGC151 (ARG1575-2) against the full length clone RGC151 (RHA419; GU085221) identified a sequence similarity of only 85% and a deletion of 24 bp in the sequence of ARG1575-2.

3.3 Transcriptome based studies

3.3.1 Resistance gene candidates are differentially expressed

Samples of HA342 and ARG1575-2 inoculated or non-inoculated with *P. halstedii* were harvested at different time points and markers RGC151, RGC52a, and RGC52b were used to analyze the expression level of the mapped resistance gene candidates by semi-quantitative RT-PCR. The sunflower elongation factor EF-1α (Radwan et al. 2005a), a housekeeping gene, was used as standard to control the amount of cDNA (Fig. 3-16). The expression of EF-1α in all samples of HA342 was weaker than in the samples of ARG1575-2. *P. halstedii* elongation factor TEF1 (Radwan et al. 2005a) showed the success of inoculation of HA342 and ARG1575-2 with *P. halstedii* (Fig. 3-16). *P. halstedii* was observed in hypocotyls and cotyledons of HA342 and ARG1575-2, respectively, with increasing intensity during the time curve. Two non-inoculated samples of HA342 (HA342, cotyledons, day 12 after treatment) and ARG1575-2 (ARG1575-2, hypocotyls, day six after treatment) also showed a weak expression level of TEF1. The plants were grown in the same climate chamber to provide the same conditions for all plants. Therefore, contamination of the non-inoculated samples could not be excluded, but the phenotypes of the seedlings differed clearly between the treatments (Fig. 3-15). Inoculated seedlings of HA342 and ARG1575-2 were reduced in length compared to the control.

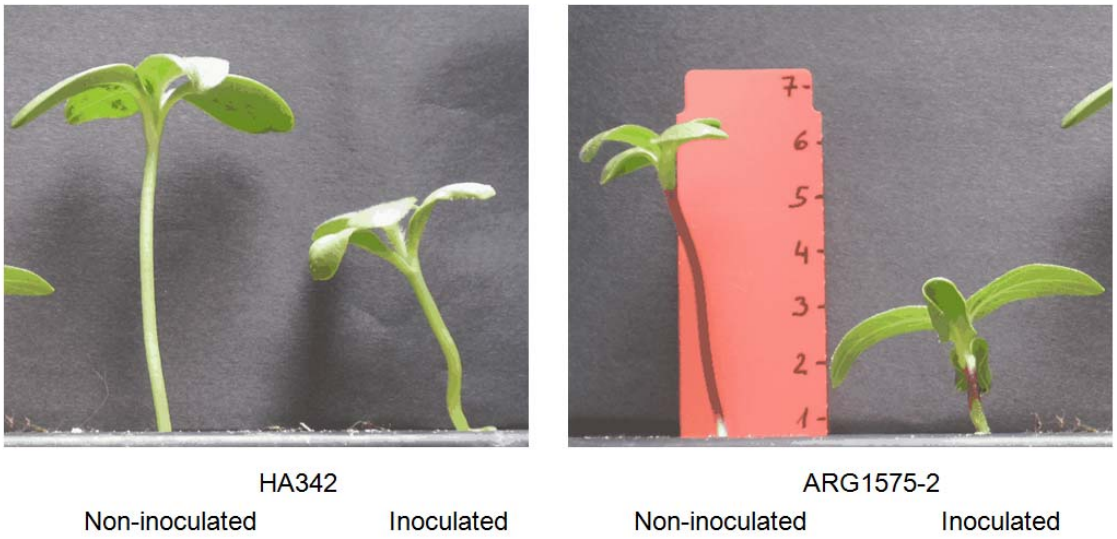


Figure 3-15: Fifteen days old seedlings of HA342 and ARG1575-2 after inoculation with *P. halstedii* (race 730) or H_2O, cultivated at 18°C and 16h light in a climate chamber. From left to right: HA342, non-inoculated, HA342, inoculated, ARG1575-2, non-inoculated, ARG1575-2, inoculated. Size in cm is indicated on the label, right side.

Neither RGC52a nor RGC52b showed expression in HA342 and ARG1575-2. Expression could be observed for RGC151 in the susceptible line HA342 and in the resistant line ARG1575-2 (Fig. 3-16) and differences in the expression patterns were identified. RGC151 was weak but constitutively expressed in the non-inoculated samples of hypocotyls and cotyledons in both, HA342 and ARG1575-2. After inoculation, a higher, but also consistent expression level was detected in the cotyledons of ARG1575-2 in comparison to the non-inoculated ARG1575-2 samples. An increased expression level of RGC151 was also observed in the susceptible line, but was delayed in comparison to ARG1575-2. Similar expression levels in cotyledons were observed six days after inoculation.

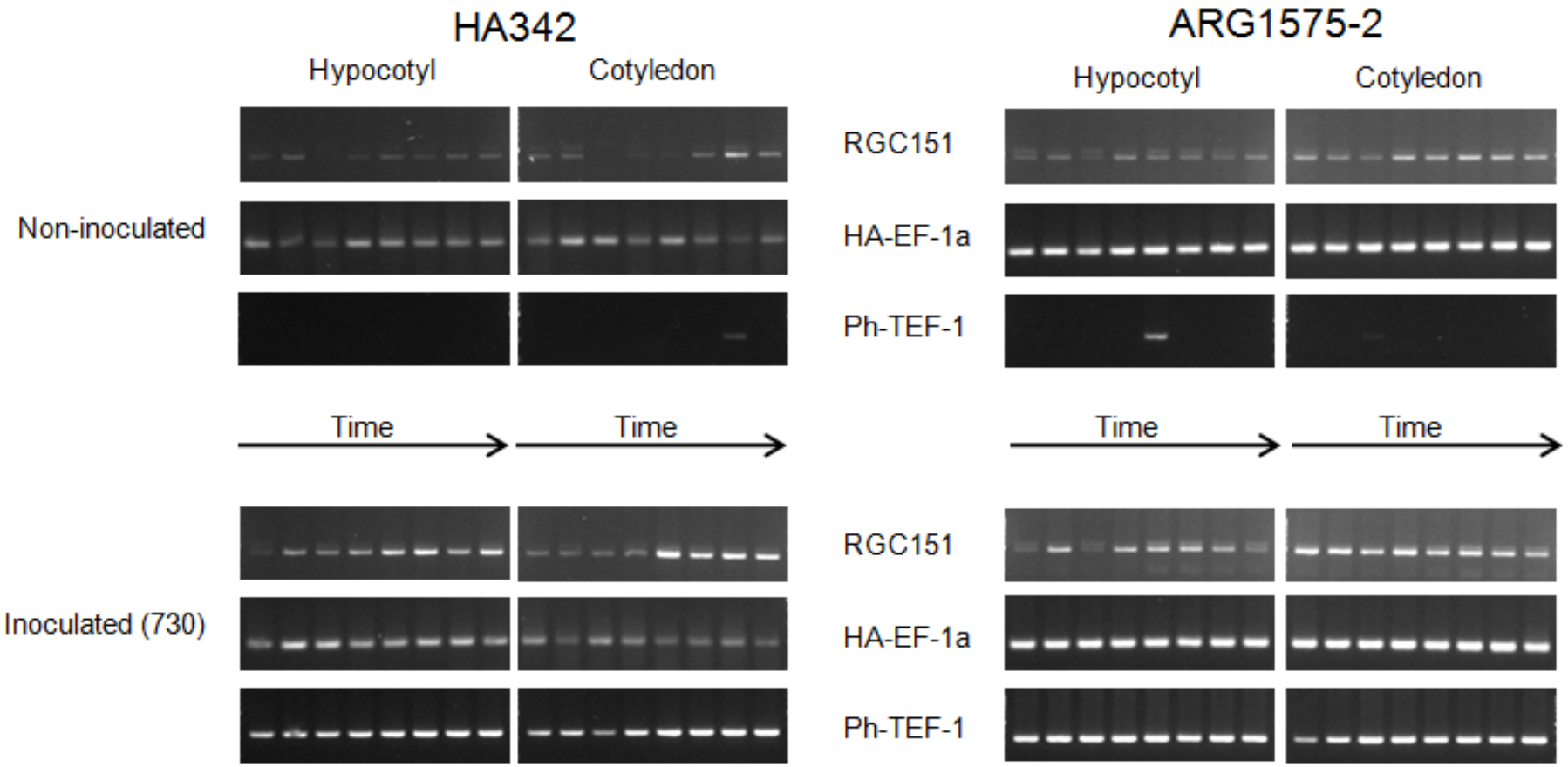


Figure 3-16: Expression pattern of RGC151, the sunflower elongation factor HA-EF-1α and the *P. halstedii* elongation factor Ph-TEF1. On the left side the susceptible line HA342 and on the right side the resistant line ARG1575-2 is illustrated. On the top the expression pattern of the non-inoculated samples are presented and on the bottom the pattern of the inoculated (race 730) samples. The timeline indicates the samples (0, 0.5, 1.5, 3, 6, 9, 12, and 15 d after inoculation) described in chapter 2.1.3.

3.3.2 cDNA-AFLP identified differentially transcribed fragments of the general defense response

The expression patterns of transcript-derived fragments (TDFs) were monitored by the cDNA-AFLP technique following the protocol of Vuylsteke et al. (2007). Hypocotyls and cotyledons of the susceptible line HA342 and the resistant line ARG1575-2 were analyzed separately. Half of the plant material was inoculated with race 730 and half of it was non-inoculated. Samples were taken at six different time points within 15 days after treatment. The restriction enzymes *BstY*I as a rare cutter and *Mse*I as a frequent cutter were chosen. The observed fragments ranged from 50 to 1,000 bp. The number of observed fragments below 200 bp was too high for evaluation and fragment intensity was very weak above 800 bp. Therefore, only fragments between 200 and 800 bp were considered in the analysis (Fig. 3-17). The gel images were visually evaluated. Fragments of genes were counted that showed differentially expression between the treatments for a given line (Table 3-11). Differences between the lines were not considered, because they can occur due to allelic polymorphism. Twenty five fragments were differentially expressed between the non-inoculated and the inoculated samples of ARG1575-2. The susceptible line HA342 showed 117 differentially expressed fragments between the control and the inoculated samples. Twelve of the differentially expressed TDFs were common between the two lines.

Table 3-11: Number of differentially expressed cDNA-AFLP fragments observed in hypocotyls, cotyledons or both tissues of the susceptible line HA342 and the resistant line ARG1575-2.

		HA3432	ARG1575-2	Common fragments
Upregulated in	hypocotyl	3	5	2*
	cotyledon	9	3	1
	both tissues	103	13	9
Downregulated in	hypocotyl	0	2	0
	cotyledon	1	1	0
	both tissues	1	1	0
Total		117	25	12

* Fragments are differentially expressed in both tissues of HA342

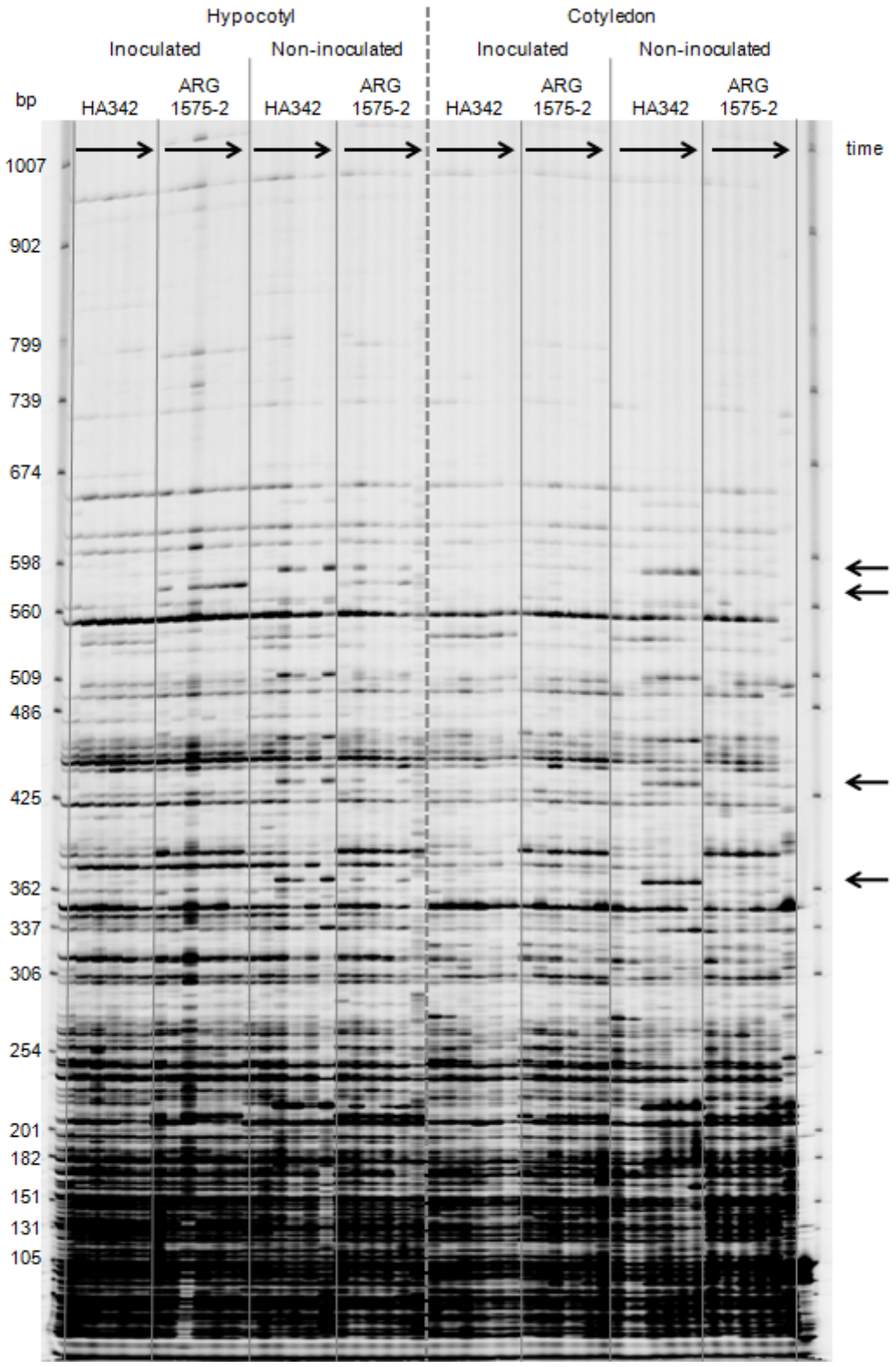


Figure 3-17: cDNA-AFLP gel image obtained with the primer combination *BstY*I-CC/*Mse*I-A. Inoculated and non-inoculated samples (cotyledons and hypocotyls) of the susceptible line HA342 and the resistant line ARG1575-2 were analyzed at six time points over a time course of 15 days to identify differentially expressed fragments. Size standards are left and right in lane. Differentially expressed fragments are highlighted with an arrow on the right side.

Five fragments out of 25 of ARG1575-2 were exemplarily chosen for sequencing. One TDF was down regulated (Fig. 3-18a) and one was upregulated in hypocotyls (Fig. 3-18b), one was upregulated in cotyledons (Fig. 3-18c) and two fragments were upregulated in both tissues (Fig. 3-18d). To enable the isolation of the fragments from the polyacrylamid gel, complexity of the fragment pattern was reduced by using a second selective base at the *Mse*I primer which allowed the exact isolation of the desired fragment. The TDFs were cut at two time points of the time curve where the highest expression level was observed.

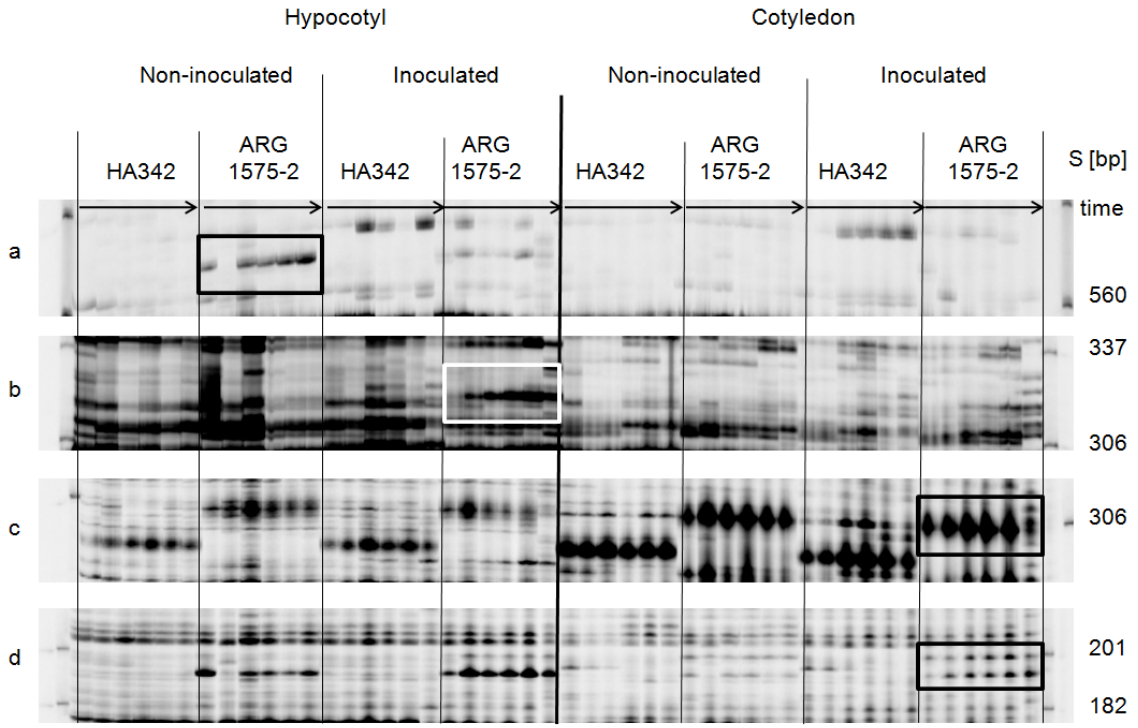


Figure 3-18: cDNA-AFLP gel images. Inoculated and non-inoculated samples were analyzed at six time points over a time course of 15 days to identify differentially expressed fragments. Cotyledons and hypocotyls of the susceptible line HA342 and the resistant line ARG1575-2 were analyzed separately. Isolated fragments are highlighted with a box. On the right side the fragment size of the standard is indicated in bp. a) Fragments derived from primer combination *BstY*I-CC/*Mse*I-A. b) Fragments derived from primer combination *BstY*I-CG/*Mse*I-T. c) Fragments derived from primer combination *BstY*I-TC/*Mse*I-T. d) Fragments derived from primer combination *BstY*I-TT/*Mse*I-T, both fragments in the box were isolated.

The five differentially expressed fragments from ARG1575-2 were excised from a polyacrylamid gel and reamplified by PCR. Isolated fragments may not only consist of a single product. For this reason the isolated fragment was cloned and at least three clones of each fragment were sequenced. A total of eight different sequences were obtained from clones of the five fragments as summarized in Table 3-12 (for sequences, see 8.2.3). The sequences were compared with the NCBI database. The majority of sequences showed homology with known nucleotide and protein sequences: dirigent-like protein, beta tubulin genomic sequences, tRNA-Ser (trnS) gene, P subunit of glycine decarboxylase multi-enzyme, and glutathione S-transferase. cDNA-AFLP sequences were blasted against the contig-wise BAC sequences to see whether there is an overlap, but no hits were observed.

Primers were designed based on the sequence of *Bst*YI-CC/*Mse*I-AC_1 (dirigent like protein) and BSA was performed to identify whether the gene maps to the target region of Pl_{ARG}. For *Bst*YI-CC/*Mse*I-AC_1 (dirigent like protein, CG2) a polymorphism was identified between HA342/BS and ARG1575-2/BR and the locus was mapped on LG 1, but outside the target region. The decreased expression of the sequence in the hypocotyls after inoculation could not be confirmed by RT-PCR (Fig. 3-19).

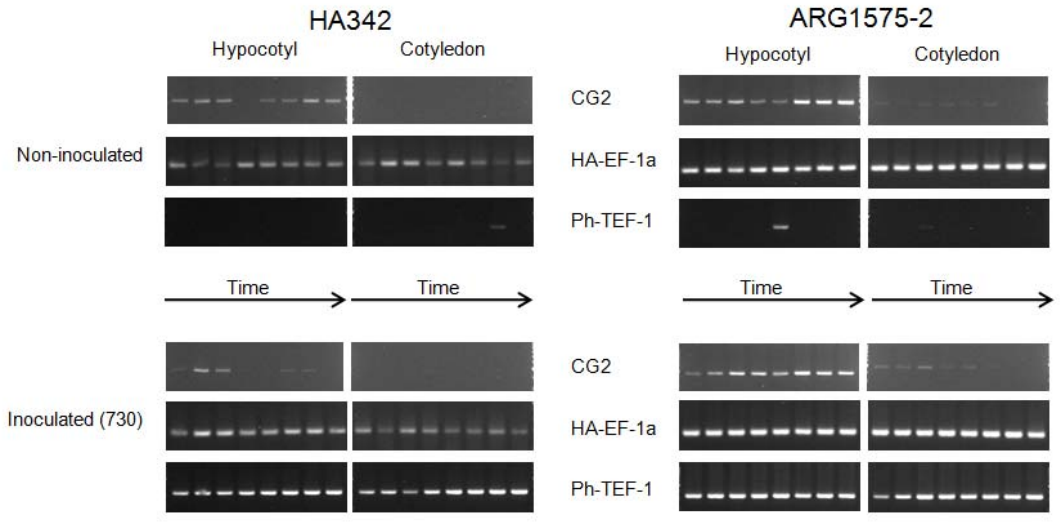


Figure 3-19: Expression pattern of dirigent like protein (CG2), the sunflower elongation factor HA-EF-1α and the *P. halstedii* elongation factor Ph-TEF1. On the left side the susceptible line HA342 and on the right side the resistant line ARG1575-2 is illustrated. On the top the expression pattern of the non-inoculated samples are presented and on the bottom the pattern of the inoculated (race 730) samples.

Table 3-12: Summary of sequences derived from differentially expressed cDNA-AFLP fragments of ARG1575-2 and their homology with nucleotide or protein sequences in the National Centre for Biotechnology Information (NCBI) database.

Sequence name	Length [bp]	Accession	Homology with tblastx	Organism	Coverage	E-value
*Bst*YI-CC/*Mse*I-AC_1	547	FJ600365.1	Dirigent-like protein 2 (DL2) mRNA, complete cds	*Gossypium hirsutum*	89%	3.00 e^{-59}
*Bst*YI-CC/*Mse*I-AC_2	520	EF461059.1	Clone ZVG20_HA89-R ZVG20 locus putative beta-tubulin	*Helianthus annuus*	90%	2.00 e^{-91}
*Bst*YI-CC/*Mse*I-AC_3	550	NM_106548.1	Transferase family protein (AT1G78990) mRNA	*Arabidopsis thaliana*	69%	1.00 e^{-27}
*Bst*YI-CG/*Mse*I-TA_1	192	EZ145581.1	Contig4987, mRNA sequence	*Artemisia annua*	99%	1.00 e^{-24}
*Bst*YI-TC/*Mse*I-TC_1	273	EU440304.1	Isolate Z67 tRNA-Ser (trnS) gene, partial sequence	*Zaluzania triloba*	88%	2.00 e^{-45}
*Bst*YI-TC/*Mse*I-TC_2	274	Z25857.1	mRNA for P subunit of glycine decarboxylase multi-enzyme	*Flaveria pringlei*	99%	1.00 e^{-53}
*Bst*YI-TT/*Mse*I-TA_1	169	EF088687.1	Glutathione S-transferase (GST2), complete cds	*Vitis vinifera*	73%	4.00 e^{-13}
*Bst*YI-TT/*Mse*I-TA_2	164	EF088687.1	Glutathione S-transferase (GST2), complete cds	*Vitis vinifera*	75%	6.00 e^{-12}

3.3.3 Bulked segregant transcriptome analysis - sequencing, preassembly processing, and assembly

The cDNA libraries of the resistant (BR) and susceptible bulks (BS) were developed to establish a resource for identifying PI_{ARG} candidate genes on the expression level. This approach is focused on constitutively expressed genes, as it can be assumed for many RGCs. Therefore, seedlings were grown in the greenhouse without inoculation and plant material was harvested at one time point (see 2.3.1). A susceptible and resistant bulk was generated, each containing 16 individuals which are homozygous in the target region (Table 8-13, 8-14). cDNA was normalized and sequenced. The goal of the experiment was to identify contigs in the BRBS assembly which carry only sequences from BR or to identify sequences which are polymorphic between the two bulks, but not within the bulks.

In total, 1,015,101 reads were generated by 454 sequencing for cDNA bulks BR and BS. Reads of the two bulks could be distinguished by barcode tag analysis resulting in 499,441 reads for BR and 500,665 reads for BS corresponding to 98.5 % of the reads. The remaining 1.5 % could not be distinguished due to data quality. A median read length of 349 bp was observed. After 454 adapter and tag trimming, a median of 327 bp and an average read length of 309 bp were received.

Three assemblies were run for the 454 cDNA reads using the program Newbler (Roche). Firstly, reads of the resistant and susceptible bulks were analyzed separately and afterwards all reads were assembled together (BRBS). The BR assembly contains 45,147 contigs with an average of 495 bp and 38,988 singletons. The BS assembly contains 45,982 contigs with an average contig length of 499 bp and 35,523 singletons (Table 3-13). The analysis of the combined cDNA reads of BR and BS resulted in 61,287 contigs with an average length of 545 bp and 42,598 singletons. Singletons were excluded from further analysis and contigs larger than 500 bp were extracted from the three data files of BR, BS, and BRBS (Table 3-13). This set of large contigs was used for open reading frame prediction.

Table 3-13: Overview of the assembly of 454 cDNA sequence reads. An assembly was performed for the reads from the resistant bulk (BR) and the susceptible bulk (BS). A third assembly with reads of both bulks was implemented (BRBS). Some analyses were only performed on the large contig (>500 bp) data set.

	All contigs			Large contigs (>500 bp)		
	BR	BS	BRBS	BR	BS	BRBS
# reads	499,441	500,665	1,000,106			
# contigs	45,147	45,982	61,287	15,403	16,200	24,004
# singletons	38,988	35,523	42,598			
Contig length [bp]						
min	89	93	69	500	500	500
max	11,793	8,286	8,861	11,793	8,286	8,861
mean	495	499	545	709	701	758
median	414	421	430	834	824	912
N50	563	564	668	847	831	954
Ø Coverage in contigs	10	10	15.6			

To investigate the number of cDNAs highly homologous to sunflower ESTs in the public database, a sequence similarity search of the new sequenced cDNA to the public sunflower EST dataset was performed using blastn. The data were obtained with an e-value cutoff of 1.00 e^{-10}. Sequence similarity with publicly available *Helianthus* ESTs was observed for 71%, 73%, and 65% of the assembled contigs of BR, BS, and BRBS, respectively. A blastn analysis of the large BR and BS contigs was performed against each other. Sequences were counted as similar if a minimum overlap of 100 bp and an identity larger than 95% existed. For the BR assembly 5,346 and for the BS assembly 5,977 unique contigs were observed. The number of reads derived from BR and BS was counted in the whole BRBS assembly to identify contigs containing only reads from one of the bulks. In the BRBS assembly 5,926 (10%) contigs had only reads from BS and 5,990 (10%) contigs had only reads from BR.

3.3.4 Bulked segregant transcriptome analysis identifies candidate genes for PI_{ARG}

It was shown that the HICF contigs01, 03, 06 and 07 mapped to the target region of PI_{ARG} (Fig. 3-4). These sequences were used for proof of concept to show that BSTA can be successfully applied to identify genes from the target interval. A blastn analysis of the BRBS assembly as a query was performed against the HICF contig-wise assembly of the BAC sequences. cDNA contigs with a minimum similarity of 99% to the predicted genes were analyzed for SNPs. In total, 15 cDNAs were identified which showed similarity to twelve predicted genes from the target region. These contigs were

visualized by the program Tablet v1.10.03.04 and were evaluated on the read level to identify SNPs. Table 8-15 shows the results of the 15 cDNAs from the predicted genes in the target region. In Fig. 3-20 the results of SNP detection are illustrated for the predicted gene coding for dynamin which was detected on HICF contig01_c1. The cDNA sequences of BRBScontig46810 and BRBScontig26395 are in agreement with the predicted intron-exon assignment of the gene annotation pipeline. This demonstrates that the results of the gene prediction as well as the assembly of the BAC and BRBS bulk are of good quality. In total, ten SNPs were clearly identified between the BR and BS reads of the two BRBS contigs, four are exemplarily shown (Fig. 3-20).

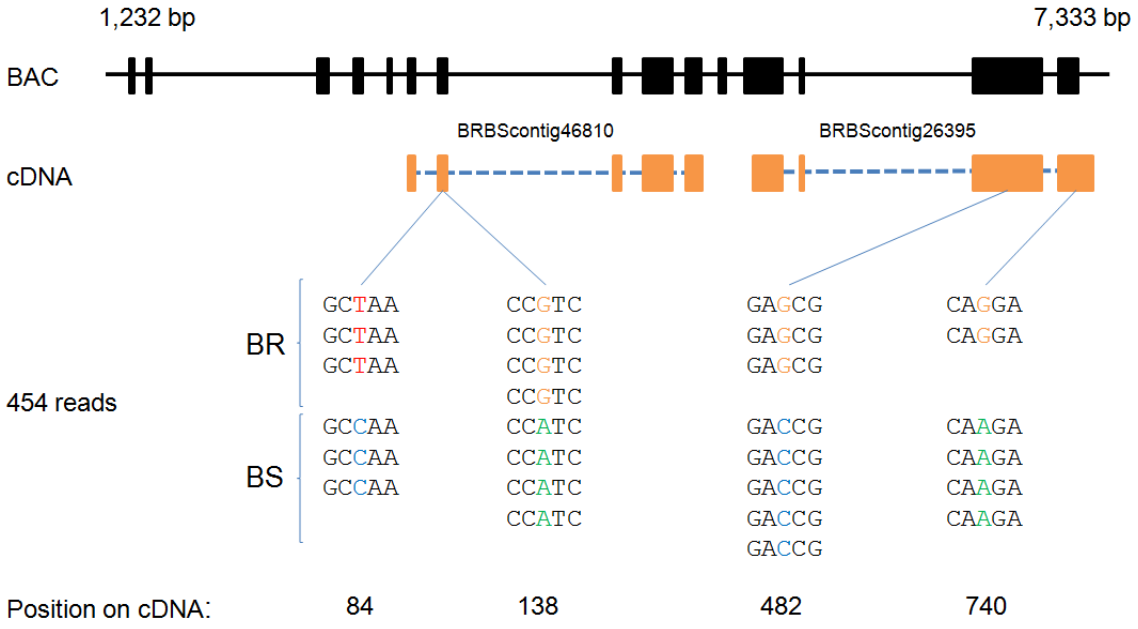


Figure 3-20: BAC sequence of HICF contig01_c1 (1,232 bp - 7,333 bp) with the intron (line)-exon (box) assignment of the predicted dynamin gene. cDNA sequence (orange boxes) confirmed the predicted exons. SNPs were observed between the resistant bulk (BR) and the susceptible bulk (BS). The sequences at four SNP positions are shown exemplarily.

It can be concluded that genes from the target region can be identified successfully by the BSTA approach. Therefore, a set of BRBS contigs was established for further analysis of sequences which show similarity to resistance genes: (1) Open reading frames for the large contigs of the BRBS assembly were predicted with EMBOS getorf and a Hidden Markov Model (HMM) search against the PFAM protein family database was performed. For 14,529 of 24,004 BRBS contigs a PFAM domain was identified. Contigs with PFAM domains typical for resistance gene candidates were chosen (Table 3-14). In total, 962 BRBS contigs were identified with the selected PFAM domains.

(2) The resistance gene reference data set (Table 8-12) was used to identify resistance candidate genes in the whole BRBS assembly by blast analysis. In total, 325 BRBS contigs showed similarity to the known resistance genes of the reference set. These were chosen for further analysis.

Figure 3-21: Number of contigs of the BRBS assembly which show homology to resistance gene candidates based on PFAM annotation and blast analysis against a reference data set of known disease resistance genes. Contigs confirmed by both analyses were evaluated in more detail.

The results of the PFAM domain search and the blast analysis were compared. In the intersection of the two analyses 274 BRBS contigs were identified (Fig. 3-21, Table 3-14). The 274 BRBS contigs were visually evaluated using Tablet v1.10.03.04. In total, 265 BRBS contigs carried reads of the BR and the BS bulk. Two and seven BRBS contigs were identified which carried only sequences from BR or BS, respectively. The two contigs containing only BR sequences could be potential candidate genes for PI_{ARG}. Five BRBS contigs were identified which carried SNPs between the two bulks, but not within the bulks (Table 8-16). The potential candidate genes are shown in Table 3-15 (sequences, see 8.2.4). Five of the seven candidate genes showed blast hits to the resistance gene class RLP, one showed a hit to RLP and RLK, and one showed similarity to *Mlo*. For six of the identified contigs a LRR PFAM domain (PF00560.26) and for one contig a *Mlo* PFAM domain (PF03094.8) was predicted.

Table 3-14: PFAM accession and PFAM name of domains identified in contigs of assembly BRBS and intersection with contigs identified by blast analysis against a resistance gene reference set.

PFAM accession	PFAM name	# contigs PFAM	# contigs intersection
PF01582.13	TIR domain	10	10
PF00931.15	NB-ARC domain	37	30
PF00560.26	Leucine Rich Repeat	232	124
PF07723.6	Leucine Rich Repeat	2	0
PF07725.5	Leucine Rich Repeat	2	1
PF08263.5	Leucine rich repeat N-terminal domain	28	4
PF00069.18	Protein kinase domain	462	64
PF07714.10	Protein tyrosine kinase	160	22
PF03094.8	*Mlo* family	16	10
PF05659.4	*Arabidopsis* broad-spectrum mildew resistance protein RPW8	1	1
PF07014.5	Hs1pro-1 protein C-terminus	2	2
PF03083.9	MtN3/saliva family	10	6
Total		962	274

Table 3-15: Summary of candidate gene sequences identified by bulked segregant transcriptome analysis sequences and homology with nucleotide or protein sequences in the National Centre for Biotechnology Information (NCBI) database.

Contig	Bulk	Accession	Homology with tblastx	Organism	Coverage	E-value	Ref.seq.	PFAM number
Contig00945	BRBS	EU888322.1	Receptor-like protein kinase	*Glycine max*	99%	5.00 e^{-78}	RLP	PF00560.26
Contig02834	BRBS	XM_002535149.1	Receptor protein kinase CLAVATA1 precursor	*Ricinus communis*	99%	7.00 e^{-98}	RLK/RLP	PF00560.26
Contig02873	BRBS	NM_116535.1	Clone pENTR221-At4g03010 leucine-rich repeat	*Arabidopsis thaliana*	96%	9.00 e^{-160}	RLP	PF00560.26
Contig07153	BR	XM_002516487.1	Serine-threonine protein kinase, plant-type	*Ricinus communis*	99%	9.00 e^{-78}	RLP	PF00560.26
Contig15072	BRBS	EU888315.1	Clone 8070606 leucine-rich repeat protein	*Glycine max*	83%	5.00 e^{-76}	RLP	PF00560.26
Contig16592	BRBS	XM_002533289.1	Protein MLO	*Ricinus communis*	86%	3.00 e^{-85}	other	PF03094.8
Contig34117	BR	FJ708626.1	Leucine-rich repeat receptor-like protein kinase (LRR-RLK)	*Arabidopsis thaliana*	98%	9.00 e^{-56}	RLP	PF00560.27

3.3.5 Two bulked segregant transcriptome analysis derived candidate genes map on linkage group 1

Primer pairs were designed on *de novo* assembled contig sequences of the candidate genes in the BRBS assembly. For the seven candidate genes amplicons were generated from HA342, ARG1575-2, BR and BS. A length polymorphism was observed on the agarose gel between HA342 and ARG1575-2 for BRBScontig16592. BR showed both fragments indicating that the gene is not closely linked to the target interval. Thus, the sequence was excluded as candidate gene. Re-sequencing and mapping of BRBScontig00945, BRBScontig02834, BRBScontig07153 and BRBScontig34117 also resulted in exclusion of these genes as candidates. Amplicons of BRBScontig00945 were heterozygous within the two bulks. Re-sequencing of BRBScontig02834 did not confirm the detected SNPs. No SNPs were detected for the BR-bulk derived BRBScontig07153 between HA342 and ARG1575-2 and thus mapping was not possible. SNPs were identified for BRBScontig34117, but the gene did not map on LG 1. Re-sequencing of BRBScontig02873 and BRBScontig15072 confirmed the identified SNPs between the two bulks. Within the bulks the sequences were homozygous. Four SNPs were successfully used to develop CAPS markers for the two contigs which were employed to map the two candidate genes. BRBScontig02873 and BRBScontig15072 mapped both on LG 1. BRBScontig02873 mapped at the top of LG 1 outside the target interval. However, BRBScontig15072 was mapped in the target region 0.1 cM above PI_{ARG} (Fig. 3-22).

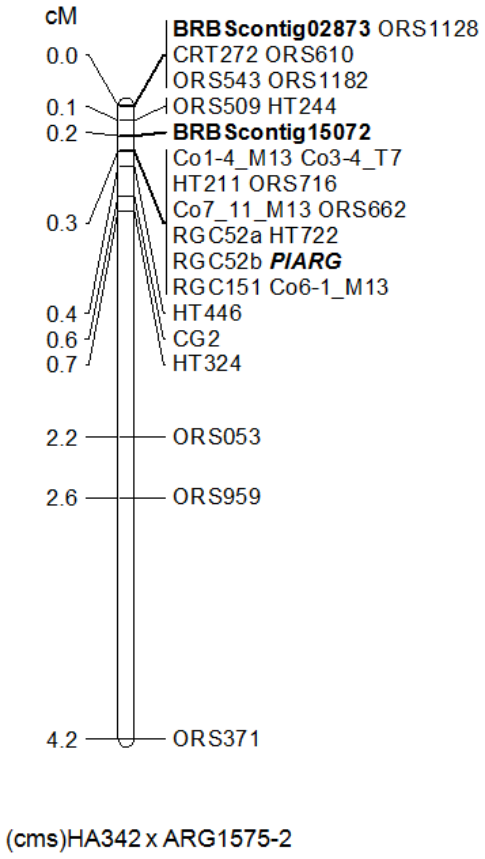


Figure 3-22: Mapping of candidate genes on linkage group (LG) 1 of (cms)HA342 × ARG1575-2 constructed with 2,145 F_2 individuals. Markers with an asterisk were screened in all individuals, while all other markers were screened in recombinant lines only. PI_{ARG} and two candidate genes are shown in bold. On the left cumulative distances are shown in centiMorgan.

4 Discussion

4.1 Relevance of wild species for sunflower breeding

Cultivated sunflower has an extremely narrow genetic base which has left the crop potentially vulnerable to diseases (Seiler 1992). One of the major diseases is downy mildew causing serious yield losses. To protect sunflower against the pathogen *P. halstedii* resistance breeding is the most effective way. Wild species have often been used to broaden the genetic background of cultivated crops to increase oil content and quality (Seiler 2007), tolerance to abiotic stresses (Miller and Seiler 2003) or herbicides (Al-Khatiba et al. 1998), or to introduce cytoplasmatic male sterility and restorer genes for developing hybrids (Leclercq 1969). Most commonly wild species are used as donors of disease resistance genes (Jan et al. 2002; Jan et al. 2004a; Jan et al. 2004b; Kuhl et al. 2001; Ling et al. 2004).

A novel source of resistance to *P. halstedii* from the wild species *H. argophyllus*

The dominant monogenic locus Pl_{ARG} originated from the wild species *H. argophyllus* and is an outstanding source of resistance, because of broad-spectrum resistance against all known races of *P. halstedii*. The resistance was first described by Seiler (1991) in ARG1575-2. Classical genetic analysis and first molecular studies on Pl_{ARG} were performed by Röcher (1999) to identify molecular markers for MAS based on RAPD and AFLP markers. Röcher (1999) concluded that Pl_{ARG} is independently inherited from Pl_2 and Pl_6 on LG 8. Dußle et al. (2004) confirmed this observation by mapping Pl_{ARG} to LG 1 using SSR markers. In the present study the exact position of Pl_{ARG} was determined based on 2,145 F_2 individuals of (cms)HA342 x ARG1575-2 and closely linked as well as cosegregating markers were identified. The target interval could be narrowed to 0.3 cM which is an excellent starting point for MAS and molecular cloning of Pl_{ARG}.

Structural and functional characterization of the Pl_{ARG} locus is of high importance, because monogenic resistance genes are often not durable. Parlevliet (2002) reported that the presence of many major resistance genes and the occurrence of hypersensitive response, both of which are typical for *Pl* loci, are characteristic for non durable resistance. Many major dominant genes for resistance to downy mildew have been described (Pl_1-Pl_{15}). Radwan et al. (2005b) showed that the resistance mechanism of Pl_8 is associated with a hypersensitive response in the hypocotyl. Hypersensitive response was also observed for Pl_{ARG} in this study. Therefore,

strategies have to be worked out to conserve the broad function of the resistance gene *PI_{ARG}*. Till now *PI_8* and *PI_{ARG}* confer resistance against all known races of downy mildew, whereas *PI_6* is an example where the resistance has been overcome by races 304 and 314 (Vear et al. 2007). A few studies discussed how to extend the durability of *PI* loci. Vear et al. (2008a), Tourvieille de Labrouhe et al. (2008) and Sakr (2010) suggested a combination of monogenic *PI* loci and quantitative resistance against downy mildew. McDonald and Linde (2002) recommended pyramiding major resistance genes in hybrid cultivars or growing cultivar mixtures containing genotypes with different major resistance genes. To apply these strategies, molecular markers and knowledge of the genetic and functional basis of resistance are required. One goal of this project was the mapping of *PI_{ARG}* and the identification of molecular markers for MAS which can be used in molecular breeding or for a map based cloning approach. Sequence information of candidate genes of *PI_{ARG}* will allow the comparison with already known *PI* clusters like *PI_{14}* on LG 1, *PI_1*, *PI_2*, *PI_6*, *PI_7* on LG 8, and *PI_5*, *PI_8* on LG 13.

4.2 Challenges using wild species for plant breeding

Suppressed recombination due to alien genome introgression

Fine mapping of the *PI_{ARG}* locus in population (cms)HA342 × ARG1575-2 identified large blocks of cosegregating markers despite the high number of segregating individuals. A block of five cosegregating markers (ORS610, ORS543, ORS1128, CRT272, and ORS1182) was observed above *PI_{ARG}* and several markers were identified which cosegregated with *PI_{ARG}*. Since it was assumed that recombination could be reduced in (cms)HA342 × ARG1575-2 a large second mapping population $NDBLOS_{sel}$ × KWS04 was established with the goal of increasing mapping resolution in the target region. It was found that the cosegregating markers above *PI_{ARG}* span a map distance of 12.5 cM in $NDBLOS_{sel}$ × KWS04 based on 528 F_2 individuals. Comparison of the marker interval ORS610 – ORS053 between the two linkage maps showed that it was possible to increase mapping resolution by a factor of ten in $NDBLOS_{sel}$ × KWS04. The target interval (HT244 – HT446) of 0.3 cM observed in (cms)HA342 x ARG1575-2 was determined to be 1.8 cM in $NDBLOS_{sel}$ x KWS04 based on 2,780 F_2 individuals. These results indicate that a suppressed recombination occurred in (cms)HA342 × ARG1575-2 which can be explained by the introgression of *H. argophyllus* in ARG1575-2. The degree of suppressed recombination was not uniformly distributed across LG 1. Suppressed recombination was mainly observed above *PI_{ARG}*. Thus, a target interval of 1.8 cM is still a good basis for a map-based cloning approach. This assumption is supported by the results of Meyer et al. (2009) who successfully

cloned *Rpp4* in soybean starting the map based cloning approach with a SSR marker linked 1.9 cM to the gene of interest.

Marker allele comparison of line ARG1575-2 and its recurrent parent cmsHA89 clarified that the entire LG 1 (ORS610 – ORS371) in ARG1575-2 originates from the wild species *H. argophyllus*. Therefore, with regard to LG 1 (cms)HA342 × ARG1575-2 is an interspecific cross between silverleaf and cultivated sunflower. No recombination occurred during the crossing of ARG1575-2 with cmsHA89 and during the two backcross steps with cmsHA89. During the development of line ARG1575-2 resistance to downy mildew was not tested before the BC_2F_5 generation (Seiler et al. 1991). Therefore, the retention of the complete LG 1 from *H. argophyllus* in ARG1575-2 was not expected since the remaining linkage groups (LG 2-17) contain a mixture of *H. annuus* and *H. argophyllus* alleles close to the expected donor genome proportion of 12.5 %. This raises the question whether the inheritance of the whole LG 1 occurred by chance during the development of ARG1575-2. ARG1575-1 also originated from the cross between cmsHA89 and *H. argophyllus* (Acc. 1575) but carries alleles similar to cmsHA89 in the target interval (data not shown). In phenotypic tests no multiple race resistance was observed for ARG1575-1 (Seiler et al. 1991). Other crosses from the original *H. argophyllus* population did not contain Pl_{ARG} either, indicating that the discovery of this locus was probably random and Pl_{ARG} occurred at a very low frequency (Gerald Seiler, personal communication).

The results of this study regarding suppressed recombination are in agreement with previous studies with wild species. Suppressed recombination is often observed in populations that carry wild genome introgressions as in the case of *Mla* introduced into cultivated barley from *H. spontaneum* (Wei et al. 1999), *Run1*, the grapevine powdery mildew resistance gene introduced from *Muscadinia rotundifolia* into *Vitis vinifera* (Barker et al. 2005), and *Mi* and *Tm2-a*, both introduced from *Lycopersicon peruvianum* into *L. esculentum* (Ganal et al. 1989; Kaloshian et al. 1998; Messeguer et al. 1991; Seah et al. 2004). One reason for suppressed recombination may be the occurrence of chromosomal inversions (Seah et al. 2007; Seah et al. 2004). Suppressed recombination was also observed around the centromers (Yan et al. 2005). Comparative genetic linkage maps of *H. annuus*, *H. petiolaris*, *H. anomalus*, *H. deserticola* and *H. paradoxus* were established to study karyotypic evolution and a high rate of chromosomal rearrangements was observed within the genus *Helianthus* (Burke et al. 2004; Lai et al. 2005b). Heesacker et al. (2009) performed comparative mapping between intraspecific crosses of *H. annuus* and *H. argophyllus* and identified 10 collinear chromosomes, 9 chromosomal rearrangements, 3 putative segmental

duplications, and 2 putative whole chromosome duplications. LG 1 was collinear between *H. annuus* and *H. argophyllus*; therefore, large chromosomal rearrangements are not likely to be the reason for suppressed recombination, but reduced homology or small inversions which were not detected by genetic mapping may be a plausible explanation. Interestingly, in the intraspecific mapping population of *H. argophyllus* recombination was also suppressed in the Pl_{ARG} region (Heesacker et al. 2009); therefore, the reduced homology between silverleaf and cultivated sunflower does not seem to be the sole cause of suppressed recombination. Heesacker et al. (2009) assumed that the chromosomal architecture could be different between *H. annuus* and *H. argophyllus*. Thus, further analysis of the LG 1 chromosome architecture of both species may deliver explanations for suppressed recombination.

Reduced seed set has been frequently observed in progenies of interspecific crosses (Chetelat and Meglic 2000; Lai et al. 2005b). Fertility reduction and distorted segregation was not observed in subpopulation HA342 × ARG1575-2. Thus, the *H. argophyllus* introgression does not seem to severely influence fertility and segregation ratios in the target region. Reduced seed set and distorted segregation was only observed in the presence of the sterile PET1 plasma in subpopulation cmsHA342 × ARG1575-2. ARG1575-2 must carry a restorer gene for the PET1 cytoplasm, otherwise no seed production would be possible in the progenies. For 25% of the F_2 individuals no seed set was observed. This fits the expected 3:1 segregation ratio of a single dominant restorer gene. Since the recurrent parent cmsHA89 possesses no restorer gene for the PET1 cytoplasm, the restorer gene must originate from *H. argophyllus*. The reduced seed set and the reduced number of plants homozygous for the ARG1575-2 allele in cmsHA342 × ARG1575-2 could be due to incomplete restoration of the PET1 cytoplasm based on unfavorable effects of the restorer gene, or other genes that may be unfavorable for seed production or viability in homozygous condition. The restorer gene is not closely linked to Pl_{ARG}, because only 64 of 209 F_2 individuals which carry the alleles of ARG1575-2 in the target region had no seed production (data not shown). Abratti et al. (2008) described a monogenic restorer gene for the PET1 cytoplasm originating from *H. argophyllus*. They mapped the *Rf3* gene on LG 7 using the population RHA340 × ZENB8. *H. argophyllus* derived fragments are present on LG 7 as shown by SSR markers. For *Rf3* no insufficient restoration has been described, but it remains to be analyzed if *Rf3* is the restorer gene in ARG1575-2.

Reduction of linkage drag through marker-assisted backcrossing

The introduction of deleterious genes along with the beneficial gene is often observed in backcross breeding using unadapted donor lines. This phenomenon is called linkage drag (Zeven et al. 1983). The results of this study revealed that the donor segment of *H. argophyllus* on LG 1 is very large and possibly comprises the whole *H. argophyllus* chromosome. For the development of ARG1575-2 two generations of backcrossing were performed (Seiler et al. 1991) resulting in no recombination event on LG 1. It was found that two inbred lines RHA419 and RHA420 both derived from a cross of RHA373 x ARG1575-2 also still carry a very large segment originating from *H. argophyllus* on LG 1. Similar cases were reported for backcross breeding where phenotypic analyses were performed, but selection occurred without marker assisted background selection (Ballini et al. 2007; Young and Tanksley 1989). Molecular markers can be used to reduce quickly and efficiently the amount of linkage drag associated with the introgression (Frisch et al. 1999; Young and Tanksley 1989). Thus, to apply Pl_{ARG} in plant breeding further marker assisted backcrossing of ARG1575-2 is necessary to reduce the introgression to a minimum amount of DNA originating from *H. argophyllus*. This requires recombination and thus large numbers of progenies need to be tested. In this study closely linked as well as cosegregating SSR and RGC markers to Pl_{ARG} were identified which can be efficiently used for foreground selection to reduce the size of the donor segment. Further reductions can be achieved with cost-efficient high-throughput SNP markers which are under development (Bachlava et al. 2010b; Pauquet et al. 2011) and which can also be used for background selection to reduce the additional donor segments of *H. argophyllus* from the non target chromosomes in cultivated sunflower to a minimum. Future map-based cloning studies with introgression lines should use crossing partners in the normal plasma and the donor segments should be reduced to a minimum to avoid problems caused by non desired donor fragments.

4.3 Structural characterization of the Pl_{ARG} locus

Segregation studies to fine map the target region Pl_{ARG}

Vear et al. (1997) gave the first evidence that major *Pl* loci are not single genes, but clusters of genes. They performed segregation studies with F_3 progenies of the cross HA335(Pl_6) x H52. The Pl_6 locus is effective against the races 100, 300, 700, 703 and 710. However, progeny tests of this cross indicated that two F_3 families were resistant to races 700, 703 and 710, but segregated with regard to races 100 and 300.

Explanation for this observation was a recombination in the gene cluster. The approach of Vear et al. (1997) was assigned to this study to elucidate the structure of Pl_{ARG}. F_2 individuals with a recombination between the closely flanking markers ORS509/HT244 and HT446 were selfed to identify progenies which are homozygous for both flanking markers. These progenies were tested with four different races of *P. halstedii* (100, 330, 710, and 730). Five of 15 recombinant lines had sufficient seeds and were analyzed in more detail. Four of the families were resistant against race 730 and one family was susceptible. In contrast to the study described before no recombination within the putative gene cluster was observed and thus the five recombinant lines showed the same resistance pattern for races 100, 330 and 710 like it was observed for race 730. Thus, using this strategy it could not be demonstrated that Pl_{ARG} is a cluster of genes. Several reasons may explain the observed result: One explanation could be that Pl_{ARG} is a single resistance gene mediating resistance against several races of *P. halstedii*, similar to *RPP13*. *RPP13* is a single resistance gene that protects *Arabidopsis thaliana* against different isolates of the biotrophic oomycete *Pernospora parasitica* (now known as *Hyaloperonospora arabidopsis*) (Bittner-Eddy et al. 2000). Pl_{ARG} may differ in that respect from the previously analyzed *Pl* loci, which clustered on LG 1 (Pl_{14}) (Bachlava et al. 2010a), LG 8 (Pl_1, Pl_2, Pl_6 and Pl_7) (Mouzeyar et al. 1995; Roeckel-Drevet et al. 1996; Vear et al. 1997) and LG 13 (Pl_5 and Pl_8) (Bert et al. 2001; Radwan et al. 2003). The second and more likely theory is that Pl_{ARG} is a cluster and the observed recombinations in the five lines occurred only below or above the cluster. No observed recombination event within the resistance gene cluster may be due to a lack of recombination in the resistance gene cluster (Seah et al. 2007), due to extreme physical proximity of clustered resistance genes or due to large physical distances between the closely linked markers ORS509/HT244 and HT446. Only five of 15 recombinants could be analyzed and thus there was a low chance to find recombination in the cluster. In further analyses, additional recombinant lines from an extended mapping population should be analyzed. To elucidate which hypothesis is correct one suitable approach is map-based cloning of Pl_{ARG} to clarify the structure of resistance genes Pl_{ARG}.

RGCs as starting points for map based cloning

In the target region of Pl_{ARG} three RGC markers RGC151, RGC52a and RGC52b were mapped and it was shown that the three markers cosegregate with Pl_{ARG}. The majority of resistance genes show typical structures and can be assigned to one of the four major resistance gene classes which were already described in the introduction: NBS-LRR, LRR-TM, LRR-TM-PK, and PK. Only few resistance genes like *RPW8* cannot be

integrated into one of the four classes and therefore belong to the fifth class. The identified RGCs from the Pl_{ARG} target region are members of the NBS-LRR class. Till now for all *Pl* loci (Pl_{14} (LG 1), Pl_1, Pl_2, Pl_6, Pl_7 (LG8), Pl_5, Pl_8 (LG13)) which were described in more detail large clusters of RGCs were observed which belong to the TIR-NBS-LRR or CC-NBS-LRR type (Bachlava et al. 2010a; Radwan et al. 2008; Slabaugh et al. 2003). This strongly supports the conclusion that RGCs cosegregating with Pl_{ARG} are good resistance gene candidates for Pl_{ARG}. For this reason a first screening of a BAC library was performed with probes derived from the cosegregating RGCs.

The publicly available BAC library HA_HBa was screened since it has excellent genome coverage of 7.3 – 8.5x and large insert sizes of 125 kb in average. These parameters are important for the success of physical mapping, positional cloning, and genome sequencing approaches based on complex genomes (Luo and Wing 2003). The library HA_HBa does not carry the functional resistance locus Pl_{ARG} but the library is very suitable to explore the genetic structure of the target region, to estimate physical distances and it may be possible to clone candidate genes for Pl_{ARG}. Previous studies support this assumption. A BAC library of a susceptible genotype was successfully used to clone the *Rpp4* gene that confers resistance against the Asian soybean rust caused by the fungus *Phakopsora pachyrhizi* (Meyer et al. 2009). Ingvardsen et al. (2010) identified four new candidate genes for the resistance locus *Scm2* against the sugarcane mosaic virus in maize using the public BAC library of the susceptible line B73.

Genetic mapping of HICF contig03 (RGC151) and contig07 (RGC52) confirmed that both contigs cosegregate with Pl_{ARG} in (cms)HA342 x ARG1575-2. The establishment of the second population $NDBLOS_{sel}$ x KWS04 turned out to be very valuable to map BAC endsequences (BES) with a higher genetic resolution. This strategy enabled the determination of the correct marker order in the target region. In population $NDBLOS_{sel}$ x KWS04 HICF contig03 (RGC151) and contig07 (RGC52) were mapped very close together at a distance of 0.02 cM but did not overlap physically. This led to the hypothesis that both contigs may be part of one cluster carrying resistance gene candidates. The identified contigs are flanked by HT211 and ORS662 which have a genetic distance of 0.2 cM. Assuming a sunflower genome size of 3500 Mb and a map length of 1400 cM (Tang et al. 2002) a theoretical physical distance of 500 kb is expected between HT211 and ORS662. Thus, the maximal extension of the Pl_{ARG} locus can be estimated to be less than 500 kb. The theoretical physical distance is a good precondition for a map based cloning approach but a single screening step was

not sufficient to bridge the interval between these two markers.

A second screening of the BAC library was performed with probes derived from BES of HICF contig07 (RGC52) as well as from HT211 and ORS662 to establish a minimum tiling path between these two markers. Due to the unspecific nature of the probe designed from BES of RGC52 no additional BAC clones could be identified. Further BAC clones were obtained with HT211 and ORS662 but they showed no overlap with HICF contig03 (RGC151) or contig07 (RGC52). Further chromosome walking is required to bridge the gap between HT211 and ORS662 and to define the exact physical distance between these two markers as a boundary of the PI_{ARG} locus.

BAC sequence analysis predicts resistance gene candidates

Analysis of the sequenced BAC clones was performed and gene models for four resistance gene candidates on three HICF contigs could be predicted. HICF contig03 carries one resistance gene candidate and HICF contig07 carries two resistance gene candidates and both were mapped to LG 1. HICF contig02 is located on LG 16 and carries one resistance gene. Thus, three of the four predicted genes are possible candidates for PI_{ARG} because they were mapped to the target PI_{ARG} region. Phylogenetic analysis classified the newly identified resistance gene candidates as belonging to the TIR-NBS-LRR class. This demonstrated the presence of a cluster of NBS-LRR genes in the target region, and supports the hypothesis that PI_{ARG} may be a complex locus.

The gene models for RGC52SW23 (LG 1), RGC52SW4 (LG 1), and RGC52SW1 (LG 16) contained interruptions in the open reading frame encompassing the NBS region for all three genes. One explanation is that the predicted genes represent pseudogenes. The occurrence of pseudogenes has been frequently described for RGC sequences (He et al. 2004; Liu and Ekramoddoullah 2004; Michelmore and Meyers 1998; Wicker et al. 2007). The assumption that RGC52SW23 (LG 1), RGC52SW4 (LG 1) and RGC52SW1 (LG 16), might be pseudogenes is supported by the fact that no expression was observed in RT-PCR in ARG1575-2 and HA342. Pseudogenes probably serve as reservoirs of potential variations. They are believed to be advantageous because they may allow recombination and gene conversion between alleles or paralogs of functional *R* genes (Michelmore and Meyers 1998). The high amino acid sequence similarity of 93 – 96% between the three RGCs suggests that they originated from tandem duplications.

RGC151SW1 is a candidate for *PI_{ARG}*

Contrary to the RGC52-like sequences, RGC151SW1 (LG 1) had an intact open reading frame and a clear intron-exon assignment could be predicted. Five exons were predicted containing TIR, NBS and LRR motifs. RGC151SW1 is a promising candidate for *PI_{ARG}*. Recently, Radwan et al. (2011) published the full length clone of RGC151 and alignment of the genomic DNA of RGC151SW1 and the full length clone of RGC151 confirmed the predicted intron-exon assignment. RT-PCR showed that the resistant line ARG1575-2 seems to react faster after inoculation with *P. halstedii* than the susceptible line HA342. These results are in agreement with the findings of Tao et al. (2003) who observed quantitative and kinetic differences in expression of genes which are involved in the defense response between compatible and incompatible plant-pathogen interactions. These findings do not agree with Radwan et al. (2011). They used real-time PCR to detect the relative gene expression of RGC151 and no significant enhancement of RGC151 was observed after inoculation with race 300 of *P. halstedii* in a time course of 1.5 -15 days after inoculation. However, it has to be taken into consideration that the time course (1.5 – 15 versus 0 – 15 days), the race of *P. halstedii* (300 versus 730) and the method (Real time PCR versus RT-PCR) were not the same between the experiment of Radwan et al. (2011) and this study. Radwan et al. (2011) observed a type I reaction for RHA419 after inoculation with race 300. Whereas, Röcher (1999) described the type II reaction for ARG1575-2 after inoculation with race 770 and type II resistance was also observed in this study after inoculation with race 730 (see 2.2). This different observation might be explained by a cluster of resistance genes which reacts variably depending on the race of *P. halstedii*.

Comparison of RGC151 restriction fragments from cmsHA89 and ARG1575-2 via Southern blot analysis showed that multiple copies of RGC151 exist in the genome of cmsHA89 and ARG1575-2. Three copies similar to RGC151 were observed in cmsHA89 and six were detected in ARG1575-2. All fragments observed in cmsHA89 were also observed in ARG1575-2. This means that the additional copies observed in ARG1575-2 must be located in the fragments derived from *H. argophyllus*. It was shown by mapping and BAC clone sequencing that at least one copy of RGC151 is located in the target region of *PI_{ARG}* on LG 1. The location of the other copies is not known. However, the whole LG 1 of ARG1575-2 originates from *H. argophyllus*, whereas only 12.7% of the remaining genome is derived from *H. argophyllus*. As a consequence one may assume that the probability is high that the additionally observed copies of ARG1575-2 are located on LG 1. It can be concluded that a cluster of RGC151 homologs exists on LG 1 which has more copies in the introgression of

ARG1575-2 originating from *H. argophyllus* than in cultivated sunflower. This hypothesis is supported by the results of Radwan et al. (2011) and Meyers et al. (2009). Radwan et al. (2011) isolated one full length clone of RGC151. Alignment of the 550 bp RGC151 amplicon derived from ARG1575-2 (this study) and the full length cDNA clone of RGC151 showed sequence identity of only 85%. Thus, several copies of the RGC151 homolog exist in ARG1575-2. Meyers et al. (2009) identified three *Rpp4* candidate genes of the CC-NBS-LRR class in the susceptible soybean genotype and five candidates in the resistant genotype.

Comparison of the predicted amino acid sequences of RGC151SW1 and of the full length clone RGC151 of the resistant line RHA419 carrying PI_{ARG} showed that several amino acid changes exist. Thus, ARG1575-2 may carry the functional allele of the resistance gene.

No further resistance gene candidates were predicted on HICF contigs01, 05, and 06 derived from HT211 and ORS662. It can be assumed that these sequences lay outside the PI_{ARG} resistance gene cluster. However, sequence information from HICF contigs01 and 06 can be used for further chromosome walking.

A newly identified RGC family

The phylogenetic tree of the NBS region shows that the newly identified resistance gene candidates RGC52SW23, RGC52SW4, RGC52SW1, and RGC151SW1 belong to the TIR-NBS-LRR class. However, the new resistance gene candidates are distantly related to the sunflower RGCs isolated by others. This indicates that a new TIR-NBS-LRR resistance subfamily has been found. One reason why this subfamily has not been identified so far might be that the majority of RGCs were identified by using degenerate primers (He et al. 2004; Leister et al. 1996) and the minority was identified by mining the sunflower EST resource (Radwan et al. 2008). The sequences of the degenerate primers which have been used to identify RGCs previously differ from the sequence of the P-loop and GLPL motif of the newly identified RGCs. Therefore, they are not suitable to amplify RGCs from the PI_{ARG} target region. The differences in the P-loop and GLPL motif explain why a previous RGC approach for PI_{ARG} failed to identify RGCs in the target region of PI_{ARG} using degenerate primers designed by Leister et al. (1996) (data not shown). Previous studies showed that the choice of degenerate primers has crucial influence on the RGC families isolated (Baldi et al. 2004; Hattendorf and Debener 2007). It can be postulated that a larger spectrum of RGCs can be identified using degenerate primers that have been designed on the newly identified resistance genes. The proposed approach can be used to explore the number and

distribution of resistance gene candidates similar to the type of the newly identified resistance gene candidates.

The identification of a new resistance gene candidate subfamily is of high relevance for understanding the organization and function of resistance genes. One may assume that combining major *Pl* loci of different types extends the durability of downy mildew resistance. Until now four *Pl* loci have been characterized in more detail: Resistance gene candidates of the TIR-NBS-LRR subfamily have been identified closely linked to Pl_{ARG} on LG 1 and to Pl_1, Pl_2, Pl_6, Pl_7 on LG 8 (Radwan et al. 2008; Slabaugh et al. 2003). Results of the present study revealed phylogenetic differences between RGCs on LG 1 and LG 8. Resistance gene candidates of the CC-NBS-LRR subfamily were detected in close proximity to Pl_{14} on LG 1 and to Pl_5, Pl_8 on LG 13 (Bachlava et al. 2010a; Radwan et al. 2003; Radwan et al. 2008). In a previous study it was shown that after five years of sunflower cultivation fewer changes in virulence of *P. halstedii* populations were observed in plots grown with hybrids which combine two *Pl* loci (Pl_6 and Pl_8) than in plots grown with hybrids carrying only one *Pl* locus (Pl_6 or Pl_8) (Tourvieille de Labrouhe et al. 2010). Thus, it is highly recommended for plant breeders to combine *Pl* loci of the TIR-NBS-LRR and CC-NBS-LRR classes.

4.4 Alternative approaches to identify candidate genes for Pl_{ARG}

Complementary approaches were performed in parallel to the map-based cloning approach to identify candidate genes for Pl_{ARG}. A non-targeted genome-wide transcriptome analysis based on the cDNA-AFLP method was applied to identify differentially expressed genes. With the targeted genome-wide bulked segregant transcriptome analysis (BSTA) candidate genes were identified via SNP detection and presence / absence detection of transcripts.

Non-targeted genome-wide transcriptome analysis

The RGC Ha-NTIR11g belongs to the CC-NBS-LRR subclass and maps to the cluster of Pl_5/Pl_8. (Radwan et al. 2005a). The gene is constitutively expressed at a low level in the non-inoculated resistant genotype, but is specifically induced after inoculation with *P. halstedii*. This supports the conclusion that Pl_{ARG} could also be differentially expressed between the non-inoculated and the inoculated plants and was the starting basis for further analyses.

Only fragments which were differentially expressed between the non-inoculated and the inoculated samples of each line were analyzed in the present study to avoid

isolation of fragments which exhibit differences due to allelic polymorphisms. Using a total of 32 different primer combinations, 25 differentially expressed fragments were observed for the resistant line ARG1575-2 and 117 for the susceptible line HA342, respectively, after inoculation with *P. halstedii* race 730. The overexpression of fragments from the susceptible line is in agreement with findings of Fung et al. (2008). Thus, via reprogramming of the transcriptome the susceptible genotype activates several defense mechanisms to repel the pathogen whereas in the resistant genotype only few transcriptional changes occur which are sufficient for protection.

Sequence analysis of differentially expressed fragments of the resistant line ARG1575-2 revealed that mainly genes of the general defense were isolated: dirigent like protein, putative beta tubulin, tRNA-Ser gene, P subunit of glycine decarboxylase, and glutathione S-transferase. In previous studies it was shown that the dirigent like protein plays an important role in plant disease resistance response (Zhu et al. 2007). Although it is closely linked to the PI_{ARG} region in population (cms)HA342 × ARG1575-2 it does not cosegregate and thus is not a candidate for PI_{ARG}.

The P subunit of glycine decarboxylase is responsible for the synthesis of nitric oxide (NO) and was up-regulated in the cotyledons of ARG1575-2. NO is a key molecular signal that is involved in the regulation of several physiological processes and plays a significant role in plant resistance to pathogens by triggering resistance-associated cell death during the hypersensitive response (Romero-Puertas et al. 2004). After inoculation with *P. halstedii*, necrotic lesions could be observed on the cotyledons of ARG1575-2 which is typical for hypersensitive response. Therefore, it can be assumed that the isolated differentially expressed fragment play a role in the synthesis of NO to activate hypersensitive response as a defense response against *P. halstedii*. Several studies indicate that NO protects cells from reactive oxygen species (ROS)-mediated cellular damage and cytotoxicity by increasing the level of cyto-protective proteins including the glutathione S-transferase (Wendehenne et al. 2004) for which transcription was also increased in ARG1575-2. Glutathione S-transferase plays a role in detoxification (Marrs 1996). A rapid and systemic induction of glutathione S-transferase was observed by Radwan et al. (2005b) in the incompatible interaction between *P. halstedii* and the resistant genotype QIR8 (PI_8). Thus, the results of the cDNA-AFLP analysis suggest that the defense mechanism against *P. halstedii* observed for PI_8 may be similar for PI_{ARG}.

In conclusion the cDNA-AFLP approach is suitable for investigating and understanding the general defense mechanism of the PI_{ARG} resistance. Identification of candidate

genes may be possible like it was shown by Chapman et al. (2009) for *Pch2* in wheat but the cDNA-AFLP approach is restricted to differentially transcribed fragments which must carry the restriction sites of the chosen enzymes and the candidate genes must be differentially expressed after inoculation.

Targeted genome-wide transcription analysis identifies resistance gene candidates

The 454 sequencing platform is one of several next generation sequencing technologies and is the method of choice for transcriptome sequencing in cases where no reference sequence is available (Barakat et al. 2009; Jarvie and Harkins 2008; Sun et al. 2010). The GS FLX Titanium technology enables read lengths of 300 - 500 bp and has very low rates of sequence errors, making it perfectly suited for identification of SNPs (Kircher and Kelso 2010). Therefore, the GS FLX Titanium technology was chosen to sequence the transcriptome of susceptible and resistant lines. The BSA approach (Michelmore et al. 1991) was applied to the cDNA level to identify candidate genes for Pl_{ARG}. This new approach was called bulked segregant transcriptome analysis (BSTA). Using this approach and a sufficient sequence coverage all expressed genes of the target region may be identified. In this study it has been shown that BSTA is a straight forward approach to identify expressed genes in a target interval. BSTA is applicable as long as the following conditions are fulfilled: the analysis of a monogenic trait in a predefined target interval, and the availability of bulks that exclusively differ in the target interval.

For transcriptome sequencing non-inoculated tissues of eight days old seedlings (hypocotyls, cotyledons and leaves) were harvested. One assumption in this approach was that Pl_{ARG} is constitutively expressed in non-inoculated seedlings. This is in agreement with observations of Tan et al. (2007). They analyzed the expression pattern of approximately 170 NBS-LRR-encoding and related genes in *Arabidopsis* and showed that most of these genes were expressed at low levels with a variety of tissue specificities. To ensure that candidates are equally represented the cDNA was normalized in the present study. Normalized cDNA libraries are used mostly to discover new genes transcribed at relatively low levels (Shcheglov et al. 2007).

Candidate genes may be only expressed in the resistant genotype Pl_{ARG} like it was described for the *N* gene in tobacco (Levy et al. 2004) but not in the susceptible genotype. Thus, contigs which contain only sequences from the BR bulk are potential resistance gene candidates and have to be validated via linkage mapping. Sequences containing only reads of the BS bulk were excluded as candidate genes because Pl_{ARG}

is a locus mediating dominant resistance against *P. halstedii*. In contrast, susceptibility factors are known to cause a recessive resistance phenotype after mutation like *pmr6 i*n *Arabidopsis* which mediates resistance against the powdery mildew pathogen *Erysiphe cichoracearum* (Vogel et al. 2002).

The BSTA approach has been successfully used to identify seven new candidate genes for Pl_{ARG}. The genes carry typical domains of resistance genes like LRR, LRR-TM, and LRR-TM-PK. Two contigs carry sequences derived from the BR bulk. Validation of the two BR bulk derived candidate genes was performed by re-amplification, re-sequencing, and mapping and did not confirm these two genes as candidates. BRBScontig00945, BRBScontig02834, and BRBScontig16592 could not be confirmed either by amplification and re-sequencing. They were false positive candidate genes due to insufficient coverage. However, the two BRBS contigs with higher sequence coverage mapped to LG 1. BRBScontig02834 mapped slightly outside the target region and BRBScontig15072 mapped between the flanking markers, but did not cosegregate with Pl_{ARG}. One reason for this observation might be that the established bulks BR and BS have a lack of recombination above the target region. BR contains four of 16 lines with a recombination event above Pl_{ARG} whereas BS carries no line with recombination above Pl_{ARG}. These results point out that one 454 run was not sufficient to generate consensus sequences which represent all lines of each bulk. However, the BSTA approach identified two sequences on LG 1. With the identification of BRBScontig15072 a closely flanking marker was identified which narrows the target interval to 0.2 cM in population (cms)HA342 x ARG1575-2 and facilitates a next screening round for chromosome walking in the target region.

The BSTA approach can be improved by deeper sequencing to achieve a higher coverage per sequence. Higher coverage results in a higher reliability of the detected SNPs and to fewer false positives in case of presence / absence variation in BR or BS. A pre-selection of contigs would not be necessary if SNP detection could be done with an automatic pipeline. However, prediction accuracy is often not very good resulting in false negative and false positive SNPs. The automated alignment and SNP detection from re-sequencing data where no reference genome sequence is available is an on-going challenge (Wegrzyn et al. 2009).

In summary, the cDNA approach is a promising approach to identify expressed genes in a target region. In this study it is of particular interest, because establishing the minimal tiling path of BAC contigs in the Pl_{ARG} target region was not possible because of the specific problems associated with interspecific crosses, such as the reduced

recombination and the lack of a BAC library for the resistance donor ARG1575-2.

4.5 Evaluation of different approaches to identify candidate genes for *PI_{ARG}*

In the search for candidate genes for *PI_{ARG}* three complementary approaches were applied in this study: map based cloning, cDNA-AFLP, and the BSTA approach. All three approaches aimed to contribute to the elucidation of the genetic structure of the resistance locus *PI_{ARG}*. The three methods will be evaluated below. The approaches differ in various aspects such as specificity, required marker density at the beginning of the project, amount of generated sequence information, and influence of suppressed recombination on the success of the approach (Table 4-1).

Table 4-1: Comparison of three different approaches to identify candidate genes.

Approach	Specificity	Required marker density	Generated sequence information	Restriction by suppressed recombination
Map based cloning	Specific for target region	High	Limited to identified BACs	Yes
cDNA-AFLP	Genome-wide	Low	Whole transcriptome but limited by choice of restriction enzymes, differentially expressed genes	No
BSTA	Genome-wide, specific for target region	Intermediate	Whole transcriptome	No

Map based cloning

The map based cloning approach is a classical forward genetic approach which was successfully used in several cases to clone plant resistance genes (Feuillet et al. 2003; Meyer et al. 2009; Milligan et al. 1998). Prerequisites are a segregating population, a high marker density to narrow the target interval and the availability of a BAC library.

One advantage of the map based cloning approach is the high specificity to the target region (Table 4-1). Identified BAC clones provide information of the target region. At the beginning of this project over thousand genome-wide sunflower SSR markers were available (Tang et al. 2002; Yu et al. 2003) and allowed a good saturation of the *PI_{ARG}* target region with molecular markers. In the future an even higher marker density will be available for sunflower through the detection of SNPs via transcriptome sequencing (Bachlava et al. 2010b; Pauquet et al. 2011). SNP markers can be used to narrow

down the target interval to a minimum depending on the resolution of the segregating population.

Several BAC libraries with a high coverage are available for sunflower. A BAC library carrying the functional allele of the resistance gene PI_{ARG} is the first choice for a map based cloning approach to identify candidate genes. However, the construction is cost intensive and exceeded the budget of this project. Using an existing *H. annuus* BAC library allowed the structural characterization of the PI_{ARG} target interval on LG 1 as a first step towards map based cloning of the resistance locus.

One constraint of the map based cloning approach can be suppressed recombination which has been observed in interspecific crosses or in pericentromeric regions (Pillen et al. 1996; Wei et al. 1999). Despite suppressed recombination candidate genes were identified by using RGC markers as starting point for map based cloning and the second mapping population $NDBLOS_{sel}$ x KWS04 allowed the mapping of BES in a higher resolution.

In summary, adequate molecular tools are available in sunflower to pursue the map based cloning approach and it is a straight forward approach to identify candidate genes in the target region. Establishing a minimum tiling path can be achieved very fast in crop species with small genome size using a high resolution mapping population or in species where BAC-based physical maps become available in the course of whole-genome sequencing projects. One successful example is the cloning of the scab resistance *Rvi15* (*Vr2*) in apple where a single BAC clone was sufficient to span the resistance locus (Galli et al. 2010). In crop species with large genomes such as sunflower chromosome walking can be time consuming and difficult but it is not impossible like it was shown for *Lr10* in wheat (Feuillet et al. 2003; Stein et al. 2000).

cDNA-AFLP

The cDNA-AFLP technique is a genome wide approach which can be used to identify candidate genes even if no prior sequence information or molecular markers exist or if suppressed recombination occurs in the target interval (Table 4-1). However, the cDNA-AFLP approach can not cover the entire transcriptome due to the use of a limited number of suitable restriction enzymes and the restricted analysis of differentially expressed fragments in the preferred length of 100 – 500 bp (Vuylsteke et al. 2007). Candidate genes can be identified like it was shown by Chapman et al. (2009) for *Pch2* in wheat but the results of the present and previous studies pointed out that the approach is much more suitable for characterization of the general defense

response. The appearance of cost efficient next generation sequencing technologies enables the application of new approaches like BSTA which produce sequence information of the transcriptome without restriction (Table 4-1).

Bulked segregant transcriptome analysis

For the BSTA approach a segregating population is needed as a basis for establishing bulk samples. Bulks are built according to phenotypic and genotypic information. Identification of closely flanking markers helps to narrow down the target region to a minimum. Therefore, an intermediate to high marker density is useful (Table 4-1). Interesting candidate genes from the target interval can be detected by searching for SNPs which are observed between the two bulks, but not within the bulks. Bioinformatic methods progress and the availability of a reference genome of sunflower (http://www.genomecanada.ca/medias/PDF/EN/Genomics-of-Sunflower.pdf) will enable the automated discovery of reliable SNPs and thus will speed up the identification of genes from the target region. BSTA is a novel approach which generates genome wide sequence information (Table 4-1). The whole transcriptome can be covered by an adequate sequence coverage which can be reached by deep sequencing. Thus, the BSTA approach has no restriction in sequence information and a large sequence resource is produced which delivers information of sunflower ESTs. One further advantage of the BSTA approach is that sequence information of both parents is generated. The only prerequisite to identify the candidate gene is that the gene of interest is expressed. For resistance genes of the NBS-LRR type it was shown that they are constitutively expressed (Tan et al. 2007) and thus fulfill the requirement.

The BSTA approach is even suitable to identify candidates in regions of suppressed recombination like it occurs in centromere near regions, interspecific crosses, or crosses with introgression lines. Thus, for identifying candidate genes of PI_{ARG} this approach was of high importance and can be recommended for similar studies.

4.6 Conclusions and outlook

Genetic mapping of PI_{ARG} (Fig. 4-1, step 1) was the prerequisite for characterization of the target region. Closely flanking and cosegregating SSR markers were identified which can be used for MAS and as starting points for map based cloning approaches. Fine mapping of PI_{ARG} in the (cms)Ha342 x ARG1575-2 population comprising 2,145 F_2 individuals and comparative mapping with population $NDBLOS_{sel}$ x KWS04 comprising 2,780 F_2 individuals revealed suppressed recombination on LG 1. Suppressed recombination is probably caused by the introgression from the wild species *H.*

argophyllus which has been often observed in crosses with wild species. Complementary approaches were performed like the map based cloning approach, the genome-wide cDNA-AFLP approach, and BSTA (Fig. 4-1, step 2) to identify candidate genes for PI_{ARG} (Fig. 4-1, step 3). Sequence analysis of the BAC clone sequences revealed three candidate genes of the TIR-NBS-LRR class which might belong to a cluster of resistance genes. Currently, the most promising candidate is RGC151SW1. Further work is necessary to validate if the resistance gene candidate RGC151SW1 mediates resistance against *P. halstedii* (Fig. 4-1, step 4). Southern blot analysis showed that several RGC151-like copies exist in the resistant line ARG1575-2. In terms of resistance gene identification a logical next step would be to continue chromosome walking to allow an exact estimation of physical distances and of the RGC cluster size and structural organization in the target region. The cDNA-AFLP approach is very suitable to characterize the general defense against *P. halstedii* and elucidates that the resistance mechanism between PI_8 and PI_{ARG} might be similar. The BSTA is a new approach which enables the identification of candidate genes in the target region based on whole transcriptome sequences and is unaffected by suppressed recombination. BSTA allowed the delimitation of the target region to 0.2 cM.

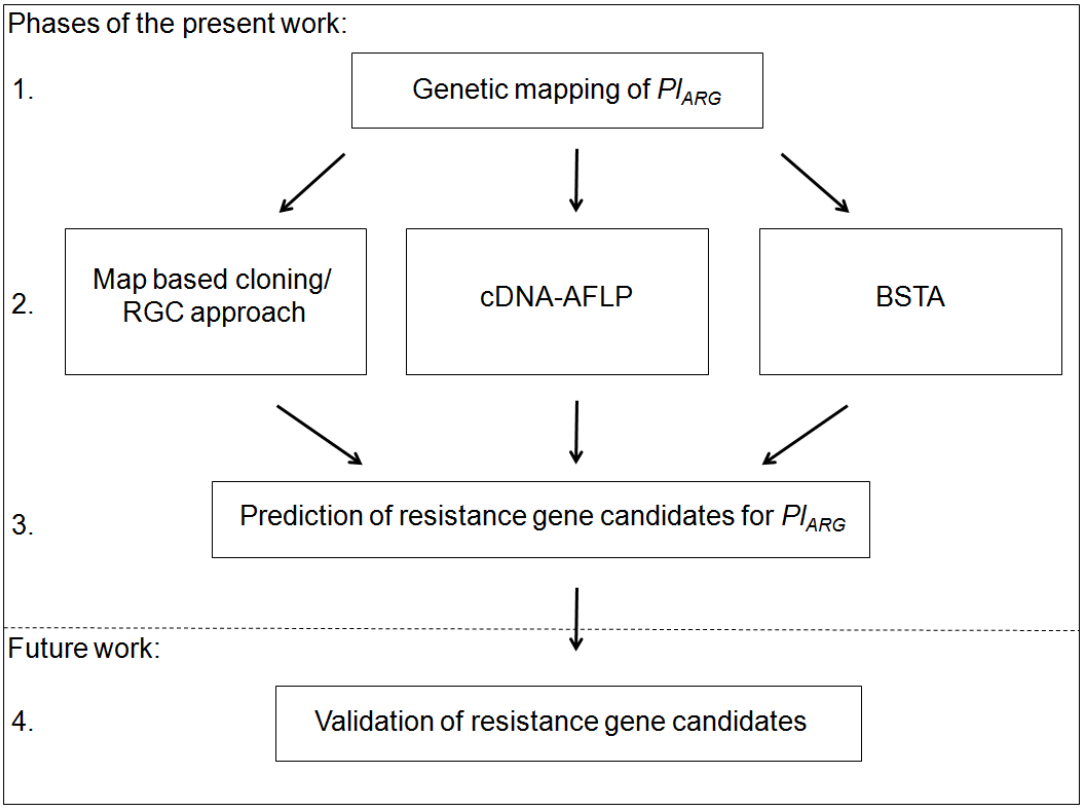


Figure 4-1: Phases of the present work and outlook for future work.

For validation, the function of several cloned resistance genes was proved using a complementation test by transferring the allele from the resistance donor into a susceptible genotype (Dodds et al. 2001; Foster et al. 2009; Liu et al. 2007; Martin et al. 1993; Ni et al. 2010). Sunflower is a recalcitrant species which is very difficult to transform. Sunflower transformation via *Agrobacterium tumefaciens* or particle bombardment has been described in the literature, but transformation rate was low and depends on the genotype (Bidney et al. 1992; Ikeda et al. 2005; Laparra et al. 1995; Mohamed et al. 2006; Sankara Rao and Rohini 1999; Weber et al. 2003). Thus, the classical complementation approach via transformation is difficult. An efficient alternative could be the transient expression of genes which has been successfully used for gene complementation (Halterman et al. 2001; Yahiaoui et al. 2004), RNAi experiments (Miki and Shimamoto 2004), and the functional assessment of defense-related genes (Schweizer et al. 1999).

Sequencing of the sunflower genome is the goal of the Genome Canada project (http://www.genomecanada.ca/medias/PDF/EN/Genomics-of-Sunflower.pdf). A hybrid approach was used which incorporates both whole genome shotgun (WGS) sequencing using the Solexa platform and 454 Life Science platform with the generation of high-density genetic and physical maps that can serve as scaffolds for the linear assembly of WGS sequences (Kane et al. 2011). Thus, a reference sequence for sunflower will be available in the near future. This will enable new possibilities in sunflower research:

- A detailed *in silico* analysis of resistance gene distribution in sunflower can be performed on a genome wide level as it was done in *Arabidopsis* (Meyers et al. 2003).
- Whole genome resequencing from multiple individuals permits a more comprehensive identification of single nucleotide polymorphisms (SNPs) (Jackson et al. 2006) and of structural rearrangements in the genus *Helianthus*.
- Whole genome resequencing is still cost intensive and requires intensive bioinformatic analyses. An alternative are target-enrichment methods in which genomic regions are selectively captured from a DNA sample before sequencing (Mamanova et al. 2010).
- Further candidate genes for Pl_{ARG} can be identified by transferring the approach of Schneeberger et al. (2009) to sunflower. They identified a causal mutation by sequencing DNA bulks of *Arabidopsis* using the program SHOREmap.

The results in the present study indicate that Pl_{ARG} might belong to a resistance gene

cluster of the TIR-NBS-LRR class. Availability of a sunflower reference sequence will provide the basis for further approaches to elucidate the sequence information from the PI_{ARG} target region.

5 Summary

Sunflower (*Helianthus annuus* L.) is a New World crop which originates from North America. It is nowadays the second largest hybrid crop after maize and one of the major oilseed crops after soybean, rapeseed, cottonseed, and groundnut cultivated worldwide in 24 million ha. *Plasmopara halstedii* (Farl.) Berl. & de Toni is a soil-, seed-, and wind-born pathogen causing downy mildew in sunflower. Downy mildew is a common sunflower disease responsible for significant yield losses. It can be controlled by fungicides and cultivation of resistant hybrids. Several major *Pl* loci (Pl_1-Pl_{15}, Pl_{ARG}) have been described, which confer resistance to one or more races of *P. halstedii*. In the last decades, new races of *P. halstedii* were discovered in the cultivation areas of sunflower. Several races developed tolerance to metalaxyl and mefenoxam, the only effective fungicides available.

The resistance locus Pl_{ARG} was introgressed from the wild species *H. argophyllus* and mediates resistance to all known races of *P. halstedii*. Till now no race of *P. halstedii* is known which overcomes the resistance of Pl_{ARG}. Therefore, Pl_{ARG} is a new source of resistance against *P. halstedii* which is very valuable for introduction into breeding material. Molecular breeding simplifies the introduction and pyramiding of resistance genes. For this purpose investigation of the structure and functionality of *Pl* loci helps to understand the defense mechanism and to manage the resistance genes effectively and durably in plant breeding.

To elucidate the molecular structure of the Pl_{ARG} locus several approaches were implemented. (1) The Pl_{ARG} gene was fine mapped in the mapping population (cms)HA342 x ARG1575-2 comprising 2,145 F_2 individuals. Cosegregating resistance gene candidates RGC151, RGC52a and RGC52b as well as closely linked markers ORS509, HT244, and HT446 were identified which flank the Pl_{ARG} locus at a distance of 0.1 cM and 0.2 cM, respectively. These markers are excellent to be used for marker assisted selection. Comparative mapping with the *H. annuus* population $NDBLOS_{sel}$ × KWS04 revealed a reduced recombination for linkage group (LG) 1 probably due to reduced homology between *H. annuus* and *H. argophyllus*. (2) Recombinant lines were tested with four different races of *P. halstedii* to investigate if Pl_{ARG} is a cluster of resistance genes. No differences in the resistance pattern were observed. Thus, further studies were necessary to explore whether Pl_{ARG} is a cluster of resistance genes. (3) Resistance gene candidates (RGC) which cosegregate with Pl_{ARG} were chosen for screening the sunflower BAC library HA_HBa (HA383). Three candidate genes of the TIR-NBS-LRR resistance class were identified on two BAC

contigs. Annotation, RT-PCR and copy number analysis indicated that RGC151SW1 is a promising candidate and might be one member of a larger resistance gene cluster. (4) The cDNA-AFLP approach highlighted differences in the defense of the susceptible line HA342 and the resistant line ARG1575-2. The method is suitable to analyze the pathogen – host interaction. (5) The new bulked segregant transcriptome analysis (BSTA) approach was conducted in combination with next generation sequencing technologies to identify further candidate genes for PI_{ARG}. Seven candidate genes were identified and two genes were mapped to LG 1. Marker BRBScontig15072 reduced the target interval to 0.2 cM. Thus, it was shown that this approach is suitable for identifying candidate genes and markers in a target region. The experiments of this study and progress in sunflower genome sequencing are the basis for the final cloning of the resistance locus PI_{ARG} in the near future.

6 Zusammenfassung

Die Sonnenblume (*Helianthus annuus* L.) ist eine der wenigen Kulturpflanzen aus der Neuen Welt, die ihren Ursprung in Nordamerika hat. Sie ist neben Mais eine der bedeutendsten Hybridpflanzen und ist eine der wichtigsten Ölpflanzen neben Soja, Raps, Baumwolle und der Erdnuss. Weltweit wird die Sonnenblume auf 24 Millionen Hektar kultiviert. *Plasmopara halstedii* (Farl.) Berl. & de Toni ist ein boden-, samen- und windbürtiges Pathogen, das den Falschen Mehltau verursacht. Der Falsche Mehltau ist eine weit verbreitete Sonnenblumenkrankheit, die zu hohen Ertragseinbußen führen kann. Die Krankheit wird durch den Einsatz von Fungiziden und den Anbau von resistenten Sorten kontrolliert. Eine Vielzahl von Resistenzloci (Pl_1-Pl_{15}, Pl_{ARG}) wurden bisher beschrieben, die die Sonnenblume gegen eine oder mehrere Rassen von *P. halstedii* schützen. Es sind jedoch Rassen bekannt, die Toleranzen gegen die einzig verfügbaren Fungizide Metalaxyl und Mefenoxam ausgebildet haben.

Pl_{ARG} wurde aus der Wildart *H. argophyllus* in die Kultursonnenblume eingeführt und schützt die Pflanze gegen alle bekannten *P. halstedii* Rassen. Bisher wurde von keiner Rasse berichtet, die die Resistenz Pl_{ARG} durchbrochen hat. Pl_{ARG} gehört somit zu einer neuen Resistenzquelle gegen *P. halstedii* und ist für die Züchtung von resistenten Sorten von großem Interesse. Der Einsatz von molekularen Techniken in der Züchtung vereinfacht die Einführung und die Pyramidisierung von Resistenzgenen in das Zuchtmaterial. Hierfür sind Kenntnisse über die Struktur und Funktionalität der Resistenzgene notwendig, um diese effektive und langfristig nutzen zu können.

Um die molekulare Struktur des Resistenzlocus Pl_{ARG} aufzuklären, wurden verschiedene Ansätze durchgeführt. (1) Pl_{ARG} wurde in der Kartierungspopulation (cms)HA342 x ARG1575-2 mit 2145 F_2 Individuen feinkartiert. Cosegregierende Resistenzgenkandidaten RGC151, RGC52 und RGC52b sowie eng gekoppelte Marker wurden identifiziert. ORS509, HT244 und HT446 flankieren Pl_{ARG} in einem Intervall von 0,2 cM und 0,1 cM oberhalb und unterhalb des Resistenzlocus. Diese Marker sind geeignet für die markergestützte Selektion von Pl_{ARG}. Durch vergleichende Kartierung mit der Population $NDBLOS_{sel}$ × KWS04 wurde unterdrückte Rekombination auf der Kopplungsgruppe 1 identifiziert, die vermutlich durch die geringe Homologie zwischen *H. annuus* und *H. argophyllus* verursacht wurde. (2) Rekombinante Linien wurden mit vier verschiedenen *P. halstedii* Rassen analysiert, um zu ermitteln, ob es sich bei Pl_{ARG} um ein Cluster von Resistenzgenen handelt. Es konnten jedoch keine Unterschiede in der Resistenzreaktion gegenüber den verschiedenen Rassen beobachtet werden, sodass weitere Ansätze notwendig waren, um zu klären, ob Pl_{ARG} aus einem Cluster

von Genen besteht, (3) Resistenzgenkandidaten, die mit *Pl*$_{ARG}$ cosegregieren, wurden ausgewählt zum Sichten der BAC Bibliothek HA_HBa (HA383). Drei Kandidatengene der TIR-NBS-LRR Klasse wurden auf zwei verschiedenen Contigs identifiziert. Annotation, RT-PCR und eine Analyse der Kopienzahl identifizierten RGC151SW1 als einen vielversprechenden Kandidaten für *Pl*$_{ARG}$, das zu einem großen Resistenzgenencluster gehören könnte. (4) Der cDNA-AFLP Ansatz identifizierte Unterschiede in der Abwehr von *P. halstedii* zwischen der anfälligen Linie HA342 und der resistenten Linie ARG1575-2. Diese Methode erwies sich als geeignet, um die Wirt-Pathogen Interaktion zu analysieren. (5) In einem Pilotprojekt wurde die neu entwickelte, „bulked segregant transcriptome analysis“ (BSTA) Methode in Kombination mit Sequenziermethoden der neusten Generation durchgeführt. Auf diese Weise wurden sieben Kandidatengene identifiziert, von denen zwei Gene auf der Kopplungsgruppe 1 kartierten, die als eng gekoppelte Marker genutzt werden können. Durch den Einsatz von BRBScontig15072 konnte das Zielintervall auf 0,2 cM reduziert werden. Somit konnte erfolgreich gezeigt werden, dass dieser Ansatz für die Identifizierung von Kandidatengenen und Markern in der Zielregion geeignet ist. Die Ergebnisse dieser Arbeit und die Fortschritte in der Sonnenblumen-Genomsequenzierung sind die Grundlage für die Klonierung des Resistenzgenes *Pl*$_{ARG}$ in naher Zukunft.

7 References

Abratti G, Bazzalo ME, León A (2008) Mapping a novel fertility restoration gene in sunflower. Proc 17th Int Sunflower Conf, Córdoba, Spain, pp 617-621

Al-Chaarani GR, Roustaee A, Gentzbittel L, Mokrani L, Barrault G, Dechamp-Guillaume G, Sarrafi A (2002) A QTL analysis of sunflower partial resistance to downy mildew (*Plasmopara halstedii*) and black stem (*Phoma macdonaldii*) by the use of recombinant inbred lines (RILs). Theor Appl Genet 104:490-496

Al-Khatiba K, Baumgartner JR, Peterson DE, Currie RS (1998) Imazethapyr resistance in common sunflower (*Helianthus annuus*). Weed Sci 46:403-407

Albourie J-M, Tourvieille J, Tourvieille de Labrouhe D (1998) Resistance to metalaxyl in isolates of the sunflower pathogen *Plasmopara halstedii*. Eur J Plant Pathol 104:235-242

Ansorge WJ (2009) Next-generation DNA sequencing techniques. New Biotechnol 25:195-203

Arumuganathan K, Earle ED (1991) Plant genomes: Databases, sizes, and nomenclature. Plant Mol Biol Rep 9:187-189

Ashfield T, Danzer JR, Held D, Clayton K, Keim P, Saghai Maroof MA, Webb DM, Innes RW (1998) *Rpg1,* a soybean gene effective against races of bacterial blight, maps to a cluster of previously identified disease resistance genes. Theor Appl Genet 96:1013-1021

Baack EJ, Whitney KD, Riesberg LH (2005) Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species. New Phytol 167:623-630

Bachlava E, Radwan O, Abratti G, Tang S, Gao W, Heesacker A, Bazzalo M, Zambelli A, Leon A, Knapp S (2010a) Downy mildew (Pl_8 and Pl_{14}) and rust (R_{Adv}) resistance genes reside in close proximity to tandemly duplicated clusters of non-TIR-like NBS-LRR-encoding genes on sunflower chromosomes 1 and 13. Theor Appl Genet 122:1211-1221

Bachlava E, Taylor CA, Bowers JE, Tang S, Zhang D, Farmer AD, Ayyampalayam S, Huntley JJ, May GD, Knapp SJ (2010b) SNP discovery for the development of high-throughput genotyping arrays using next-generation transcriptome sequences in sunflower. Plant & animal genomes XVIII conference, San Diego, CA, USA

Baldi P, Patocchi A, Zini E, Toller C, Velasco R, Komjanc M (2004) Cloning and linkage mapping of resistance gene homologues in apple. Theor Appl Genet 109:231-239

Baldo A, Norelli J, Farrell R, Bassett C, Aldwinckle H, Malnoy M (2010) Identification of genes differentially expressed during interaction of resistant and susceptible apple cultivars (*Malus* x *domestica*) with *Erwinia amylovora*. BMC Plant Biol 10:1-10

Ballini E, Berruyer R, Morel J-B, Lebrun M-H, Nottéghem J-L, Tharreau D (2007) Modern elite rice varieties of the 'Green Revolution' have retained a large introgression from wild rice around the *Pi33* rice blast resistance locus. New Phytol 175:340-350

Barakat A, DiLoreto D, Zhang Y, Smith C, Baier K, Powell W, Wheeler N, Sederoff R, Carlson J (2009) Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. BMC Plant Biol 9:51

Barker CL, Donald T, Pauquet J, Ratnaparkhe MB, Bouquet A, Adam-Blondon AF, Thomas MR, Dry I (2005) Genetic and physical mapping of the grapevine powdery mildew resistance gene, *Run1*, using a bacterial artificial chromosome library. Theor Appl Genet 111:370-377

Bassam BJ, Caetano-Anollés G, Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196:80-83

Bent AF, Mackey D (2007) Elicitors, effectors, and *R* genes: The new paradigm and a lifetime supply of questions. Ann Rev Phytopathol 45:399-436

Berry ST, Leon AJ, Hanfrey CC, Challis P, Burkholz A, Barnes SR, Rufener GK, Lee M, Caligari PDS (1995) Molecular marker analysis of *Helianthus annuus* L. 2. Construction of an RFLP linkage map for cultivated sunflower. Theor Appl Genet 91:195-199

Bert PF, Tourvielle de Labrouhe D, Philippon J, Mouzeyar S, Jouan I, Nicolas P, Vear F (2001) Identification of second linkage group carrying genes controlling resistance to downy mildew (*Plasmopara halstedii*) in sunflower (*Helianthus annuus* L.). Theor Appl Genet 103:992-997

Bertero de Romano A, Romano C, Bulos M, Altieri E, Sala C (2010) A new gene for resistance to downy mildew in sunflower. Proc Int Symp "Sunflower breeding on resistance to diseases", Krasnodar, Russia, pp 142-147

Bidney D, Scelonge C, Martich J, Burrus M, Sims L, Huffman G (1992) Microprojectile bombardment of plant tissues increases transformation frequency by *Agrobacterium tumefaciens*. Plant Mol Biol 18:301-313

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

Bouzidi MF, Badaoui S, Cambon F, Vear F, Tourvieille de Labrouhe D, Nicolas P, Mouzeyar S (2002) Molecular analysis of a major locus for resistance to downy mildew in sunflower with specific PCR-based markers. Theor Appl Genet 104:592-600

Bouzidi MF, Franchel J, Tao Q, Stormo K, Mraz A, Nicolas P, Mouzeyar S (2006) A sunflower BAC library suitable for PCR screening and physical mapping of targeted genomic regions. Theor Appl Genet 113:81-89

Brahm L, Rocher T, Friedt W (2000) PCR-based markers facilitating marker assisted selection in sunflower for resistance to *downy mildew.* Crop Sci 40:676-682

Burke JM, Lai Z, Salmaso M, Nakazato T, Tang S, Heesacker A, Knapp SJ, Rieseberg LH (2004) Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. Genetics 167:449-457

Chacón O, González M, López Y, Portieles R, Pujol M, González E, Schoonbeek H-J, Métraux J-P, Borrás-Hidalgo O (2010) Overexpression of a protein kinase gene enhances the defense of tobacco against *Rhizoctonia solani*. Gene 452:54-62

Chapman NH, Burt C, Nicholson P (2009) The identification of candidate genes associated with *Pch2* eyespot resistance in wheat using cDNA-AFLP. Theor Appl Genet 118:1045-1057

Chetelat RT, Meglic V (2000) Molecular mapping of chromosome segments introgressed from *Solanum lycopersicoides* into cultivated tomato (*Lycopersicon esculentum*). Theor Appl Genet 100:232-241

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

Delmotte F, Giresse X, Richard-Cervera S, M'Baya J, Vear F, Tourvieille J, Walser P, Tourvieille de Labrouhe D (2008) Single nucleotide polymorphisms reveal multiple introductions into France of *Plasmopara halstedii,* the plant pathogen causing sunflower downy mildew. Infect Genet Evol 8:534-540

DeWoody JA (2004) Universal method for producing ROX-labeled size standards suitable for automated genotyping. Biotechniques 37:348-352

Ding Y, Johnson MD, Colayco R, Chen YJ, Melnyk J, Schmitt H, Shizuya H (1999) Contig assembly of bacterial artificial chromosome clones through multiplexed fluorescence-labeled fingerprinting. Genomics 56:237-246

Dixon MS, Jones DA, Keddie JS, Thomas C, Harrison K, Jones J (1996) The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell 84:451-459

Dodds PN, Lawrence GJ, Ellis JG (2001) Six amino acid changes confined to the leucine-rich repeat ß-strand/ß-turn motif determine the difference between the *P* and *P2* rust resistance specificities in flax. Plant Cell 13:163-178

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13 - 15

Dußle CM, Hahn V, Knapp SJ, Bauer E (2004) Pl_{Arg} from *Helianthus argophyllus* is unlinked to other known downy mildew resistance genes in sunflower. Theor Appl Genet 109:1083-1086

Eckey C, Korell M, Leib K, Biedenkopf D, Jansen C, Langen G, Kogel K-H (2004) Identification of powdery mildew-induced barley genes by cDNA-AFLP: functional assessment of an early expressed MAP kinase. Plant Mol Biol 55:1-15

Eitas TK, Dangl JL (2010) NB-LRR proteins: pairs, pieces, perception, partners, and pathways. Curr Opin Plant Biol 13:472-477

EPPO/CABI (1997) *Plasmopara halstedii*. In: Smith I, McNamara D, Scott P, Holderness M (eds) Quarantine Pests for Europe, 2nd edn. CABI International, Wallingford, UK, p 1425 pp

Feng J, Vick B, Lee M-K, Zhang H-B, Jan C (2006) Construction of BAC and BIBAC libraries from sunflower and identification of linkage group-specific clones by overgo hybridization. Theor Appl Genet 113:23-32

Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B (2003) Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc Natl Acad Sci USA 100:15253-15258

Flor H (1971) Current status of the gene-for-gene concept. Annu Rev Phytopathol 9:275-296

Foster SJ, Park T-H, Pel M, Brigneti G, Åšliwka J, Jagger L, van der Vossen E, Jones JDG (2009) *Rpi-vnt1.1*, a *Tm-2*^{2} homolog from *Solanum venturii*, confers resistance to potato late blight. Mol Plant Microbe In 22:589-600

Frisch M, Bohn M, Melchinger AE (1999) Comparison of selection strategies for marker-assisted backcrossing of a gene. Crop Sci 39:1295-1301

Fung RWM, Gonzalo M, Fekete C, Kovacs LG, He Y, Marsh E, McIntyre LM, Schachtman DP, Qiu W (2008) Powdery mildew induces defense-oriented reprogramming of the transcriptome in a susceptible but not in a resistant grapevine. Plant Physiol 146:236-249

Galli P, Patocchi A, Broggini GAL, Gessler C (2010) The *Rvi15* (*Vr2*) apple scab resistance locus contains three TIR-NBS-LRR genes. Mol Plant Microbe In 23:608-617

Ganal MW, Young ND, Tanksley SD (1989) Pulsed field gel electrophoresis and physical mapping of large DNA fragments in the *Tm-2a* region of chromosome 9 in tomato. Mol Gen Genet 215:395-400

Gebhardt C (1997) Plant genes for pathogen resistance - variation on a theme. Trends Plant Sci 2:243-244

Gentzbittel L, Vear F, Zhang YX, Bervillé A, Nicolas P (1995) Development of a consensus linkage RFLP map of cultivated sunflower (*Helianthus annuus* L.) Theor Appl Genet 90:1079-1086

Gentzbittel L, Mouzeyar S, Badaoui S, Mestries E, Vear F, Tourvieille de Labrouhe D, Nicolas P (1998) Cloning of molecular markers for disease resistance in sunflower, *Helianthus annuus* L. Theor Appl Genet 96:519-525

Gentzbittel L, Abbott A, Galaud J, Georgi L, Fabre F, Liboz T, Alibert G (2002) A bacterial artificial chromosome (BAC) library for sunflower, and identification of clones containing genes for putative transmembrane receptors. Mol Genet Genomics 266:979-987

Gedil MA, Slabaugh MB, Berry S, Johnson R, Michelmore R, Miller J, Gulya T, Knapp SJ (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

Gulya T (1985) Registration of five disease-resistant sunflower germplasms. Crop Sci 25:719–720

Gulya TJ (1996) Everything you should know about downy mildew testing but were afraid to ask. Proc 18th Sunflower Research Workshop, Fargo, ND, USA, pp 39-48

Gulya TJ (2007) Distribution of *Plasmopara halstedii* races from sunflower around the world. Proc 2nd Int Downy Mildew Symp "Advances in Downy Mildew Research", Olomouc, Czech Republic, pp 121-134

Gulya TJ, Sackston WE, Viranyi F, Masirevic S, Rashid KY (1991) New races of the sunflower downy mildew pathogen (*Plasmopara halstedii)* in Europe and North and South America. J Phytopathol 132:303-311

Halterman D, Zhou F, Wei F, Wise RP, Schulze-Lefert P (2001) The *MLA6* coiled-coil, NBS-LRR protein confers *AvrMla6*-dependent resistance specificity to *Blumeria graminis* f. sp. *hordei* in barley and wheat. Plant J 25:335-348

Hammond-Kosack KE, Jones JDG (1997) Plant disease resistance genes. Annu Rev Plant Phys 48:575-607

Han CS, Sutherland RD, Jewett PB, Campbell ML, Meincke LJ, Tesmer JG, Mundt MO, Fawcett JJ, Kim U-J, Deaven LL, Doggett NA (2000) Construction of a BAC contig map of chromosome 16q by two-dimensional overgo hybridization. Genome Res 10:714-721

Hattendorf A, Debener T (2007) Molecular characterization of NBS-LRR-RGAs in the rose genome. Physiol Plantarum 129:775-786

He L, Du C, Covaleda L, Xu Z, Robinson AF, Yu JZ, Kohel RJ, Zhang H-B (2004) Cloning, characterization, and evolution of the NBS-LRR-encoding resistance gene analogue family in polyploid cotton (*Gossypium hirsutum* L.). Mol Plant Microbe In 17:1234-1241

Heesacker AF, Bachlava E, Brunick RL, Burke JM, Rieseberg LH, Knapp SJ (2009) Karyotypic evolution of the common and silverleaf sunflower genomes. Plant Genome 2:233-246

Huang S, coauthors (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet 41:1275-1281

Hunger S, Di Gaspero G, Möhring S, Bellin D, Schäfer-Pregl R, Borchardt DC, Durel C-E, Weber M, Weisshaar B, Salamini F, Schneider K (2003) Isolation and linkage analysis of expressed disease-resistance gene analogues of sugar beet (*Beta vulgaris* L.). Genome 46:70-82

Ikeda M, Matsumura M, Kamada H (2005) Suitability of small and branching sunflower varieties for molecular genetic experiments and their transformation by *Agrobacterium* infection. Plant Biotechnol J 22:97-104

Ingvardsen C, Xing Y, Frei U, Lübberstedt T (2010) Genetic and physical fine mapping of *Scmv2*, a potyvirus resistance gene in maize. Theor Appl Genet 120:1621-1634

Jackson S, Rounsley S, Purugganan M (2006) Comparative sequencing of plant genomes: Choices to make. Plant Cell 18:1100-1104

Jan CC, Fernández-Martínez JM, Ruso J, Munoz-Ruz J (2002) Registration of four sunflower germplasms with resistance to *Orobanche cumana* race F. Crop Sci 42:2217-2218

Jan CC, Quresh Z, Gulya TJ (2004a) Registration of seven rust resistant sunflower germplasms. Crop Sci 44:1887-1888

Jan CC, Tan AS, Gulya TJ (2004b) Registration of four downy mildew resistant sunflower germplasms. Crop Sci 44:1887

Jarvie T, Harkins T (2008) Transcriptome sequencing with the genome sequencer FLX system. Nat Methods 5:1-3

Jones JDG, Dangl JL (2006) The plant immune system. Nature 444:323-329

Jones DA, Thomas CM, Hammond-Kosack KE, Balint-Kurti PJ, Jones JD (1994) Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. Science 266:789-793

Kaloshian I, Yaghoobi J, Liharska T, Hontelez J, Hanson D, Hogan P, Jesse T, Wijbrandi J, Simons G, Vos P, Zabel P, Williamson VM (1998) Genetic and physical localization of the root-knot nematode resistance locus *Mi* in tomato. Mol Gen Genet 257:376-385

Kane NC, Gill N, King MG, Bowers JE, Berges H, Gouzy J, Bachlava E, Langlade NB, Lai Z, Stewart M, Burke JM, Vincourt P, Knapp SJ, Rieseberg LH (2011) Progress towards a reference genome for sunflower. Botany 89:429-437

Kinman M (1970) New developments in the USDA and state experiment station sunflower breeding programs. Proc 4th Int Sunflower Conf, Memphis, TN, pp 181-183

Kircher M, Kelso J (2010) High-throughput DNA sequencing - concepts and limitations. Bioessays 32:524-536

Kolkman JM, Slabaugh MB, Bruniard JM, Berry S, Bushman BS, Olungu C, Maes N, Abratti G, Zambelli A, Miller JF, Leon A, Knapp SJ (2004) Acetohydroxyacid synthase mutations conferring resistance to imidazolinone or sulfonylurea herbicides in sunflower. Theor Appl Genet 109:1147-1159

Kolkman JM, Berry ST, Leon AJ, Slabaugh MB, Tang S, Gao WX, Shintani DK, Burke JM, Knapp SJ (2007) Single nucleotide polymorphisms and linkage disequilibrium in sunflower. Genetics 177:457-468

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugen 12:172-175

Kuhl JC, Hanneman RE, Havey MJ (2001) Characterization and mapping of *Rpi1*, a late-blight resistance locus from diploid (1EBN) Mexican *Solanum pinnatisectum*. Mol Genet Genomics 265:977-985

Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 9:299-306

Lai Z, Livingstone K, Zou Y, Church SA, Knapp SJ, Andrews J, Rieseberg LH (2005a) Identification and mapping of SNPs from ESTs in sunflower. Theor Appl Genet 111:1532-1544

Lai Z, Nakazato T, Salmaso M, Burke JM, Tang S, Knapp SJ, Rieseberg LH (2005b) Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. Genetics 171:291-303

Laparra H, Burrus M, Hunold R, Damm B, Bravo-Angel A, Bronner R, Hahne G (1995) Expression of foreign genes in sunflower (*Helianthus annuus* L.) — evaluation of three gene transfer methods. Euphytica 85:63-74

Lawrence GJ, Finnegan EJ, Ayliffe MA, Ellis JG (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

Lawson WR, Goulter KC, Henry RJ, Kong GA, Kochman JK (1998) Marker-assisted selection for two rust resistance genes in sunflower. Mol Breeding 4:227-234

Leclercq P (1969) Cytoplasmic male sterility in sunflower. Ann Amelior Plant 19:99-106

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

Levy M, Edelbaum O, Sela I (2004) Tabacco mosaic virus regulates the expression of its own resistance gene *N*. Plant Physiol 135:2392-2397

Ling HQ, Qiu JW, Singh RP, Keller B (2004) Identification and genetic characterization of an *Aegilops tauschii* ortholog of the wheat leaf rust disease resistance gene *Lr1*. Theor Appl Genet 109:1133-1138

Liu JJ, Ekramoddoullah AKM (2004) Isolation, genetic variation and expression of TIR-NBS-LRR resistance gene analogs from western white pine (*Pinus monticola* Dougl. ex. D. Don.). Mol Genet Genomics 270:432-441

Liu X, Lin F, Wang L, Pan Q (2007) The in silico map-based cloning of *Pi36*, a rice coiled-coil nucleotide-binding site leucine-rich repeat gene that confers race-specific resistance to the blast fungus. Genetics 176:2541-2549

Lu YH, Melero-Vara JM, Garcia-Tejada JA, Blanchard P (2000) Development of SCAR markers linked to the gene *Or5* conferring resistance to broomrape (*Orobanche cumana* Wallr.) in sunflower. Theor Appl Genet 100:625-632

Luo M, Wang Y, Frisch D, Joobeur T, Wing R, Dean R (2001) Melon bacterial artificial chromosome (BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon Fusarium wilt (Fom-2). Genome 44:154-162

Luo M-C, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, Dvorak J (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82:378-389

Luo M-C, Wing R (2003) An improved method for plant BAC library construction. Plant Functional Genomics 236:3-19

Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ (2010) Target-enrichment strategies for next-generation sequencing. Nat Meth 7:111-118

Marrs KA (1996) The function and regulation of Glutathione S-transferase in plants. Annu Rev Plant Phys 47:127-158

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432-1436

McDonald BA, Linde C (2002) Pathogen population genetics, evolutionary potential, and durable resistance. Annu Rev Phytopathol 40:349-379

Messeguer R, Ganal M, de Vicente MC, Young ND, Bolkan H, Tanksley SD (1991) High resolution RFLP map around the root knot nematode resistance gene (*Mi*) in tomato. Theor Appl Genet 82:529-536

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J 20:317-332

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

Meyer JDF, Silva DCG, Yang C, Pedley KF, Zhang C, van de Mortel M, Hill JH, Shoemaker RC, Abdelnoor RV, Whitham SA, Graham MA (2009) Identification and analyses of candidate genes for *Rpp4*-mediated resistance to Asian soybean rust in soybean. Plant Physiol 150:295-307

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8:1113-1130

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828-9832

Micic Z, Hahn V, Bauer E, Schön CC, Melchinger AE (2005) QTL mapping of resistance to *Sclerotinia* midstalk rot in RIL of sunflower population $NDBLOS_{sel}$ x CM625. Theor Appl Genet 110:1490-1498

Miki D, Shimamoto K (2004) Simple RNAi vectors for stable and transient suppression of gene function in rice. Plant Cell Physiol 45:490-495

Miller JF, Zimmerman DC, Vick BA, Roath WW (1987) Registration of sixteen high oleic sunflower germplasm lines and bulk population. Crop Sci 27:1323

Miller JF, Gulya TJ (1991) Inheritance of resistance to race 4 of downy mildew derived from interspecific crosses in sunflower. Crop Sci 31:40-43

Miller JF, Gulya TJ, Seiler GJ (2002) Registration of five fertility restorer sunflower germplasms. Crop Sci 42:989-991

Miller JF, Seiler GJ (2003) Registration of five oilseed maintainer (HA 429-HA 433) sunflower germplasm lines. Crop Sci 43:2313-2314

Miller JF, Gulya TJ, Vick BA (2006) Registration of Imidazolinone herbicide-resistant maintainer (HA 442) and fertility restorer (RHA 443) oilseed sunflower germplasms. Crop Sci 46:483-484

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

Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, Marshall D (2010) Tablet - next generation sequence assembly visualization. Bioinformatics 26:401-402

Mohamed S, Boehm R, Schnabl H (2006) Stable genetic transformation of high oleic *Helianthus annuus* L. genotypes with high efficiency. Plant Sci 171:546-554

Molinero-Ruiz ML, Dominguez J, Melero-Vara JM (2002a) Races of isolates of *Plasmopara halstedii* from Spain and studies on their virulence. Plant Dis 86:736-740

Molinero-Ruiz ML, Melero-Vara JM, Dominguez J (2002b) Inheritance of resistance to race 330 of *Plasmopara halstedii* in three sunflower lines. Plant Breeding 121:61-65

Molinero-Ruiz ML, Melero-Vara JM (2003) First report of resistance to metalaxyl in downy mildew of sunflower caused by *Plasmopara halstedii* in Spain. Plant Dis 87:749

Molinero-Ruiz ML, Melero-Vara JM, Domínguez J (2003) Inheritance of resistance to two races of sunflower downy mildew (*Plasmopara halstedii*) in two *Helianthus annuus* L. lines. Euphytica 131:47-51

Mouzeyar S, Tourvieille De Labrouhe D, Vear F (1994) Effect of host-race combination on resistance of sunflower, *Helianthus annuus* L., to downy mildew *Plasmopara halstedii*. J Phytopathol 141:249-258

Mouzeyar S, Roeckel-Drevet P, Gentzbittel L, Philippon J, Tourvieille de Labroughe D, Vear F (1995) RFLP and RAPD mapping of the sunflower *Pl1* locus for resistance to *Plasmopara halstedii* race 1. Theor Appl Genet 91:733-737

Mulpuri S, Liu Z, Feng J, Gulya T, Jan C-C (2009) Inheritance and molecular mapping of a downy mildew resistance gene, Pl_{13} in cultivated sunflower (*Helianthus annuus* L.). Theor Appl Genet 119:795-803

Ni X, Tian Z, Liu J, Song B, Xie C (2010) Cloning and molecular characterization of the potato RING finger protein gene *StRFP1* and its function in potato broad-spectrum resistance against *Phytophthora infestans*. J Plant Physiol 167:488-496

Özdemir N, Horn R, Friedt W (2004) Construction and characterization of a BAC library for sunflower (*Helianthus annuus* L.). Euphytica 138:177-183

Parlevliet J (2002) Durability of resistance against fungal, bacterial and viral pathogens; present situation. Euphytica 124:147-156

Pashley CH, Ellis JR, McCauley DE, Burke JM (2006) EST databases as a source for molecular markers: Lessons from *Helianthus*. J Hered 97:381-388

Pauquet J, Lucante N, Claudel C, Langlade N, Coque M, Romestang M, Blanchard P, Dugas O, Grezesbesset B (2011) High-throughput SNP discovery in sunflower using mRNAseq technology. Plant & Animal Genomes XIX Conference, San Diego, CA, USA

Pérez Vich B, Berry ST (2010) Molecular Breeding. In: Hu J, Seiler G, Kole C (eds) Genetics, Genomics and Breeding of Sunflower CRC Press, Pullman, Washington, USA, pp 221-252

Pillen K, Ganal M, Tanksley S (1996) Construction of a high-resolution genetic map and YAC-contigs in the tomato *Tm-2a* region. Theor Appl Genet 93:228-233

Quagliaro G, Vischi M, Tyrka M, Olivieri AM (2001) Identification of wild and cultivated sunflower for breeding purposes by AFLP markers. J Hered 92:38-42

Radwan O, Bouzidi MF, Vear F, Philippon J, Tourvieille de Labrouhe D, Nicolas P, Mouzeyar S (2003) Identification of non-TIR-NBS-LRR markers linked to the *Pl5*/*Pl8* locus for resistance to downy mildew in sunflower. Theor Appl Genet 106:1438-1446

Radwan O, Bouzidi MF, Nicolas P, Mouzeyar S (2004) Development of PCR markers for the *Pl5*/*Pl8* locus for resistance to *Plasmopara halstedii* in sunflower, *Helianthus annuus* L. from complete CC-NBS-LRR sequences. Theor Appl Genet 109:176-185

Radwan O, Mouzeyar S, Nicolas P, Bouzidi MF (2005a) Induction of a sunflower CC-NBS-LRR resistance gene analogue during incompatible interaction with *Plasmopara halstedii*. J Exp Bot 56:567-575

Radwan O, Mouzeyar S, Venisse JS, Nicolas P, Bouzidi MF (2005b) Resistance of sunflower to the biotrophic oomycete *Plasmopara halstedii* is associated with a delayed hypersensitive response within the hypocotyls. J Exp Bot 56:2683-2693

Radwan O, Gandhi S, Heesacker A, Whitaker B, Taylor C, Plocik A, Kesseli R, Kozik A, Michelmore RW, Knapp SJ (2008) Genetic diversity and genomic distribution of homologs encoding NBS-LRR disease resistance proteins in sunflower. Mol Genet Genomics 280:111-125

Radwan O, Bouzidi MF, Mouzeyar S (2011) Molecular characterization of two types of resistance in sunflower to *Plasmopara halstedii*, the causal agent of downy mildew. Phytopathology 101:970-979

Rahim M, Jan CC, Gulya TJ (2002) Inheritance of resistance to sunflower downy mildew races 1, 2 and 3 in cultivated sunflower. Plant Breeding 121:57-60

Roath WW, Miller JF, Gulya T (1987) Registration of three oilseed sunflower germplasm pools ND-BLPL2, ND-BLOS, and ND-RLOS. Crop Sci 27:373

Röcher T (1999) Molekulargenetische Untersuchungen zur Resistenz der Sonnenblume (*Helianthus annuus* L.) gegen den Erreger des Falschen Mehltaus (*Plasmopara halstedii*). Dissertation, Justus-Liebig-Universität Gießen, Gießen

Roeckel-Drevet P, Gagne G, Mouzeyar S, Gentzbittel L, Philippon J, Nicolas P, Tourvieille de Labrouhe D, Vear F (1996) Colocation of downy mildew (*Plasmopara halstedii*) resistance genes in sunflower (*Helianthus annuus* L.). Euphytica 91:225-228

Romero-Puertas MC, Perazzolli M, Zago ED, Delledonne M (2004) Nitric oxide signalling functions in plant-pathogen interactions. Cell Microbiol 6:795-803

Rozen S, Skaletsky HJ (2000) Primer3 on the www for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: Methods in molecular biology. Humana Press, Totowa, NJ, pp 365-386

Sackston W (1981a) The sunflower crop and disease: progress, problems, and prospects. Plant Dis 65:643-648

Sackston WE (1981b) Downy mildew of sunflower. In: Spencer DM (ed) The downy mildew. Academic Press, New York, pp 545-575

Sackston WE (1990) A proposed international system for designating races of *Plasmopara halstedii*. Plant Dis 74:721-723

Sakr N (2010) Can we enhance durable resistance against *Plasmopara halstedii* (sunflower downy mildew)? J Plant Protection Res 50:15-21

Sankara Rao K, Rohini VK (1999) Agrobacterium-mediated transformation of sunflower (*Helianthus annuus L*.): A simple protocol. Ann Bot 83:347-354

Schilling EE (2006) *Helianthus*. In: Committee FoNAE (ed) Flora of North America North of Mexico. Oxford University Press, New York and Oxford, pp 141-169

Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jorgensen J-E, Weigel D, Andersen SU (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat Meth 6:550-551

Schweizer P, Pokorny J, Abderhalden O, Dudler R (1999) A transient assay system for the functional assessment of defense-related genes in wheat. Mol Plant Microbe In 12:647-654

Seah S, Yaghoobi J, Rossi M, Gleason CA, Williamson VM (2004) The nematode-resistance gene, *Mi-1*, is associated with an inverted chromosomal segment in susceptible compared to resistant tomato. Theor Appl Genet 108:1635-1642

Seah S, Telleen A, Williamson V (2007) Introgressed and endogenous *Mi-1* gene clusters in tomato differ by complex rearrangements in flanking sequences and show sequence exchange and diversifying selection among homologues. Theor Appl Genet 114:1289-1302

Seiler GJ (1992) Utilization of wild sunflower species for the improvement of cultivated sunflower. Field Crop Res 30:195-230

Seiler GJ (2007) Wild annual *Helianthus anomalus* and *H. deserticola* for improving oil content and quality in sunflower. Ind Crop Prod 25:95-100

Seiler G (2010) Utilization of wild *Helianthus* species in breeding for disease resistance. Proc Int Symp "Sunflower breeding on resistance to diseases", Krasnodar, Russia, pp 37-51

Seiler G, Jan CC (2010) Basic information. In: Hu J, Seiler G, Kole C (eds), Genetics, Genomics and Breeding of Sunflower CRC Press, Pullman, Washington, USA, pp 1-50

Seiler GJ, Christie BR, Choo TM (1991) Registration of 13 downy mildew tolerant interspecific sunflower germplasm lines derived from wild annual species. Crop Sci 31:1714-1716

Shcheglov A, Zhulidov P, Bogdanova E, Shagin D (2007) Normalization of cDNA libraries. In: Buzdin A, Lukyanov SA (eds), Nucleic Acids Hybridization Modern Applications. Springer Netherlands, pp 97-124

Slabaugh MB, Huestis GM, Leonard J, Holloway JL, Rosato C, Hongtrakul V, Martini N, Toepfer R, Voetz M, Schell J, Knapp SJ (1997) Sequence-based genetic markers for genes and gene families: single-strand conformational polymorphisms for the fatty acid synthesis genes of *Cuphea*. Theor Appl Genet 94:400-408

Slabaugh MB, Yu J-K, Tang S, Heesacker A, Hu X, Lu G, Bidney D, Han F, Knapp SJ (2003) Haplotyping and mapping a large cluster of downy mildew resistance gene candidates in sunflower using multilocus intron fragment length polymorphisms. Plant Biotechnol J 1:167-185

Soderlund C, Longden I, Mott R (1997) FPC: a system for building contigs from restriction fingerprinted clones. Comput Appl Biosci 13:523-535

Soldatov K (1976) Chemical mutagenesis in sunflower breeding. Proc 7th Int Sunflower Conf, Krasnodar, USSR, pp 352-357

Song W-Y, Wang G-L, Chen L-L, Kim H-S, Pi L-Y, Holsten T, Gardner J, Wang B, Zhai W-X, Zhu L-H, Fauquet C, Ronald P (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270:1804-1806

Spring O, Zipper R, Heller-Dohmen M (2006) First report of metalaxyl resistant isolates of *Plasmopara halstedii* on cultivated sunflower in Germany. J Plant Dis Protect 113:224

Stein N, Feuillet C, Wicker T, Schlagenhauf E, Keller B (2000) Subgenome chromosome walking in wheat: A 450-kb physical contig in *Triticum monococcum* L. spans the *Lr10* resistance locus in hexaploid wheat (*Triticum aestivum* L.). Proc Natl Acad Sci USA 97:13436-13441

Steuernagel B, Taudien S, Gundlach H, Seidel M, Ariyadasa R, Schulte D, Petzold A, Felder M, Graner A, Scholz U, Mayer K, Platzer M, Stein N (2009) *De novo* 454 sequencing of barcoded BAC pools for comprehensive gene survey and genome analysis in the complex genome of barley. BMC Genomics 10:547

Sun C, Li Y, Wu Q, Luo H, Sun Y, Song J, Lui EM, Chen S (2010) *De novo* sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. BMC Genomics 11:262

Tan X, Meyers B, Kozik A, West M, Morgante M, St Clair D, Bent A, Michelmore R (2007) Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in *Arabidopsis*. BMC Plant Biol 7:56

Tang S, Yu JK, Slabaugh MB, Shintani DK, Knapp SJ (2002) Simple sequence repeat map of the sunflower genome. Theor Appl Genet 105:1124-1136

Tang S, Kishore VK, Knapp SJ (2003) PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower. Theor Appl Genet 107:6-19

Tang S, Hass CG, Knapp SJ (2006) *Ty3*/*gypsy*-like retrotransposon knockout of a 2-methyl-6-phytyl-1,4-benzoquinone methyltransferase is non-lethal, uncovers a cryptic paralogous mutation, and produces novel tocopherol (vitamin E) profiles in sunflower. Theor Appl Genet 113:783-799

Tao Y, Xie Z, Chen W, Glazebrook J, Chang H-S, Han B, Zhu T, Zou G, Katagiri F (2003) Quantitative nature of *Arabidopsis* responses during compatible and incompatible interactions with the bacterial pathogen pseudomonas syringae. Plant Cell 15:317-330

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl Acids Res 22:4673-4680

Tourvieille de Labrouhe D, Gulya T, Masirevic S, Penaud A, Rashid KY, Viranyi F (2000a) New nomenclature of races of *Plasmopara halstedii* (sunflower downy mildew). Proc 15th Int Sunflower Conf, Toulouse, France, pp 61-65

Tourvieille de Labrouhe D, Lafon S, Walser P, Raulic I (2000b) Une nouvelle race de *Plasmopara halstedii*, agent du mildiou du tournesol. Oleagineux 7:404-405

Tourvieille de Labrouhe D, Serre F, Walser P, Roche S, Vear F (2008) Quantitative resistance to downy mildew (*Plasmopara halstedii*) in sunflower (*Helianthus annuus*). Euphytica 164:433-444

Tourvieille de Labrouhe D, Bordat A, Tourvieille J, Mestries E, Walser P, Sakr N, Ducher M, Delmotte F, Vear F (2010) Impact of major gene resistance management for sunflower on fitness of *Plasmopara halstedii* (downy mildew) populations. Oleagineux 17:56-64

Toyoda K, Collins N, Takahashi A, Shirasu K (2002) Resistance and susceptibility of plants to fungal pathogens. Transgenic Res 11:567-582

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. Nat Genet 42:260-263

Van Ooijen JW (2006) JoinMap4, software for the calculation of genetic linkage maps in experimental populations. Kayazama, B.V., Wageningen, Netherlands

Vear F (1974) Studies on resistance to downy mildew in sunflowers (*Helianthus annuus* L.). Proc 6th Int Sunflower Conf, Bucharest, Rumania, pp 297-302

Vear F (2004) Breeding for durable resistance to the main diseases of sunflower. Proc 16th Int Sunflower Conf, Fargo, USA pp 15-28

Vear F, Leclercq P (1971) Deux nouveaux gènes résistance au mildiou du tournesol. Ann Amelior Plant 21:251-255

Vear F, Gentzbittel L, Philippon J, Mouzeyar S, Mestries E, Roeckel-Drevet P, Tourvieille de Labrouhe D, Nicolas P (1997) The genetics of resistance to five races of downy mildew (*Plasmopara hastedii*) in sunflower (*Helianthus annuus* L.). Theor Appl Genet 95:584-589

Vear F, Tourvieille de Labroughe D, Miller J (2003) Inheritance of the wild-range downy mildew resistance in the sunflower line RHA419. Helia 26:19-24

Vear F, Serre F, Roche S, Walser P, Tourvieille de Labrouhe D (2007) Recent research on downy midew resistance useful for breeding industrial-use sunflowers. Helia 30:45-54

Vear F, Serre F, Jouan-Dufournel I, Bert PF, Roche S, Walser P, Tourvieille de Labrouhe D, Vincourt P (2008a) Inheritance of quantitative resistance to downy mildew (*Plasmopara halstedii*) in sunflower (*Helianthus annuus* L.). Euphytica 164:561-570

Vear F, Serieys H, Petit A, Serre F, Boudon J-P, Roche S, Walser P, Tourvieille de Labrouhe D (2008b) Origins of major genes for downy mildew resistance in sunflower. Proc 17th Int Sunflower Conf, Córdoba, Spain, pp 125-130

Vick BA, Hu J (2010) Future prospects. In: Hu J, Seiler G, Kole C (eds), Genetics, Genomics and Breeding of Sunflower CRC Press, Pullman, Washington, USA, pp 313-326

Virányi F (2008) Research progress in sunflower diseases and their management. Proc 17th Int Sunflower Conf, Córdoba, Spain, pp 1-12

Vogel JP, Raab TK, Schiff C, Somerville SC (2002) *PMR6*, a pectate lyase-like gene required for powdery mildew susceptibility in *Arabidopsis*. Plant Cell 14:2095-2106

Vranceanu V, Stoenescu F (1970) Immunity to sunflower downy mildew due to a single dominant gene. Problème Agricole 22:34-40

Vranceanu VL, Pirvu N, Stoenescu FM (1981) New sunflower downy mildew resistance genes and their management. Helia 4:23-27

Vuylsteke M, Peleman JD, van Eijk MJT (2007) AFLP-based transcript profiling (cDNA-AFLP) for genome-wide expression analysis. Nat Protocols 2:1399-1413

Wang X, Liu W, Chen X, Tang C, Dong Y, Ma J, Huang X, Wei G, Han Q, Huang L, Kang Z (2010) Differential gene expression in incompatible interaction between wheat and stripe rust fungus revealed by cDNA-AFLP and comparison to compatible interaction. BMC Plant Biol 10:9

Weber S, Friedt W, Landes N, Molinier J, Himber C, Rousselin P, Hahne G, Horn R (2003) Improved *Agrobacterium*-mediated transformation of sunflower (*Helianthus annuus* L.): assessment of macerating enzymes and sonication. Plant Cell Rep 21:475-482

Wegrzyn JL, Lee JM, Liechty J, Neale DB (2009) PineSAPE sequence alignment and SNP identification pipeline. Bioinformatics 25:2609-2610

Wei F, Gobelman-Werner K, Morroll SM, Kurth J, Mao L, Wing R, Leister D, Schulze-Lefert P, Wise RP (1999) The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. Genetics 153:1929-1948

Wendehenne D, Durner J, Klessig DF (2004) Nitric oxide: a new player in plant signalling and defence responses. Curr Opin Plant Biol 7:449-455

Whitham S, Dinesh-Kumar SP, Choi D, Hehl R, Corr C, Baker B (1994) The product of the tobacco mosaic virus resistance gene *N*: Similarity to toll and the interleukin-1 receptor. Cell 78:1101-1115

Wicker T, Yahiaoui N, Keller B (2007) Contrasting rates of evolution in *Pm3* loci from three wheat species and rice. Genetics 177:1207-1216

Wieckhorst S, Bachlava E, Dußle C, Tang S, Gao W, Saski C, Knapp S, Schön CC, Hahn V, Bauer E (2010) Fine mapping of the sunflower resistance locus Pl_{ARG} introduced from the wild species *Helianthus argophyllus*. Theor Appl Genet 121:1633-1644

Xiao S, Ellwood S, Calis O, Patrick E, Li T, Coleman M, Turner JG (2001) Broad-spectrum mildew resistance in *Arabidopsis thaliana* mediated by *RPW8*. Science 291:118-120

Yahiaoui N, Srichumpa P, Dudler R, Keller B (2004) Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. Plant J 37:528-538

Yan H, Jin W, Nagaki K, Tian S, Ouyang S, Buell CR, Talbert PB, Henikoff S, Jiang J (2005) Transcription and histone modifications in the recombination-free region spanning a rice centromere. Plant Cell 17:3227-3238

You FM, Huo N, Gu YQ, Luo M-C, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD (2008) BatchPrimer3: A high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9:253

Young ND, Tanksley SD (1989) RFLP analysis of the size of chromosomal segments retained around the *Tm-2* locus of tomato during backcross breeding. Theor Appl Genet 77:353-359

Yu J-K, Tang S, Slabaugh MB, Heesacker A, Cole G, Herring M, Soper J, Han F, Chu W-C, Webb DM, Thompson L, Edwards KJ, Berry S, Leon AJ, Grondona M, Olungu C, Maes N, Knapp SJ (2003) Towards a saturated molecular genetic linkage map for cultivated sunflower. Crop Sci 43:367-387

Zeven A, Knott D, Johnson R (1983) Investigation of linkage drag in near isogenic lines of wheat by testing for seedling reaction to races of stem rust, leaf rust and yellow rust. Euphytica 32:319-327

Zheng J, Svensson J, Madishetty K, Close T, Jiang T, Lonardi S (2006) OligoSpawn: a software tool for the design of overgo probes from large unigene datasets. BMC Bioinformatics 7:7

Zhu L, Zhang X, Tu L, Zeng F, Nie Y, Guo X (2007) Isolation and characterization of two novel dirigent-like genes highly induced in cotton (*Gossypium barbadense* and *G. hirsutum*) after infection by *Verticillium dahliae*. J Plant Pathol 89:41-45

Zimmer DE, Kinman ML (1972) Downy mildew resistance in cultivated sunflower and its inheritance. Crop Sci 12:749-751

8 Supplementary Material

8.1 Supplementary Tables

8.1.1 Laboratory

Table 8-1: Enzymes

Enzyme	Provider
*Bam*HI	New England Biolabs, Ipswich, MA, USA
*Bst*YI	New England Biolabs, Ipswich, MA, USA
DNA Ligase (T4)	New England Biolabs, Ipswich, MA, USA
DNA Polymerase I	Fermentas, St. Leon Roth, Germany
DNAse	Fermentas, St. Leon Roth, Germany
*Hind*III	Fermentas, St. Leon Roth, Germany
*Mse*I	New England Biolabs, Ipswich, MA, USA
*Not*I	New England Biolabs, Ipswich, MA, USA
RNase A	Qiagen GmbH, Hilden, Germany
RNase H	USB Europe GmbH, Staufen, Germany
*Rsa*I	Fermentas, St. Leon Roth, Germany
SuperScript® II Reverse Transcriptase	Invitrogen, Carlsbad, CA, USA
Taq DNA polymerase and buffer w/wo $MgCl_2$ and 25 mM $MgCl_2$ solution.	MP Biomedicals Europe, Illkirch, France
*Taq*I	Fermentas, St. Leon Roth, Germany
*Xba*I	New England Biolabs, Ipswich, MA, USA

Table 8-2: Kits

Kits	Provider
BigDye® Terminator v1.1 Cycle Sequencing Kit	Applied Biosystems Inc, Foster City, CA, USA
High Pure Plasmid Isolation Kit	Roche Diagnostics GmbH, Mannheim, Germany
MinElute Gel Extraction Kit	Qiagen GmbH, Hilden, Germany
NucleoSpin® Extract II Kit	Macherey-Nagel GmbH & Co. KG, Düren, Germany
NucleoSpin® RNA Plant	Macherey-Nagel GmbH & Co. KG, Düren, Germany
QIAquick PCR Purification Kit	Qiagen GmbH, Hilden, Germany
RevertAid™ H Minus First Strand cDNA Synthesis Kit	Fermentas, St. Leon Roth, Germany
RNeasyPlant Mini	Qiagen GmbH, Hilden, Germany
TOPO® TA Cloning® Kit (electro)	Invitrogen, Carlsbad, CA, USA

Table 8-3: Chemicals, electrophoresis components and DNA size standards

Material	Provider
Chemicals	
Acetic acid ROTIPURAN® ($C_2H_4O_2$)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Adenosine 5′-triphosphate disodium salt hydrate ($C_{10}H_{14}N_5Na_2O_{13}P_3 \cdot xH_2O$)	Sigma-Aldrich Chemie Gmbh, Munich, Germany
Agar-agar	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Binde silane	Sigma-Aldrich Chemie Gmbh, Munich, Germany
Boric acid (H_3BO_3)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Cetrimonium bromide ($C_{19}H_{42}BrN$)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Chloroform ($CHCl_3$)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Dextran ($C_6H_{10}O_5)_xOH$	VWR International GmnbH, Darmstadt, Germany
Diethylpyrocarbonate ($C_6H_{10}O_5$)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Dynabeads® M-280 Streptavidin	Invitrogen, Carlsbad, CA, USA
Ethanol absolute (C_2H_6O)	VWR International GmnbH, Darmstadt, Germany
Ethidium bromide ($C_{21}H_{20}BrN_3$)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Ethylenediamine tetraacetic acid ($C_{10}H_{16}N_2O_8$)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Formaldehyde ROTIPURAN® (CH_2O)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Formamide (CH_3NO)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Hi-Di™ Formamide	Applied Biosystems Inc, Foster City, CA, USA
Hydrochloric acid (HCl_{AQ})	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Isoamyl Alcohol ($C_5H_{12}O$)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Isopropyl alcohol (C_3H_8O)	VWR International GmnbH, Darmstadt, Germany
NAD grade I, free acid	Roche Diagnostics GmbH, Mannheim, Germany
Peptone from casein (Tryptone), pancreatic, granulated	VWR International GmnbH, Darmstadt, Germany
Polyvinylpyrrolidone ($(C_6H_9NO)_n$)	Sigma-Aldrich Chemie Gmbh, Munich, Germany

Table 8-3 (continued): Chemicals, electrophoresis components and DNA size standards

Materials	Providers
Chemicals	
POP-7™ Polymer for 3130/3130xl Genetic Analyzers	Applied Biosystems Inc, Foster City, CA, USA
Primers	Biomers.net GmbH, Ulm, Germany
Primers	Metabion International AG, Planegg-Martinsried, Germany
RainEx	Shell Car Care, Manchester, UK
Silver nitrate ($AgNO_3$)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Sodium chloride (NaCl)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Sodium hydroxide (NaOH)	Merck KGaA, Darmstadt, Germany
TEMED	Bio-Rad Laboratories GmbH, Munich, Germany
Tris ultra quality PUFFERAN® ($C_4H_{11}NO_3$)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
X-Gal ($C_{14}H_{15}BrClNO_6$)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Electrophoresis components	
PeqGOLD pulsed field agarose	PEQLAB Biotechnologie GMBH, Erlangen, Germany
Universal agarose	Bio&SELL e.K., Feucht bei Nürnberg, Germany
SERDOGEL SSCP 2 x concentrate	SERVA Electrophoresis GmbH, Heidelberg, Germany
Sequa Gel™ XR	Biozym Scientific GmbH, Hess. Oldendorf, Germany
DNA size standards	
GeneRuler 100bp DNA ladder	Fermentas, St. Leon Roth, Germany
Lambda DNA	Fermentas, St. Leon Roth, Germany
MidRange II PFG marker	New England Biolabs, Ipswich, MA, USA

Table 8-4: Buffers and solutions

Buffers and solutions	Composition		
Tris-HCl	1	M	Tris
		pH	8.0
Ethylenediamine tetraacetic acid (EDTA)	500	mM	EDTA
		pH	8.0
TE	10	mM	Tris-HCl
	1	mM	EDTA
		pH	8.0
1x TBE	89	mM	Tris
	89	mM	Boric acid
	2.84	mM	EDTA
		pH	8.3
1x TAE	40	mM	Tris
	20	mM	Acetic acid
	1	mM	EDTA
		pH	7.6
CTAB-extraction buffer	100	mM	Tris/HCl
	20	mM	EDTA
	1400	mM	Sodium chloride
	2	%	CTAB
	1	%	Sodium bisulfite
CIA	96	%	Chloroform
	4	%	Isoamylalcohol
Acetat mix	3	M	Sodium acetat
	10	M	Ammonium acetat
Silver stain fixer (SSCP)	780	ml	*A. bidest.* H_2O
	210	ml	Ethanol
	10	ml	Acetic acid
Silver nitrate solution (SSCP)	1000	ml	Silver stain fixer
	2	g	Silver nitrate
Developer (SSCP)	924	ml	*A. bidest.* H_2O
	75	ml	10 N NaOH
	1	ml	Formaldehyd (37%)
Silver stain fixer (cDNA-AFLP)	10	%	Glacial acetic acid
Silver nitrate solution (cDNA-AFLP)	1500	ml	*A. bidest.* H_2O
	2.25	ml	Formaldehyd (37%)
	1.5	g	Silver nitrate
Developer (cDNA-AFLP)	1500	ml	*A. bidest.* H_2O
	45	g	Sodium carbonate
	300	µl	Sodium thiosulfate (10 mg/ml)

Table 8-4 (continued): Buffers and solutions

Buffers and solutions	Composition		
LB-medium solid	10	g	Peptone from casein (Tryptone)
	5	g	Yeast extract
	10	g	Sodium chloride (NaCl)
	15	g	Agar-agar
add after autoclaving	3	ml	IPTG (20 mg/ml)
	3	ml	X-Gal (20 mg/ml)
	1.25	ml	Ampicillin (40 mg/ml)
LB-medium liquid	10	g	Peptone from casein (Tryptone)
	5	g	Yeast extract
	10	g	Sodium chloride (NaCl)
add after autoclaving	3	ml	IPTG (20 mg/ml)
	3	ml	X-Gal (20 mg/ml)
	1.25	ml	Ampicillin (40 mg/ml)
Suspension buffer	50	mM	Glucose
	10	mM	EDTA, pH 8.0
	25	mM	Tris-HCl, pH 8.0
	3	µl	RNase A (10 mg/ml)
Lysis buffer	0.2	N	Sodium hydroxide (NaOH)
	1	%	SDS
Neutralizing buffer	60	ml	5M Potassium acetat
	11.5	ml	Acetic acid
	28.5	ml	*A. bidest.* H_2O
Loading dye	500	ml	Formamid
	150	mg	Bromphenolblue
	150	mg	Xylencyanol
	10	ml	EDTA, pH 8.0

8.1.2 Primers

Table 8-5: List of publicly available markers which were screened for polymorphism in (cms)HA342xARG1575-2 and NDBLOS$_{sel}$xKWS04

		Polymorphism				
		(cms)HA342xARG1575-2		NDBLOS$_{sel}$xKWS04		
Name	**Type**	**Inheritance**	**Mapping**	**Inheritance**	**Mapping**	**Reference**
CRT194	SSR	dominant	no	dominant	no	Tang et al. 2003
CRT272		codominant	yes	codominant	yes	Tang et al. 2003
HT324		codominant	yes	monomorph	no	S.J. Knapp p.c.
HT446		codominant	yes	codominant	yes	S.J. Knapp p.c.
HT722		codominant	yes	dominant	no	S.J. Knapp p.c.
ORS53		codominant	yes	codominant	yes	Tang et al. 2003
ORS365		monomorph	no	monomorph	no	Tang et al. 2002
ORS371		codominant	yes	n.a.	no	Tang et al. 2002
ORS509		codominant	yes	monomorph	no	Tang et al. 2002
ORS543		codominant	yes	monomorph	no	Tang et al. 2002
ORS606		monomorph	no	codominant	no	Tang et al. 2003
ORS610		codominant	yes	codominant	yes	Tang et al. 2002
ORS662		codominant	yes	codominant	yes	Tang et al. 2002
ORS675		monomorph	no	codominant	yes	Tang et al. 2002
ORS710		codominant	no	n.a.	no	Tang et al. 2002
ORS716		codominant	yes	monomorph	no	Tang et al. 2002
ORS959		codominant	yes	codominant	no	Tang et al. 2002
ORS1039		codominant	no	monomorph	no	Tang et al. 2002
ORS1182		codominant	yes	codominant	yes	Tang et al. 2002
ORS1128		codominant	yes	codominant	yes	Tang et al. 2002
HT121	SNP	monomorph	no	n.a.	no	Lai et al. 2005a
HT206		monomorph	no	n.a.	no	Lai et al. 2005a
HT211		codominant	yes	codominant	yes	Lai et al. 2005a
HT242		codominant	no	n.a.	no	Lai et al. 2005a
HT244		codominant	yes	codominant	yes	Lai et al. 2005a
HT261		monomorph	no	n.a.	no	Lai et al. 2005a
RGC52a	RGC	codominant	yes	dominant	no	Radwan et al. 2008
RGC52b		codominant	yes	n.a.	no	Radwan et al. 2008
RGC151		codominant	yes	dominant	no	Radwan et al. 2008

n.a.: not analyzed

p.c.: personal communication

Table 8-6: Primers designed on BAC end sequences (BES) of the identified BAC clones. Amplicon size, methods used for analyzing the amplicons, population and remarks. Abbreviations are explained at the end of table.

Marker name	BES	Forward primer (5'-3')	Reverse primer (5'-3')	Amplicon size	Method [1)]	Population[2)]	Remarks
Co1-1_T7	HBA100B21_T7	ACTAACGATGTGTGGCATGG	CGTGTATAGTCGGGTTGAAAAA	664	d		
Co1-2_M13	HBA027E16_M13	CATTGTCTTTACCTTCTTTTATAGGG	AGAAAGCTCGGCTCGAAAT	209	a	HxA, NxK	failed
Co1-2_T7	HBA027E16_T7	AGCTCGGCTCAAACTCGAT	GCTGGCTCGATAAGGTTTACA	365	a	HxA, NxK	several fragments
Co1-3_M13	HBA102G02_M13	AGCTCGGCTCAAACTCGAT	AAGAAAGCTCGGCTCGAAAT	473	a	HxA, NxK	monomorph
Co1-4_M13	HBA426J17_M13	TGTTGATACGGATGATGAGGA	GATCCCCCACGTACAGAATG	555	b	HxA	dominant
Co1-4_T7	HBA426J17_T7	AGTTTTCTCTCACCGGCAAA	CTCGTGGGTAGTGCGGATAG	504	a	HxA, NxK	unspecific
Co1-5_M13	HBA028B12_M13	AAATCCAAATGCCAAGCAAG	CGAAACTCGTCTCTGCCTGT	486	a	HxA, NxK	monomorph
Co1-5_T7	HBA028B12_T7	GGTTTTCACCCTCCACATTG	CGCTTTTTGTTGCATACGTG	483	a	HxA, NxK	monomorph
Co3-1_T7	HBA269A09_T7	TTGGATCAAGTTCAAGGTGCT	CCGTGGGAACAACATAACATC	421	a	HxA, NxK	failed
Co3-2_M13	HBA323D15_M13	TCAAGAGCTACCAAATTCTCCA	TGGCACCATGAGGAGTGTT	487	a	HxA, NxK	codominant/dominant
Co3-2_T7	HBA323D15_T7	TCTCACTTTCACCGCACAGT	TGAGCCTTTTCTCAGCATCA	498	a	HxA, NxK	dominant
Co3-3_M13	HBA222M11_M13	GGTCATTCCTTGCTTCTCCA	TGTGGGAAAGGTCAGTGTCA	283	a	HxA, NxK	failed
Co3-3_T7	HBA222M11_T7	GTGGACTTGGGATTTTGCTG	TCTGACATTCACATCGGTTCA	209	d		
Co3-4_T7	HBA202H01_T7	TTTACCGCACGAAAAGGAAT	TCAATTAAAAATGCAAAATAACCA	617	b, a	HxA, NxK	dominant
Co3-5_M13	HBA054C15_M13	CTGGTGGACTTGGTGTACGA	AGCGTGACCGGAAAATTCTA	622	a	HxA H, NxK	monomorph/codominant
Co3-6_T7	HBA195O14_T7	CTTAGAGCCAGCGATCAAGG	TGCATCCAAGCGAAAAATTC	496	a	HxA, NxK	monomorph
Co3-7_M13	HBA164F24_M13	CGGAGTTTTGGAGCATTTGT	CGTTGTTACGGGCAGTTTTT	487	a	HxA, NxK	unspecific
Co3-7_T7	HBA164F24_T7	TCCCTCCTTCATTCCTAACTTG	GGAAGTACCATCGGCGTATC	548	a	HxA, NxK	dominant
Co3-8_T7	HBA177H05_T7	AGCCAAAACTGTCGATTTCTC	ATCAGACCTGCCGATGAAAC	501	a	HxA, NxK	monomorph
Co3-10_M13	HBA354P09_M13	ACTTTGCCACCATCAACCAT	GGTCCACCTCTTTTCAACCA	501	a	HxA, NxK	monomorph
Co5-1_M13	HBA225G08_M13	CCAAAGGTAGGGTCGATGAT	CCAGTGCCTCAAGAAGAGGA	510	a	HxA, NxK	monomorph
Co5-1_T7	HBA225G08_T7	ATCGTCGGCGTGGTAATC	CATGGAAAAACCCAGTTCAGA	513	a	HxA, NxK	monomorph
Co5-2_M13	HBA226B04_M13	ACGGTTAATGGGTCGGATAA	TGTTCGGTTCTTGAAAGCAT	503	a	HxA, NxK	unspecific
Co5-2_T7	HBA226B04_T7	AATCACCGTTGTATGTTGTTACTGTT	ATGCCACGAAAAATGTCCAC	689	a	HxA, NxK	monomorph

Table 8-6 (continued): Primers designed on BAC end sequences (BES) of the identified BAC clones. Amplicon size, methods used for analyzing the amplicons, population and remarks. Abbreviations are explained at the end of table.

Marker name	BES	Forward primer (5'-3')	Reverse primer (5'-3')	Amplicon size [bp]	Method [1)]	Population[2)]	Dominance
Co5-3_T7	HBA109P18_T7	TCAACTGGAGGGAGGATGAG	TTGCCTTTGTACCCGTAAGC	498	a	HxA, NxK	monomorph/dominant
Co6-1_M13	HBA213N01_M13	TGTTATGCAGTTCCCACGAG	CATGAGGTCGTTTTTCAAGG	517	a	HxA, NxK	dominant
Co6-1_T7	HBA213N01_T7	TCTTTCAAGTTTGGCGTTTT	GCTGCATGAGATGGACAAAT	495	a	HxA, NxK	unspecific
Co6-2_M13	HBA173M22_M13	GGTGCAACTGGTGTAGCAAA	TTCGCTTGAAAATGTTGTCG	472	a	HxA, NxK	monomorph
Co6-2_T7	HBA173M22_T7	TGTTCATGTGATCTTGTTGATTATG	TGACGTTTCCCACGTTACAA	493	a	HxA, NxK	dominant
Co7-1_M13	HBA522P18_M13	GGTATGACAAAAAGCGGCATA	CATCTTCCTTGAATGCCATTAG	159	d		
Co7-2_T7	HBA477O03_T7	CTCACTCCTAAGAGGGCTTATCACA	TAAGGGTATTCCGCTTCTGATGTC	678	d		
Co7_3_M13	HBA477O03_M13	GGGTTGGGACATAGTCACTAAACCTA	CCAAACCATACACCCCTTGTCTA	397	b, a	HxA, NxK	dominant, monomorph
Co7-4_T7	HBA170G12_T7	AGGGACTAATTGCGTCAAACTG	GAGGACCGTTAGGATCGTTTCTT	760	c	HxA, NxK	monomorph
Co7-5_M13	HBA170G12_M13	GACTCTTTGCTCTGGAGCTCATTT	CTCGAAGAGACACCAATAACGACA	375	a	NxK	monomorph
Co7-6_T7	HBA10N15_T7	CATCCTTGAGTCCATCAACATCAC	AGAATGACCCAAACCCTTTGTC	701	c	HxA, NxK	monomorph
Co7_7_T7	HBA90K07_T7	CAGTCTTTCAGACAGGCGTTATTG	ATTATGTGGACAGTGGGAGTGC	811	a	HxA, NxK	monomorph
Co7_8_M13	HBA90K07_M13	ACGGTGTCACTCATCGTCTTTCTAC	GTCGACCTCATCGTATAGCTCAATC	557	c	HxA, NxK	monomorph
Co7-9_M13	HBA10N15_M13	TCAATTACGTCACGGTACTCCCTAC	GTTGCCCTAAACGAGACTTCTAAGG	467	a	HxA, NxK	monomorph
Co7_10_T7	HBA399D04_T7	GTCGTCATCAGACATGGTGCTATAC	CTTTCAAGGTTATCCTTCGAGGTC	652	a	HxA, NxK	monomorph
Co7_11_M13	HBA399D04_M13	TTGACCCGCACACACTTTACAT	GTGCAGCTTGTCAGACTTCTTTG	502	b	HxA, NxK	dominant
Co7_12_M13	HBA327M01_M13	GGACATCATAAAGCACATCACG	CCATGCTCAGAATATCCTGTCTGT	220	d		
Co7-13_T7	HBA522P18_T7	CAAGGACCTTTCTTGCAATTC	GCAAAGAAATCTGTCCATACAAGTG	100	a	HxA, NxK	monomorph
Co7_14_T7	HBA327M01_T7	CTTGCAGATGGGTAGGTTTCTAGTC	TGACCCTTAATTGGCTTGTACG	464	c	HxA, NxK	monomorph

1) Method: a:SSCP, b: agarose gel, c) cloned and sequenced, d: failed
2) Population: HxA: (cms)HA342xARG1575-2, NxK: NDBLOS$_{sel}$xKWS04

Table 8-7: Primers for candidate genes designed on *de novo* assembly BRBS of cDNA derived from a susceptible (BS) and resistant bulk (BR). Marker names and primer pair sequences are indicated.

Marker name	Forward primer (5'-3')	Reverse primer (5'-3')
BRBScontig00945	AAACAGCTTCAACGGTAGTATTCC	TAGCTAAATCAGTCAACCACAACG
BRBScontig02834	CCGACGACTTTCAAAACTTGTATC	GCTTTAAGCTCTTCAAGTTGCCTAT
BRBScontig02873	CTGCAATTTTAGACCCGGTAGAT	GGTTCAAGCTCAGGTCAAGATAGT
BRBScontig07153	GCTTGGAATTTGTCTAGATTGGAG	ATCCATCTTGTGGAATGAGGTTAG
BRBScontig15072	CACAAACCGTGTCATAGACTTTTC	AGCCTTCTGATCGATACTTGAGTT
BRBScontig16592	GGGAACAAAGCTAGAGCACATAAT	GGCAGTTGAACCATCATGG
BRBScontig34117	TCCAAACAGTCTCAGTAGTTGGAG	AGCAAGACTTGGTATCCTGCTAAG

8.1.3 BAC clones/BAC end sequences

Table 8-8: High Information Content Fingerprinting (HICF) contigs. BAC clone, overgo probe name, BAC end sequence (BES) information which were used to anchore the HICF contigs to the sunflower map. Abbreviations are explained at the end of table.

HICF contig	BAC clone	Probe	BES	GeneBank Accn	Length [bp]	GC [%]	LG
Contig01	HBA027E16	HT211	HBA027E16_M13	JM495484	263	44.1	1
			HBA027E16_T7	JM495485	598	39.3	1
	HBA028B12	HT211	HBA028B12_M13	JM495486	587	42.3	1
			HBA028B12_T7	JM495487	587	45.7	1
	HBA100B21	HT211	HBA100B21_M13	JM495488	663	44.7	1
			HBA100B21_T7	JM495489	773	22.8	1
	HBA102G02	HT211	HBA102G02_M13	JM495490	663	42.2	1
			HBA102G02_T7	JM495491	429	47.3	1
	HBA426J17	HT211	HBA426J17_M13	JM495473	751	35.8	1
			HBA426J17_T7	JM495474	691	46.7	1
	HBA436D05	HT211	HBA436D05_M13	JM495475	710	31.4	1
			HBA436D05_T7	JM495476	668	46.3	1
Contig02	HBA111C11	RGC52	HBA111C11_M13	FI274071	622	36.7	16
			HBA111C11_T7	FI274355	773	32.0	16
	HBA225I18	RGC52	HBA225I18_M13	FI274085	299	26.4	16
			HBA225I18_T7	FI274373	708	38.6	16
	HBA289M18	RGC52	HBA289M18_M13	FI274094	603	29.5	16
			HBA289M18_T7	FI274382	740	43.5	16
	HBA289P24	RGC52	HBA289P24_M13	FI274095	239	36.8	16
			HBA289P24_T7	n.a.	n.a.	n.a.	16
	HBA330K07	RGC52	HBA330K07_M13	FI274099	386	26.7	16
			HBA330K07_T7	FI274387	767	41.6	16
	HBA341C01	RGC52	HBA341C01_M13	FI274103	451	29.3	16
			HBA341C01_T7	FI274391	602	38.2	16
	HBA408I22	RGC52	HBA408I22_M13	FI274112	489	28.2	16
			HBA408I22_T7	FI274401	728	36.8	16
	HBA415D01	RGC52	HBA415D01_M13	FI274113	493	39.6	16
			HBA415D01_T7	FI274403	677	43.7	16
	HBA485J08	RGC52	HBA485J08_M13	n.a.	n.a.	n.a.	16
			HBA485J08_T7	n.a.	n.a.	n.a.	16
	HBA490H21	RGC52	HBA490H21_M13	FI274125	533	30.2	16
			HBA490H21_T7	FI274415	694	29.8	16
	HBA59H15	RGC52	HBA59H15_M13	FI274062	532	45.9	16
			HBA59H15_T7	FI274346	761	38.5	16
Contig03	HBA054C15	RGC151	HBA054C15_M13	HN153247	704	37.9	1
			HBA054C15_T7	n.a.			1
	HBA164F24	RGC151	HBA164F24_M13	HN153248	652	49.2	1
			HBA164F24_T7	HN153249	649	31.9	1
	HBA177H05	RGC151	HBA177H05_M13	HN153250	560	42.1	1
			HBA177H05_T7	HN153251	663	41.8	1

Table 8-8 (continued): High Information Content Fingerprinting (HICF) contigs. BAC clone, overgo probe name, BAC end sequence (BES) information which were used to anchore the HICF contigs to the sunflower map. Abbreviations are explained at the end of table.

HICF contig	BAC clone	Probe	BES name	GeneBank Accn	Length [bp]	GC %	LG
	HBA195O14	RGC151	HBA195O14_M13	HN153252	328	27.4	1
			HBA195O14_T7	HN153253	681	48.2	1
	HBA202H01	RGC151	HBA202H01_M13	HN153254	174	41.4	1
			HBA202H01_T7	HN153255	679	19.7	1
	HBA222M11	RGC151	HBA222M11_M13	HN153256	347	51.6	1
			HBA222M11_T7	HN153257	364	33.0	1
	HBA269A09	RGC151	HBA269A09_M13	HN153258	730	40.4	1
			HBA269A09_T7	HN153259	538	37.2	1
	HBA323D15	RGC151	HBA323D15_M13	HN153260	705	36.0	1
			HBA323D15_T7	HN153261	660	35.6	1
	HBA354P09	RGC151	HBA354P09_M13	HN153262	702	35.8	1
			HBA354P09_T7	HN153263	540	41.9	1
Contig05	HBA109P18	ORS662	HBA109P18_M13	JM495492	774	38.4	n.a.
			HBA109P18_T7	JM495493	655	46.3	n.a.
	HBA225G08	ORS662	HBA225G08_M13	JM495480	745	38.5	n.a.
			HBA225G08_T7	JM495481	768	32.6	n.a.
	HBA226B04	ORS662	HBA226B04_M13	JM495482	782	40.5	n.a.
			HBA226B04_T7	JM495483	784	32.7	n.a.
Contig06	HBA173M22	ORS662	HBA173M22_M13	JM495494	695	35.4	1.
			HBA173M22_T7	JM495477	672	25.7	1
	HBA213N01	ORS662	HBA213N01_M13	JM495478	624	34.9	1
			HBA213N01_T7	JM495479	656	26.7	1
Contig07	HBA10N15	RGC52	HBA10N15_M13	FI274058	544	39.2	1
			HBA10N15_T7	FI274340	803	40.3	1
	HBA170G12	RGC52	HBA170G12_M13	FI274080	434	46.3	1
			HBA170G12_T7	FI274367	772	42.4	1
	HBA327M01	RGC52	HBA327M01_M13	FI274098	319	48.0	1
			HBA327M01_T7	FI274386	648	35.5	1
	HBA399D04	RGC52	HBA399D04_M13	FI274109	622	33.0	1
			HBA399D04_T7	FI274398	746	40.1	1
	HBA477O03	RGC52	HBA477O03_M13	FI274121	426	43.7	1
			HBA477O03_T7	FI274411	768	42.2	1
	HBA522P18	RGC52	HBA522P18_M13	FI274129	220	36.4	1
			HBA522P18_T7	FI274419	149	37.6	1
	HBA90K07	RGC52	HBA90K07_M13	FI274069	634	45.7	1
			HBA90K07_T7	FI274353	834	45.1	1

GeneBank_Accn: GeneBank Accession
LG: linkage group

n.a.: not analyzed

Table 8-9: Summary of sequencing raw data derived from 29 BAC clones which were generated on a half picotiterplate on a GS FLX platform and were assembled into six High Information Content Fingerprinting (HICF) contigs. BAC insert sizes were determined by pulsed field gel electrophoresis. Abbreviations are explained at the end of table.

HICF contig	BAC clone	Tag[1)]	# reads	bp	Av_len [bp][2)]	*E. coli* ratio[3)]	Vec ratio[4)]	# clean reads	Clean reads [bp]	BAC insert size [kbp]	Sequence coverage [kbp]
Contig 01	HBA027E16	TACTATG	11,792	4,135,503	351	1%	6%	11,208	3,866,219	115	34
	HBA028B12	TCTGATG	12,824	4,450,825	347	1%	7%	12,095	4,120,734	122	34
	HBA100B21	ATGCATG	11,221	3,880,380	346	1%	6%	10,640	3,612,967	120	30
	HBA102G02	ACTGTCG	5,023	1,751,532	349	1%	5%	4,782	1,643,049	252	7
	HBA426J17	TACGTCG	8,182	2,898,395	354	1%	8%	7,695	2,666,497	203	13
	HBA436D05	TGTCTCG	11,889	4,132,633	348	2%	6%	11,233	3,827,610	125	31
Contig 02	HBA111C11	TATCATC	11,351	3,960,896	349	1%	5%	10,911	3,760,797	160	24
	HBA289M18	TGAGTGC	20,467	7,101,533	347	2%	6%	19,272	6,543,548	130	50
	HBA289P24	TCGATGC	12,248	4,269,727	349	1%	5%	11,660	4,012,491	130	31
	HBA341C01	TAGTCGC	6,591	2,276,455	345	2%	6%	6,201	2,101,611	110	19
	HBA485J08	TGTACGC	18,238	6,519,688	357	2%	5%	17,398	6,113,458	135	45
Contig 03	HBA054C15	TAGATAG	14,763	5,176,221	351	1%	6%	14,033	4,862,914	125	39
	HBA164F24	ATGTGAG	28,793	9,925,997	345	1%	7%	27,309	9,208,706	110	84
	HBA177H05	AGTCGAG	15,792	5,591,095	354	2%	5%	15,031	5,245,696	n.a.	n.a.
	HBA195O14	TGCAGAG	21,246	7,428,296	350	1%	6%	20,254	6,968,453	130	54
	HBA202H01	TCGTCAG	13,454	4,728,601	351	1%	6%	12,707	4,405,899	125	35
	HBA222M11	TGTGCAG	19,042	6,716,698	353	2%	6%	17,950	6,216,558	120	52
	HBA269A09	TCGTGTC	13,775	4,855,675	352	1%	6%	13,038	4,525,894	110	41
	HBA323D15	TGCTCTC	15,465	5,382,268	348	2%	7%	14,559	4,962,921	110	45
	HBA354P09	AGATATC	17,319	5,999,535	346	1%	6%	16,521	5,624,960	125	45
Contig 05	HBA109P18	AGCTGCG	8,151	2,756,388	338	0%	6%	7,756	2,571,880	105	24
	HBA225G08	TCATGCG	12,306	4,244,661	345	1%	6%	11,779	3,983,091	125	32
	HBA226B04	ATCGACG	7,447	2,596,252	349	1%	5%	7,125	2,446,476	140	17

Table 8-9 (continued): Summary of sequencing raw data derived from 29 BAC clones which were generated on a half picotiterplate on a GS FLX platform and were assembled into six High Information Content Fingerprinting (HICF) contigs. BAC insert sizes were determined by pulsed field gel electrophoresis. Abbreviations are explained at the end of table.

HICF contig	BAC clone	Tag[1)]	# reads	bp	Av_len [bp][2)]	*E. coli* ratio[3)]	Vec ratio[4)]	# clean reads	Clean reads [bp]	BAC insert size [kbp]	Sequence coverage [kbp]
Contig 06	HBA173M22	AGAGTAG	19,609	6,957,064	355	1%	5%	18,758	6,534,105	150	44
	HBA213N01	TCACTAG	6,312	2,212,475	351	1%	6%	6,026	2,078,582	118	18
Contig 07	HBA090K07	ACGTAGC	6,851	2,383,511	348	2%	7%	6,411	2,172,489	110	20
	HBA327M01	TACGAGC	11,150	3,933,950	353	3%	5%	10,466	3,605,583	225	16
	HBA399D04	ATAGAGC	4,347	1,484,312	341	1%	12%	3,956	1,302,628	60	22
	HBA522P18	TCACAGC	10,895	3,829,907	352	1%	6%	10,385	3,588,002	240	15

1) Tag: Barcodes for tagging different BACs

2) Av_len: Average length [bp]

3) *E. coli* ratio: Ratio of reads derived from *E. coli* DNA

4) Vec ratio: Ratio of reads derived from vector sequences

Table 8-10: Summary of best MIRA assembly results for 29 BAC clones which were assembled separately and insert size determined by pulsed field gel electrophoresis. Abbreviations are explained at the end of table.

Contig	BAC	n[1]	n_100[2]	n_N50[3]	Min_len[4] [bp]	Median_len[5] [bp]	Av_len[6] [bp]	N50[7] [bp]	Max_len[8] [bp]	MIRA parameter	BAC insert size [bp]	n>1kb[9]	Sum>1kb[10]	Diff[11]
Contig 01	HBA027E16	40	40	3	209	547	3,254	12,962	45,907	BPH_24_HSS_3	115,000	10	114,657	0.00
	HBA028B12	74	67	3	103	458	2,162	13,000	49,012	BPH_9_HSS_8	122,000	13	122,798	0.01
	HBA100B21	39	38	2	117	575	3,409	13,196	56,754	BPH_21_HSS_3	120,000	9	114,894	-0.04
	HBA102G02	157	153	13	114	614	1,444	3,611	26,195	BPH_14_HSS_1	252,000	47	168,458	-0.33
	HBA426J17	114	111	3	104	426	1,226	10,460	53,884	BPH_7_HSS_7	203,000	9	93,254	-0.54
	HBA436D05	53	52	2	118	501	2,815	13,213	72,,298	BPH_9_HSS_8	125,000	11	127,840	0.02
Contig 02	HBA111C11	45	45	3	323	576	3,900	22,080	61,179	BPH_16_HSS_1	160,000	10	156,943	-0.02
	HBA289M18	77	73	2	139	605	2,308	31,245	53,330	BPH_9_HSS_5	130,000	10	131,909	0.01
	HBA289P24	24	23	2	129	536	6,147	45,020	51,529	BPH_9_HSS_4	130,000	6	133,288	0.03
	HBA341C01	30	30	2	133	569	4,290	23,223	51,467	BPH_25_HSS_11	110,000	8	117,847	0.07
	HBA485J08	27	27	1	102	507	5,280	119,967	119,967	BPH_16_HSS_4	135,000	5	131,744	-0.02
Contig 03	HBA054C15	57	56	1	301	596	2,745	85,503	85,503	BPH_25_HSS_1	125,000	9	127,751	0.02
	HBA164F24	104	103	3	104	551	2,508	16,239	80,016	BPH_7_HSS_6	110,000	16	216,089	0.96
	HBA177H05	129	121	3	107	496	1,886	27,018	69,854	BPH_21_HSS_1	n.d.	18	179,699	n.d.
	HBA195O14	50	47	1	172	610	3,273	81,412	81,412	BPH_18_HSS_3	130,000	7	131,407	0.01
	HBA202H01	53	53	3	161	619	2,781	13,070	51,674	BPH_7_HSS_6	125,000	12	125,130	0.00
	HBA222M11	107	102	3	102	438	1,542	13,065	53,875	BPH_9_HSS_6	120,000	10	118,838	-0.01
	HBA269A09	45	45	2	183	563	2,977	22,556	51,393	BPH_22_HSS_3	110,000	8	112,910	0.03
	HBA323D15	46	46	1	186	613	2,850	65,661	65,661	BPH_8_HSS_5	110,000	9	111,045	0.01
	HBA354P09	64	64	2	136	595	2,471	33,360	57,986	BPH_18_HSS_1	125,000	12	130,317	0.04
Contig 05	HBA109P18	31	31	1	234	608	3,693	60,028	60,028	BPH_23_HSS_1	105,000	6	101,464	-0.03
	HBA225G08	48	48	2	210	620	3,004	28,931	61,817	BPH_17_HSS_8	125,000	11	124,839	0.00
	HBA226B04	41	41	3	200	686	3,808	15,471	43,549	BPH_8_HSS_8	140,000	15	142,586	0.02
Contig 06	HBA173M22	61	59	1	123	502	2,990	99,187	99,187	BPH_23_HSS_7	150,000	10	152,572	0.02

Table 8-10 (continued): Summary of best MIRA assembly results for 29 BAC clones which were assembled separately and insert size determined by pulsed field gel electrophoresis. Abbreviations are explained at the end of table.

Contig	BAC	n[1]	n_100[2]	n_N50[3]	Min_len[4] [bp]	Median_len[5] [bp]	Av_len[6] [bp]	N50[7] [bp]	Max_len[8] [bp]	MIRA parameter	BAC insert size [bp]	n>1kb[9]	Sum>1kb[10]	Diff[11]
Contig 06	HBA213N01	69	68	5	339	638	2,223	8,330	35,632	BPH_8_HSS_6	118,000	19	123,432	0.05
Contig 07	HBA090K07	35	34	2	251	716	3,475	24,970	41,730	BPH_18_HSS_12	110,000	10	104,862	-0.05
	HBA327M01	127	123	5	104	781	2,041	5,746	94,826	BPH_16_HSS_4	225,000	44	208,976	-0.07
	HBA399D04	26	25	1	165	497	2,642	35,140	35,140	BPH_25_HSS_13	60,000	6	56,407	-0.06
	HBA522P18	73	73	4	256	1058	3,540	9,162	104,628	BPH_10_HSS_10	240,000	38	241,273	0.01

1) n: Number of subcontigs

2) n_100: Number of subcontigs larger 100 bp

3) n_N50: Number of subcontigs which cover more than 50% of the BAC clone

4) Min_len: Minimum length [bp] of subcontigs larger 100 bp

5) Median_len: Median length [bp] of subcontigs

6) Av_len: Average length [bp] of subcontigs

7) N50: N50 value [bp]

8) Max_len: Maximum length [bp] of subcontigs

9) n>1kb: Number of subcontigs lager than 1 kb

10) Sum>1kb: Sum length of all subcontigs lager than 1kb

11) Diff: Difference between size of assembled BACs and insert size determined by pulsed field gel electrophoresis

Table 8-11: Annotation of the high information content fingerprinting (HICF) contig-wise assembled BACs of HICF contig01, 02, 03, 05, 06, and 07. Abbreviations are explained at the end of table.

Gene name	Completeness	Start	Stop	Orien-tation	Accession	Putative protein function	Organism	E-value
Contig01_c1_00001	Predicted gene	1,232	7,145	+	XP_002526508.1	Dynamin	*Ricinus communis*	0.00 e^{+00}
Contig01_c1_00002	Gene fragment	11,886	12,575	+	XP_002532729.1	Protein binding protein	*Ricinus communis*	3.07 e^{-29}
Contig01_c1_00003	Gene fragment	27,348	31,285	-	BAB10790.1	Retroelement pol polyprotein-like	*Arabidopsis thaliana*	0.00 e^{+00}
Contig01_c1_00004	Gene fragment	41,381	42,583	+	AAT38724.1	Putative retrotransposon protein	*Solanum demissum*	8.36 e^{-129}
Contig01_c1_00005	Gene fragment	45,361	46,422	+	ABI34342.1	Polyprotein, putative	*Solanum demissum*	7.00 e^{-67}
Contig01_c1_00006	Predicted gene	54,895	55,221	-	AAW57784.1	Putative polyprotein	*Oryza sativa*	6.93 e^{-22}
Contig01_c1_00007	Gene fragment	58,853	62,189	+	AAX92979.1	Retrotransposon protein, putative, Ty3-gypsy sub-class	*Oryza sativa*	1.67 e^{-129}
Contig01_c1_00008	Gene fragment	67,759	68,670	-	AAO45752.1	Pol protein	*Cucumis melo*	1.53 e^{-103}
Contig01_c1_00009	Predicted gene	68,812	72,197	-	ABA98922.2	Retrotransposon protein, putative, Ty3-gypsy subclass	*Oryza sativa*	0.00 e^{+00}
Contig01_c13_00001	Gene fragment	967	3,360	+	AAO45752.1	Pol protein	*Cucumis melo*	0.00 e^{+00}
Contig01_c13_00002	Predicted gene	5,497	6,213	-	XP_002271693.1	Similar to retrotransposon protein, putative, Ty3-gypsy sub-class	*Vitis vinifera*	5.14 e^{-84}
Contig01_c13_00003	Gene fragment	12,837	15,059	+	AAP94599.1	Putative copia-type pol polyprotein	*Zea mays*	5.93 e^{-127}
Contig01_c13_00004	Predicted gene	18,688	19,752	-	AAX95143.1	Retrotransposon protein, putative, Ty3-gypsy sub-class putative, Ty3-gypsy subclass	*Oryza sativa*	3.12 e^{-23}
Contig01_c13_00005	Gene fragment	24,808	25,284	+	AAL59229.1	Gag-pol	*Zea mays*	9.83 e^{-08}
Contig01_c13_00006	Gene fragment	25,728	28,844	+	XP_002271693.1	Similar to retrotransposon protein, putative, Ty3-gypsy sub-class	*Vitis vinifera*	0.00 e^{+00}
Contig01_c3_00001	Predicted gene	520	7,115	-	XP_002526507.1	Malic enzyme, putative	*Ricinus communis*	0.00 e^{+00}

Table 8-11 (continued): Annotation of the high information content fingerprinting (HICF) contig-wise assembled BACs of HICF contig01, 02, 03, 05, 06, and 07. Abbreviations are explained at the end of table.

Gene name	Completeness	Start	Stop	Orien-tation	Accession	Putative protein function	Organism	E-value
Contig01_c3_00002	Predicted gene	22,568	27,142	+	AAT38724.1	Putative retrotransposon protein, identical	*Solanum demissum*	0.00 e^{+00}
Contig01_c10_00001	Gene fragment	274	1,038	+	AAL79340.1	Putative 22 kDa kafirin cluster; Ty3-Gypsy type	*Oryza sativa*	5.21 e^{-11}
Contig01_c10_00002	Gene fragment	7,132	7,395	+	AAO45752.1	Pol protein	*Cucumis melo*	8.65 e^{-38}
Contig01_c10_00003	Gene fragment	8,049	8,576	+	XP_002271693.1	Similar to retrotransposon protein, putative, Ty3-gypsy sub-class	*Vitis vinifera*	1.93 e^{-67}
Contig01_c14_00001	Predicted gene	16,138	16,830	-	ABI34354.1	Retrotransposon gag protein	*Solanum demissum*	3.04 e^{-73}
Contig01_c2_00001	Gene fragment	123	682	+	XP_002526507.1	Malic enzyme, putative	*Ricinus communis*	1.35 e^{-36}
Contig01_c2_00002	Predicted gene	908	1,880	-	AAY87906.1	Caleosin B	*Sesamum indicum*	1.86 e^{-38}
Contig01_c2_00003	Gene fragment	6,234	8,331	+	AAX95295.1	Reverse transcriptase (RNA-dependent DNA polymerase), putative	*Oryza sativa*	2.77 e^{-20}
Contig01_c2_00004	Gene fragment	12,223	12,393	+	XP_002526508.1	Dynamin, putative	*Ricinus communis*	3.71 e^{-12}
Contig01_c4_00001	Predicted gene	540	5,667	+	ABA95229.1	Retrotransposon protein, putative, Ty3-gypsy subclass	*Oryza sativa*	0.00 e^{+00}
Contig01_c7_00001	Predicted gene	3,161	3,561	+	NP_181989.1	RAB6A; GTP binding / protein binding	*Arabidopsis thaliana*	4.71 e^{-16}
Contig01_c12_00001	Predicted gene	8,564	9,037	-	XP_002529704.1	Mitochondrial carrier protein	*Ricinus communis*	1.54 e^{-20}
Contig01_c8_00001	Gene fragment	7,505	8,350	-	AAT38724.1	Putative retrotransposon protein, identical	*Solanum demissum*	5.64 e^{-95}
Contig01_c8_00002	Gene fragment	8,384	10,375	-	ABA98459.1	Retrotransposon protein, putative, Ty3-gypsy subclass	*Oryza sativa*	0.00 e^{+00}
Contig01_c15_00001	Gene fragment	4,274	4,396	+	n.d.	No hit		
Contig01_c33_00001	Gene fragment	4,398	4,520	+	n.d.	No hit		

Table 8-11 (continued): Annotation of the high information content fingerprinting (HICF) contig-wise assembled BACs of HICF contig01, 02, 03, 05, 06, and 07. Abbreviations are explained at the end of table.

Gene name	Completeness	Start	Stop	Orien-tation	Accession	Putative protein function	Organism	E-value
Contig01_c33_00002	Gene fragment	4,658	6,175	+	XP_002271693.1	Similar to retrotransposon protein, putative, Ty3-gypsy sub-class	*Vitis vinifera*	7.00 e^{-67}
Contig01_c6_00001	Gene fragment	891	1,397	+	XP_002327066.1	Predicted protein	*Populus trichocarpa*	5.20 e^{-26}
Contig02_c2_00001	Gene fragment	2,448	4,365	+	XP_002271693.1	Similar to retrotransposon protein, putative, Ty3-gypsy sub-class	*Vitis vinifera*	4.38 e^{-69}
Contig02_c2_00002	Gene fragment	16,991	20,374	+	ABF67921.1	Ji1 putative pol protein	*Zea mays*	0.00 e^{+00}
Contig02_c2_00003	Predicted gene	26,650	29,069	+	AAL79340.1	Putative 22 kDa kafirin cluster; Ty3-Gypsy type	*Oryza sativa*	5.73 e^{-72}
Contig02_c2_00004	Gene fragment	36,284	36,415	-	n.d.	No hit		
Contig02_c2_00005	Gene fragment	36,738	38,772	-	XP_002271693.1	Similar to retrotransposon protein, putative, Ty3-gypsy sub-class	*Vitis vinifera*	1.56 e^{-51}
Contig02_c2_00006	Gene fragment	40,063	41,011	-	AAR84410.2	Glyceraldehyde 3-phosphate dehydrogenase	*Daucus carota*	6.66 e^{-13}
Contig02_c2_00007	Gene fragment	43,911	44,165	-	Q6YZX6.1	Putative aconitate hydratase, cytoplasmic	*Oryza sativa*	3.52 e^{-08}
Contig02_c2_00008	Gene fragment	55,613	56,233	+	AAT38744.1	Putative gag-pol polyprotein, identical	*Solanum demissum*	5.96 e^{-85}
Contig02_c2_00009	Gene fragment	56,270	56,392	+	n.d.	No hit		
Contig02_c2_000010	Predicted gene	64,173	64,944	+	ABA93883.1	transposon protein, putative, Pong sub-class, expressed	*Oryza sativa*	7.00 e^{-42}

Table 8-11 (continued): Annotation of the high information content fingerprinting (HICF) contig-wise assembled BACs of HICF contig01, 02, 03, 05, 06, and 07. Abbreviations are explained at the end of table.

Gene name	Completeness	Start	Stop	Orien-tation	Accession	Putative protein function	Organism	E-value
Contig02_c2_000011	Predicted gene	65,637	67,696	+	ABF81449.1	TIR-NBS type disease resistance protein	*Populus trichocarpa*	3.94 e^{-55}
Contig02_c2_000012	Gene fragment	69,422	72,923	+	XP_002512273.1	Leucine-rich repeat-containing protein, putative	*Ricinus communis*	1.36 e^{-15}
Contig02_c2_000013	Gene fragment	73,915	74,193	+	AAF66824.1	Poly(A)-binding protein	*Nicotiana tabacum*	3.70 e^{-07}
Contig02_c2_000014	Predicted gene	74,426	89,785	-	ABR67407.1	Integrase	*Cucumis melo*	2.07 e^{-114}
Contig02_c2_000015	Gene fragment	90,812	91,205	+	AAF66824.1	Poly(A)-binding protein	*Nicotiana tabacum*	1.52 e^{-07}
Contig02_c5_00001	Gene fragment	25,393	25,554	+	BAB01023.1	Helicase-like protein	*Arabidopsis thaliana*	1.56 e^{-06}
Contig02_c1_00001	Gene fragment	7,368	8,621	-	AAT38724.1	Putative retrotransposon protein, identical	*Solanum demissum*	8.27 e^{-135}
Contig02_c1_00002	Gene fragment	8,871	9,983	-	ABI34339.1	Polyprotein, 3'-partial, putative	*Solanum demissum*	5.18 e^{-134}
Contig02_c1_00003	Gene fragment	11,705	11,863	-	n.d.	No hit		
Contig02_c1_00004	Predicted gene	15,387	16,167	-	AAX95481.1	Integrase core domain, putative	*Oryza sativa*	1.37 e^{-42}
Contig02_c1_00005	Gene fragment	22,040	22,870	+	ABA97250.1	Retrotransposon protein, putative, Ty3-gypsy subclass	*Oryza sativa*	2.14 e^{-55}
Contig02_c1_00006	Predicted gene	25,230	26,060	-	AAT38724.1	Putative retrotransposon protein, identical	*Solanum demissum*	3.15 e^{-53}
Contig02_c6_00001	Gene fragment	41	1,537	+	AAT38744.1	Putative gag-pol polyprotein, identical	*Solanum demissum*	2.00 e^{-165}
Contig02_c6_00002	Gene fragment	1,562	1,888	+	AAT38724.1	Putative retrotransposon protein, identical	*Solanum demissum*	7.73 e^{-22}
Contig02_c4_00001	Gene fragment	1,135	1,986	+	ABW74585.1	Pol polyprotein	*Boechera divaricarpa*	3.69 e^{-45}
Contig02_c4_00002	Predicted gene	3,263	4,102	+	ABF94034.1	Retrotransposon protein, putative, unclassified	*Oryza sativa*	5.59 e^{-49}

Table 8-11 (continued): Annotation of the high information content fingerprinting (HICF) contig-wise assembled BACs of HICF contig01, 02, 03, 05, 06, and 07. Abbreviations are explained at the end of table.

Gene name	Completeness	Start	Stop	Orien-tation	Accession	Putative protein function	Organism	E-value
Contig03_c2_00001	Predicted gene	3,291	7,837	+	ABM55240.1	Retrotransposon protein	*Beta vulgaris*	0.00 e^{+00}
Contig03_c2_00002	Gene fragment	20,805	22,349	-	ABA94813.1 ABA98459.1	Retrotransposon protein	*Beta vulgaris*	0.00 e^{+00}
Contig03_c2_00003	Gene fragment	28,050	28,199	+	n.d.	No hit		
Contig03_c2_00004	Gene fragment	28,717	28,974	+	AAX95912.1	Retrotransposon protein, putative, Ty3-gypsy sub-class	*Oryza sativa*	5.37 e^{-07}
Contig03_c2_00005	Gene fragment	29,039	31,596	+	AAT38724.1	Putative retrotransposon protein, identical	*Solanum demissum*	0.00 e^{+00}
Contig03_c2_00006	Gene fragment	36,679	36,972	-	XP_002332150.1	Argonaute protein group	*Populus trichocarpa*	1.45 e^{-21}
Contig03_c2_00007	Gene fragment	36,973	37,182	-	XP_002332150.1	Argonaute protein group	*Populus trichocarpa*	5.98 e^{-08}
Contig03_c2_00008	Gene fragment	38,066	38,272	-	XP_002303389.1	Predicted protein	*Populus trichocarpa*	3.42 e^{-16}
Contig03_c2_00009	Gene fragment	44,197	47,214	+	XP_002517696.1	TMV resistance protein N, putative	*Ricinus communis*	1.05 e^{-154}
Contig03_c2_000010	Gene fragment	56,510	59,020	+	AAT38724.1	Putative retrotransposon protein, identical	*Solanum demissum*	4.20 e^{-135}
Contig03_c2_000011	Gene fragment	66,762	67,822	-	NP_172814.1	F-box family protein	*Arabidopsis thaliana*	4.66 e^{-31}
Contig03_c2_000012	Predicted gene	77,101	78,123	+	AAO45751.1	Gag-protease polyprotein	*Cucumis melo*	5.78 e^{-18}
Contig03_c2_000013	Gene fragment	82,808	82,993	+	n.d.	No hit		
Contig03_c2_000014	Gene fragment	83,496	85,701	+	AAQ56407.1	Putative gag-pol polyprotein	*Oryza sativa*	2.00 e^{-113}
Contig03_c1_00001	Gene fragment	13,194	16,049	-	AAM94290.1	Putative TNP-like transposable element	*Sorghum bicolor*	0.00 e^{+00}
Contig03_c1_00002	Predicted gene	16,999	18,474	-	AAP44598.1	Putative retrotransposon gag protein	*Oryza sativa*	3.15 e^{-36}
Contig03_c1_00003	Predicted gene	32,368	32,778	-	XP_002447124.1	Hypothetical protein SORBIDRAFT_06g029070	*Sorghum bicolor*	1.29 e^{-45}

Table 8-11 (continued): Annotation of the high information content fingerprinting (HICF) contig-wise assembled BACs of HICF contig01, 02, 03, 05, 06, and 07. Abbreviations are explained at the end of table.

Gene name	Completeness	Start	Stop	Orien-tation	Accession	Putative protein function	Organism	E-value
Contig03_c1_00004	Gene fragment	36,528	36,886	+	Q0DVX2.2	Full=DEAD-box ATP-dependent RNA helicase 50	*Oryza sativa*	1.66 e^{-17}
Contig03_c1_00005	Gene fragment	40,243	42,540	-	AAU90262.1	Integrase core domain containing protein	*Oryza sativa*	5.77 e^{-115}
Contig03_c1_00006	Gene fragment	51,277	51,597	+	CAC42080.1	Polyamine oxidase	*Hordeum vulgare*	7.26 e^{-31}
Contig03_c1_00007	Gene fragment	51,918	52,810	+	CAC42080.1	Polyamine oxidase	*Hordeum vulgare*	3.35 e^{-70}
Contig03_c3_00001	Gene fragment	6,764	6,955	+	ABA93883.1	Transposon protein, putative, Pong sub-class, expressed	*Oryza sativa*	8.98 e^{-15}
Contig03_c3_00002	Gene fragment	8,402	10,958	-	NP_001078550.1	SRT2 (SIRTUIN 2); DNA binding / NAD or NADH binding / NAD-dependent histone deacetylase/ zinc ion binding	*Arabidopsis thaliana*	1.28 e^{-137}
Contig03_c3_00003	Predicted gene	14,933	19,490	-	XP_002273073.1	Hypothetical protein	*Vitis vinifera*	1.73 e^{-154}
Contig03_c3_00004	Predicted gene	24,605	29,107	-	AAN04919.1	Putative polyprotein	*Oryza sativa*	0.00 e^{+00}
Contig03_c21_00001	Predicted gene	5,216	9,192	-	BAA22288.1	Polyprotein	*Oryza australiensis*	0.00 e^{+00}
Contig03_c21_00002	Gene fragment	19,906	20,433	-	n.d.	No hit		
Contig03_c18_00001	Gene fragment	5,773	9,943	+	ABA98459.1	Retrotransposon protein, putative, Ty3-gypsy subclass	*Oryza sativa*	0.00 e^{+00}
Contig03_c24_00001	Gene fragment	3,509	5,245	+	BAA22288.1	Polyprotein	*Oryza australiensis*	0.00 e^{+00}
Contig03_c24_00002	Gene fragment	9,049	12,942	+	BAA22288.1	Polyprotein	*Oryza australiensis*	0.00 e^{+00}
Contig03_c5_00001	Predicted gene	125	2,244	-	AAP53852.1	HAT family dimerisation domain containing protein	*Oryza sativa*	6.33 e^{-51}
Contig03_c5_00002	Gene fragment	4,182	4,340	+	n.d.	No hit		

Table 8-11 (continued): Annotation of the high information content fingerprinting (HICF) contig-wise assembled BACs of HICF contig01, 02, 03, 05, 06, and 07. Abbreviations are explained at the end of table.

Gene name	Completeness	Start	Stop	Orien-tation	Accession	Putative protein function	Organism	E-value
Contig03_c13_00001	Gene fragment	3	735	+	XP_002510330.1	Ubiquitin-protein ligase, putative	*Ricinus communis*	2.29 e^{-17}
Contig03_c13_00002	Predicted gene	6,054	7,554	+	XP_002510330.1	Ubiquitin-protein ligase, putative	*Ricinus communis*	1.62 e^{-35}
Contig03_c7_00001	Gene fragment	5,408	9,373	+	ABD63142.1	Retrotransposon gag protein	*Asparagus officinalis*	0.00 e^{+00}
Contig03_c8_00001	Predicted gene	24	1,040	+	AAU44318.1	Putative polyprotein	*Oryza sativa*	3.10 e^{-62}
Contig03_c8_00002	Predicted gene	10,004	11,229	-	AAR00617.1	Hypothetical protein	*Oryza sativa*	4.89 e^{-49}
Contig03_c9_00001	Predicted gene	2,900	4,685	+	NP_172814.1	F-box family protein	*Arabidopsis thaliana*	1.31 e^{-28}
Contig03_c9_00002	Gene fragment	7,374	7,906	+	XP_002268894.1	Similar to SPPA (signal peptide peptidase); protease IV/ serine-type endopeptidase	*Vitis vinifera*	4.71 e^{-45}
Contig03_c33_00001	Gene fragment	4,039	5,059	+	ABA94311.1	Retrotransposon protein, putative, Ty1-copia subclass	*Oryza sativa*	3.33 e^{-19}
Contig03_c33_00002	Gene fragment	6,892	7,788	-	BAA22288.1	Polyprotein	*Oryza australiensis*	3.73 e^{-118}
Contig03_c6_00001	Gene fragment	2,250	2,468	-	NP_180147.2	Protein kinase family protein	*Arabidopsis thaliana*	2.84 e^{-14}
Contig03_c6_00002	Predicted gene	4,599	5,552	-	ABW81060.1	GagPol3	*Arabidopsis lyrata*	5.00 e^{-44}
Contig03_c31_00001	Gene fragment	1,768	4,063	-	ABD63142.1	Retrotransposon gag protein	*Asparagus officinalis*	2.21 e^{-175}
Contig03_c17_00001	Gene fragment	2,562	2,768	+	XP_002303389.1	Predicted protein	*Populus trichocarpa*	3.42 e^{-16}
Contig03_c17_00002	Gene fragment	3,648	3,857	+	XP_002332150.1	Argonaute protein group	*Populus trichocarpa*	5.98 e^{-08}
Contig03_c17_00003	Gene fragment	3,858	4,151	+	XP_002332150.1	Argonaute protein group	*Populus trichocarpa*	1.45 e^{-21}
Contig03_c69_00001	Predicted gene	3,436	3,573	-	ABN08586.1	Retrotransposon gag protein, , putative	*Medicago truncatula*	1.85 e^{-05}
Contig03_c40_00001	Predicted gene	90	1,028	-	YP_002405951.1	Transposase, IS4 family, IS10	*Escherichia coli*	0.00 e^{+00}
Contig03_c25_00001	Predicted gene	294	728	-	XP_002519358.1	Leucine-rich repeat-containing protein, putative	*Ricinus communis*	9.02 e^{-26}

Table 8-11 (continued): Annotation of the high information content fingerprinting (HICF) contig-wise assembled BACs of HICF contig01, 02, 03, 05, 06, and 07. Abbreviations are explained at the end of table.

Gene name	Completeness	Start	Stop	Orien-tation	Accession	Putative protein function	Organism	E-value
Contig05_c1_00001	Gene fragment	3,174	3,533	+	XP_002481393.1	Mutator-like element transposase, putative	*Talaromyces stipitatus*	5.95 e^{-21}
Contig05_c1_00002	Predicted gene	22,828	27,154	-	AAT38724.1	Putative retrotransposon protein, identical	*Solanum demissum*	0.00 e^{+00}
Contig05_c1_00003	Predicted gene	35,596	37,139	-	AAP52358.2	Retrotransposon protein, putative, Ty3-gypsy subclass	*Oryza sativa*	6.02 e^{-73}
Contig05_c1_00004	Gene fragment	45,299	47,062	-	AAY99339.1	Pol-polyprotein	*Silene latifolia*	2.50 e^{-130}
Contig05_c1_00005	Predicted gene	54,495	55,043	-	XP_002279786.1	Similar to binding	*Vitis vinifera*	8.19 e^{-40}
Contig05_c2_00001	Gene fragment	19,233	19,865	-	EAY99823.1	Hypothetical protein OsI_21814	*Oryza sativa*	4.70 e^{-09}
Contig05_c2_00002	Gene fragment	32,999	34,501	+	AAK50072.1	At1g02330/T6A9_12	*Arabidopsis thaliana*	1.95 e^{-34}
Contig05_c3_00001	Predicted gene	8,094	8,807	-	XP_002523598.1	DNA binding protein, putative	*Ricinus communis*	4.63 e^{-32}
Contig06_c2_00001	Gene fragment	4,709	5,122	+	XP_002271693.1	Similar to retrotransposon protein, putative, Ty3-gypsy sub-class	*Vitis vinifera*	2.00 e^{-43}
Contig06_c2_00002	Predicted gene	30,026	31,492	+	XP_002318869.1	Predicted protein	*Populus trichocarpa*	9.82 e^{-91}
Contig06_c2_00003	Predicted gene	32,154	33,461	-	ABN08132.1	Putative non-LTR retroelement reverse transcriptase	*Medicago truncatula*	1.00 e^{-30}
Contig06_c2_00004	Predicted gene	34,518	36,413	+	XP_002510197.1	Conserved hypothetical protein	*Ricinus communis*	4.86 e^{-81}
Contig06_c2_00005	Predicted gene	45,888	46,580	-	XP_002436807.1	Hypothetical protein SORBIDRAFT_10g009190	*Sorghum bicolor*	6.06 e^{-12}
Contig06_c2_00006	Predicted gene	48,405	49,634	+	ABQ08937.1	Transposase	*Oryza sativa*	3.52 e^{-87}
Contig06_c2_00007	Gene fragment	52,951	53,769	-	ABN05720.1	RNA-directed DNA polymerase (Reverse transcriptase)	*Medicago truncatula*	3.74 e^{-66}

Table 8-11 (continued): Annotation of the high information content fingerprinting (HICF) contig-wise assembled BACs of HICF contig01, 02, 03, 05, 06, and 07. Abbreviations are explained at the end of table.

Gene name	Completeness	Start	Stop	Orien-tation	Accession	Putative protein function	Organism	E-value
Contig06_c2_00008	Gene fragment	55,119	55,319	+	n.d.	No hit		
Contig06_c2_00009	Gene fragment	72,545	73,102	+	AAL79340.1	Putative 22 kDa kafirin cluster; Ty3-Gypsy type	*Oryza sativa*	1.49 e^{-05}
Contig06_c2_000010	Predicted gene	83,168	84,178	-	NP_001078504.1	MAPKKK21; ATP binding / protein kinase/ protein serine/threonine kinase	*Arabidopsis thaliana*	4.12 e^{-87}
Contig06_c2_000011	Gene fragment	96,011	96,337	+	ABA95205.1	Retrotransposon protein, putative, Ty1-copia subclass	*Oryza sativa*	3.00 e^{-12}
Contig06_c2_000012	Predicted gene	96,757	98,997	+	XP_002276920.1	Similar to Gag-protease-integrase-RT-RNaseH polyprotein	*Vitis vinifera*	0.00 e^{+00}
Contig06_c1_00001	Gene fragment	18,432	18,788	-	XP_002319038.1	Histone 2	*Populus trichocarpa*	6.31 e^{-48}
Contig06_c1_00002	Gene fragment	22,788	23,081	-	ABS78654.1	At4g02390-like protein	*Arabidopsis lyrata*	1.18 e^{-06}
Contig07_c1_00001	Predicted gene	5,762	8,890	-	AAV25234.1	Putative polyprotein	*Oryza sativa*	4.01 e^{-109}
Contig07_c1_00002	Gene fragment	10,049	12,946	+	CAN69053.1	Hypothetical protein	*Vitis vinifera*	3.33 e^{-76}
Contig07_c1_00003	Gene fragment	27,076	28,098	+	XP_002329457.1	Tir-nbs-lrr resistance protein	*Populus trichocarpa*	1.00 e^{-33}
Contig07_c1_00004	Predicted gene	30,303	31,816	-	AAT85031.1	Putative polyprotein	*Oryza sativa*	1.00 e^{-53}
Contig07_c1_00005	Gene fragment	38,846	39,067	-	n.d.	No hit		
Contig07_c1_00006	Predicted gene	46,003	46,425	-	ABG37650.2	CCHC-type integrase	*Populus trichocarpa*	6.00 e^{-15}
Contig07_c1_00007	Gene fragment	49,015	49,990	+	NP_001119319.1	Disease resistance protein (TIR-NBS-LRR class), putative	*Arabidopsis thaliana*	4.78 e^{-15}
Contig07_c1_00008	Gene fragment	51,677	52,107	+	AAF66824.1	Poly(A)-binding protein	*Nicotiana tabacum*	5.01 e^{-10}
Contig07_c1_00009	Gene fragment	53,164	57,959	+	T02087	gag/pol polyprotein	*Zea mays*	0.00 e^{+00}

Table 8-11 (continued): Annotation of the high information content fingerprinting (HICF) contig-wise assembled BACs of HICF contig01, 02, 03, 05, 06, and 07. Abbreviations are explained at the end of table.

Gene name	Completeness	Start	Stop	Orien-tation	Accession	Putative protein function	Organism	E-value
Contig07_c1_000010	Gene fragment	65,135	67,864	-	ABA97807.1	Retrotransposon protein, putative, Ty3-gypsy subclass	*Oryza sativa*	0.00 e^{+00}
Contig07_c1_000011	Gene fragment	67,907	68,221	-	XP_002271693.1	Retrotransposon protein	*Vitis vinifera*	7.71 e^{-12}
Contig07_c1_000012	Gene fragment	74,942	78,687	-	AAN05371.1	Putative retroelement	*Oryza sativa*	2.51 e^{-148}
Contig07_c1_000013	Predicted gene	87,437	87,838	-	AAP54205.2	Retrotransposon protein, putative, Ty3-gypsy subclass	*Oryza sativa*	2.35 e^{-26}
Contig07_c1_000014	Gene fragment	95,108	95,425	-	AAT40532.1	Hydrolase, alpha/beta fold family protein	*Solanum demissum*	3.98 e^{-15}
Contig07_c1_000015	Gene fragment	104,949	116,538	+	XP_002519358.1	Leucine-rich repeat-containing protein, putative	*Ricinus communis*	4.91 e^{-60}
Contig07_c1_000016	Gene fragment	116,770	117,024	-	XP_002280852.1	Hypothetical protein	*Vitis vinifera*	8.20 e^{-13}
Contig07_c1_000017	Gene fragment	117,193	117,405	-	AAO20078.1	Putative polyprotein	*Oryza sativa*	5.00 e^{-16}
Contig07_c3_00001	Gene fragment	7,621	7,872	-	n.d.	No hit		
Contig07_c3_00002	Gene fragment	11,428	11,982	-	CAJ38367.1	PolyA-binding protein	*Plantago major*	9.16 e^{-12}
Contig07_c3_00003	Gene fragment	12,416	12,655	+	AAR01752.1	Putative transcription elongation factor, 5'-partial	*Oryza sativa*	7.16 e^{-24}
Contig07_c2_00001	Gene fragment	5,510	6,166	+	XP_002271693.1	Similar to retrotransposon protein, putative, Ty3-gypsy sub-class	*Vitis vinifera*	4.09 e^{-86}
Contig07_c5_00001	Predicted gene	1,238	5,362	+	ADB85321.1	Putative retrotransposon protein	*Phyllostachys edulis*	2.00 e^{-13}
Contig07_c6_00001	Gene fragment	120	809	+	BAB10790.1	Retroelement pol polyprotein-like	*Arabidopsis thaliana*	6.44 e^{-69}
Contig07_c24_00001	Gene fragment	170	3,297	-	AAM94552.1	Putative polyprotein	*Oryza sativa*	3.94 e^{-147}
Contig07_c28_00001	Gene fragment	33	809	-	AAO45751.1	Gag-protease polyprotein	*Cucumis melo*	8.28 e^{-08}
Contig07_c16_00001	Predicted gene	261	4,923	+	AAT38724.1	Putative retrotransposon protein, identical	*Solanum demissum*	0.00 e^{+00}

Table 8-11 (continued): Annotation of the high information content fingerprinting (HICF) contig-wise assembled BACs of HICF contig01, 02, 03, 05, 06, and 07. Abbreviations are explained at the end of table.

Gene name	Completeness	Start	Stop	Orien-tation	Accession	Putative protein function	Organism	E-value
Contig07_c9_00001	Gene fragment	1,117	6,437	+	AAM15783.1	Putative DNA topoisomerase III	*Oryza sativa*	3.15 e^{-173}
Contig07_c10_00001	Gene fragment	1,740	4,912	-	NP_201197.1	DNA topoisomerase III alpha, putative	*Arabidopsis thaliana*	1.03 e^{-53}
Contig07_c49_00001	Predicted gene	62	277	-	AAO45752.1	Pol protein	*Cucumis melo*	6.06 e^{-26}
Contig07_c49_00002	Gene fragment	296	1,457	-	XP_002271693.1	Similar to retrotransposon protein, putative, Ty3-gypsy sub-class	*Vitis vinifera*	5.97 e^{-100}
Contig07_c27_00001	Gene fragment	3,279	4,725	-	XP_002510394.1	Acyltransferase, putative	*Ricinus communis*	1.86 e^{-172}
Contig07_c48_00001	Gene fragment	3,635	4,546	-	AAO45752.1	Pol protein	*Cucumis melo*	5.18 e^{-106}
Contig07_c51_00001	Predicted gene	211	3,644	+	ABM55240.1	Retrotransposon protein	*Beta vulgaris*	0.00 e^{+00}
Contig07_c41_00001	Predicted gene	3,008	3,271	-	AAV31171.1	Putative polyprotein, identical	*Solanum tuberosum*	1.00 e^{-30}
Contig07_c61_00001	Gene fragment	14	3,531	-	BAH80025.1	Putative retrotransposon protein	*Oryza sativa*	0.00 e^{+00}
Contig07_c54_00001	Predicted gene	1,143	2,407	-	XP_002512792.1	Xyloglucan endotransglucosylase/hydrolase protein A precursor, putative	*Ricinus communis*	3.70 e^{-138}
Contig07_c32_00001	Gene fragment	2,464	3,214	-	XP_002321313.1	Predicted protein	*Populus trichocarpa*	9.30 e^{-26}
Contig07_c12_00001	Predicted gene	1,342	2,247	+	NP_201197.1	DNA topoisomerase III alpha, putative	*Arabidopsis thaliana*	1.60 e^{-57}
Contig07_c58_00001	Predicted gene	41	2,488	+	AAY99339.1	Pol-polyprotein	*Silene latifolia*	9.08 e^{-172}
Contig07_c35_00001	Gene fragment	31	2,681	+	XP_002510387.1	Cytosolic purine 5-nucleotidase, putative	*Ricinus communis*	3.91 e^{-47}
Contig07_c4_00001	Gene fragment	1,302	1,973	-	ABA95230.1	Retrotransposon protein, putative, unclassified	*Oryza sativa*	4.00 e^{-58}

Table 8-11 (continued): Annotation of the high information content fingerprinting (HICF) contig-wise assembled BACs of HICF contig01, 02, 03, 05, 06, and 07. Abbreviations are explained at the end of table.

Gene name	Completeness	Start	Stop	Orien-tation	Accession	Putative protein function	Organism	E-value
Contig07_c4_00002	Gene fragment	2,346	2,615	-	CAN75051.1	Hypothetical protein	*Vitis vinifera*	3.88 e^{-20}
Contig07_c79_00001	Gene fragment	702	1,928	+	XP_002512693.1	Zinc finger protein, putative	*Ricinus communis*	2.61 e^{-47}
Contig07_c38_00001	Gene fragment	109	1,680	-	ABM55240.1	Retrotransposon protein	*Beta vulgaris*	0.00 e^{+00}
Contig07_c38_00002	Gene fragment	1,778	2,368	+	NP_180929.1	DEAD box RNA helicase, putative	*Arabidopsis thaliana*	3.11 e^{-20}
Contig07_c60_00001	Gene fragment	1,738	2,307	-	AAO45751.1	Gag-protease polyprotein	*Cucumis melo*	1.31 e^{-11}
Contig07_c85_00001	Gene fragment	1,801	2,229	-	CAN74600.1	Hypothetical protein	*Vitis vinifera*	3.96 e^{-18}
Contig07_c78_00001	Predicted gene	181	1,697	+	AAT38724.1	Putative retrotransposon protein, identical	*Solanum demissum*	2.70 e^{-157}

n.d.: not detecte
+: plus strand
-: minus strand

8.1.4 Reference genes

Table 8-12: Resistance gene reference data set

RGC class	SequenceID	Species Name	GeneID	GenBank Acc.no. (nt)	GenBank Acc.no. (aa)
CNL	Ath.RPP13	*Arabidopsis thaliana*	RPP13	NM_114520	NP_190237
	Ath.RPP8	*Arabidopsis thaliana*	RPP8	NM_123713	NP_199160
	Ath.Rps2	*Arabidopsis thaliana*	Rps2	NM_118742	NP_194339
	Ath.RPS5	*Arabidopsis thaliana*	RPS5	NM_101094	NP_172686
	Cch.Bs2	*Capsicum chacoense*	Bs2	AF202179	AAF09256
	Les.Hero	*Lycopersicon esculentum*	Hero	AJ457052	CAD29729
	Les.Mi1.2	*Lycopersicon esculentum*	Mi1.2	AF039682	AAC67238
	Lsa.Dm3_RGC2B	*Lactuca sativa*	Dm3_RGC2B	AH007213	AAD03156
	Osa.Pi9	*Oryza sativa*	Pi9	DQ285630	ABB88855
	Osa.Pi-ta	*Oryza sativa*	Pi-ta	AY196754	AAO45178
	Zma.Rp1-D	*Zea mays*	Rp1-D	AF107293	AAD47197
	Han.PI8	*Helianthus annuus*	PI8	AY490793	AAT08955
	Osa.Pi36	*Oryza sativa*	Pi36	DQ900896	ABI64281
	Tae.Lr10	*Triticum aestivum*	Lr10	AY270157	AAQ01784
	Hvu.MLA13	*Hordeum vulgare*	MLA13	AF523683	AAO16014
	Ata.Cre1	*Aegilops tauschii*	Cre1	AY124651	AAM94164
	Ata.Cre3	*Aegilops tauschii*	Cre3	AF052641	AAC05834
	Gma.Rps1-k-1	*Glycine max*	Rps1-k-1	EU450800	AAX89382
	Gma.Rps1-k-2	*Glycine max*	Rps1-k-2	EU450800	AAX89383
	Hvu.Mla12	*Hordeum vulgare*	Mla12	AY196347	AAO43441
	Hvu.MLA1	*Hordeum vulgare*	MLA1	GU245961	ACZ65507
	Hvu.Mla6	*Hordeum vulgare*	Mla6	AJ302293	CAC29242
	Les.I-2	*Lycopersicon esculentum*	I-2	AF118127	AAD27815
	Osa.Pi2	*Oryza sativa*	Pi2	DQ352453	ABC94599
	Osa.Piz-t	*Oryza sativa*	Piz-t	DQ352040	ABC73398
	Osa.XA1	*Oryza sativa*	XA1	AB002266	BAA25068
	Sbu.Rpi-blb1	*Solanum bulbocastanum*	Rpi-blb1	AY336128	AAP86601
	Stu.R3a	*Solanum tuberosum*	R3a	AY849382	AAW48299
	Tae.Pm3	*Triticum aestivum*	Pm3	AY325736	AAQ96158
	Les.Sw-5	*Lycopersicon esculentum*	Sw-5	AY007366	AAG31013
	Les.Tm-2	*Lycopersicon esculentum*	Tm-2	AF536200	AAQ10735
	Les.Tm-2a	*Lycopersicon esculentum*	Tm-2a	AF536201	AAQ10736
	Lpi.Prf	*Lycopersicon pimpinellifolium*	Prf	AF220602	AAF76308
	Sde.R1	*Solanum demissum*	R1	AF447489	AAL39063
	Ath.HRT	*Arabidopsis thaliana*	HRT	AF234174	AAF36987
	Ath.RPM1	*Arabidopsis thaliana*	RPM1	NM_111584	NP_187360
	Sac.Rx2	*Solanum acaule*	Rx2	AJ249448	CAB56299
	Sbu.Rpi-blb2	*Solanum bulbocastanum*	Rpi-blb2	DQ122125	AAZ95005
	Stu.Gpa2	*Solanum tuberosum*	Gpa2	AF195939	AAF04603
	Stu.Rx	*Solanum tuberosum*	Rx	AJ011801	CAB50786
	Ath.RCY1	*Arabidopsis thaliana*	RCY1	AB087829	BAC67706
	Hvu.MLA10	*Hordeum vulgare*	MLA10	AY266445	AAQ55541
	Osa.PIB.5	*Oryza sativa*	PIB	AB013449	BAA76282

Table 8-12 (continued): Resistance gene reference data set

RGC class	SequenceID	Species Name	GeneID	GenBank Acc.no. (nt)	GenBank Acc.no. (aa)
TNL	Ath.RPP1	*Arabidopsis thaliana*	RPP1	NM_114316	NP_190034
	Ath.RPP4	*Arabidopsis thaliana*	RPP4	NM_117790	NP_193420
	Ath.RPP5	*Arabidopsis thaliana*	RPP5	NM_117798	NP_193428
	Ath.Rps4	*Arabidopsis thaliana*	Rps4	NM_123893	NP_199338
	Ath.RRS1	*Arabidopsis thaliana*	RRS1	NM_001085246	NP_001078715
	Ath.SSI4	*Arabidopsis thaliana*	SSI4	AY179750	AAN86124
	Les.Bs4	*Lycopersicon esculentum*	Bs4	AY438027	AAR21295
	Lus.L6	*Linum usitatissimum*	L6	U27081	AAA91022
	Lus.M	*Linum usitatissimum*	M	U73916	AAB47618
	Lus.P2	*Linum usitatissimum*	P2	AF310960	AAK28805
	Ngl.N	*Nicotiana glutinosa*	N	U15605	AAA50763
	Stu.Gro1.4	*Solanum tuberosum*	Gro1.4	AY196151	AAP44390
	Stu.RY-1	*Solanum tuberosum*	RY-1	AJ300266	CAC82812
RLK	Ath.EFR	*Arabidopsis thaliana*	EFR	NM_122055	NP_197548
	Ath.ER_Erecta	*Arabidopsis thaliana*	ER_Erecta	NM_128190	NP_180201
	Ath.FLS2	*Arabidopsis thaliana*	FLS2	NM_124003	NP_199445
	Ath.PEPR1	*Arabidopsis thaliana*	PEPR1	NM_105966	NP_177451
	Ath.RFO1	*Arabidopsis thaliana*	RFO1	NM_106616	NP_178085
	Hvu.RPG1	*Hordeum vulgare*	RPG1	DQ854803	ABK51311
	Lpi.Pto	*Lycopersicon pimpinellifolium*	Pto	U02271	AAC48914
	Osa.xa21	*Oryza sativa*	xa21	AB212799	BAE93934
	Osa.Xa26	*Oryza sativa*	Xa26	DQ426646	ABD84047
RLP	Ath.RPP27	*Arabidopsis thaliana*	RPP27	AJ585978	CAE51863
	Les.Cf-5	*Lycopersicon esculentum*	Cf-5	AF053993	AAC78591
	Les.LeEIX1	*Lycopersicon esculentum*	LeEIX1	AY359965	AAR28377
	Les.LeEIX2	*Lycopersicon esculentum*	LeEIX2	AY359966	AAR28378
	Les.Ve1	*Lycopersicon esculentum*	Ve1	AF272367	AAK58682
	Les.Ve2	*Lycopersicon esculentum*	Ve2	AF365929	AAK58011
	Lhi.Cf-4	*Lycopersicon hirsutum*	Cf-4	AJ002235	CAA05268
	Lhi.Cf4A	*Lycopersicon hirsutum*	Cf4A	AJ002235	CAA05269
	Lpi.Cf-2	*Lycopersicon pimpinellifolium*	Cf-2	U42444	AAC15779
	Lpi.Cf-9	*Lycopersicon pimpinellifolium*	Cf-9	AJ002236	CAA05274
	Lpi.Cf9B	*Lycopersicon pimpinellifolium*	Cf9B	AJ002236	CAA05273
	Pvu.PGIP	*Phaseolus vulgaris*	PGIP	X64769	CAA46016

Table 8-12 (continued): Resistance gene reference data set

RGC class	SequenceID	Species Name	GeneID	GenBank Acc.no. (nt)	GenBank Acc.no. (aa)
other	Ath.RPW8.1	*Arabidopsis thaliana*	RPW8.1	AF273059	AAK09266
	Ath.RPW8.2	*Arabidopsis thaliana*	RPW8.2	AF273059	AAK09267
	Ath.RTM1	*Arabidopsis thaliana*	RTM1	NM_100456	NP_172067
	Ath.RTM2	*Arabidopsis thaliana*	RTM2	NM_120571	NP_568144
	Can.Bs3	*Capsicum annuum*	Bs3	EU078684	ABW82012
	Can.Bs3-E	*Capsicum annuum*	Bs3-E	EU078683	ABW82011
	Cme.At1	*Cucumis melo*	At1	AY066012	AAL47679
	Cme.At2	*Cucumis melo*	At2	AF461048	AAL62332
	Hvu.Mlo	*Hordeum vulgare*	Mlo	Z83834	CAB06083
	Les.Asc-1	*Lycopersicon esculentum*	Asc-1	AF198177	AAF67518
	Osa.Xa13	*Oryza sativa*	Xa13	DQ421396	ABD78944
	Osa.xa27	*Oryza sativa*	xa27	AY986491	AAY54163
	Osa.Xa5	*Oryza sativa*	Xa5	AY643716	AAV53715
	Zma.Hm1	*Zea mays*	Hm1	NM_001112450	NP_001105920
	Zma.Hm2	*Zea mays*	Hm2	EU367521	ABY68564
	Bpr.Hs1	*Beta procumbens*	Hs1	U79733	AAB48305

8.1.5 Bulked segregant transcriptome analysis

Table 8-13: Graphical genotypes and composition of the resistant bulk (BR) containing 16 progenies in (cms)HA342xARG1575-2 and their marker scores.

Bulk R								**Target region**									
	F2/F3 progeny	CRT 272	ORS 543	ORS 1128	ORS 1182	ORS 610	ORS 509	ORS 662	ORS 716	PI_{ARG}	RGC 151	HT 722	HT 446	HT 324	ORS 053	ORS 959	ORS 371
1	4001-016	B	B	B	B	B	B	B	B	B	B	B	B	B	H	H	H
2	4004-033	B	B	B	B	B	B	B	B	B	B	B	B	B	H	H	H
3	4005-112	H	H	H	H	B	B	B	B	B	B	B	B	B	B	B	B
4	4005-155	B	B	B	B	B	B	B	B	B	B	B	B	H	H	H	H
5	4006-012	H	H	H	H	B	B	B	B	B	B	B	B	B	B	B	B
6	4006-072	B	B	B	B	B	B	B	B	B	B	B	B	B	H	H	H
7	4007-083	B	B	B	B	B	B	B	B	B	B	B	B	B	H	H	H
8	4007-096	B	B	B	B	B	B	B	B	n.d.	B	B	B	B	H	H	H
9	697	B	B	B	B	B	B	B	B	n.d.	B	B	B	B	H	H	H
10	781	B	B	B	B	B	B	B	B	B	B	B	B	B	H	H	H
11	1071	B	B	B	B	B	B	B	B	B	B	B	B	B	H	H	H
12	1092	B	B	B	B	B	B	B	B	n.d.	B	B	B	B	H	H	H
13	2473	B	B	B	B	B	B	B	B	B	B	B	B	B	H	H	H
14	2304	B	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A
15	2105	A	A	A	A	A	A	B	B	B	B	B	B	n.d.	B	B	B
16	2313	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B

Yellow A: (cms)HA342 allele
Grey H: heterozygous
Green B: ARG1575-2 allele
n.d.: not determined

Table 8-14: Graphical genotypes and composition of the susceptible bulk (BS) containing 16 progenies in (cms)HA342xARG1575-2 and their marker scores.

Bulk S								Target region									
	F2 progeny	CRT 272	ORS 543	ORS 1128	ORS 1182	ORS 610	ORS 509	ORS 662	ORS 716	PI_{ARG}	RGC 151	HT 722	HT 446	HT 324	ORS 053	ORS 959	ORS 371
1	4006-056	A	A	A	A	A	A	A	A	n.d.	A	A	A	H	H	H	H
2	4007-006	A	A	A	A	A	A	A	A	n.d.	A	A	A	H	H	H	H
3	4007-034	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H
4	4007-126	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H	H
5	4008-120	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H
6	4008-195	A	A	A	A	A	A	A	A	n.d.	A	A	A	A	H	H	H
7	4010-007	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H
8	4010-034	A	A	A	A	A	A	A	A	n.d.	A	A	A	A	H	H	H
9	4010-199	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H
10	390	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H
11	1010	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H
12	1343	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H
13	1731	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H
14	2127	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H
15	2245	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H	H
16	2394	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H

Yellow A: (cms)HA342 allele
Grey H: heterozygous
n.d.: not determined

Table 8-15: Predicted genes and contig sequence of the BRBS assembly. SNP position, allele and read coverage of the resistant bulk (BR) and susceptible bulk (BS). Numbers of observed reads for each allele are indicated in brackets. Abbreviations are explained at the end of table.

BAC subcontig	Contig name	SNP position	BR	BS
contig01_c1_00001	BRBScontig26395	449	G(5)	A(4)
		482	G(3)	C(5)
		719	T(2)	G(4)
		740	G(2)	A(4)
		826-828	GGC(2)	TGA(4)
		853	T(2)	G(3)
		865	A(2)	G(3)
		865-867	ATC(2)	GTT(3)
	BRBScontig46810	84	T(3)	C(3)
		138	A(4)	G(4)
contig01_c3_00001	BRBScontig55568	38	G(9)	C(9)
		57	G(25)	A(32)/G(3)
		75	C(36)/A(3)	A(37)
		316	G(31)/A(2)	A(32)
		378	T(19)/C(2)	C(30)/T(1)
		382	A(27)	G(24)/A(1)
		541	G(3)	A(4)
		838	G(4)/C(1)	C(8)
		952	C(3)/T(1)	T(5)
		1,024	G(3)/C(1)	C(7)
		1,081	G(2)	A(3)
		1,679	G(8)/T(1)	T(5)
		1,720	A(26)/G(1)	G(22)
		1,743	C(25)/T(1)	T(22)
		1,794	A(19)/C(1)	C(20)
		1,797	T(19)/C(1)	C(20)
		1,863	C(15)/A(1)	A(19)
contig01_c7_00001	BRBScontig25525	143	C(7)	T(5)
		152	T(7)	C(5)
		279	T(4)	C(8)
		377	T(6)	C(9)
		420	A(5)	C(9)
		653	C(8)	T(15)
		713	T(8)	G(15)
		719	G(7)	A(15)

Table 8-15 (continued): Predicted genes and contig sequence of the BRBS assembly. SNP position, allele and read coverage of the resistant bulk (BR) and susceptible bulk (BS). Numbers of observed reads for each allele are indicated in brackets. Abbreviations are explained at the end of table.

BAC subcontig	Contig name	SNP position	BR	BS
contig03_c3_00002	BRBScontig28758	6	A(5)	T(2)
		11	C(5)	T(2)
		69	A(5)	T(1)
		78	T(5)	C(1)
		102	G(4)	C(1)
		266	A(2)	T(1)
		842	C(3)	T(7)
		1,013	A(5)	C(6)
contig03_c3_00003	BRBScontig38888	453	T(5)	C(2)
		577	A(2)	T(1)
contig03_c3_00003	BRBScontig33656	393	G(2)	A(3)
		402	G(2)	A(3)
		486	A(2)	C(1)
contig03_c13	BRBScontig23377	32	C(3)	T(5)
		81	C(3)	T(3)
		93	G(3)	A(3)
contig05_c3_00001	BRBScontig47802	SNPs n.d.*		
contig06_c2_00004	BRBScontig03342	SNPs n.d.*		
	BRBScontig34449	SNPs n.d.*		
contig06_c2_000010	BRBScontig12826	404	T(3)	C(3)/T(1)
contig07_c1_000015	BRBScontig24208		-	only BS
	BRBScontig35838		-	only BS
contig07_c54_00001	BRBScontig11750	SNPs n.d.*		

*not detected

Table 8-16: Contig name of the BRBS assembly, SNP position and allele of the resistant bulk (BR) and susceptible bulk (BS) of candidate genes identified by cDNA bulk analysis. Numbers of observed reads are indicated in brackets.

Contig name	SNP position	Allele BR	Allel BS
BRBScontig00945	362	T (1)	C (3)
	413	G (1)	T (3)
BRBScontig02834	25	G (3)	A (2)
	125	G (2)	T (2)
BRBScontig02873	8	C (5)	T (4)
	34	A (6)	G (4)
	42	G (6)	A (4)
	54	C (6)	G (4)
	58	A (6)	G (4)
	62	T (6)	C (4)
	72	A (6)	G (4)
	73	T (6)	C (4)
	80	A (6)	C (4)
	117	A (6)	G (4)
	156	C (7)	T (4)
	237	A (3)	G (4)
	252	T (3)	C (4)
	276	G (2)	T (3)
	336	A (2)	G (2)
	666	G (5)	A (5)
	741	T (7)	C (6)
	793	G (6)	A (5)
	843	T (6)	G (5)
	877	T (4)	A (4)
	889	G (4)	A (4)
BRBScontig15072	62	C (2)	T (9)
	110	T (2)	C (9)
	113	A (2)	G (11)
	128	C (2)	A (11)
	147	C (2)	T (11)
	223	T (2)	C (10)
	275	C (2)	T (12)
	299	C (2)	G (12)
	353	A (3)	G (13)
	376	A (3)	G (13)
	428	T (3)	C (12)
	449	A (3)	G (10)
BRBScontig16592	165	C (3)	T (3)
	216	G (5)	A (4)
	225	T (5)	C (4)
	231	A (6)	T (4)
	255	A (7)	G (6)

Table 8-16 (continued): Contig name of the BRBS assembly, SNP position and allele of the resistant bulk (BR) and susceptible bulk (BS) of candidate genes identified by cDNA bulk analysis. Numbers of observed reads are indicated in brackets.

Contig name	SNP position	Allele BR	Allel BS
BRBScontig16592	279	C (7)	T (6)
	303	T (7)	G (6)
	435	C (4)	T (5)
	479	C (3)	G (5)
	511	G (3)	A (2)
	517	T (3)	G (2)
	519	G (3)	T (2)

8.2 Sequence resources

8.2.1 BAC derived resistance gene candidate sequences

>RGC151SW1_03_c2_44150_48500
TAACCAAAACCAACCATTTACTTCACACTCTATTTTCTTTACTCACAATGGTGGTTCTTACTGAAATTAC
AGAACAATCTTCTTCTTCAACTACCTTATCAACTTACGATTATGATGTATTTTTAAGCTTTAGAGGTGAG
GATACTCGCCATAGTTTCACCGATCATCTTCTCAAAGCACTGCGTGAAGCCACCATCGACACCTTCTTCG
ATGATGCAGAGATCCAGATCGGAGACTTTCTAAAACCGGAATTGGAGAAGGCAATCAAGTCATCTAGGGC
TTCTATTATCGTGCTGTCCAAGAATTATGCTTCTTCAACCTGGTGCCTGGATGAACTGGCACTCATCATG
GAGCAGAGAAGGACCTCCGGACATATTGTTTTTCCGATCTTCTATCATGTTAAGCCTTCAGATGTCAGGA
AGCAACGAAACAGTTTTGGAGATGCAATGGCAGAGCATAAACAGAGGATGGAAACAGAGTTAAATTCAGA
GAAAAGAAGTGAATGGGCTCACAAGATAGAGAAATGGAAGAAAGCCCTAACAGAAGTTGCTGATATGAAA
GGGGAGGAAGCCAATGACAGGTAATTTCATTGCTTCTGAGATGGAACCCTAGATTTTGAGAGTTGTTTTT
TTTTCCGAATCAAAGAGCCTGTGAAAATGCCTTTCCATGTTCAAATTGTGATTAGAAAATATCTAAATAC
TCATGAATTGTTTGATTATGCTATGATGAGTTTATATTAATCACTTTACTTGTAACTTCTTGTTGTGATT
GGAATCGGTTCCAGGCGAGAGACAATACTTATTGAAGGAATTGTTAAAGAGATAAGTAGCAGATTAGATC
TATACAAAAGGAGTGAAATCCCACATATAATTGGAATGGAGTCTTCGATAGGAATGATCACTTCATTCTT
GAAAGACAGATCCCAACAAAGCACTAAAATTCTCACAATATGGGGAATGGCTGGGATCGGAAAGACATAT
TTAGCAAACTATGTCTTTCAGTTGCACTATCTCGACTTTGAAAGAAGCTGCTTTCTTGAAGATATCGAAA
GGAGGTGTACACAGCCAAATGGATTGCTTGATTTACAAAAACAGCTCCTTAAAGATATTCAAGCAAGAAG
TTGGATGGACATCGACAATGTTAATGCGGGAACCTTGAAGATTGAGAAATCGCTTTTAAGAAAAAGAATT
CTTCTTGTTCTTGATGGAATAAATAAGTTTGAGCAGTTGGATGTGTTAATTGGGACAAAAGGTCTTCATC
CGGGAAGCAAAATCATAATAACATCTAAGAACGAATCGTTAACAGAAAAATGTAGGCTGTTTCAAACGCA
AGTTCCACCAAAGCATATGAAGCACTGCCTGCAAGGCTTAGATGATGAGGAATCTCTTCAGCTTTTAAGT
AGGCATGCCTTTGGATGCAATGAACCCAACAAAGGTGGCAATAAGCAATCAATAAAGGTTGCTAAATATT
GCCAAGGGCATCCATTGGCTCTTAAAGTTTTGGGCAGTTCTTTCCGTCGTGAAGACGCTACATGGGAGGA
TATCTTAGAATCACTGGGGAAAGAAATTAATCCAGATATTAAAAAGGTCCTGAAAATTAGTTATGATACT
TTGCCATCTGAAAAAGACAAGGAACTGTTTAAATACATTGCTTGTTTATTCGTTGGCGAAGATAGAAAGT
TCACTGAAGACATACTAAAAGCTTGTGATATAGGCAAATCATCTGGGATAAAGGTTCTTATCAACAGATG
TCTTCTCACTTTAGAATCATCTGATAAACTGACAATGCATCAACTTCTTCAAGACATGGGGAAAGATATA
GTTCATCAAGAATCACTGAAGAAGCCATGGAAGCGTAGTATATTATGGCACCATGAGGAGTGTTTGGATG
TGTTGCAAAATAGACAGGTAATCATGTATACCTTTTCTGTGTACTTTGATTAATTTGTATCATGTATAGA
CAACATTGTTAACTATTTGTTATTTCTTAGGGAACAACAACAACTCAAGGACTGGCCCTTGACATGAGAA
CATTTGAGAATGAAATATGGAAAGAACCAAGTAGTGCCAATATGCAAATGTTTGGATTTAGATCATATCC
ATCATTCAGTTGTCTTTGGTGGTTATTTGGTTGGGTTTCGGGAATGTGTTCCTCATCTCGTAAAACAAAA
GGTGACTTTGAGACACTTGCTTTAAGTGAAATGCGTAATTTAAAACTACTCCAACTGAACTATGTACAAC
TCAGTGGATCCTACAAGAATTTTCCTCATGGTATAAGATGGTTATGCATGCATGGATTCCCTTTGAGTTA
TATTCCTTCAGACTTAGAAATGGAGAATTTGGTAGCTCTTGACTTGTCTAATAGCAAACTACAACAGCTT
TGGAAGAAGCCCAAGGTACTAAAATATCAAGCTTCTAGTTATCTTGTTTAAGTATCTTTGCAACACTAAC
ATGCATCTTAATTTTGTTTATATGTTAATTTAACAGCTGCTGAGGTCACTGAAGTTTCTTAATCTGAGTA
GTTGTCATGAGCTTCTTAGAGTGGGTCACTTCTCTGGGCTCCCTCTACTTGAAAGGTTGACACTCACAAG
ATGTACAAGTCTACTTGAGGTCTGTGAGTCAATTGGTAATTACTGTCAAAAACTTGAAGTCATTGACCTT
AGTAAATGCAATAAGTTGAAGGAACTTCCAAGAAGCATACGAAAGTTAAAGAACCTTACTCATCTTCTAA
TAGATGGTTGTTCAAATCTTGGTGAATTTCCAGCTGACAGTGTTAACATGAAGTCTCATGGGTCTTCTTC
CTCTACCGTGGTACCAAGAACTCCCGAGTCCTTTGCGAATTCTTTACCAAGATCACTAGTAACGTTGTCT
CTTAAACATTGTAATTTGTATAATGAATCATTTCCTATGGATTTTAGTAACCTTCCAATGTTGAAGAATT
TGTTTTTGGATGGAAATCCAATGGATTCCATGCCCGATTGTGTGAGAAGCCTTAGTAGGCTGGAGACACT
CAGTTTTACTTGGTGTTGGAAGCTGAAAACAGTCTTGTGTGCTCCAATCCAATTAAAGCGGTTGGATGTA
TACTTTTGCAACTCATTGGAGAAAATAAAATTTCATCCAGAAAAGTCAGCTGTACCTAGAGTTTTTTATC
AGACAAGTTTTAACCTGACTGAGATTCAGCACGTATTCAAAATACAAGCTTTATCAGAAATTGATGAAGA
GGTATTATGCAGTTTGGGTTGGATTAATATAGCATATTTGAACCCCTGCTGGTTGTCGAAGGCCAATGAT
ATGGGATTGTACTATTTGGAACAAAGGATACTTCCAGCACAGGTGGATGTCTCTTTATATATCTTTATAT
TTCACCTTTTAATATGCATTTAAAAGCTAATAAAACTTTGATGTGACACTCAGATGCTTTATGAACATGG
AATATTCAGCACATATGTTCAAGGGCAAGAGGTTCCAAAGTGGTTCACACATAGAAGCAGTGGGTCATTA
TTTACCTTGCAATCATCTCCTGAAAAAGGTAAGATTAAAGGGCTAAATGTGTGCATTGTGCATACGATTT
CAGGCAAGGAAGTTGGTCCTTCAAGAATCAAAATAACGAATCTCACAAAGAAGTCCTCCTGGACATACCA
GCCTATGATGTATCTTTTTCCATATAATAGCGCGTTTGGATTTGGTGATAAAGCGGTAGTATGGTTAAGC
CATTGGATGTTTGGGAAAAATGAGTTTGAAGATGGTGATAAAGTTCGCATTGACTTTAGTGTAAAACATT
ATATGAGTTTGTCAGGTCTTAGGTATGGTGGTGAAGGACCAGAATATGCAAATGTAAGGGAGTATGGTAT
TAGCCCCGTGTATGATGATGATGATGGTGGTGGTGGTGGAAAGCAGAAAGAGGATCCATTGGGTTATTAC

AAGTCATGGAAGCACATCATTGGTGGGGATCTCTCTGCTTCTGAAGTCTCCCCAAGTCACTACCTTCTAC
AACCCTACCACATTTGGTTTTTGAAACACAGGATCCACCAAAGCATATGAAGCACTGGCTGCAAGGCTTA
GATGATGAGGAATCTCTTCAGCTTTTAAGTTGGCATGCCTTTGGATGCAATGAACCCAATGAAGGTAACA
AAAAGCAGTCAATAAAGGTCGTTGAATATTGTCAAGGGCATCCATTGGCTCTTAAAGTTTTGGGCTGTTC
TTTGCGTAGTGAAGACGCTACATGGGAGGATATCTTAGAATCACTAGGGAAAGAAATTAGTCCAGATTTT
AAAAAAGTCCTGAAAATAAGTTTTGATGCTTTGCCATCAGAAAAAGACAAGGAACTGTTTAAACACATTG
CTTGTTTATTT

>RGC52SW1_02_c2_65600_73700
GTCACTCAAACCTTCCTCTAATAATTCATAGTTATACATGGTTGTTCTCACTGAACATGAATCTTTACCT
GGATCCTCATCTTCATCATCAGCTCATGATCATAATCAGAGATATGATATATTTTTAAGTTTCAGAGGTG
CCGACACTCGTAATGGTTTCACTGATCACCTCTACAACACCCTTTTGGAAGCCGGTATTAACACGTTTTT
GGATGATGAAGAGATTGAAACTGGAGAGCCACTGAAACCAGAACTGGAGAGTGCGATTAAGTCATCCAGA
GCTTCTATTATTGTGCTGTCCAAGAATTATGCTTCCTCCACATGGTGCCTAGATGAATTGGTATTAATCC
TTGAGCAACAGAGGAGCTTTAACCAAATTGTTATCCCCATCTTCTATGATGTGGAGCCCACGGATGTCAG
GAAGCAACAAAGCAGCTTAGGGGAAGCAATGTCGAAGCATAAAGAGAGGATGGAGCTGGAAACCAATGCA
CAGAAAAGAAGTGAACGGGGTCGAAAGATGGAGTTATGGGAGAGGGCACTCACACAAGTTGCTGATTTAA
AAGGACAAGTTGCAAATGGCAGGTAAATCACTTACTTAAATTACTTAGATGTCAAGCTATTAGAATCTTT
TCATTACACAACAAAATTCTGTACATGCTAATTGATGAAGCATCATTTAGCATTGACATAAAGACAGACA
TATAAACATTCTGTACTTGGTACACGTATGACACGTTTTTAGACCATGGGGTATGGTTCCCCCTTCTCCC
ATGGCCACCTAGGCTCCACATCACCTTCTAGTCCACATCATCACTACTCCCCCATATGTTGGGTGTGGTT
CCCTTCATTCCCCTTTCCCCTTCCATCACAACATCACCTTCATTCTCTTCGGGACGATCGGGAAGGAGTT
ATCGAGTGCACCGATCACCGCTCCAGGGGGGGGGGGGTGTTCCCGATCGGGGAAGACCTCCACATCAGCC
TCCCGATCCTTTCTCCCGAACCACACCCCCACGTCTTATAACAATAGATTTAACGTGGATCCATATTTAA
ATTATTCTTCTTTTTTCTAATTTTAAAAGAGGTGGGGTTTCTGGGGTCTTATTTATTATGCATTTAATAA
TCAGGGTATTTTGATTCTTTCATGAAAACTAATAGAAAAATGTAACATAAAAGGTGTAAAAATTAAAATC
ATAATAGACTTAAATAGTTAGTGAAAACAGAGGGACTGACATGTACAAGTTGATTTAAGAGAGAGACGCA
AAGGATTTTTGGTTTCGATCTTCTATTATCAGGTTGTCTAGTTTTCACATTGCTATTTCATATTTCTCAT
TTAGTGATCAATTCTCAGGAAGGAGACAAAGCTGATTGAAGAAGTTGTTACGAACATCCATCGTAGATTA
GGTGGACCCTTAAGTAATACTCTACCATCCCTTACGGGGAGGAACCATGACATTGAATTCATTAATTCGT
GGTTGACAGATGGATCCTGTCATACTGTTGACATTCTTACAATTGTAGGCATCAGTGGGATTGGAAAAAC
ATCTTTGGCTAAGTATGTTTTCATGTTACCCTTTGGTAAATTCCAGAAAAGCAGCTTTATCGAAGGTATC
AATACAAGATGTGAAGAACGCCATAATGGATTGCTTGATTTACAAAAGCAACTTCACGGAGACATTTCAA
AAAAAACTCGACTTACATGTTAATGATGTCAACGAGTACACCTCTAAAATTGAGGATGTGTTATCTCGTA
AAAAGGTGTTTATAGTGTTTGATGACATTGGTAGTCTTGACCAGCTGAACGCTTTACATGGAAAAAAAGG
TTTACATCCAGGAAGCAAAATCATTATTACAACCAAGGATTCATCTTTAACAGAAAGGTGTGACTTATTC
AAGACACAAGTTCATCCCAAGCATACTGAGGTCTTACTTGATGGGTTAGATGAGTATGAATCACTAAAGT
TATTATGCATTCATGCATTCCAGAGCCATACGCCCAAAGAAGGTTATAAAGATGTCTCAATAGAGCTTGC
GAAATATTGTGATGGACATCCATTGGCTCTAGAAGTTCTGGTGGGTCTCTACATAAGCGAGATGTAGCAT
ATTGGGAAGAGCGCATAAAGGGGCTAAAGAAAGAACCTCTTTCTGGTATAGAAAAGGCAAAAAGGGCATT
ACAATTAAGCTTTGACACTTTGTCATCCGAAAATGATAAGGAGCTGTTTAAGCATATTGCTTGTTTCTTT
GTGGGAGAAGATAGGGATCTGACTGAAACAATATTAAAAACATGTGCTATAAACACAACATCTGGGATTA
CAAATCTCTGTTGCCCAGAATCTTGATAACGAACAATGAATATCACAATGAACGAATAAACGTAAACTAA
CCTTTATTTCTGAATAAATACAAATCGTGAATATGTACAAGTGAATGAATGAGAGTAGCAATTACAAATG
AAATCAATCTATCCTATTTATAGTTTCCTACGGGTACGAGTGAATAGACGTGTCTCTTCGTGAAAAGTCA
TGTCTCTTGGATGTGTGGTTGTGATTGGAACTTGAAAAGGTGTGTTGATGTCGTCCGCATGATCAACAGT
CGCGTCCTTTCCTTGTTGCCTTGATACCGGTTCGGAAATGGTGTGTGTAGGAGACACACCCATTCGGTTA
TTGAGGGAAGTTTCAAACTTCCATCTTGTCAACTTTGGTCCTTCAACTTTGTAACCGCCACTTAATAATA
AATCTAAACTATCTATCTACTAATTACATAATGAACCCTAATACTTATATAGGATGTGAAGAGATTATGT
TTTCAATCTCATGGAAAAATGCTTTCTTCGTATTAGATGGAATAATAAATTGTTGATGCATTCATTGATT
CAAGAGATGGGAAGAGATTTAGTACGTTAAGAATCACTTAACAAGCCATGGGAGAGGAGTCGATTATGGT
GTCACGAGGAGTCATTCGATGTGTTGAAACGAAGAAAAGTAAGAGGCCATTTATATTAAAGTAGTAATGC
ATGTTCTGAGTATTATTATTGTCTATATACATGTATATAACCTTTATATATTATTTTGTTTGTATTCCTT
TTCTTCGTAGGGAACCCAAAATACTCTAGGCCTTGCCCTTGACATGAGGATGCTTGACAAAAAGAAGTTG
CGTGGATCATTTGAGCTGAAAACAGAGTCATTTAGTCAGATGAATAATCTAAAGCTACTACAACTCCATT
ATGTGCAATTAAATGAGTGTTTCCGCAATTTTCCAGAGGAGAGGCGTTAAGATGGTTGTCTATGCATGGG
TTCCCTTCAAAGTCTATTCATTTAGACCTACCAATGGAGAATTTGGTTGTTCTTGACATGTCTTGCAGCA
ATATTGAATCTTTTGACATGTCATACAGCGATCCACAACCACCTGCGAAGAAGCAAAAGGTAAGATTTTA
GATATTTGTATATGAGATTCCGAATAATCAATAATCCGCTTGGATAAACTTTGAATGGTGATGATAAAGC
TATTACAATGTAGTATTTTAACCCAAGAATTTCCTAGGCTTACATGTAATCACAATCAGATGATTCAGAT
TCACACTCAGGTCTTAGGTATAAAACCAAACACCGACACTAGGATCCTATGGTCTCACGTAAATTTCAAT

AATTTAACTTGACTGATGACACTAGAATCGCCACACAAGATAACGCTTATTATTCTTTCATATTGAGTAA
ATGCATGTTTTACTTATTTTAACTTATCTCTGTTTCACCAGTTGGTTGGATCATGCTCAAAACATGAACC
ATTGCTTGGATCATTGAAGATTCTAAATCTGAGTTTCTGTGAACATGTCAACAGCTTCATAGTGTTGGTG
GTTTTTTTGAACTACCTGCACTTGAGAAGTTAATAGTTAGAAACTGCATAAGTTTGATAGAGGTGTGTGA
ATCAGTCGAGCAATGTGTTGAACTTGTCCACATTGATCTTGAAGATCAAAGAATAATCAAGTCCATCCGA
ATGAATCCGGTTTTGGGCCGGTAGTTATTTATGTTTATTAGTTTTTGCTTAATGGGTCAAAACAGTTTAC
TGTTTTATTATTTGTGTTTAGAACTTGGGCCGGAGGCCACGTGATGTATATGGGTCAAACAATTAGTAGT
CCAGTAGGTTTAAAGGTTTGGGTCATATGAGGTTTGGGCCATGTTACATGAACCGGCGTGAGTGAGAAGG
TATAACCGCTTATACGGTTATTGAGGGAAGAGACGTACCCCTCTGTTGATCGTCGCACCCTTTCACTCAC
ACCTATTCGTTCGGTATAAATACTTGGCTACATCCTCATTCTATGATATATCTGTAATGAAAGTTTATGT
TTGTTAGAACTCAAGTTCGTTCAGATTTGTTTACGGTTAGTGTTATTATCATTCAGCAAGTTCTGTTAGT
TTGATTCATGATGAAGTATTGTAAAGGTTGTGGTTTGTTCGAAAGTGATTGTGTATGTGGTCACAACTCA
ATCCATTCCGACTGGTCCTGGGTTTCCGATGACGATGATACCAACATTGTTATCAGAGCCAAAGGTTTAA
CAGGAACAGGAAAGGGTTGTGACAATGATGGAAACCATTTACGAGGAGACAAGGGAAAATCGGTTGAGAC
CGACTACGGTGTAGATGTTGTTCGTAGTGGAAAGTGCGGCCGAGCGGCGGAAGAGATGAATGAGTACGCG
TAGGGGAAGCCAACCATACGAACCGGGAAGTCACCGGTGAAGAAGTTTCCAAAGAAAATGGTTCCAAAAG
GGTTTGTGAAGAAGTCCGGGATGAAGGCAAGTGCACCGGTGGGAAGACCAAGGAAACCGGGAATCCGTTG
CTTCAAGTGCAAGCAATTTGGTCATATCGCCTCGATTTGCCCAAGTAAGTATGTTAATTTATCGCCATTA
CCGGTTGAAAGTGATTATATGGTTAAGGATACATGTGGCAGTAAATGGGATTCGATTTGGGTTGTTGATC
CTACATACAAAACACACATGACGGGAAACTGGAATGTGTTCAAATGTTTCAAAAGGCATTTTGGGTTAGT
GACAAACGAGAATAGGAAGGATTTCTCGTTTGTGCATGGTATTGGAGAAGTTAGAGTCCCGGTGGATGGT
AAAGACAAAACAATCCCATGTGTAAGTTACGCTCCTAGTCTAGACAAGAACGTGTTAAGTCTAGAACAAC
TGTTGTTTCAAGGAATTGAAACAATTACAATGGGGGATACATGTTTACTAAAGAAGATGTTTGGAAGTCG
CTCGAAAGGGTTTGATATTTATGAAGATAAGTCGGAGGTCGACTTAGAACAAGATTATCTCAACACGTTT
TATGATAATCTTGGTGTTGATAACGGGTACAAGAAAGAAAAGAAGGAATTTAAAGAATGTTTGGAAGACT
ACTACGAAAGGGAATTCGAAAAGGAAAGGAACAAGAAAAAGGCGTACGGTGAAAGTTCGAGGGCAAGAAA
GAAAGATATCGTTTACTCCAAGAAAGCAAAGAAGGCGTTTTTAATATACATCGAAGGTGAAAAGGATAAT
CCGTATGAGTCAAAAGAAACGCTCCGGGAACGAGCGTTAATTCTTTCACAACTCGAGGAAGACGTTATCG
AAAAGAACATGATCATTGAAAGTTGTGTAGAACCGAAAGATCTCATAACTTTGCACGAAGAATTAAAGAA
TCATCAAAGGTTCTTCGATGCTCCATTCAAAGAGATTTTGATTTGGTTCATTTCAGATTTTCTTGGAATC
GTATGTAAAAAGGCAATGCCACCCGAGTTGATTGATGGCCGTGAACTTAGTCTAATCTTGTTACATCGAA
TTGTAAAGCTTAACGGGGGATTCAAGGAAGTAATGGAAAAGAATTTATGGGATGTGATTGCGGCTAAGTA
TGGTTATGAGCCGGATGATGCATACGAAGTAAAGGTTGCCAACATTTACTACTTGGAACTAGTCGAATGG
TATTTCGACTTCATGAGGAAAGAGCGTGGGAAAAAGGCAAACATCATGGCTGGAAATTCTAACAACAACG
GTTGGCTCGAAGATTCAAGTGATGACGAAGTGGTGGTCAAGATAGAGGTCACCAACAACTGCAAGGAGTA
ATCAAGGCTGAAGATTCGGATGTTCTTGCTTAGGGCGGTGAATGAAGATCAAAGAATAATCAAGTCCATC
CGAATGAATCCGGTTTCGGGCCGGTAGTTATTTATGTTTATTAGTTTTTGCTTAATGGGTCAAAATAGTT
TACTGTTTTATTATTTGTGTTTAGAACTTGGGCCGGAGGCCACGTGATGTATATGGGTCAAACAATTAGT
AGTCCAGTAGGTTTAAAGGTTTGGGACATATGAGGTTTGGGCCATGTTACATGAACCGGCGTGAGTGAGA
AGGTATAACCGCTTATACGGTTATTGAGGGAAGAGACGTACCCCTCTATTAATCGTCGCACCCTTTCACT
CACACCCATTCGTTCGGTATAAATACTTGGCTCCATCCTCATTCTATAATATATCTGTAATGCAAGTTTA
TGTTTGTTAGAAGTTCAAGTTTGTTCAGATTTGTTTACGCTTAGTGTTATTATCATTCAGCAAGTTCTGT
TAGTTTGATTCATGATGAAGTATTGTAAAGGTTGTGGTTTGTTCGAAAGTGATTGTGTATGTGGTCACAA
CTCAATCCGTTCTGACTGGTCCTGGGTTTCCGATGACGATGATACCAACAGATCTGAGTTACTGCTACAA
GCTTAAAAAGCTTCCAATATCCCTAGGCAAGTTACAAAAGGTTCAAACACTATTGCTAAATGGTTGTAAT
TTATCTGAATCTGCCATTATGGATTCATCATATCTTAGCATAAGCTCACAAACCACTTCATCAGCCATTA
TAAGGGAGGCTATACCAAGTGATTACAAGTTCTTTATGACTTTTTTACCAAGCTGGTTGAGAATTTTGTC
ACTTGTAGATAATAATTTGTCAAACGAATCCTTTCCCATGGATTTGAGTTGTCTAGCCATGTTAAAAGAA
TTATATCTAGATGGAAATCCAATTGTTTCACTACCGAATTGTGTGGGAACCCTTCCAAGGATTGAGAAAC
TTAGCATTAACCGCTGTAATGATGTGATATCAGTTGAGCATCTTCCACGTACCCTAAGAGTATTCTCCAT
GTACACAAATATTCTTCGATCTGCAACTAGGAAAATTAAATTTGATCCAGAACTTCTTCCACTCAAATTA
GGAGGAATTCCCTTGTTGGCATCACATTTGCCGGTTGAAATTGAAGGCATAATCAAAATTCAACCAATGG
CAGATATTGAGGAAAAGTTATTACCTAGTTTGGGCTGGAAAAGGTCAGACTTTATTAAAGAAGGACACTT
GTTAACTCATGACATCCAGGTATCCTCTATTTAATTGTCTCTTTTGTGTGTACAGGTGTATTTTATGAAA
GGTGATTAATGTTGTATTTATTTTGCAGATGCATTATGAATTTGGAATATTCAACACCATTTATTGGTGC
AAAGGGATGCCGGAGTGGATTAGGTGTAGAAGTTGGGGCTCATCAATATCATTTATCATCCCATCATCTA
CTAAGAAGCTCCGAGGATTGAATTTCTGTTGTGTGAATTGGCCTCCATTTACAAGTGATTCTTTTTCCAT
GCCGACGATCAAAACTAGTAATATTACAAAAAAGCAAACCAGGATTTATAATCATTATAGTGGACACGCC
TGGGTACGTGGAAAGGAGTGCTTTAGTTGGTTAAGCCATTGGATGTTTGGGCCAAATGAGATGAAAGCTA
GTGATAACATTACAATAGAGTGTGACCGTAAAATGGAGTGTGGGGTTGGTGTTGTGTATGATGATGGTAG
TATGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGATGTGTTGAGTTATTACAAGTCA

TGGAATCACATCATTGGTGGAGACCTCTCTGCTTTTCAACTTACAACAGGAGAATATTACCTAGGTAACT
GGAAATTTACACAGGACTCCCATGAAGAAATTTATGCTCCTTTCATTGGTGAAGAAGGCAGCAATAAGGG
TAAATGTCGTCTCTCTATATAACATTTAGCCAAGTCTTGTGTATTACTATG

>RGC52SW23_07_c1_27500_51500
TTTCACAGCTTATCTACTCATCTCTTCACTCAACCCTTCCAATAGTTTATATTCACAATGGTTGTTCTGA
CTGAACATGAATCTTTACCTGGATCCTCATCTTCATCAGCAGCTCATGATCATAATCATAGATATGATGT
ATTTTAAGTTTCAGAGGTGCCGACACTCGTAATGGTTTCACCGATCACCTCTACAACACCCTTTTGGATG
CCCGTATTAACACGTTTTTGGATGATGAAGAGATTGAAACTGGAGAGCCACTGAAACCAGAACTGGAGAG
TGCGATTAAGTCATCCAGAGCTTCTATTATTGTGCTGTCCCAGAATTATGCTTCCTCCACATGGTGCCTA
GATGAATTGGTATTAATCCTTGAGCAACAGAGGAGCTTTAACCAAATTGTTATCCCCATTTTCTATCATG
TGGAGCCCACGGATGTCAGGAAGCAACAAAGCAGCTTCGGGGAAGCAATGTGGAAGCATAAAGAGAGGAT
GGAGGCAGAAACCAATGCAGAGAAAAGAAGTGAATGGGGTCGAAAGATGGAGTTATGGGGGAGGGCACTC
ACACAAGTTGCTGATTTAAAAGGAAAAGATGCAAAGAACAGGTAAATCACTTAGTTAGAGGTGTTTGGTG
TTGCGTTTTCAAAATAGATTATCCGTTTTTAAAACTGCGTTTTGAAAAAGCATGTAAGTACATGCTTCTC
CAAAAAGGCAGATAATCAATTTTTCATCCAAACATTTTGATAGGTTTACGGAGAAAACTTACAAACTCCC
AAGAACACAAGAGAAGAAGAGAGACATAAACTAAAGTTTCTTTCAATAATTTTCTCATTATAGTTGCTAA
AAACACAAAGCCTTATATAGGCATACAATGCAAGTATGGAAATAAACTAACCACTAAATATGCTATCCAC
TAACTACTGAAAATATGCAAACAATAGAAATAATAAAGGGAAACACAAAGTGACGTGCAAACTAAATGAT
GGTGCATGAGAACGTGTGATGGTTGTTGTGCTTCATGGGTTTGTTTACGTGGGTTACGTGATCGTATCAA
TGCCACGCCCTCGAAAACGTCCTTGTCCTCAAGGACGGTATGAAATCCCATGGGGGTCGCCATTCGAGTC
GTACCACAGCATTAGGATCGGTCTCAACATAGAGGCACGATGGGAACCATGGTGGTCTCCATTCCCTTCT
GTCGATGGCGTTTGGGTCTGCATCACTGAACTTGGAGCACCGGTTTTCGAATGAGACATGGCCATCGAGC
AAGACGAATGTATACAACACCGGTATGGAGGCCTTGTTGGTGAGAATGGCAGCTTGAGGGTAGATGGAAT
CTAATGGTGACTGTTGTGGCGATGATGTAGGAATCGGTGGTGAAGCGACCCTTCTAAGAGACGATGTCGG
TGGTATTCGCGGTTCTTGGCATGGAAGTGACGGTGGTGGTACGTGGCTAAGCGGTGGTGGTCGAAAACGA
TGAAGCATTGCAATAGTGAAAGACTTCAAGGTAAGGTGTTCCGGTCCTCGATACCATGAATTGAACCAAG
AAAAAGGTTCATCATCGAACGAGTCAACAACATAAGGAAACCGTTCTTCCTCATAAATTGCATAAAAGGA
AAAGAACCGTTCAGCATCACAAATCCACCGGGCAGGATTGTAACCGGTGAAGCGAGATAGATACGTGGCG
TCCTCGAGTTCTTCCAAAGCCATGGAGGCTAAGGCTCAATGAAAGCACCAAATGATAGGTTTATGGAGAA
AACCTACAAACTCCCAAAAACACAAGAGAAGAAGAGAGACATAAACTAAAGTTTCTTTCAATAATTTTCT
CATTATAGTTGCTAAAAACACAAAGCCTTATGTAGGCATACAATGCAAGTATGGAAATAAACTAACCACT
AAATATGCTATCCACTAACTACTGAAAATATGCAAACAATAGAAATAATAAATGGAAACACAAAGTGACG
TGCAAACTAACGGTGTGTTCCCGAGTTAAAAAACCCTCCCCTTCCCTCCCCTCCCCTCCCCTCCCCTCCA
TGTCTTTCCAAATTGGAAAGACTATTTTCAGGCAAAAAAAACCCCATTTTCCCTCCCCTCCCCCTGTTAA
GTGGATCTCTGGAACACCATTTTCCCTCCCCTCCCCTCCCCTCCCCCTGTTAAGTAGATCTCTGGAACAC
AGTGTAAATGATGGTGCATGAGAACGTGTGATGGTTGTTGGGCTTCATGGGTTTGTTTACGTGGGTTACG
TGATCGTATCACATTTTTTAGATTATTTATGTTTTACAAACGCAATAATCAAATAATCACTTCAAAACGT
AATCCCAAACACCCTCTTAAATTACTTAGATGTCAAGCTATTAGAATCTTTTCATTACACAACAAAATTT
TGTACATACATGATAATTGATGAAGCGTCATTTAGCTGAAGTGAAAATAAACAACTCTTAAATATGAAGA
CTAAAAGTGTACTGCAAAGAAAAGAAAAGGTAGGTTACTTTATTCATTCCATAGCTAGCTGGCTATACAA
AGGTAGTTTACAAGAGAACGTGGGAATGCAGAAAATAAGGAATAATACAAATACTACCTAAGGCCACTCG
GTGTAGGTTATAAATGCTTGATTGTAAATAATAAACACATCAAATACAATGGCTTTAAATAGCCACTACA
GGTCTTGAATTCTTGATACAATACATGAATGGAAAGGAAAAATACTAAGGCTATGGTGAAGACTAAATAT
GTGGAGAAGAATAATAATAAAATGCTATCTAAAAATAACTCTAGCATTTAAACAGAAACTACAATCATGA
GAATCAAATATTATTTTCTTTGATATGCCCCTGCAAGATGGATGATCCATCGGAGACTCCAATCTTGAAC
CGTAACCGTTGGAAGGGACCCCGACCCAGTGGTTTGGTAAGCATGTCAGCAAGTTGGTCGGCCGAGTTGA
CATGAAGGACCTTTCGTGAACCATTAGTGACTTTCTCTCGAACAAAGTGGTAATCGAGTGCAACATGTTT
CATACACGAGTGGTAAACTGGATTTGCACATACATAAGTAGCACCGGCATTGTCAAAAAGTAATGTGGGG
GTTTTAGCCATCTTCAAACCAAGTTCCAATAAGAGGTTCAGGACCCAAGAAAGTTCAGCGACACCATTTG
CTAATGCCTTGTACTCCGCCTCCGTGGAAGATCTGGATACCGATTTTTGACGCGATGAGCACCAGGAGAT
AATGTTAGAACCAATATATATGATATATGCAGTAGTTGAACGGCCCCATCGGTAATGCCACCCCAATCGG
AGTCGGTAAAAGCAGTGAGAATGATGGGAGAACCACGGTTTGGAAATAAGCCGTGGTGGATTGTATCTTT
GAGGTAACGAAGGACTCATTTTAAGGCTTACCAATAGAGTTGAGTGGGGTTGTGCATGTATTGAGAGAGT
TTGTTAAGGCAAATGAGATATTGGGTCGTGTAAAGGCAAGGTATTGAAGAGAACCCACTAATTGACGATA
AGGTGTAGGATCAACATGAGGCGAACCATCAACAAGAGAGAGTGGTTCAGTGGAGTTGAGAGGGGTGGTA
ACATGTTTAGCACCATCCATTTTTCAACTTGGTGAGAATATTTTGGATATGAGCATGCTGGGATAGAAAT
AAACCATTAGTAGTGGATATAACTTCTATACCAAGAAAATGATGTAGCATGCCAAGAAAGATCTTTGATA
GAAATTTTTTTGGATAATGAGTGAACGATGTGTTGGACAAAGATATCACTACTACCGGTTAGACAATGTC
ATAAACATATACCAATAAATATGAGATGATCGCATCATGGCTATAGATGAATAGAGATGGACCCGTTATG

GATTTGCGAAAACCAAGTTCAAGAAGGAAGTTTGTGAGATCCATGTACCAAGCACGTGGTGCTTGCTTGA
GACTATAAAGGGATTTCCTCAAGTGACAAAGATGATTAGGTGAATCCGGATTTTCATAACCCGGGGGTTG
GATTATGTAAACATCTTCATGCAAAGTGCCATGTAAGACGGCGTTGGTGACATCGAGTTGCCGAAGTGGC
CAATTTTGAGATAACGCAATGGCAAGAACCATACGGATTGTGATGAGTTTGGTAATCGGACTAAATGTCC
CAAAGTAATCCTTCCCATATTCTTGGAGAAAGCCTTTAGCGAAGAGACGCGCCTTGTATTTCTCAATAGA
CCCATCTGGACGTCGTTTGATACGAAATACCCATTTGCAACCGATAGGTTTGTGATGTGAGGGAGGGACT
AGTTCCCACGTGGCATTCGTTATTAAAACATTAAACTTAATATCCATGGCTTTACGCCAGTCCGGGTCTT
TTAAGCTTGATTATGAGTGGATGGTTCGAGTGTGGGAGGAATTGGATGAGAAGTGGTTTGATTAACGAAG
GCTGAGTTGTAATATTTTGGGTTTGGTTTGCGGGTGCGGGTAGGTGGTGGAGGAGGTGTAGGAATCGATG
AGGTGGTGGTTGATTCGGTGGAAGAGTCAGAGGAGGTTGAGGAGGAAGTGGGCTCGGTAGAAGTAGGTGT
GGTGCGTTGTGTAGAGATAAACGATTGGGCCTGGTGGTGGGAGGATGGGACATGTTGGAATGGAGGAATA
AGCTGGTTGGGCGAGATGGGAATGCAGGCGATTGGGTTGCCACAGTTGGGCCTGTGGGTTCGGTGGAAGG
CTGGTTGGGTTGAAGGCCCGTAGAAGTGGCAGAAAACTGTTGGGTCGGCGGCGATGAAGTGGTGTTTGAT
AGGGTTGTAGACGGGCTGGATGATGGTGTGGAATTGAGTGATGGAAGAGCAGTAGATTTTTGGGGAAAGG
GATGATATTAGGCTTGCGAGGGACTTAGAAATTCATCAACACTGGGGTTTGCTAAGGATGGTGAGAGATT
TTGTTGGAAATTCAGACTCTATGAATTCCACATAACGAGAGTGATATAATCGTCGGTTATTATAATCATA
GCATTTGTAAGCCGATTTTGATGGCGAGTAGCCTAAAAATACACATCTTTTCGAACGTGGATTGTCACAG
CCCCCGTCCCGTACCCCGGTTCTGGCGCCGCGAGAGAGATAACAGTGGTACCGGTGTTTATTAATTAGCA
GCAGAAATTTCATCAGGCCCGGATGTCAGGAAAATAATATCAGAGTTTCAAAACACCAAATTATTTTATT
TTAGGGTTATTAGATTCTATCACCCTCAACTATTGTCTATTGGCCGCTGCCACCCCCAACTATCACTTTG
ACGCCCATCACCCCCAACTTAACACTTGGTGTGTTCTGTCACCACGTCGTTAACCGATCTCTAACTTTTG
AAGCTGTTACTACACTTTTTGGGGTGTCATAGGGATCTTGGGAAGGTTTTAGGGATCTTACAACACCCCC
AAATGTATAATAACAGGATCAAAAGTTAGTGATCTGGTGACAGAACACACTAAGTGTTAAGTTGAGGGTG
ACGGGAGTCAAAGTGATAGTTGGGGGTGGCAGCGGCCAATGACCAATAGTTAAGGGTGATAAAATCCAAT
AACCCTTTATTTTATAATCGTGGGGTAAACCCAGTATTTACAGAAAATGAATTTCACTGGGGAAATCCTT
TAATTATAGAAAACATGTTTTATTTCTAATTTACTGAGCCACTTTATTTAAGCTGAAGTGCTTCACGACA
CTTTTCCTTTCTTCACAGTAAATCACTTGAAACGTGTTTTAGAAAGATTTGTGTCACACCCCGACCACGT
GGAACAACAAAACGTGGCGGAAACGTCGGGGAGTGTTGTAACAGAATCATTGTTTCATAACACATGAAAA
TTGAAATATTGTTTTATTGATTAAATGAACTCATTGTTCTATTACAATCACATAAGTACTACTATGATCT
TCCAAGTTTTAAAGTTGCTAAGGCACGAGTCCATCCTAAATGTAGCATGTATCAACAATCATCTCAAGTC
AACACCTGAAACATGTGTAAAAATAGGTACGTCAGCATAAAAATGCCTGTGAGATACATAGGTTTTGTGA
AAATGGGATTCATGACTTGAGTTTAAAGAAAACGTTTAATAACAGTTAGTCGGGGGGGGGGGGTTATTGC
ACGGACCTCCCGGCCGCTGGAGTGATCCCACTCCACAAGCTAGCAACGTCCCAATGCTGCCTGGCAGTGA
ATACCCCATCCCGAGCTGAGTCCGTTTCCCTCTGGGAAGAAGGTGGCAGAATACCCCCTGCCTGGACTCG
AACCCGGGTCTCTGTGAAGAGTAGAGGTGGGACTGGCCAACTGGGACACCCCAGTTGGTTTTAATAACAG
TTAGTCATGAACCTTGTAATTTGCTTAGTCTTGTAAATCATTTGAAAAAACGATAAGATTGGATGATATG
TATAAATAAGAGAACACTGTATTGTTAAATGGATAACCATGTAAAAAATGTGTTTGTATAAGATAAGTCG
TTTGTAAAACAATGTCTCGTGAAAAGTGTGTTATTTGTATAAAATGCCATATGTCTCAAATTGAAATGAT
TCAAATAACGCTACGATATGTAATACCATACAAACACTTATATATAGGAAGTACCAGCGGCGTATCCACC
ATGCTTGTATCATATTACACACGCCTCGTTACTTAAGTCACTTACGCAAACCAAACCACCAATATAGAAT
GTTCATGTTTAAACCAATAGTCCAAGGTTATTGTATAACTCAAGTTAAATGTAAACCACGTAATGTATAA
ACAAAAGTTATGTATGGTAAGTGAAATGATGTAAGGTTAATCGAATAACCTTATAGCAGCATATTAACAG
TATACTTTGGATTTGTAAATAACATATTAAAAATATGTTTTGGTTTGTTGATAAAATGCTTATGTTGGTT
CTGTACAATACCACACATGAGAGTATATCAAACTATACTTTTTGGGAGTATATCAAAATATACTTTAAAG
GTGCGATTTAAGAATTCATTCAGACCTAGTGCATCAAACTACACCTTTGAGACTATAAGGATACCACAAA
GGAAATCCCTATTTGGATGGAGAAACAAATTGGTTTCAGACAATCCATGACAACAGGAATAAGGTGTCTA
GTCCTATAGCGCTATATCATACTACCAATCGGTCTGGTGAACAAAATAAGTACTATTCCAACGATTACTT
TTGTAATCTTTGAGAAATCAGTGTTTAAACCATAGATAGTCATTTAGTTTGTCAAAAGATAAATACAAAC
ATGTATAATCAATTATACTTTGAAATCATGAAAATAACAGATAGGTACACATGCTTCACCCCAAAACAGT
TTGGAAAACAGTAAAGAGGGGTTCAATGTACTCACCTGAGATTGCTTTGGAGTTCTTGTATAATAACCAG
ATAAAGCTAAAGATCACGGGATATCAAACGGCACCTAATAGGTAGCTATGTTAATATACCGGACCAAAAT
GGAAGGATCGGATAGTACGCGGGTTCGTAAACCAAACGAGTATGGAGACTCGTGTAATATGGTTTAACAA
AGCCTACATACTAAAATGAAACTTAACCAAAGTGCTTACGGTCCATCACGACCCGTTTAGGTAGCTTATG
CTACCTTACGCGTCGTTCGCGTAGAACGCGTCTGGAACGCCTAACATCGTGACCACAAGGTATAACCTCG
GGAGGTTATAGCTATGGTCACCTAATGTGTTTGGTCAGATCCTAATGATCGACTGAATGGGTCGGGTTCG
AAAGTATAAGCGATGGTTTAGATCGCTTACCTTACGACCCTATATAAGCACTAAACTAAAAGTGACGAGC
TAAGCATGTTAGAACATGCTTAACTAAGTTTAGAAAACAGGTTTGGCATCAAAACAAACGGCTTTGATGC
TCACGAGTAGTTTGGTTACAAAATACGCAAGAATGCGCATTTTGGCCGAAACTACGACTCGTCACTGAGC
CTAGATAACGTGGTGACCAGTAGGTGTAGTCACTACGGACTATAACCATCGTGATCACGCTCACGTTATG
AAGTTCCATGAACTTCGCATCGACCATAAGCTGGTCAATGCAGAAAGTGAACAAAACGTTGACTTTCGGA
CTCGAAAAGCGATAAAACAACGAAAGAATACTTACGAAGGGTCCCCGAATGCTAATCTAGATCAGAGAAA

CTCAGGTATGAAGAATACTTCAACTTAGAGCTTTTAGAGCTGATTCTTGTGATTTTTACACAAAAGGGGG
GGGGGGTATTTATAGGAAAAGTGGAACCGTTAGGATCGTTTATCGAATATCGTGCCGCGATCTCGTGCGT
ACACTTGTCAAAAATGTGGTGGTTACTGAAAATGGCCTTTGGCCTTTGATTGGGTGAAAGGGCATTGCCC
CTTGATCGAACGAAAGGCCAACTGCATTGATTAGAAACATTGAATCTGGTCTGACAGTCTGACGCGGCCC
GCCTCAACATTTAGGCAAAACTCACGCGGGCCGCCTGAAGCTTCTGATCTGCACATACTTTCAAAATTGG
CAGTTTTGGTCCCTGTTGCGTATTTAAGCCATTTCCGACACTTCTAAGGCCCGTAAAGCCAACTTTAAGG
CCCTAAAATGATGCCTAAACATTGTGGACATGAAACATGCTCAAAAATGTTCCGGATGTTGGTTCGTTTG
GTCGTACGAACGCGATGTTCGCTTAATTACGACGGAATGCGCATAAGCGCGAAAAACGATCCAAATGACG
CGACGAATGGATTTTCTCATGCTAAACACTAAGGCATAATATAAGGATGCTTACATAAATTTTTGGATGT
CCGGATGTATTCAGAACGTAAGTTATGCGCGAAAGTGCAAACTTGTGCAGTTTTTGACACTTTTAGTCCC
TGAATGATCCAAAAGTTTGTTTTAGCATACCAAACACCTCAAAGCCTATTTCTAAGCTATGTAAAGGATA
TTTATGGTATGTTTAACTTATGGACATGTTCCGGAATGTTTGTTACAGTTCAAATTGGCATACTTTCGCA
GTTTGTCAACTTTAGTCCATGTAAGCGAATTAACTTGTTTTTGCCATACCAAAGCCTTCGAAACTTATTT
CTAAGTTATGTAAAGGTTATTTAAGGTATGTTAAGTGTATGATGGTCTTCCGGAGTATTTGTCGCACTAA
ACTGAGTACGTTTACGCACTAGTTTGCGTATAATTCCCCAGAAAGTGTTGTAGAGTTTGAAATTAAACAA
GAATTGATATGTGCAAAACATACACATATTTATACAAATCCCAAATATGAAATAAAACATTTTATTGGTT
TGGCATTCGTTTGATGGTTGGAGTGACACAGGTGTCACAGTCTCCCCTATTTTAGGAAATTTCGTCCCGA
AATTAAATTCTAGAGGAAACTTGTGAAGACAGCTGTGACGGTTTAACCTTCTAATAGATCCCTTGAACCC
TTAAGTAAAACTCGGGGCCATGTTAGGACTCTTAGCAAATTTGTTAACCCTCAAGCGAATTCCTAGTAAT
AATAAAACATGGAATTTTGAACTACGAATAGAAGAATGTTTCTTATGAAGACTTATCATAGGAAACTTGG
TGTGTGATTGAGATCAGAATTGGGGGAGCGTGAGAACAAATGACATTGGTCGAGGTCCGAGACTTCTCCC
GTATCATTATTCTTGTCACTAGGTCTTAACGCAAGATGAAGCTTCCTGTGGCTCCTGTGCACTGAATTCC
ATAATTATGTAGGCCTCCATAATTACGTAATTCCTTGCACAGTCCGCACAGTCCATTTCATATTGTGTAG
GAAAAATTACCTAAATAAACATGAAGCGATTCGACTAGACCACCAAAGGATCCTTTTAATTAGATCAACG
AACGATCTTTTTAACTAGATCAACACGTGATCTTCTTAACTAGACCGTCGCACGATCTCTTTAAATAGAC
CACTTAAGGGATCATTTTTAGTTATATCATCACATGATATTTCTAATCAGATTTTCAAAATCAATCTGTT
TAAGCAGACCATTCAAGAGGTCCAATTGTAGAATAATACTTTATAACCAGAGGGACTTAACTACAATGTT
GAAATCCATTAAGGGTTTAAGATAGTATCTTTGGATATCCTATTACGAATCTCTCGTATGAGATATGAAA
GAGTCTATGAGGATTTGGATGGATCACATACAACCACAAAACATGTAAGAAAACATATAAGCACAAAATT
AGTTAACTTGATTCTTTATTTTGTTATCCTTAGGTATGGATATTAAAGCACACTAAGAATCCAATCTATG
GATTTCATTTGGCACTTTAATGATTTATAAAGATCTATCTATAAAGATAGAGGGTTTCTCAATAAGAGTT
TAACTTTGATTCTTAAAAGAAACGAGAGTTTAAGGTTCTTGGGGGGATCACAAGTGATCATAATATGACA
GAACCATACGAGTAGTTTGTCTTTGGGTTGACATCATAAGGTTTCATCAAGCATTCATACAAACGCATAA
ATTTTGCAGAAATAAAACAATAAATCTTTATTTATGTAACAACGGAATTGTCTTGAAGTGTCAAAAGATA
ACAACCAAGGAAACACAAAACACTAATTGTTCCCTGCTGCATCATTGTTCATATTTGCCTCATTGATGTT
GAACACTCGTCCACGCGCTGGAGGTATGTTAACAAGCCTTGGGCAATTGTTCCTGTAGTGGGTAAGGTCA
CCACAGTTATAACATGACCCTGGAGGGAACCGTGCTTGAACAGCAGCAGGGTTTGCAGCAGCTTGATTTG
CATAACGACAAACATTGATCAAATGTCCCAACCGTCCACAGTGGGTACATTTGTGACACTGCTCTTGCTC
TAGATGATGACGGTTGCACTTGTTGCACAAAGGTGCGGTACCAACATAGTTTTGTCTTTTACGAGGTTGT
GCTGGTTGGTTTGGTGCAACTTATTTGTCTTGAGCAGTTACAGCGTAGCTTTGAGAGGCTTTGCGCTTTT
TCTTCTTGGACACCTCAGCACTTTCTTCCTTTGGTTCCACATGAGCAGTCTTGTCAGATGGTCTTAAGTC
GCCCTTCCGAAACAGCTTGCGCTTCCTGATTTGGGATTCAGTCAGAGTGGCAGCCAATTCAATGGCTTGT
CTGACAGTAGTAGGGTTGCTGCCAGTAACGATATCCTGAACTACGTCAGGAAGCCCATCAATGTATTTCT
CAATCGCCTTATCCAGAGGCGTAACCATAGTAGGGCATAACAAGCTCAATTCCTCATAGCGATCAGTATA
GGCTCGATGTTCCCCACTGTCTTGCTTTAAATCATCAAATTCCCTTTCTAGAGCCCTAAGCTCATGACGA
GGACAGAATTCCTTCATCATAAGAGCCCTAAGCTCGGCCCAAGTCTGTGCTAATGCAACCTCTGCACCGC
GGTCTCTCATAACCCCGTTCCACCACGTAAGAGCCCTCTTCTGAAACACACTCGAAGAAAACTCGACCTT
GCGATTTTCAGGACACTGAACGTGACGGAAAGTATTCTCAATGCTCTCGAACCATTGAAGAAGCCCAGTT
GCTCCCTCAGAACCACTAAACTTGAGTGGTTTAGCCGAGTTAAAAATCTTGAAATTGCATGGAGCATTGT
TGTTGTTGTTGTTGATGGCTTAGGTTATTTGAGCATAAAGATTTGGGAAGGCAGCAGCCATTTGCTGCGC
GATAATTTCCGCCAGTTTGGCATTCGCCATTTGATTGTCGCGTCGAGGAGGCATTCTAAAAGAGGAAACA
TGAGAGGAAACGAGTGAGAGAATTAGACGAAAAGAATGAGATGAAACATATCTGATGAAAACAAGGATAG
TGGTCATTTGTTTGTTACTACAAAGCAACAAACGACCTTGGTGATTAATCAAAGTAAATAGGCCAAAAAT
AATGTATCGCGAAGACATGCTCGCCTATAAGTGGACACTCACCCCAAGAGTTCCCAGGTAAGAGTGACTG
GTCCGATTATGTGGATTTGTACGAACACTCTAGCCTTAGACAGAAAACTCAGGGTACAGGCATCCACTCT
TCCGGATTGCACGTGTTCACATTATTTAGACCAACTTTGACGAGAGTTTTGAAAACTCAAAGGGGTTCAA
AACCTTATAACAGAGGGTTCAAAACCTAGTAATCAATCATCCTAGAACAGATGATTAGTTTTCAAAGCGG
TTTTGAAATTTATGTTCTTGTTGCGGTCGTCGCCTAAGGATAGGTGACGGTGTTGTTTAATGACTAAACG
TAAGTAAACTCGCGTTAGGGTCCTAGGAAGGTTATAGACTAGGTCAAAGCATTACTAATAACCTAATTCC
CTATAACCATGAGCTCTGATACCAACTTTTCTGTCACACCCCGACCACGTGGAACAACAAAACGTGGCGG
AAACGTCGGGGAGTGTTGTAACAGAATCATTGTTTCATAACACATGGAAATTGAAATATTGTTTCATTGA

TTAAATGAACTCATAGTTCTATTACAATCACATAAGTACTACTATGATCTTCCAAGTTTTAAAGTTGCTA
AGGCACGAGTCCATCCTAAATGTAGCATGTATCAACAATCATCTCAAGTCAACACCGGAAACATGTGTAA
AAATAGGTACGTCAGCATAAAAATGCCTGTGAGATACATAGGTTTTGTGAAAATGGGATTCATGACTTGA
GTTTAAAGAAAACGTTTAATAACAGTTAGTCATGAACCTTGTAATTTGCTTTGTCTTGTAAATCATTTGA
AAAACGATAAGATTGGATGATATGTATAAATAAGAGAACACTGTATTGTTAAACGGATAACCATGTAAAA
ATGTGTTTGTATAAGATAAGTCGTTTGTAAAACAATGTCTCGTGAAAAGTGTGTTATTTGTATAAAATGC
CATATGTCTCAAATTGAAATGATTCAAATAACGCTACGATATGTAATACCATACAAACACTTATATATAG
GAAGTACCAGCGGCGTATCCACCATGCTTGTATCATATTACACACGCCTCGTTACTTAAGTCACTTACGC
AAACCAAACCACCAATATAGAATGTTCATGTTTAAACCAATAGTCCAAGGTTATTGTATAACTCAAGTTA
AATGTAAACCACGTAATGTATAAACAAAAGTTATGTATGGTAAGTGAAATGATGTAAGGTTAATCGAATA
ACCTTATAGCAGCATATTAACAATATACTTTGGATTTGTAAATAACATATTAAAAATATGTTTTGGTTTG
TTGATAAAATGCTTATGTTGGTTCTGTACAATACCACACATGAGAGTATATCAAACTATACTTTTTGGGA
GTATATCAAAATATACTTTAAAGGTGCGATTTAAGAATTCATTCAGACCTAGTGCATCAAACTACACCTT
TGAGACTATAAGGATACCACAAAGGAAATCCCTATTTGGATGGAGAAATAAATTGGTTTCAGACATTCCA
TGACAACAGGAATGGGGTGTCTAGTCCTATAGCGCTATATCATACTACCAATCGGTCTGGTGAACAAAAT
AAGTACTATTCCAACGATTACTTTTGTAATCTTTGAGAAATCAGTGTTTAAACCATAGATAGTCATTTAG
TTTGTCAAAAGATAAATACAAACATGTATAATCAATTATACTTTGAAATCATGAAAATAACAGATAGGTA
CACATGCTTCACCCCAAAACAGTTTGGAAAACAGTAAAGAGGGGTTCAATGTACTCACCTGAGATTGCTT
TGGAGTTCTTGTATAATAACCAGATAAAGCTAAAGATCACGGGATATCAAACGGCACCTAATAGGTAGCT
ATGTTAATATACCGGACCAAAATGGAAGGATCGGATAGTACGCGGGTTCGTAAACCAAACGAGTATGGAG
ACTCGTGTAATATGGTTTAACAAAGCCTACATACTAAAATGAAACTTAACCTAAGTGCTTACGGTCCATC
ACGACCCGTTTAGGTAGCTTATGCTACCTTACGCGTCGTTCGCGTAGAACGCGTCTGGAACGCCTAACAT
CGTGACCACAAGGTATAACCTCGGGAGGTTATAGCTATGGTCACCTAATGTGTTTGGTCAGATCCTAATG
ATCGACCGAATGGGTCGGGTTCGAAAGTATAAGCGATGGTTTAGATCGCTTACCTTACGACCCTATATAA
GCACTAAACTAAAAGTGACGAGCTAAGCATGTTAGAACATGCTTAACTCAGTTTAGAAAACAGGTTTGGC
ATCAAAACAAACGGCTTTGATGCTCACGAGTAGTTTGGTTACAAAATACGCAAGAATGCGCATTTTGGCC
GAAACTACGACTCGTCACTGAGCCTAGATAACGTGGTGACCAGTAGGTATAGTCACTACGGACTATAACC
ATCGTGATCACGCTCACGTTATGAAGTTCCATGAACTTCGCATCGACCATAAGCTGGTCAATGCAGAAAG
TGAACAAAACGTTGACTTTCGGACTCGAAAAGCGATAAAACAACGAAAGAATACTTACGGAGGGTCCCCG
AATGCTAATCTAGATCAGAGAAACTCAGGTATGAAGCAATACTTCAACTTAGAGCTTTTAGATCTGATTC
TTGTGATTTTTACAGAAAGGGGGGGGGGTATTTATAGGAAAAGTGGAACTGTTAGGATCGTTTATCGAAT
ATCGTGCCGCGATCTCGTGCGTACACTTGTCAAAAATGTGGTGGTTACTGAAAATGGCATTTGGCCTTTG
ATTGGGTGAAAGGGCATTGCCCCTTGATCGAACGAAAGGCCAACTGCATTGATTAGAAACATTGAATCTG
GTCTGACAGTCTGACGCGGCCCGCCTCAACATTTAGGCAAAACTCACGCGGGCCGCCTGAAGCTTCTGAT
CTGCACATACTTTCAAAATTGGCAGTTTTGGTCCCTGTTGCGTATTTAAGCCATTTCCGACACTTCTAAG
GCCCGTAAAGCCAACTTTAAGGCCCTAAAATGATGCCTAAACATTGTGGACATGAAACATGCTCAAAAAT
GTTCCGGATGTTGGTTCGTTTGGTCGTACGAACGCGATGTTCGCTTAATTACGACGACGGAATGCGCATA
AGCGCGAAAAACGATCCAAATGACGCGACGAATGGATTTTTCTCATGCCAAACACTAAGGCATAATATAA
GGATGCTTACATAAATTTTTGGATGTCCGGATGTATTCAGAACGTAAGTTATGCGCGAAAGTGCAAACTT
GTGCACTTTTTGACACTTTTAGTCCCTGAATGATCCAAAAGTTTGTTTTAGCATACCAAACACCTCAAAG
CCTATTTCTAAGCTATGTAAAGGATATTTATGGTATGTTTAACTTATGGACATGTTCCGGAATGTTTGTT
ACAGTTCAAATTGGCATACTTTCGTAGTTTGTCAAGTTTAGTCCCTGTAAGCGAATTAACTTGTTTTTGC
CATACCAAAGCCTTCGAAACTTATTTCTAAGTTATGTAAAGGTTATTTAAGGTATGTTAAGTGTATGATG
GTCTTCCGGAGTATTTGTCGCACTAAACTGAGTACGTTTACGCACTAGTTTGCGTATAATTCCCCAGAAA
GTGTTGTAGAGTTTGAAATTGAACTAGAATTGATATGTGCAAAACATACACATAATTATACAAATCCCAA
GTATGAAATAAAACATTTTATTGGTTTGGCATTCGTTTGATGGTTGGAGTGACACAGGTGTCACAATTTG
GTCAGTGAAAAATAGTGGCGAGTGCATTCAATTTGTACAAAAGACACATTGTTATAATATACAGTATTAA
GAGCTATTACAATGTTTATATCAACCAATTACCCCAAATCGGTATTTGTCACTCGACTGGATCTGTGACT
ATGGTCAAATCACACATTGGCTAACCTGTTGTCCAATGGTGAAGCTTATGAAGTAGTGTATACAAAACCC
CACATACCGGCAGTAATTGTAGATTAGAAAGACTTAATCACTGTAATTACAAATTTGAAAACAAGTAGGG
TTTTGAGAACAATTTGATAAAAAAGAGAATGACTCACATTGCAGTTTTAGACTATCAAAGCTTGCCTATT
GGTTAAGCTGATAAACCTAATTAAATATACAATGCAAACGAACTAGGTTATTAACTTAATACAACATTTA
CGATAATCACGATATCGAAACCCTCACAACGAATGATAAAGTATAGCCAAATATTTGCAGCACATAAACA
TCCATTGGATCGGTTTTAATCGATAGGATATAGTACCGTAGCAGTGATCGAGTTATTACCCTGATATCGT
AGCAGAGTTCTGGTACTTCGGGTATATTTGGGGTATATTATTTGTGCATGATGTTAGTGAAACAAAAGGA
ATTCGAGCGAATTGAACTGTCGAATTGAAGCTCAATTTATAGCATTTTCACAGGTTTCTCGCGCCCCGCG
GCTGGCCTAGGGCTGTCTGTCGCGCCCTGCGAGGGGCACTCTAGCTACGGCTAGGGCTGCCTTGTCAACT
TGTAGCCTTGACACATGAACAACATGTGTCAATCAACGTATCTGGTCAACATGGGGTTTGTCGCGCCCCA
CGACCCACTCATATAAGCTTGTCGCGGCCCGCGACATGTAGGGAAATTTATTTTATTTGATTTTTTATAA
AAAATAAGGTTATACGGGTTCGGGTTTTCACTTACGAAGAGTTTTAGGATGTTTTGTGGGTTGTTACATG
TATGAAGTTTGAATTTTGCATACGAGCGAAGCCATGGATAGCAAAGGCATCCAAACGGTTTTAATTTGGT

GTATAATGGAGCTTTCTTGTAGAGGATGGAATATGGACTTTGAAAATTTAAGATTGGTGTAGGGAGACGA
TTAATGAGATACACCGCGGTTTGAAAAGCATGTGTCCAAAATGAGAGAGGTAATTTAGCAAAGTGGAGAA
GAACTAAGCCGGTTTCCACAATGTGACGATGACGGTGTTCGGCCACGCCATTTTGTTCAGGAGTATGTGG
AGGTGTAGTGAAGTGAGAATTACCTTGAGCTTGTAGATAAGGAATTAAACCAATATATTTTCCACCGTTG
TCACTTAAAAGAGAAATTATTGGTGTTTGGAAAAACTTTTCTACAAGCAATTTAAATTGAGGAAACAGGA
TAGACACGTCAGATTTGTGTTTTATGGGATATAGCCAAATGTATTCGGTATGAAAGTCAAATAAAATCAC
ATAATAACTGAAGCCGTCAATAGACTGTTCAACTAGTCCCCAAACATCCAAATAAATCAATTGTAAGGGT
TTTGTAGTAGTGTAAATGAGTTTATGCCAAATGAATCTTTGTGACTTTTATTGGTTGAACATGAATTACA
ATGAAAATTCGGAAAATTAAGTTGTAAACCTAAACACTTGGTTATTTTCTTGAGAATCCGATTGAAAGGG
TGCCCAAGTTTGTGATGCCAATCGAGAGGCGTTATGGTTGAAACCACATTTATTTGCGGTGACGATGGAG
GATTTGCATAGTAGACATCAAGGTAATCCTCTCCTCGCATTAGACGTGCCCCCGTGCGAAGATCCTTGAC
AAAGAAATGTGTGGGAAAGAATTCCACAGAAACTTGATTAGTTTTACATAATTTTGCAACGGAAATTAAT
TTGTTTTAAAATGTGGAACACATAAAACGTACTTTAAAGAAAGGGAACCATTGGACGTTTTAATATATTG
TTGACCAATGTGTGTAATAGAGAGAGTGTTCCCATTGCCAAGGATGATTTCATCCGGGCTACCATATTCG
GATACGGAGTGGAAAGACGCACAATCGGGATCAACATGATGGGATGCTCCAGTATCAAACATGAGAGGCG
AGGTGGTGACATTAGCAGTTGGTGTATTTTTTATTGGTTATTGAACCTCCGGTGGTTACTTGATTTTCTC
TAAAGGAATCGGGCAAGTTTCCGACATTCTTTGGCTTCATGTCCTGCAAAGTTGCAATATTGACAATAAA
TATTGTTACGGTTTTCACGATTACCATTATTTGAACCATTGGACCACTGATTGCGAGATTTGGTGTTGAC
ATTGCCATAGTTTCCACGACGTCCGATTCTAGCTTGTCTTTGAGTATAATTGGCTGTGGCCATAATTGGT
TATGGTGAGATGGCTGTAACAGTGAGTTCTCTTTCATAATCCACAAGCTTTTCAAACAATTCATGAAAAG
TAATTTGGGATCCCAGTAGGCGTAAAGAAGTAGCAATGGGTTTGAAATCCTAACCCAACTGACAAAGTAT
ATGTATCATGAAATCTTTATCTTTCACTGGACTTTGAACTAGAGCGAGGTCATCGGCAATCAATTTCATC
TCACGGAGATACCCAGCAACCAAGGCCGGCCTTGGGGGTGGGCGGGGAGGACCACCGGCCAGGGCCCGAT
ATTTTGAGGGCACTTTTTTTATGAAAAAAACTCGATATGTATATATAAAATATTTTTTTAATATGGTACA
CTCATACAGATACCAACATAGGCCCACATACGTAACTAATCTTATGGTTTAGATTAGTTATTGACCCCTT
TGGCATTAGGTTTAGACCTTTAGTGAGCCCAATACCTAATTTCGTTAAGGACTAAGGGCACATTTTCTCG
AGCTCGAACAGGGTACACGAATTCTAAGGGCCGGCCCTGCCAGCAACGAGATGGTTTTCTCTTGGATTGG
AAGCAAGTTTCGATTTCAACGGTATAATTCTAGAACGAGATGAGTTTGCGTAGCTCGTTACGAGCCGATC
CCAAGCCTGTCGGGACGTTTCAGCCGATGCGATTAACGGTTGAATTGCGGGTGAGCAGCTGCCAAGAAGA
GCAACTAAAATCACCTGGTCTTGCCAATACCAGAGATAGTAATCGAGATTAGGTTTAGTTCGTTCCGCAT
CAAGGAATTCAGCGGGTGCCGTTTTGGTTCCTTTAATGAAGTGGACCAGGTCAATACCGATGAGTGCTGA
TTGAACGTGTTTGCGCCAGATTGAGAAGTTTCGGGTGGTGAGGGTGATGTTGAAATTGGTAGAGGCATTG
AGTTGAACAACAGTTTGTGGGGCCATGGTGATGATCAGAATTGGTGGTGACCGGAAAAAAGAAACGGACA
GGATCAAGACTCATTAGAGTGCAATGCTCTTAATACCATATAAATGCTTGATTGTATTATCAAAAGGCAT
ATTGAACACATCGAATACAATGGCTTTAAATAGCCAGTACAGATCTTGATACAATACATGAATGGAAAGG
AAAATACTAAGGCTTTGGTGAAGGCTAAATATGTGGAGAAAAATAATAATAAAATGCTATCTAAAAATAA
CTCTAGCATATAAACAGAAACAACAATCATGAGAATCAAATATTATTTTCCTTAATATGGGGGTTGTCCA
GGCCCGGCGATCAAATACTGACACTCATTATAACAATTATAATACATGTTTTACATCTATCAAATAACAT
AAAAACGATATCGGTTCCGGTGTTGTTTAGTCGGTATTGGTTTAGTTCGCCACGATATCGATACACTAAC
AACACTCGATACAAATTACAATACCAAAAAGTATTATATTGTAGATACAACTCTATATTCAACGAATTAT
ATACCTCCACATAGAGAAAACCATAATGCTGAGTATAGTACTGATACCGTATTGCTTGTCGGGACCTGTT
TGGTTCGGTTCGTGTCAATAAAGAACAAGAAAATCTATATCTTTTATCGAATCGCAATGGTATATTCTAA
GTTTTGTAAAACACTATATTGTACTACCAAAGAAAACCGAAAAATAATAGATTACATAAAAAAAGATAAA
CAAATGTTATACATGTTCCGTCCTGAATTTATTCGAACAATGTTGGTTTAGTTCCTCACTGTAGAAAACT
AAAAAAGTCAACTGTATTTGACTCGTTCGAATAGAAGTTTTGTCGAAACATGGATCAAAGTAAGCATATT
GCAACAGTATAATCAGAATCAATCTTAATAAACATGATATACTATATTACTAAAAAGATGTGACATGACT
TTTATAGCATGTACATTTGTGCTAGTTAAAACAAAACGTTAAAACAAATTGAAAAGCGGGATAAAAGGGA
AATAGAAAAAAAAAAAAGAATAAGAAAAACGAGTAAAAGAATACCCGAATAAAACCTCCATTGCCCAAGAT
TCTAACCTCGAAGGATGGAAACCAAGTTGCATGCCCACTATACCAGATTACCAGTGGAGTCGCGCCTACT
TTGGATCAATTAGAAGCAGTCACTATTCATAAACCTTTTAAAATAATATACATTGTTATTAATATATTAT
TTAAAAAATAAAAGCAAACACTTAAAAAACGTAAATAACTATTCATCACCTTTTTCTAATATATATATAT
AGGTTTGTTTACGTCGCGTATTGGGGCGGAATAGAACAAATGATAATGTAAAACAAACTAGGTTTGTTAA
CGTGTTGCGTGACATGTATCATAAATAATTACTGAAAACAGAGGGACTAGCATATACAAGTTGATTTAAA
AGAGAGACGCAAAGGATAGGATAGGCTAAGCTGAATATGCCATTTTTGGTTTCGATATTCCATTATCAGG
TTGTCTAGTCTTCACATTGCTATTTCATATTTCTCATTTAATGATCAATTCTCAGGAAGGAGACAAAACT
GATTGAAGAAGTTGTTACCAACATCCATCATAGATTAGGTGGACCCTTAAGTAATACTCTACCATCCCTT
ACGGGGAGGAACCATGACATTGAACTCATTACTTCGTGGTTGACAGATGGATCTTCTCACAACGTCGACA
TTCTTACAATTGTAGGCATCAGTGGGATTGGAAAAACATCTTTGGCTAAGTATGTTTTCATGTTACACTT
TAGTAAATTCCAGAAAAGCAGCTTTATCGAAGATATCAATACAAGATGCGAAGAACGCCATAATGGATTG
CTTGATTTACAAAAGCAACTTCACGGAGACATTTCAAAAAAACTCGACTTACATGTTAATGTTGTCAACG
AGTACACCTCTAAAATTGAGGATGTGTTATCTCGTAAAAAGGTGTTTATAGTGTTTGATGACATTGGTAG

TCTTCACCAGCTGAACGCTTTACTTGGAAAAAAAGGTTTACATCCAGGAAGCAAAATCATTATTACAACC
AAGGATACATCTTTAACAGAAAGGTGTGAATTATTCAAGACACAAGTTCATCCCAAGCATACTCAGGTCT
TACTTGATGGCTTAAATAAGTATGAATCACTAACGTTTTATGCATTCATGCATTCCAGAGCCATACCCCC
AAAGAAGGTTATAAAGATGTCTCGATAGAGCTTGCGAAATATTGTGATGGACATCCATTGGCTCTAGAAG
TTCTGGGTGGGTCTCTACATAAGCGAGATGTAGCATATTGGGAAGAGCACATAAAAGAGCTAAAGAAAGA
ACCTCTTTTGGGTATAGAAAAGGTAAAAAAGGCATTACAATTGAGCTTTGACTCTTTGTCATCCGAAAAT
GATAAGGAGCTGTTTAAGCATATTGCTTGTTTCTTTGTGGGAGAAGATAGGGATCTTACCGTAACAATAT
TAAAAGCCTGTGATATAAACACGACATCTGGGATTACAAATCTCATGGAAAAATGCTTTCTTCGTGTCAG
AAGGAATAATGAATTGTCGATGCATTCATTGATTCAAGAGATGGGAAGAGATTTAGTACGTCAAGAATCA
CTTAACAAGCCATGGGAGAGGAGTCGATTATGGTGTCACGAGGAGTCATTCGATGTGTTGAAATGTAAAA
AAGTAAGAGGCCATTTATATTTTAGTAGTAATGCATGTTCTAAGTATTATTATTGTCTATATACATGTAT
ATAACCTTTATATATTATTTTGTTTGTATTCCTTTTCATCGTAGGGAACCCAAAATATTCTAGGCCTTGC
CCTTGACATGAGGATGCTTGACAAAAAGAAGTTGCGTGGATCATTTGAGCTGAAAACCGAGTCATTAAGT
CAGATGAATAATCTAATGCTACTACAACTAAATAATGTGCAATTAAATGAGTGTTTCCCCAATTTTCCGG
AGGAGTTAAGATGGTTGTGTATGCACGGGTTCCCTTCAAAGTCTATTCATTTAGACCTACCAATGGAGAA
TTTGGTTGTTCTTGACATCTCTCGCAGCAATATTGAATCTTTTGACATGTCATACAGTGACCCACAACCA
CCTGCGGAGAAGCAAAAGGTAAGATTTTAGATACTTGTATATGAGATTCTGAATAATCAATAATCCGCTT
ACACGTAATTACAATGTAGAATTTTAACCCAAAAATTGCCTAGGCTTAAACGTAATCACAATCAGATGAT
ACAGATTCACGCTCAGGTCTTAGGTATAAAACCAAACACATACACTAGGATCCTATGGTCTCGCTTAAAT
TTCAATCATTTAACATGACTGATGACACTAGAATCGCCACATAAGATGACTCTTATTATTCTTTCATATT
GAGTAAATGCATGTTCTACTTATTTTAACTTATCTATGTTTCATCAGTTGGTTGGATCATGCTCAAACGA
CGAACCGTTGCTTGGATCATTGAAGATTCTAAATCTGAGTTTCTGTGAACAGCTTCATAGTGTTGGTGGT
TTTTTTGAACTACCTGCACTTGAGAAGTTAACAGTTAAAAGGTGCATAAGTTTGATGGAGGTGTGTGAAT
CAATTGAGCAATGTGTTGAACTTGTCCACATTGATCTGAGTTACTGCCACAAGATTAAAAAGCTTCCAAT
ATCCCTAGGCAAGTTACAAAAGCTTCAAACACTATTGCTAAATGGTTGTAATTTATCTGAATCTGCCATT
ATGGATTCATCAGATCTTAGCATAAGCTCACAAACCTCTTCATCAGCCATTGTAAGGGAGGCTATACCAA
GTGATTTCAAGTTCTTTATGACTTTTTTACCAAACTCGTTGAGAATTTTGTCACTTGAAAATAATAATTT
GTCAAACGAATCATTTCCCATGGACTTGAGTTGTCTAGTCATGTTAGAGGAATTATGTTTAAATGAAAAT
CCAATTGTGTCGCTGCCCATTTGTGTGGGAACCCTTCCAAGGATTCAGAAACTTAGTATTGACAACTGCT
ATGATGTGATATCAGTTGAGCATCTTCCATGTACCCTAAGAGTATTCTCCATGCTCACATATATTGGATG
TGCGACTAGGAAAATTAAATTTGATCCAGAACTTCCTCCACTCAAATTAAGAGGAATTCCTTTATTGGCA
TCGCATTCGTCGATTGAATTTGAAGGCATAATCAAAATTGAACCAATGGCAGATGTTGAGGAAAAGTTAT
TACATAGTTTGGGCTGGAAAAGGTCAGACTTGATTAAAGGACACCTGGAATTTCGTGACCTCCAGGTATC
CTCTAATTAATTGTCTCTTGTTATTTGTGCACATGTGTTTATGAAAGGTGATTAATGTTGTATTTATTTG
GCAGATGTGTTATGAATTTGGAATATTCAGCACCTTTTATGGGGGGAAAGGGATGCCGGAGTGGATTAGG
TGTAGAAGTCGGGGCCCATCAATATCATTTATCATCCCATCATCTCCTAAGAAGCTCCAAGGATTGAATT
TCTGTTGTGTGAATTGGCCTCTATCATTAAGTTTTCCCGTGCCAACGATCAAAATTAGTAATATTACAAA
AAAGCAAACCTGGATTTATAATCCTTATTGTGAAAACATCTGGTTAGGTGGAAAGGACTACATTAGTTGG
TTAAGCCATTGGATGTTTGGGCCAAATGAGATGAAAGCTGGTGATAACATTATAATAGAGTGTTATATGG
AAGAGGAAGAGGAAGAAGATAGCTACTACATAATTGATTATAAAACAGAGTGTGGGGTTGATGTTGTGTA
TAATGATGGTAGTATGGAAGAAGATGTGTTGAGTTATTACAAGTCATGGAATCACATCATTGGTGGAGAC
CTCTCTGCTTTTCAACTTACAACAGGTGAATACTACCTATGTAACAAGGCTTTTACGCGGGATTCCACTG
ACGAATATGATCCTTTCATTGGTGAAGAAGGCAGCCATAAGGGTAAATGTGCTCTCTCTGTATAAAATTT
AGCTAAGTCTTGTGTATTACTATGTTTTGTAAGATCATAGCTAAAATAAATAACGTAAAAG

>RGC52SW4_07_c1_104000_117000
GGGAGGTGACGGTGGGCTGGTGGTGGAGAAAAGCTAAAGATGGGGGAATTGAGATAGAAAGAGAAGAGAG
CTGAGAGAGAGAAAGTGTGTGCGTGAGTGTGTTCACTTATATATATTTTAATTTAATTTAATTAAATTAA
TTTTAGAGTAAAATTACCAACTTACCCTCATGTGCAAGGCACATGACATACTTAACTGAAAATTTTGACA
AGGTTAGTGCTAAAGGACGTAAGATGTAATGAGTTTTACAAATAAAGGACTGTGACTGTAATTACTGAAG
TTAAAGGTCATCTATTGCAATCCGATACAAACACAAGGGACGAAAAGTGTAATTTATCCTAATTATAATA
AAATATTTTCTGGTCATCCAAATTTGAAACTTCAGCAATTGGACAAAAAGTAGGTGTTGACTCATCGAAG
ACTTCTTTCTTTTAAAGTTACAAGTCTGGCTTGTTTCATACCTATCTTCTTTACTCAAATCACTCAAACC
TTCCTCCAATAATTCATAGTTATAAACAATGGTTGTTCTCACTGATCTTAATCATAGATATGATGTATTT
TTAAGTTTCAGAGGTGCCGACACTCGTAACGGTTTCACCGATCACCTCTACAACACCCTTTTGGATGCCC
GTATTAACACGTTTTTAGATGATGAAGAGATGGAAACTGGAGAACCACTTAAACCAGAACTGGAGAGTGC
GATTAGGTCTTCCAGGGCTTCTATTATCGTCTTGTCTGAAAATATGCTTCTTCAACATGGTGCCTTGACG
AATTGGTATTAATCCTTGAGCAGAAGAGGAAGTTTAACCAAATCGTCATCCCCATCTTCTATCATGTGGA
GCCCACGGATGTGAGAAAGCAACAAAGCAGCTTCATGAAAGCAATGGCACAGCATAAAGAGAGGATGGAT
GCAGAAGCAGATGCAGAGAAAAGAAGTGAATGGGGTCAAAAGATGGAGTTATGGGAGAGGGCACTCACAC
AAGTTGCCAATTTAAAAGGAGTAGTTGCAAATGGCAGGTAAATCACTTAATTTTCTAGATGTCAGCCATT

GCAATCTTTTCGAGAATTCATTACAAATGGCAGGCAAAATTCTGCACATGAGGCATCAGTTAATAATGGC
ATAATATGATATATTTAGTTTGGACATAGCAAGTGTTTTATTTTAAATAAATATTACAAATAAATCCATC
TTAAACAATTTCAAGACTTTAATTAACATGGAGTAAAAGGGTGAACTTTTCAGACATAGAGTGAGCGGGT
GAACTTTTCAGAAAAAATAAAAAAGTTATGTTATTTTTTATTTAGTTAAGTTAGGTTTTTGTCACTAGTT
GAGTTCAGAACCCAAACCTTCAACTATATTAGTCACTATGGTTTCAGTTTTTAACCTTTTCGTCTTTAAA
GCTAACGCAGTCCAGGTTATTCTGTTAAGTGGATGTCCAATCCCAAAACCTATTCAGAACGAACTATTCT
GGCCTCTGACATTTGTGTCTCTTGAAGCATGAGAAACGAAATCTTGCTCTCTTTTCAATCTCCACAGTTT
GCCTCCATTTCATGAACCATCTTACTCACGTGATTTTGAAAGCCTTCCAAATGGAATCCCAATGACTCCA
AATTTCAAAGTTGCTCCCACCTCTTTCTAAGTCTTCGTTGCAATTGGGAGGGAGGTTCTTATTTAAACCT
AACCCTATATATATACATATATATATAAAGGATTGTGGAGGGAATCCTGGCCCAGTACCCTTTTGGGTAC
TCATCATCCCACCCCGCAAGGGTTATGATGCCAACAAATAGCCTCGGTCACTACCTCCATGATATGGGCA
CCACAAGGGTCGAACCCGCGTCACTACAACTTCCCATTTCAACGTGACTTTCATTGGATGCGGGTACCAC
TGGGCTATAAGCCAATTGGTTAACCCTTTATATATAGGGGAGGGTTTATGAGGGTTAACGAGAAAACTCC
TGTTTTTTAGAGAACTCAAGAAACCAACATATCTAACCTTACTAAACCCTAAACTCTAATCTTAATCCCT
AAACCCTAAAAAAAAATGAACCATAAACCCAAAAAAAACTAAATCCTAAAAGCTAAACCATTAACTCTAAA
GGCTAACGGATAAAAGCTAAAAATAAAATCCTAAAGTTAAATCCTAAGATTTAAAAATTAAACTCTAAAC
CATAAAAAAATAAACCTTAGGAAAGAAAATTCAACAAAAATAGAGTTTCTCTTAATCCTTTCTATAGATGT
ATATAGATGCATCTTAAGAAAATTCCTCTTTGTCATAAAACTCTGTCTAACCATTATTAACCCTAAAGTG
GAAACTCAAATCAATAAACCCTAAGAAATAAACCCTAAACTCTAAAGTTAAACTCTAAAAGTTAAACCCC
TATAAAGGCTAGCGTCTAAAATCTAAAAACTAAATGCTAAATCTAAACCCTAAAATTTAAACTAACCTCT
AAAATTAAACGCTAAACCATAATCCCTAAAAACTAAACCTTAAATAAAATTCAACACAAATATCAATCGT
TTTATACAGTTTTCTAACTCACGAAGAATGAATTTTCTAATGCGCCTCTCCCTATATATATTGGAGCTTC
CAATGAGAAGAAAACATTTTTTAAGAAAAGAAAGAACATTTTAAATCAATAAGAATGATTCATTTAATAT
TTGTATATAATGTTATTCTAAGGATATAAGGATAAATTTACCAAGTTATTAATTTAATGTATTCCTTTAA
ATTAGCTAACTAAATTAAACCAATAGCTCATTTTTTAAATTTACCCTTTCAACACAAACTTGATTTACAG
TTAAGTATAACTTTTGATATTTTTAGAATAAAGTGTAAAATAAATGTAGTATAAATTTTGATGTGCATAG
GATATTTGTAGATTAAATGTAGTATAACTTCCGATTTGTGTAGCATAAATATGAATATGTGTATGATAAA
GTTTTTTTGACTAATTAATTAAACTAGAATTGAGACCCGCCGCAATGCGGCGGAGATTCTTTACTTATAA
CTAAGTTGACTTAAGACCCGCACGTTATGTTAAACCTGTCAAACGAAAAAAAAAAAAGACGATGTAAAAAC
GTTCACCCATACACGCACGTTGCGTCGTGTTAACTCGCAAAATTTAGAACGAAACGTAAAAAAAGTTAAA
CCAAAGACGCACGTTGCGATGTGTTAAGTCACAAAATTTAGAACCAAGCATAAAGCGGAAAATTTGCGGA
AAATGAAAACTATAAAGGATCAAAGTTGAAAGTAAAAAAAGTTATGAGGATAGATTGCAAAAGATAAAAA
GTTTTGGGTTAAAAGCAAAAAAAGAAAAAAAAACAAATAGTTTTGGGTTAAAAGTAATTTATGAAATACT
TTTGGGTGAAAAGTAAAAAAAAAATCAATTTTTTTTTTGGAAAACCTCTAAAGCCAAGGTTACGACAACCA
TATGCATAACCATTTTTTCTTTAGAAAACCCCCAAAGCACAACCCGCTTTGCGGCGGGGCGTAAAACGGC
TTCAAATAGTATTAATGTCACACAACCGTGATCGACCACCAACACTGAGTCGACCTAGGATCTGCGTGTT
GCGACGAACCTGTCAAACATGAAAAAATAGAGGTAAAAACGTTGAACCACATATGCACGTTGCGCCATGT
TAACTTTCAAAATTTGAAACAAAGCATAAAAGTTGAACCACACCCGTACGTTGCGTCCTATGAACTCGAA
AAATTTAGAACTATACGTAACGAAACGTAAAAACGTTAAACCAAAGACGAGAAGTATAATTGACAAAAGT
TGTGAAGTTAAATTGCAAATAATGAAAAGTTTTGAGTTAAAGCTAAAAAAACAAATTTTATAGGGTTAAA
ATTACAAAAGGTAAAACCTTTGGGTTAAAACTAAAAAATCAAGTTTTTTTTTTTTTTAAAAATCCTCAAA
GCACATGTTACAACTCTATATGCATGAAAAAGTTACTTCTTTATATAGGTTTTAATTAAGAAATATTTAG
TGTAAGCAGTTATGAAATAATTAGTGTGTTAGCATTATGTTTTTTTTTGAACGGCGGATCATCTATATAA
AAATATAGAAAAGCCGGCCAGCTTGAAGCTGCTGAAAAACCCGGAATTACAACCCGATGTTCCTTTATTT
GCTCCATCGTCCTGTTAATCGACGTGTGTCGTTTGTTGAACTCTATTTCATTCCTAGCTTTCCAAATCCC
CCACATAGTGACGAGTACGATAGCTTCAATAAGTTTCTTCCACATCTTTGATCCTATCGCATTCTCCACA
ACCAAGACCAATTCGTGAAAGGAAGACGGCATGAAAGATAAAGGCAATTTCAACCAAACCAAGACTTGCC
ACCATATCCATTTCGAAAGCTGACACCGGAACAGAATGTGGTTCATAGATTCCGTCCCGATACCGCATCT
GCTGCACACAATACTAGGTAAATTAATACCTCGAGCTCTTAACGCATCTTTAGATGGATTTCTGCCATGT
GAAACTTTCCATCCGAAATCCGACACGTTAGAAGGTATCCATTTATTCCAGCGATGAACCACTCGTTGCC
CATTTTCATTCATGACCTTCGCTAACTGCTCTCTCACAAACCAAACTTAGAAACTTCCTCCTTCGCACCC
AGCTCAGATCCATTTATCAGGCCCTTTCTACGGCCTTACCCCGTTCAGCAGACCATCAGACGGCCCAACT
GATCCCATTCGTATTGTTGTTGTGGGTCTTTAGTCCATTCCCATGCCCATAAGTATATATCCCCCACTTT
GTTGTAGCACTGATCCACTTTAATGCCTTTATCTTTGGCCAATTTAAAAATGTCTGGATACGCTTCCTGT
AGGGGACCACAATCCAGCCAATTATCTAGCCAAAAAACACACCTTTACCATCCCCTATTACCGGTTTGAG
GTGTTCATTTATCCGTATATTGTAGCTCTTAAACTCCTTTTCCACGACCATAATATCTTTTCAAACACCT
AGACACGAAGCTTTGACCGGGAAGCTTTTAAACGTTCTTCCATTACCATGCATTGCAGAAATTACTTTGG
TCCAAAGAGCCGATTTGGAATCTTTAAACTTTCACCACCATTTAACGAGAAGGCCCAGATTCACACCCTG
GATATCACCCATGCCAGTTCCCCCATAAACTTTAGAATTAATCACTTTCTCCCATGCTATCCACCGGATT
CTTCCTTTCCCACCTTTATCTCCCCAAACGAACCTCTTCTAATACTTTCGAGGGTGTTAATCACTTTGAC
CGGAGCCTTAAAAAGGGACAAGTAGTAATTTGGCAAAGCGCGCCAAGAACCGCTTTAGCCAAAGTCACCC

TTCCGGCGAACAATAGGCATTTGGCTTTCCATCCTGAAAGTTTAGCGTTAAACTTATCGATAACGAGCTT
CCAATTTTTAACCCTTTTCATGTTAGCTCCGACGGAAATTCCAAGAAAAGTAAAAGGCAATATACCCATT
CTACAGTTAATTTAGTTAATTTTTTGGGCCATGGCGACCACCTTCTTACTTTCCACCCAACCCCATAGAT
ATTGCTTTTGGCAAATTGATCTTAAGACCCGAACAAAGATAGAAACCCCTAAGTATCCGGTTGAGGTTCT
TTATATTTGTTTCCGACCACTCCCCTAAAAAAATATCGTCCGCATAACAAAGGTGTGAAATAGCTGGCCC
ATTCGGTAATTTTACACCTTGAAAAACACCTCCATTCAGCGCTCTCCTCATAATGACATCTACTGCTTCC
ATAACCAAGATGAAAAGGAACTTGGATATAGGGTCTCCTTGGCGAAGCCCTCTAGTCATATGAAACTCCT
AGGTTGGAGATCCATTTACGAGTACCGACGCCCTACTAGACTTCAAACAAGCTTCCACCCACTGTCTCCA
TTTAACCGGAAACCCCATAGCCTTTTTAGTGGATAATAAAAACGACCAATTGACCGAGTCATAGGCCTTT
TCGAAGTCCACCTTAAAAATCATCACCTCCTTTTTTGTGCTCTTCGCCAAAGATATGACCTCGCTAACTA
TTAAAGGGCCATCTAGGATGGATCTTCCTTCGATGAACGCGGTTTGCACCGGAGCCACCAAAGAAGATAC
CACACGCTTCAACCGACTTGTCAAAACCTTTGCAATCATCTTGTATAGCGATCCAATCAGAGAGATTGGT
CTAATTCTTTTATTCCATTCGGGTCTCGAATTTTAGGAATAAGTGTGATAAAAGATGAATTACAACCAAT
AGGGATTATTGGTTTAGAATGAAACTGTTGAAACACTTCTCCCACTATTGGTTCGAAGTACTCCCAACAC
TTTCTTATAAATTTGAAAGGGCACCCGTTTGGGCCCTGCCCCACATTCCCACACAGCTTGTTTGATTTCT
TCTTTAGAAAAATTGGTAACCAACGATCCTGCTTGTACATCCAAAAGCCTTGAAAATCCATCTTTATCCA
TTATTGGTCGCTTCTTCGATGGTTCTCTAAATTGTTTTTTAAAAGCTGAGTTAAACTTCTTCTTAATCAC
AACAGGATCCTTCATCACAATCCTGTCAACCTGAATCGAGGTAATCCTGCTCTGAGCCTTCCTGATGTTG
ATGTTGACATAATTATGGAAAACTTGGAATTTTCATCTCCAAGTTTTGCCCACTTAATTCTCGCCTTCTG
TTTTAGGTTCAGCACTTTCTACTTTTCTAAGGTTTTTATTCAACATTTGGCAGCAACTCGTTTAGATACT
TCATCCCTGTTTAACAATCTTGTCTGAGCCAGCATATCCAGATGCCTAACCACATCCGACGCCATCTTTA
AATCAGCTCCCTCTTTGATGATACAATCAGCCCTCCATTGTTTTATTCTTGTTTTCAGATCTCTTAACAG
CTCCAAAAATTTTGTGTACGCATTTGAATCTTGTTTCCTCTTATCTACACAATGTCATTTGTTTTCTCTT
TTCTCACATTAATATTTTTTTCTCGAAAGACTCCCCTACATATGGTTAGGGTAAAATAAAACCCAATTGA
TCGAGTCGTGAAAGCTTATTTAAAAATGATATGTTTTGTACGGTTTTATTTAAATTTAAATAGCAGGAAA
TGTTATGTGAATAAGTTCGCGAGACTTCTATTAATTCTCAAATGTAGTAATTTTTGTCTATGAAAAATGT
TTTTTTCTAACTTTTTTAAAAAATGTTTTTTTTTCTAAGTTTTTTATTCAAATTCGTATATGAAGTTAGTT
AGAGGCATTAATTTATTTTTTTTTTTTTTTTGAATGGCTAAACTTGCTTGTCATGGGAATTGAACCCATG
ACCTCCACCTATTCTACACCTTATCTAACTCACCAACACCATTGGGCTACTCCCAATGGATGCATTAATT
CATGATTAACCATAATTATTCTTATATATTGTAAGAAAATGTGATGATATAAAGGGTGAAGATACGATAG
GAAGTTTATTTGGCTAGCAAGGCTAGGAAGTGATCTTGACCATCCATTTTTAGGGGCATTCTAGTCAATC
CAAGCCAATAGTTTCTCTCTCCTCCAATTCTCCCCCCCCCCCCCCATTTTTTAAACGTTAATAACTCTTT
CATACAACATTATTTTTTTATAAAAATTGCACCAAAAAACAAGCGTTTGTTATCTTTAAAACGAGTATAC
TATTGCTATATTTAAAAAAAAAATTTAAAACCCAGTTGCGTAAAACGCAATGAAAAAAAAAACCCCAGTT
ACGTAAAACGCAATGAAAAAAAAACCCTAAAAAATGACATTTTTCTAAAAAGCAATGCACCAAAAACACA
AAGAAGTGACTTATTTGTAAAACGCAATGGCCAGAAAACACAAAGAAATGTCTTATTTCTAAAATGCAAT
AGCCTAAAAACTCAAAAGAAAATTTACTTCCTAAAACGCAATGGACTAAAAACACATAAAAATGTGTTTT
ACCTAAAACGCAATTCACCAAAAACACATACAAATGTCTTATTTCTAAAAGGCAATAGCCAGAAAACACT
TAAAATGTCTGTTTTCTAAAACGCAATGGAATGAAAACACATAAAAAGTGTTTTACACGCAATACACAAA
AAACACAAAAAAATGTCTTATTTGTAAAACGCAATGGCCAGAAAACACATAAAAATGACTCATTCTAAAA
CGCAATTGCCTAAAAAACAAAAGAAAAGTGTTTTATGTAAAAAACAATGGACTGAAAACACCTAAAAAGT
GTTTTACCTAAAACGCAATGCACCAAAAACACTTCAAGAATGTGTTTTTCTAAAACGCAATGGCCAATGC
AGAACTGTTCTGTGAATTGTCTGTGCTGTGAACTGTCTAAACCAATGCAGCAGAAAATTAATTTTTCTGA
TGCATTTTTATTACAGCAATTTATTTGAAAAAAAAAATTTCTGTGCTGTGAACTGTCTAAACCAATGCAG
CTCAACCCCAGCCAGGGGCTCTGCCCCTTGGACCCCGCTACCAGGGGCGCTGCCCCCGGACCCCGCCAAG
ATCGTAAAACGCAATGACTAAATCAAAAACCCAGATCGTAAAAATGAAATTAAAGAATTTCTTACATGGA
TCGAAGTGATTTCTTCAACAATTGACGAATTTTGAATGATTGAAACACTTATCAGCCCTCGAATCGAACG
GATCGAGTGATTATCTTTCAAATCGCCGGAAAAAAACGAGATTTTGTATGAAATTAAACTGGGTTTTCTT
CAAAAAAGCTGAAGAACACGTTGATCGGGTGTTTGAATCATTGATTGGTGTTGAAAATCGCACTATAATG
TTGTGATTATTGAGATAGAAAGTGAAGAAATGGTTGAAGATGGTGGGTTTTGAAAATGGTGGGTTTTGAA
GTTACTAGGTTTTGAAAAAGGAAGAAGAAAGGGTTGGATTGACTAAAATACCTTTTTTCTTTATTTTAAA
TGTTGCCACATGTCCTAATCCTATTGCTTCCTACACAGCCAAAATAAACTTCCTATTTGATCTTTCACAT
GATATAAAAACTCAATATATTATTTTATTGATCACATCAATCTGTATAATCCCGGATTATTCATTCATTA
ATTAGATAAATAGCAACAAGTAACGTTACCATAAATAAATTAGTAAAACCAGAGGAACTCACATAAACAA
GTTGATTAAAAAGAGCGACGCAAAGGATAGGCTAAACTAAATATGCCACATTTGGTTTCGATCTTCAAAT
TGTTATATCTTTGGATCAGGTAGTCTATTTTTAAGAACGCTATTTCATATTTTTCATTTAATGATCAATT
CTCAGGAAGGAGACAAAGCTGATTGAAGAAGTTGTTATGGACATCCATCGTAGATTAGGTGGACCCTTAA
GTAATACTCTACCATCCCTTACGGGGAGGAACCATGACATTGAATTCATTACTTCGTGGTTGACAGATGG
ATCTTCTCATAGCGTCGACATTCTTACAATTGTAGGCATCAGTGGGATTGGAAAAACATCTTTGGCTAAG
TATGTTTTCATGTTGCACTTTAGTAAATTCCAGAAAAGCAGCTTTATCGAAGATATCAATACAAGATGTG
AAGAAAGCCATAATGGATTGCTTGATTTACAAAAGCAACTTCACGGAGACATTTCAAAAAAACTCGACTT

ACATGTTAATGATGTCAACGAGTACAGCTCTAAAATTGAGGATGTGTTATCTCGTAAAAAGGTGTTTATA
GTGTTGATGACATTGGTAGTCTTGACCAGCTAAACGCTTTACTTGGAAAAAAAGGTTTACATCCAGGAAG
CAAAATCATTATTACAACCAAGGATACATCTTTAACAGAAAGGTGTGACTTATTCAAGACACAAGTTCAT
CCCAAGCATACTCAGGTCTTACTTGATGGCTTAAATAAGTATGAATCACTAAAGTTTTTATGCATTCATG
CATTCCAGAGCCATACGCCCAAAGAAGGTTATAAAGATGTCTCAATAGAGCTTGCGAAATATTGTGATGG
ACATCCATTGGCTCTAGAAGTTCTGGGTGGGTCTCTACATAAGCGAGATGTAGCATATTGGGAAGAGCGC
ATAAAAGAGCTAAAGAAAGAACCTCTTTCTGGTATAGAAAAGGTAAAAAAGGCATTGCATATGAGCTTTG
AATCTTTGATGTGTGAAAATGATAAAGAATTGTTTAAGTATATTGCTTGTTTCTTTGTTGGAGAAGATAG
AGAATTGACTAAAACAATATTAGAGACATGTGGTATAAACACAACATCTGGGATTGAAAATCTCATGGAC
AAATGCTTTCTTCGTATGGGATGGAATAACAAATTGACGATGCATTCATTGATTCAAGAGATGGGAAGAG
ATTTAGTCCTTCAAGAATCACCGAACAAGCCAGGGGAGCGTAGTCGGTTATGGTGTCACGCGGAGTCATT
CGATGTGTTGAAACGAAAAAAAGTAAGAGGCCATTTATATTATAGTAGTAATGCATGTTCTAAGTATTAT
TATTGTCAATATACATGTATATAACCTTTATATATTAATGTTTGTATTCCTTTTCTTTGTAGGGAACCCA
AAATATTCTAGGCCTTGCCCTTGACATGAGGATGCTTGACAAAAAGAAGTTGCGTGGATCATTTGAGCTG
AAAACAGAGTCATTTAGTCAGATGGATAATCTAAAGCTACTACAACTCCATTATGTTCAATTAAATGAGT
GTCTCCACAATTTTCCGGAGGAGTTACGATGGTTGTGTATGCACGGGTTCCCTTCAAAGTCTTTTCATTT
AGACCTACCAATGGAGAATTTGGTTGTTCTTGACATGTCTTACAGCAATATTGAATCTTTTGACATGTCA
TACAGTGACCCACAACCACCTGCAAAGAAGCAAAAGGTAAGATTTTAGATATTTGTATACGAGATTCTGA
ATAATCAATAATCCGCTTGGATAAACTTTGAATGGTGATGATAGAGCTATTACAATGTAGTATTTTAACC
CAAGAATTGCCTAGGCTTTCACGTAATCACAATCAACTGATTCAGATTCACACTCAGGTCTTAGGTATAA
AATCAAACACAGACACTAGAATCGCCACACAAGATAGCTCTTATTATTCTTTCATATTGAGTAAATGCAT
GTTTTACTTATTTTAACTTATCTCTGTTTCACCAGTTGGTTGGATCATGCTCAAAAGAAGAACCATTGCT
TGGATCATTGAAGATTCTAAATCTGAGTTTCTGTGAACAGCTTCATAGTGTTGGTGGTTTTTTTGAACTA
CCTGCACTTGAGAAGTTAATAGTTAGATGGTGCATAAGTTTGATGGAGGTGTGTGAATCAGTCGAGGAAT
GTGTTGAACTTGTCCACATTGATCTGAGTTACTGCTACAAGCTTAAAAAGCTTCCAATATCCCTAGGCAA
GTTAAAGAAGGTTAAAACACTATCGCTAAATGGTTGTAATTTATCTGAATCTGCCATTATGGATTCATCA
GATCTTAGCATAAGCTCTCAAACCTCTTCATCAGCCATTATAAGGGGGGCTATGCCAAGTGATTTCAAGT
TCTTTATGACTTTTTTACCAAGCTCATTGAGAATTTTGTCACTTGCAAACAATAATTTGTCAAACGAATC
CTTTCCCATGGACTTGAGTTGTCTAGCCATGTTAGAGGAATTATGTTTAAATGAAAATCCAATTGTGTCA
CTGCCCATTTGTGTGGGAACCCTTCCTAGGATTCAGAAACTTAGTATTGACCACTGCTATGATGTGATAT
CAGTTGAGCATCTTCCACGTACCCTAAGAGAATTCTCCATGTTCACAAATGATCAATATGCGATTAGGAA
AATTAAATTTGATCCAGAACTTCCTCCACTCAAATTAGGAGGAATTCCTGATTGGGCGGCATCACATTCG
TCGGTTGAAATTGAAGGCATAATCAAAATTCAACCAATGGCAGATGTTGAGGAAAAGTTATTACATAGTT
TGGGCTGGAAAAAGCCAGACTTTATTAAAGGACACCTGGAAACTCATGACATCCAGGTATCCTCTATTTA
ATTGTCTCTTTTGTGTGTACATGTGTATTTTATGAAAGGTGATTAATGTTGCATTTACTTGGCAGATGTA
CTACGAATTTGGAATATTTAGCACCATTAATTGGTTGAAAGGGATGCCGGAGTGGATTAGGTGTAGAAGT
TGGGGCCCATCAATATCATTTACCATCCCATCATCTCCTAAGAAGCTTCGAGGATTGAATTTCTGTTGTG
TGAATTGGCCTCTATTTACATGTTTTCCCATGCCAACGATCAAAATTAGTAATATTACAAAAAAGCAAAC
CTGGATTTATAATCATTATAGTGGACACGCCTGGGTATCTGTGGACTGCATTAGTTGGTTAAGCCATTGG
ATGTTTGGGCCAAATGAGATGAAAGCTGGCGATATCATTATTATAGAGTGTGGCAGTAGAATGGAGTGTA
AAACGGAGTATGGGGTTGGTGCTGTATATGATGATGAAGAAGAAGATGTGTTGAGTTATTACAAGTCATG
GAATCACATCATTGGTGGAGACCTCTCTGCTTTTCAACTTACAATAGGAGAATACCTCCTATACCGTAAC
TGGTTTCTCAGTAAGCTCCATATAATTCCATATTATTGCCCTTTCATTCTATTATATATACTGATACACT
TATCATCAGGTATAGAGATAAAAATGATGTTTTTGATTATTGAGTTCATGCTTGAATACACATACATATG
CCCAATTTATAGTAAACTGAAAACAGGAATGAAAGGAAAGGCAATAAATAGCTGAGCTGGAGTCTGGGCT
GTAAAAAAGTTGTGATTGGGCTGAAAATCCTCATAATATTCTTGAATTATTCTCTTCTTTCCGACACTCC
CCCTCAAGTTGAGTAGTGGGGTCACCAATACTTAACTTGCTCAAGCAACTACTAAAAGCATTTCGGTTGA
CGGCCTTTGTAAGAATATCTGCGAGTTGATCTTTTGACGATACATATGGGAGCTCTATGATTCTTTCCTC
CAGTTTCTCTTTGATGAAATGTCGGTCTACCTCCACGTGTTTTGTTCGATC

8.2.2 Predicted amino acid sequences of BAC derived candidate genes

>RGC151SW1_aa
MVVLTEITEQSSSSTTLSTYDYDVFLSFRGEDTRHSFTDHLLKALREATIDTFFDDAEIQIGDFLKPELE
KAIKSSRASIIVLSKNYASSTWCLDELALIMEQRRTSGHIVFPIFYHVKPSDVRKQRNSFGDAMAEHKQR
METELNSEKRSEWAHKIEKWKKALTEVADMKGEEANDRRETILIEGIVKEISSRLDLYKRSEIPHIIGME
SSIGMITSFLKDRSQQSTKILTIWGMAGIGKTYLANYVFQLHYLDFERSCFLEDIERRCTQPNGLLDLQK
QLLKDIQARSWMDIDNVNAGTLKIEKSLLRKRILLVLDGINKFEQLDVLIGTKGLHPGSKIIITSKNESL
TEKCRLFQTQVPPKHMKHCLQGLDDEESLQLLSRHAFGCNEPNKGGNKQSIKVAKYCQGHPLALKVLGSS
FRREDATWEDILESLGKEINPDIKKVLKISYDTLPSEKDKELFKYIACLFVGEDRKFTEDILKACDIGKS
SGIKVLINRCLLTLESSDKLTMHQLLQDMGKDIVHQESLKKPWKRSILWHHEECLDVLQNRQGTTTTQGL
ALDMRTFENEIWKEPSSANMQMFGFRSYPSFSCLWWLFGWVSGMCSSSRKTKGDFETLALSEMRNLKLLQ
LNYVQLSGSYKNFPHGIRWLCMHGFPLSYIPSDLEMENLVALDLSNSKLQQLWKKPKLLRSLKFLNLSSC
HELLRVGHFSGLPLLERLTLTRCTSLLEVCESIGNYCQKLEVIDLSKCNKLKELPRSIRKLKNLTHLLID
GCSNLGEFPADSVNMKSHGSSSSTVVPRTPESFANSLPRSLVTLSLKHCNLYNESFPMDFSNLPMLKNLF
LDGNPMDSMPDCVRSLSRLETLSFTWCWKLKTVLCAPIQLKRLDVYFCNSLEKIKFHPEKSAVPRVFYQT
SFNLTEIQHVFKIQALSEIDEEVLCSLGWINIAYLNPCWLSKANDMGLYYLEQRILPAQMLYEHGIFSTY
VQGQEVPKWFTHRSSGSLFTLQSSPEKGKIKGLNVCIVHTISGKEVGPSRIKITNLTKKSSWTYQPMMYL
FPYNSAFGFGDKAVVWLSHWMFGKNEFEDGDKVRIDFSVKHYMSLSGLRYGGEGPEYANVREYGI

>RGC52SW1_aa
MVVLTEHESLPGSSSSSSAHDHNQRYDIFLSFRGADTRNGFTDHLYNTLLEAGINTFLDDEEIETGEPLK
PELESAIKSSRASIIVLSKNYASSTWCLDELVLILEQQRSFNQIVIPIFYDVEPTDVRKQQSSLGEAMSK
HKERMELETNAQKRSERGRKMELWERALTQVADLKGQVANGRKETKLIEEVVTNIHRRLGGPLSNTLPSL
TGRNHDIEFINSWLTDGSCHTVDILTIVGISGIGKTSLAKYVFMLPFGKFQKSSFIEGINTRCEERHNGL
LDLQKQLHGDISKKLDLHVNDVNEYTSKIEDVLSRKKGVYSVDDIGSLDQLNALHGKKGLHPGSKIIITT
KDSSLTERCDLFKTQVHPKHTEVLLDGLDEYESLKLLCIHAFQSHTPKEGYKDVSIELAKYCDGHPLALE
VLGGSLHKRDVAYWEERIKGLKKEPLSGIEKAKRALQLSFDTLSSENDKELFKHIACFFVGEDRDLTETI
LKTCAINTTSGITNLLEKCFLRIRWNNKLLMHSLIQEMGRDLVRQESLNKPWERSRLWCHEESFDVLKRR
KGTQNTLGLALDMRMLDKKKLRGSFELKTESFSQMNNLKLLQLHYVQLNECFRNFPEEEALRWLSMHGFP
SKSIHLDLPMENLVVLDMSCSNIESFDMSYSDPQPPAKKQKLVGSCSKHEPLLGSLKILNLSFCEHLHSV
GGFFELPALEKLIVRNCISLIEVCESVEQCVELVHIDLSYCYKLKKLPISLGKLQKVQTLLLNGCNLSES
AIMDSSYLSISSQTTSSAIIREAIPSDYKFFMTFLPSWLRILSLVDNNLSNESFPMDLSCLAMLKELYLD
GNPIVSLPNCVGTLPRIEKLSINRCNDVISVEHLPRTLRVFSMYTNILRSATRKIKFDPELLPLKLGGIP
LLASHLPVEIEGIIKIQPMADIEEKLLPSLGWKRSDFIKEGHLLTHDIQVINVVFILQMHYEFGIFNTIY
WCKGMPEWIRCRSWGSSISFIIPSSTKKLRGLNFCCVNWPPFTSDSFSMPTIKTSNITKKQTRIYNHYSG
HAWVRGKECFSWLSHWMFGPNEMKASDNITIECDRKMECGVGVVYDDGSMEEEEEEEEEEEEEDVLSYY
KSWNHIIGGDLSAFQLTTGEYYLGNWKFTQDSHEEIYAPFIGEEGSNKGKCRLSI

>RGC52SW23_aa
MVVLTEHESLPGSSSSSAAHDHNHRYDVFLSFRGADTRNGFTDHLYNTLLDARINTFLDDEEIETGEPLK
PELESAIKSSRASIIVLSQNYASSTWCLDELVLILEQQRSFNQIVIPIFYHVEPTDVRKQQSSFGEAMWK
HKERMEAETNAEKRSEWGRKMELWGRALTQVADLKGKDAKNRKETKLIEEVVTNIHHRLGGPLSNTLPSL
TGRNHDIELITSWLTDGSSHNVDILTIVGISGIGKTSLAKYVFMLHFSKFQKSSFIEDINTRCEERHNGL
LDLQKQLHGDISKKLDLHVNVVNEYTSKIEDVLSRKKGVYSVDDIGSLHQLNALLGKKGLHPGSKIIITT
KDTSLTERCELFKTQVHPKHTQVLLDGLNKYESLTFLCIHAFQSHTPKEGYKDVSIELAKYCDGHPLALE
VLGGSLHKRDVAYWEEHIKELKKEPLLGIEKVKKALQLSFDSLSSENDKELFKHIACFFVGEDRDLTVTI
LKACDINTTSGITNLMEKCFLRVRRNNELSMHSLIQEMGRDLVRQESLNKPWERSRLWCHEESFDVLKCK
KGTQNILGLALDMRMLDKKKLRGSFELKTESLSQMNNLMLLQLNNVQLNECFPNFPEELRWLCMHGFPSK
SIHLDLPMENLVVLDISRSNIESFDMSYSDPQPPAEKQKLVGSCSNDEPLLGSLKILNLSFCEQLHSVGG
FFELPALEKLTVKRCISLMEVCESIEQCVELVHIDLSYCHKIKKLPISLGKLQKLQTLLLNGCNLSESAI
MDSSDLSISSQTSSSAIVREAIPSDFKFFMTFLPNSLRILSLENNNLSNESFPMDLSCLVMLEELCLNEN
PIVSLPICVGTLPRIQKLSIDNCYDVISVEHLPCTLRVFSMLTYIGCATRKIKFDPELPPLKLRGIPLLA
SHSSIEFEGIIKIEPMADVEEKLLHSLGWKRSDLIKGHLEFRDLQVSSNKLSLVICAHMCYEFGIFSTFY
GGKGMPEWIRCRSRGPSISFIIPSSPKKLQGLNFCCVNWPLSLSFPVPTIKISNITKKQTWIYNPYCENI
WLGGKDYISWLSHWMFGPNEMKAGDNIIIECYMEEEEEEDSYYIIDYKTECGVDVVYNDGSMEEDVLSYY
KSWNHIIGGDLSAFQLTTGEYYLCNKAFTRDSTDEYDPFIGEEGSHKGKCALSV

>RGC52SW4_aa
MVVLTDLNHRYDVFLSFRGADTRNGFTDHLYNTLLDARINTFLDDEEMETGEPLKPELESAIRSSRASII
VLSENYASSTWCLDELVLILEQKRKFNQIVIPIFYHVEPTDVRKQQSSFMKAMAQHKERMDAEADAEKRS

EWGQKMELWERALTQVANLKGVVANGRKETKLIEEVVMDIHRRLGGPLSNTLPSLTGRNHDIEFITSWLT
DGSSHSVDILTIVGISGIGKTSLAKYVFMLHFSKFQKSSFIEDINTRCEESHNGLLDLQKQLHGDISKKL
DLHVNDVNEYSSKIEDVLSRKKGVYSVDDIGSLDQLNALLGKKGLHPGSKIIITTKDTSLTERCDLFKTQ
VHPKHTQVLLDGLNKYESLKFLCIHAFQSHTPKEGYKDVSIELAKYCDGHPLALEVLGGSLHKRDVAYWE
ERIKELKKEPLSGIEKVKKALHMSFESLMCENDKELFKYIACFFVGEDRELTKTILETCGINTTSGIENL
MDKCFLRMGWNNKLTMHSLIQEMGRDLVLQESPNKPGERSRLWCHAESFDVLKRKKGTQNILGLALDMRM
LDKKKLRGSFELKTESFSQMDNLKLLQLHYVQLNECLHNFPEELRWLCMHGFPSKSFHLDLPMENLVVLD
MSYSNIESFDMSYSDPQPPAKKQKLVGSCSKEEPLLGSLKILNLSFCEQLHSVGGFFELPALEKLIVRWC
ISLMEVCESVEECVELVHIDLSYCYKLKKLPISLGKLKKVKTLSLNGCNLSESAIMDSSDLSISSQTSSS
AIIRGAMPSDFKFFMTFLPSSLRILSLANNNLSNESFPMDLSCLAMLEELCLNENPIVSLPICVGTLPRI
QKLSIDHCYDVISVEHLPRTLREFSMFTNDQYAIRKIKFDPELPPLKLGGIPDWAASHSSVEIEGIIKIQ
PMADVEEKLLHSLGWKKPDFIKGHLETHDIQMYYEFGIFSTINWLKGMPEWIRCRSWGPSISFTIPSSPK
KLRGLNFCCVNWPLFTCFPMPTIKISNITKKQTWIYNHYSGHAWVSVDCISWLSHWMFGPNEMKAGDIII
IECGSRMECKTEYGVGAVYDDEEEDVLSYYKSWNHIIGGDLSAFQLTIGEYLLYRNWFLSKLHIIPYYCP
FILLYILIHLSSGIEIKMMFLIIEFMLEYTYICPIYSKLKTGMKGKAINS

8.2.3 cDNA-AFLP sequences from isolated transcript-derived fragments

> *Bst*YI-CC/*Mse*I-AC_1
ACCGAGTTATTGATGACAAAAAAATTTCGTCACTGTATAGTTACAAGATCAGAAATGCAACACATGAACA
TTATACTCAACAACCGCATTTCCTGTTTTATTGTCAAAACTATGTGTACTTGCTTTGACATACCCTCTAG
CAAACCGGAACACCCCACTCCCTCCGATCACCGGCATCTCCCGCACCTTCTCAAACACCGCATTCCGCCC
CAAAACCGTCAACGTACTCCCGTTATATTTCCCAGTGGAAAATGCAAAGTTCATTGCCATAATCAACCCG
ACTTCCTCCTGCGATGCCAGCCCGTAAAACCCTTGTGCTCTTCCCAATAACTTTGACCCTGCTTCGGGTT
TTTCGGTCAACGGATCATCGATCATGTTGACTTGACCAAACGCGGTGGCTGCTGTTTTGTTGTTTGCTGA
TGGTGGTATGATTTTCATGGCTGTCGGGTTGGGACCGCTAAGGATATCGTGCCAGTAGAACCGGAAGTGG
CTAAGTTTTTCGGTTCGGGTCTTCATCGGGTTTAGTTTTACGGATTTTAGAAAGGCG

> *Bst*YI-CC/*Mse*I-AC_2
CAAACAATGTGAAGTCAACCGTTTGTGTCATCCCACCGACTGGTTTGAAAATGGCGTCTACCTTCATTGG
TAACTCGACGTCTATTCAAGAAATGTTTAGGCGTGTGAGTGAGCAGTTTACGGCTATGTTCAGGAGGAAG
GCGTTTTTGCATTGGTACACTGGTGAAGGAATGGATGAGATGGAGTTCACGGAGGCCGAGAGCAACATGA
ATGATCTGGTATCTGAATATCAGCAGTATCAAGATGCCACCGCTGATGAGGAGGGAGAGTATGAGGAAGA
AGAGGAGGAAGAGGGAGATTATCAGGGTTGATGATATTATATTGCACAAGGGGTATTTGTTGATGGGGTT
TTTGGTTATTGTAGGTGTATGTGTGTTTATACTCTCAACTACCATTCGATTGGATATGAAAGCAGCAGGG
GTTTGGGTTTAGTTTGTGTTTTACAAGAAAGCATCTGTTTTTGCTAAACGAGTATGTGTAATTCTGGTTT
TGGATTATGAATATTATATGTTTTTGTGGT

> *Bst*YI-CC/*Mse*I-AC_3
ACAGGTGACTCGCACGTGTCTCTATGTAGTTCAAATGCTTCCTAGGCAAATGCATGTAGACAATCCAATC
GTCTGTTTTGACTGGACTTGACATAGTCATCACATAACAGTCCCTTCGTGCCACTGGAACATAGCATGAT
CCGAACACTAGTTTTCCCCACCCAAAATCCATCTTGTCCATGATCGAAAACCGTTGACCGGATGATACCA
TGATCGAGACCGCCTTCTCGGTTCCAGCAAATGGCCGAGACATCAAAGGTGTTGGCCTTTGATCCTCAAC
CCAATCTATCAATTCCAAGAAGTGATCCTTGTTAGCTGCACTTTTCAAGAACTTGTGAACTTCGTTTGCC
ACTGAATTCAATGGCATGTCCTTCAAATCCTGTTGTAGTATATTTCAAATTGATTTTTCTTATAAGGAGC
TGAATAGTTAGCAGACTCAAAAATAGAAATAAATAAATAAAAAAAACCTAAACATAAAACAAGTAGATGT
ATATCAAAGAGCGCAATGGTCCCGAAGAGCCGAATCATGTGACAAAAGGTGATAGACAGT

> *Bst*YI-CG/*Mse*I-TA_1
TCAGCATTGCCACCCCATCCAGTCGCATCTGCAGCCGAATCTTTGTTGCCACCATCAACCTTCGAAGCGC
CCGCATCATCCTCAACAATTTCCCCTTCTTCAATCTTCACAACAGTCGACGAAAGCCTCCGTTTGGGGGC
GGGTTGGTCTTCTTCGTTCCTATCCGCAGGTCTAACAAAACCACGTTGACGG

> *Bst*YI-TC/*Mse*I-TC_1
TCCTCCAAGCTACCAACAAATGAACTATACCATTTGATCCCTATATCGGCCAAAAATTGTACTAAAATAC
ATCATACAGGGATTTTACCACGAATCCATTGATTCTATATGATTTATTATATTACTACCCCTTTACCACT
CTATTCGTACATGGAGTCGGGGGAGAATGGAATTTTGTTTTTTTTTTTTTTTTTTTTTTTTTTACAAA
TGGGCATGCTGGATCATATATCTATCCATATATCGGGATAGATAGATTGGCCATCACACCCGA

> *Bst*YI-TC/*Mse*I-TC_2
TCTCCATTCGACAAGAAATCGCGGAGATTGAGAAAGGAACAGTTGATCTCAACAACAATGTCATCAAGGG
CGCTCCTCACCCGCCACAACTACTCATGGCTGACAAGTGGACAAAACCGTATTCTAGAGAATACGCAGCG
CACCCTGCTCCCTGGCTCCATGCTGCTAAGTTCTGGCCTACTACATGCCGTGTGGACAATGTGTATGGAG
ACCGCAACCTCATCTGTACTCTTCAGCGACCAGAGGAGTATGAAGAGAAAGCGGAAGCTACTGA

> *Bst*YI-TT/*Mse*I-TA_1
TCACCACCTTCCTGTTTTGAAGTTTTTGTTTGGAACTAAGGCGAAGAGGTTGTTCGACGCACGGCCGCAT
GTCAGCGCGTGGGCTGCCGATATCCAGTCTAGGCCTGCATGGGTGAAAGCAATCACGGCTTGAGTGATTT
TGTGGTGTGTGTGTTTGTGGCTTCATGTA

> *Bst*YI-TT/*Mse*I-TA_2
TCACCACCTTCCAATTCTGAAGTTTTTGTTTGGAACAAAAGCGAAGAGGCTGTTTGACGCAAGGCCGCAT
GTCAGCGCCTGGGCTGCCGAGATACAGTCCAGGCCTGCATGGGTGAAAGCAACCACGGCTTGATTTCATA
GTGTGTATTTGAGGCTTTCACGTA

8.2.4 Bulked segregant transcriptome analysis derived resistance gene candidate sequences

>BRBScontig00945 length=894 numreads=7
TCGGTAACCTGCGCATGCTGCGCAGACTAAGCCTCCGGTCAAACAGCTTCAACGGTAGTATTCCGGTATC
TCTGTCAAAATGTGAACTCTTACAATATGTTTTCCTTCAGTATAACTCTTTTTCCGGCGAGATACCGCCG
GAGATTTCAAATCTGAAACAGATTTTAATATTTAACGTCGCCGGAAACAATATCTCCGGCGAAGTTCCGA
AAACTTTTCCGTCAAGTCTCCGGTACCTTGATCTGTCCTCAAACTCACTCTCCGGCGAGTTACCGAACAA
CCTCTCCGACTCACCGGATCTGCAACTCGTAAACTTTTCATACAATCATCTCACCGGAGTGATTCCGGAG
AGGTATGGCGGCTTAAAAAACCTTCAGTACCTTTTACTTGATTACAATAACTTGCAAGGTACTTTACCGT
CGGCGTTAGCGAACTGTGTTTCACTGCTGCATTTTAGCGCCGACGGTAACGCAATCGGAGGTGTTATTCC
GGCCGGAATAGCTTCGCTTCCTAAGCTACAAGTTATTTCACTCTCGAATAATAATCTCACCGGTTTAATT
CCGGCGTCTATATTTTGTAATGTTTCCGTTGATAAACCGTCGATACGGAATGTTCAGTTAGGGTTTAATG
CGTTTACGGATATTATTCCTCCTGTGTCTGGTAAATGTTTTAGTGTTTTGCAAGTTTTAAATCTTTCAAC
AGAATCAGATTCATGGTGATTTTCCGTTGTGGTTGACTGATTTAGCTACATTAACGGCTTTAGATGTTTC
TGGTAATCGTTTGTCTGGTGAATTCCTGGTGAGATTGGGTGGTTGAAGCAGTTGGAGGAGTTGAAGTTGG
CGAATAATTCGTTATCCGGTTTGTTTCCTAGTGAGATTAAGGAACTGGTTAACC

>BRBScontig02834 length=762 numreads=10
ATCTATCAAACATGAATCTCACCGGAAATATCTCCGACGACTTTCAAAACTTGTATCATCTTTCCTTTCT
AAACATTTCGTCCAACGGGTTTTCGTCTACGCTGCCGAAGTCGTTTGCGAATCTGACGTCAATCGAAACG
ATTGACGTCAGTCAAAACGAGTTCGTCGGGGGAGTTCCCGGCGGGGGTTTGGAACGGGAGCTACCCGGCTG
AAGTCTGTTAATGCTTCCAGCAACAATTTCGCCGGGTACCTTCCGGAGTCCCTCGGAAGTGCCACGTCAC
TCGAGACGATTGATTTCCGAGGGAGTTTCTTTGTGGGATCAATCCCGAAGAGTTTCAAGAACCTTCAAAG
TTTAAAGTTTCTTGGACTCTCCGGGAACAATCTCACTGGACCGATCCCATCAGAGATCGGTCAGCTTTTG
TCTCTAGAGGTTATGATCATTGGTTACAATGAGTTCGAAGGCCCAATCCCAAAAGAGATTGGGAACCTCA
AAAATCTAGAGTATCTAGATTTGGCCGTCGGGACACTCAGTGGAAGTATCCCGACGGAACTGGGTGAGTT
AAAAAAACTCAGCACAATTTACATTTACCAGAACAGTTTTGATGGAGAGATTCCCGCCGAAATCGGGAAT
CTCTCCTCGCTGGTTTATTTGGATCTCTCAGATAACCAGTTTACCGGAGAGATCCCGAAACAAATAGGCA
ACTTGAAGAGCTTAAAGCTACTAAATCTGATGTGCAATCAGCTCAACGGTTCGATTCCGACT

>BRBScontig02873 length=1171 numreads=28
ATACACACCCCATCTCAAATTATCATGAAGGGAATGTTTGAGATGATTTCCTTCATGATTATATTGCTCC
AATTGCTATATACAACTTCTGCAATTTTAGACCCGGTAGATTTTCTAGCATTACAATCGATTCGAAAATC
GCTAGAAGATCTCCCCGGGTCGAATTACTTTTCATCATGGGATTTTACCTCAGACCCATGTAACTTCGCT
GGAGTTTACTGCGAAGGTGGAAAAGTGATCGTGTTGAACCTCGGTGATCCGCGAGCAGGCGCCCCTGGTC
TTTCGGGGCGCCTGCATCCGGATATCGGAAAGTTGAGTTCCCTCACGGAGTTTACAGTTGTTCCCGGGAG
GATAATGGGGGCATTGCCAGTAACACTGTCGAAGTTAAAAAACCTCCGGTTTTTAGCGGTTAGCCGGAAT
TTTATCTCGCAAGCAATACCAGCGGAGTTAGGGGAGCTTAAGCAGTTGAGAACCCTTGATCTGAGTTACA
ACCAGCTCACTGGTAGTATACCATGGGCTATTGGAAATTTGCCAGAGTTAACTAACGTTATTCTTTGTCA
TAACCATTTATCGGTATCTATCCCTCCGTTCATCTCGCAGAGTCTAACGCGGTTAGATCTGAAGCACAAC
GAGTTATCAGGCATCATAGAGCGGAACTCACTACCATCATCGCTACAATATCTATCTCTATCCTGGAACC
AACTTACGGGTCCGGTGGACCAGCTCTTGCCCAGGCTCAATCAAACTAAACTATCTTGACCTGAGCTTGA
ACCAGTTCACCGGACCCATCCCGGGTTGCGTGTTTGGTTTCCAAATCACCAACCTGCAACTTGAAAGAAA

CCTGTTTTCTGGACCGGTCCAACCGGAAGGTTCGGTAACGATCCCCACGTTGATCTTAGTCACAACATGT
TGTACGGTCAAGTGTCGCCGCTTTTTGCGACGGTTAGAATCTTTACTTGAATAATAACCGGTTCACCGGT
TCGGTTCCGACGGTTTTTGTGGACCGGTTAACGGCTGGTGATATACGATTGTTGTATTTGCAACATAATT
TTTTAACCGGTATGCCGATTAACCCCAGGACTGAAATACCTCTGAGCAGTTCGTTGTGTTTGCAGTACAA
TTGTATGGTCCTACCGGTTCAGACTCCATGCCCGATTAATGCGGGTACACA

>BRBScontig15072 length=988 numreads=27
AATATAAAAGAGCTTCACCCTCCATTTCAAATCCACAACCTTTATCATCATAGTCACTAGATATGACAAT
GCTACTTCTTTTACTTCTTCTGGCTCCGGTTGTTGTTACCGCGGAGCGTTGTCACCCAGACGACAAAGCC
GCCATCTTAAAAATAAAAGATAGCTTCTCAAACGGGCGTGAGCTCTTGCCAATGTGGACCGCAGATCGTG
ATTGTTGCAACACATTTGATTGTGACGAGACCACAAACCGTGTCATAGACTTTTCCCTCACCAATTCTAA
CCTCGCGGGTCCAATTCCGGAAGCAATCGGGGATTTGACATATCTCAAAACCCTTCGTCTTCACAAGATG
CCGTTTCTTATTGGTCAAATTCCGCGTAGCCTTACAAGACTCAAACATGTAACATTTTTCGATATAAGTT
GGACCAACGTTTCGGGCCACGTGCCCTCGTTTCTTTCTGAACTCAAAAGTGTCATGATTCTTGATCTTTC
GTTCAACAACCTATCGGGCCCAATCCCACCCTCACTAGCCACACTACCGAATATGATCGGGCTAGATCTT
AGCCGAAACCAGCTCACAGGGTCAATCCCCGAATCATTCGGTCACCTTGCGTCTCCGGCTTTACAAGGAT
TAGGTTTATCTCATAACATGCTGTCTGGTGAGATACCTACATCCATTGGCAACATGCAAATATATAGGAT
TGATGTATCAAGAAACAACTTGTCCGGGGATGCATCAATGTTGTTTGGCTCATCAAAGAATACAACCGAG
ATTGACATTTCGCGAAACAATTTCGCTTTTGATTTGTCACTGGTGACTTTCATGGTTGACACATTGGTAA
CGTTGGATATATCACATAACAAGATATACGGGGAAATTCCGTCTCAAATATTGGAAGCTTACATGTTACA
AAGTCTTAACGTTAGTTACAATCAGTTATGTGGCACGATTCCTTCACCGTGGAAACTCAAGTATCGATCA
GAAGGCTT

>BRBScontig16592 length=568 numreads=17
TGGATAGCTTTCGTTCCGTTCATCCTTCTACTTGCCGTGGGAACAAAGCTAGAGCACATAATAATTCAGT
TAGCTCATGAGGTTGCTGAGAAACATATAGCCATAGAAGGCGATTTAGTTGTGCAACCATCGGATGATCA
TTTTTGGTTTCACCGCCCCAAAATCGTGCTTTTTCTCATACATTTCATCCTCTTTCAAAATGCATTTGAG
ATCGCGTTCTTCTTTTGGATATGGGTTCAATATGGCTTTGACTCATGCATAATGGGACAAGTGCGCTACA
TTATCCCTCGGCTCATTATTGGTGTGTTTATTCAGGTTCTGTGCAGTTACAGCACTCTACCACTTTACGC
CCTTGTGACACAGATGGGGACCAATTTCAAGAAAGCAATATTTGAAGACCATATACAGGCCAAAGTTGTG
GGCTGGGCTCTAAACGTCAAGAAAAAGATGGCCAATAACGGTGGGACCAATGGCTCGAGCCATGATGGTT
CAACTGCCACCACTTCTGCCGCTACTTCTGCTGCCACATCTGCTTTTAAGACATTCAACTTGGTAAAATT
AATCAGAG

>BRBScontig07153 length=861 numreads=10
ATATGATAACAAGTTTACTGGTGAGATTCCATCCGGAATCTGGACATTATCGAATATTTCGAGTTTGAGG
TTATCTGGAAACTCATTTTCCGGTGAGCTTCCGAGTAAAGTTGCTTGGAATTTGTCTAGATTGGAGATTA
GCGACAACAAGTTTTCGGGTCGAATTCCAGAAGGGATATCTTCTTGGGGGAAACTAAATGTGTTTAAAGC
TAGTAACAATATGTTTTCTGGTGAAATTCCAGTTGATATCACTAGTTTATCCGAGTTAACCGTGCTTTAT
CTCGATGGGAATTCGCTTTCGGGTGAGCTTCCGTCGGAGTTTAAGTCGTGGAACTCGCTAACGGTTTTGA
ATCTTTCCAGAAACAAATTGTCTGGTCAGATTCCGAGTTCACTTGGTTCTTTACAAGATCTTCTTGAACT
TGATCTATCCGAAAACCAACTTTCGGGTCAGATTCCATCACAATTGGGCACTCTACGGCTTACGAGTTTG
AATCTTTCTTCGAACAAGCTCACCGGAAGAATCCCGTCCTCTTTTGATAACATGGCTTATCAAAACAGTT
TTCTTAACAACCCTGACATCTGTTCTGATTCCCATATCTCCAACCTCCAAAACTGTCATCCCAAAAACTC
ACACTCTAAAAAATTCTCACCAAAAATCATAGCCATGATCGTTGTTTTATCCGTATTTGTTCTACTCGCG
GCTGTACTGTTTGCGTTGTTTGTGTTCCGAGACTACCAAAAGAGGAAGCCGAAACGAATTCGCACTACAT
GGAAACTAACCTCATTCCACAAGATGGATTTCACAGAAAGAAACATTTTGTCATGCATGACTGAAGCCAA
TGTGATCGGTACTGGTGGTGC

>contig30846 length=782 numreads=5
AGTAACCTTCCAATGTTGAAGAAGTTGTATTTGGATGGAAATCCAATGGATTCCATGCCCGATTGTGTGA
AAAGCCTTAGTAGGCTGGAGACACTTAGTTTTTGTTGGTGTCGGAACCTGAAAACAGTCTTGTGTGCTCC
AATCCAATTAAAGCAGTTGGATATACTCTTTTGCGACTCATTGGAGAAAACAACATTTCATCCTGAAAAG
TCAGCTATACCTAGAGTTCTTTGTAACAGAAGTGTTACCCTGACTGAGATTCAACACATACTCAAAATAC
AAGCTTTATCAGAAATTGATGAAGAGGTATTATGCAGTTTGGGTTGGATTAATATAGCATATTTAAACCA
CTGCCGGTTTTCGAAGGCCTATGATTTTGGAATATACCATTTGGCAGGAAGAATACTTCCAGCACAGATG
CTTTATGAACATGGAATATTCAGCACATATCTTCAAGGGCAAGAGGTTCCAGAGTGGTTCGCTCAAAGAA
GCAACGGGTCATTCTTTACCTTGCAATCATCTCCTGAAAAAGGTAAGATTAAAGGGCTAAAATGTGTGTA
TCGTGAATACAGTTTCAAGCATGAAGGAAGTTGGCGTTTTTAAAGAATTAAAATAAGGAATCTGACAAAG
AACTCCTCCTGGACATACGAGCCTATTATGTATCTTGTTCCAGAAGATGATGCGTTTGAAGATGGTGATG
AAGTAGTTGCAGTATGGTTAAGCCATTGGATGTTTGGGAAAAATGAGCTTGAAGATGGTGATGAAGTTAG

CATTCAATTTAG

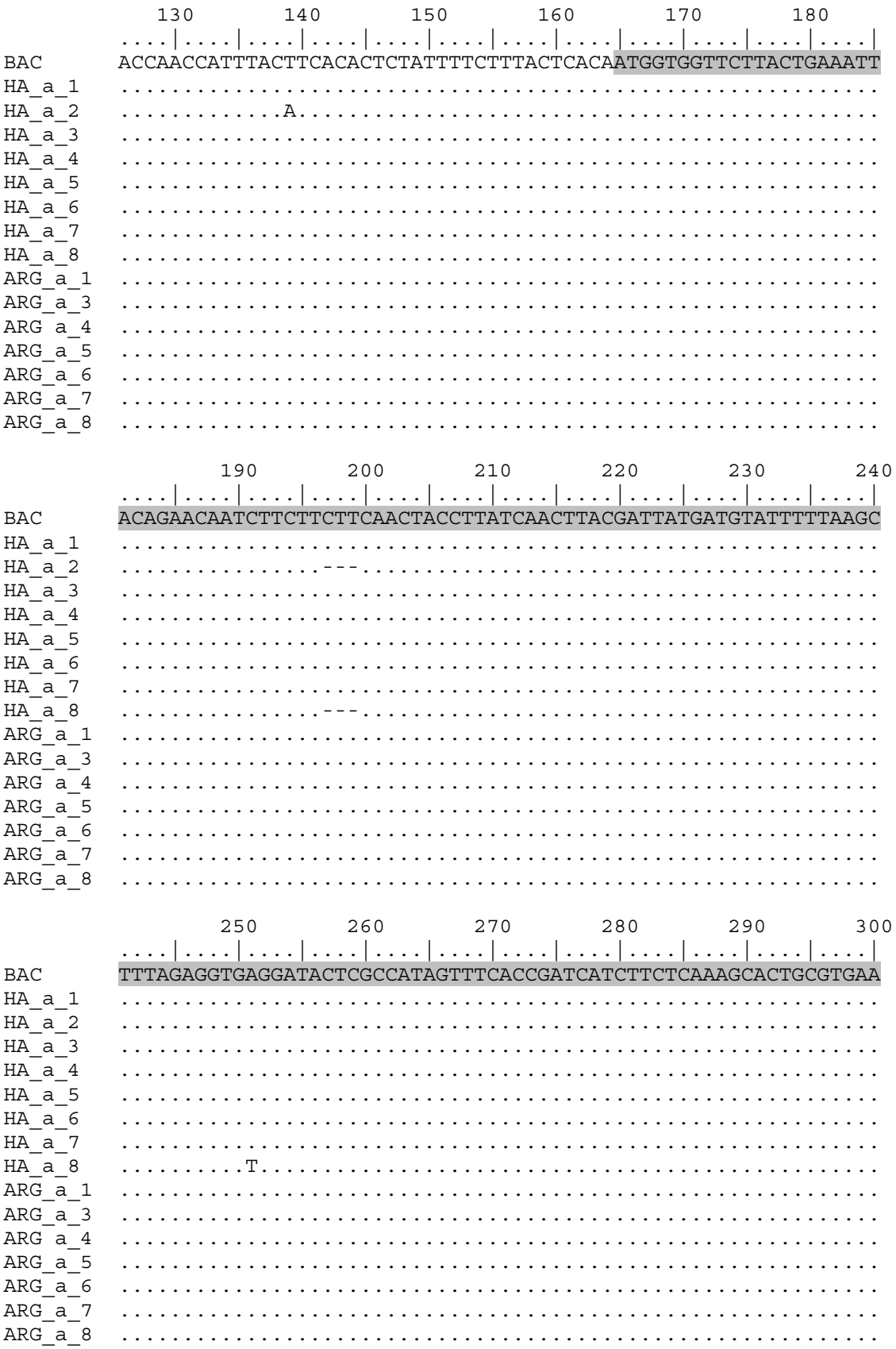
>BRBScontig34117 length=797 numreads=6
AAATGATCCAAACAGTCTCAGTAGTTGGAGACGTGGAGATCCGTGTGTATCAAATTGGACCGGGGTCTTG
TGCTTCAACACAACCCGCGATGACGGTTTTCTTCATGTCCGAGAACTGCAGCTACTGAATTTGAATCTTG
AAGGAACTTTATCTCCCGCTTTAGCGCAATTATCTTATATGGAAATATTGGATGTCATGTGGAACAATAT
AACCGGGAGTATACCAAGAGAAATCGGCCAACTTACAAATTTGAGACTTCTGTTGTTAAACGGAAACGAG
TTAACCGGTTCATTGCCAGACGAGCTCGGTTACCTCCCGAATTTGGACCGAATACAAATCGATCAAAACC
ACATATCTGGACCGATACCGAAGTCATTTGCGAATTTGAACAAGACAATGCACTTCCATATGAATAATAA
TTCGCTTAGTGGGCCGATCCCTCCTGAGATATCTAGATTGCCAATGCTTGTTCACATGCTGCTTGACAAT
AACAATTTATCGGGCCCGCTTCCTCCAGAGCTCAGTGAACTGCCAAGACTTCTTATACTCCAACTCGACA
ACAACCACTTCGATGGAACTATACCTCCCTCTTATGGCAACATGACTCGGTTGTTAAAGTTGAGTTTTGA
GGAATTGCACCTTGCAAGGACCGATTCCCGATCTTAGCAGGATACCAAGTCTTGCTTATATAGACCTCAG
TAATAACAATCTCAATGGATCCATACCCGACATGAATCTTTCGGATAACATCACAACAATAGATCTGTCA
AGGAATCAGCTTACGGGAACTATTCCT

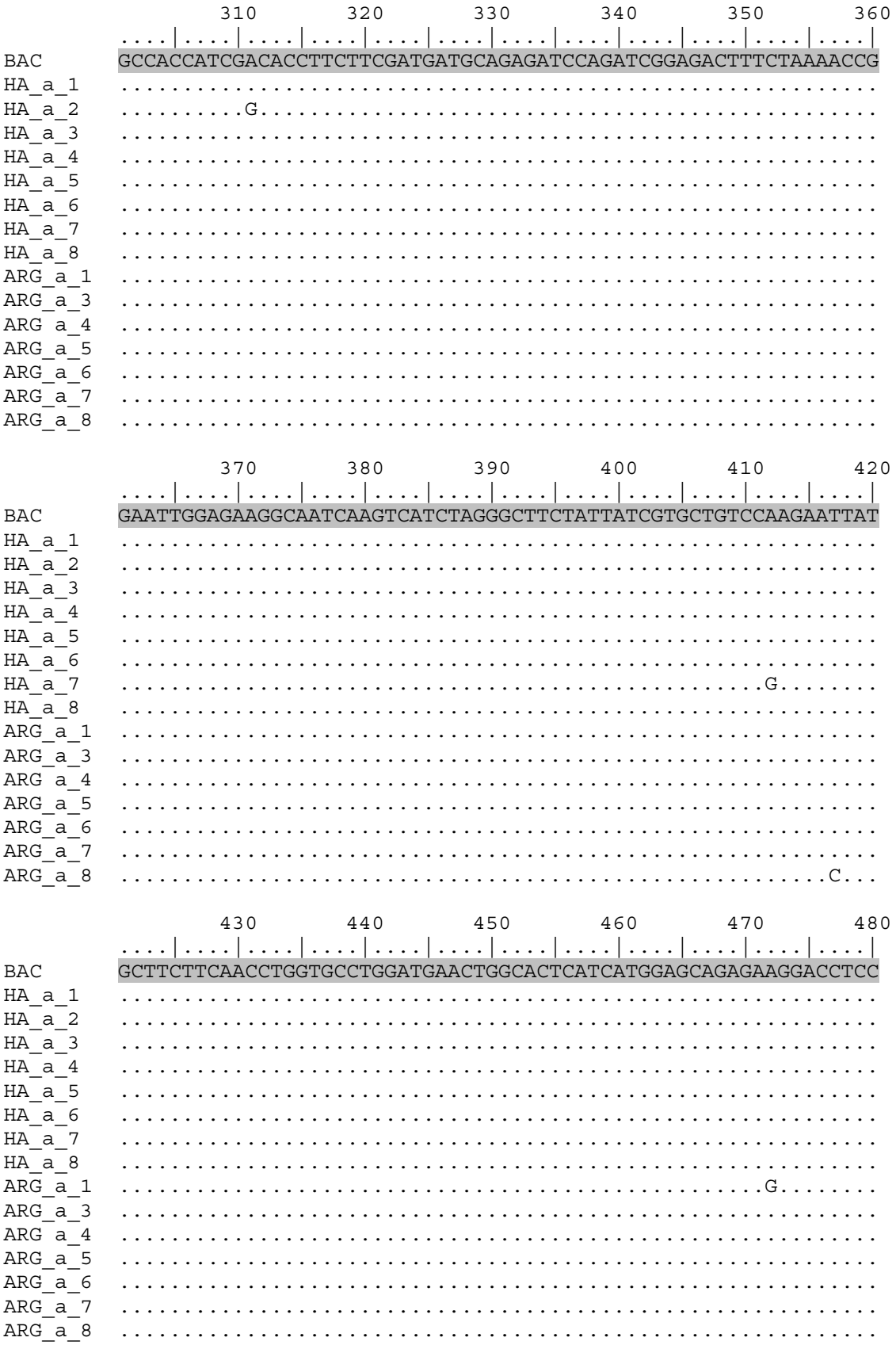
8.3 Sequence alignments

Alignment 8-1: Nucleotide alignment (Clustal W) of CG1_a for the BAC sequence Co3_c2 (44,038 bp – 44,931 bp) and sequences cloned from the susceptible line HA342 (HA) and the resistant line ARG1575-2 (ARG). The first predicted exon of resistance gene candidate TIR-NBS-LRR identified on HICF contig03_c2 is highlighted in grey on the BAC sequence. Dots represent conserved bases, letters and dashes indicate polymorphic sites, and deletion events, respectively. Primer sequences which were used to amplify the fragment are underlined.

```
                10        20        30        40        50        60
          ....|....|....|....|....|....|....|....|....|....|....|....|
BAC       GATATTTTCAAGGGGCTGTTGTATAAAATAGAAGGGAAAGTTTCCAATGAATAATATAAA
HA_a_1    ............................................................
HA_a_2    ............................................................
HA_a_3    ............................................................
HA_a_4    ............................................................
HA_a_5    ............................................................
HA_a_6    ............................................................
HA_a_7    ..........................................................C.
HA_a_8    ............................................................
ARG_a_1   ............................................................
ARG_a_3   ............................................................
ARG a_4   ............................................................
ARG_a_5   ..........................................-.................
ARG_a_6   ............................................................
ARG_a_7   ............................................................
ARG_a_8   ............................................................

                70        80        90       100       110       120
          ....|....|....|....|....|....|....|....|....|....|....|....|
BAC       AGATGACTATGACTTCGACGCCAGGGAGTCAACTTCATCTCCGATCATTTCATAACCAAA
HA_a_1    ............................................................
HA_a_2    ............................................................
HA_a_3    ............................................................
HA_a_4    ............................................................
HA_a_5    ............................................................
HA_a_6    ............................................................
HA_a_7    ............................................................
HA_a_8    ............................................................
ARG_a_1   ..................................G.........................
ARG_a_3   ..................................G.........................
ARG a_4   ..................................G.........................
ARG_a_5   ..................................G.........................
ARG_a_6   ..................................G.........................
ARG_a_7   ..................................G.........................
ARG_a_8   ..................................G.........................
```


130 140 150 160 170 180
BAC ACCAACCATTTACTTCACACTCTATTTTCTTTACTCACAATGGTGGTTCTTACTGAAATT
HA_a_1 .....
HA_a_2 .....A.....
HA_a_3 .....
HA_a_4 .....
HA_a_5 .....
HA_a_6 .....
HA_a_7 .....
HA_a_8 .....
ARG_a_1 .....
ARG_a_3 .....
ARG a_4 .....
ARG_a_5 .....
ARG_a_6 .....
ARG_a_7 .....
ARG_a_8 .....
190 200 210 220 230 240
BAC ACAGAACAATCTTCTTCTTCAACTACCTTATCAACTTACGATTATGATGTATTTTTAAGC
HA_a_1 .....
HA_a_2 .....---.....
HA_a_3 .....
HA_a_4 .....
HA_a_5 .....
HA_a_6 .....
HA_a_7 .....
HA_a_8 .....---.....
ARG_a_1 .....
ARG_a_3 .....
ARG a_4 .....
ARG_a_5 .....
ARG_a_6 .....
ARG_a_7 .....
ARG_a_8 .....
250 260 270 280 290 300
BAC TTTAGAGGTGAGGATACTCGCCATAGTTTCACCGATCATCTTCTCAAAGCACTGCGTGAA
HA_a_1 .....
HA_a_2 .....
HA_a_3 .....
HA_a_4 .....
HA_a_5 .....
HA_a_6 .....
HA_a_7 .....
HA_a_8 .....T.....
ARG_a_1 .....
ARG_a_3 .....
ARG a_4 .....
ARG_a_5 .....
ARG_a_6 .....
ARG_a_7 .....
ARG_a_8 .....


310 320 330 340 350 360
BAC GCCACCATCGACACCTTCTTCGATGATGCAGAGATCCAGATCGGAGACTTTCTAAAACCG
HA_a_1 .....
HA_a_2 ....G.....
HA_a_3 .....
HA_a_4 .....
HA_a_5 .....
HA_a_6 .....
HA_a_7 .....
HA_a_8 .....
ARG_a_1 .....
ARG_a_3 .....
ARG a_4 .....
ARG_a_5 .....
ARG_a_6 .....
ARG_a_7 .....
ARG_a_8 .....
370 380 390 400 410 420
BAC GAATTGGAGAAGGCAATCAAGTCATCTAGGGCTTCTATTATCGTGCTGTCCAAGAATTAT
HA_a_1 .....
HA_a_2 .....
HA_a_3 .....
HA_a_4 .....
HA_a_5 .....
HA_a_6 .....
HA_a_7 .....G.....
HA_a_8 .....
ARG_a_1 .....
ARG_a_3 .....
ARG a_4 .....
ARG_a_5 .....
ARG_a_6 .....
ARG_a_7 .....
ARG_a_8 .....C...
430 440 450 460 470 480
BAC GCTTCTTCAACCTGGTGCCTGGATGAACTGGCACTCATCATGGAGCAGAGAAGGACCTCC
HA_a_1 .....
HA_a_2 .....
HA_a_3 .....
HA_a_4 .....
HA_a_5 .....
HA_a_6 .....
HA_a_7 .....
HA_a_8 .....
ARG_a_1 .....G.....
ARG_a_3 .....
ARG a_4 .....
ARG_a_5 .....
ARG_a_6 .....
ARG_a_7 .....
ARG_a_8 .....

```
                490       500       510       520       530       540
           ....|....|....|....|....|....|....|....|....|....|....|....|
BAC        GGACATATTGTTTTTCCGATCTTCTATCATGTTAAGCCTTCAGATGTCAGGAAGCAACGA
HA_a_1     ............................................................
HA_a_2     ............................................................
HA_a_3     ............................................................
HA_a_4     ................G...........................................
HA_a_5     ............................................................
HA_a_6     ............................................................
HA_a_7     ............................................................
HA_a_8     ..........................................A.................
ARG_a_1    ............C...............................................
ARG_a_3    ............C...............................................G
ARG a_4    ............C...............................................
ARG_a_5    ............C...............................................
ARG_a_6    ............C...............................................
ARG_a_7    ............C...............................................
ARG_a_8    ............C...............................................

                550       560       570       580       590       600
           ....|....|....|....|....|....|....|....|....|....|....|....|
BAC        AACAGTTTTGGAGATGCAATGGCAGAGCATAAACAGAGGATGGAAACAGAGTTAAATTCA
HA_a_1     ............................................................
HA_a_2     ............................................................
HA_a_3     ............................................................
HA_a_4     ............................................................
HA_a_5     ............................................................
HA_a_6     ............................................................
HA_a_7     ............................................................
HA_a_8     ............................................................
ARG_a_1    ............................................................
ARG_a_3    ............................................................
ARG a_4    ............................................................
ARG_a_5    ............................................................
ARG_a_6    ............................................................
ARG_a_7    ............................................................
ARG_a_8    ............................................................

                610       620       630       640       650       660
           ....|....|....|....|....|....|....|....|....|....|....|....|
BAC        GAGAAAAGAAGTGAATGGGCTCACAAGATAGAGAAATGGAAGAAAGCCCTAACAGAAGTT
HA_a_1     ............................................................
HA_a_2     ............................................................
HA_a_3     ............................................................
HA_a_4     ............................................................
HA_a_5     ........T...................................................
HA_a_6     ............................................................
HA_a_7     ............................................................
HA_a_8     ............................................................
ARG_a_1    ..........................................................G.
ARG_a_3    ............................................................
ARG a_4    ............................................................
ARG_a_5    ............................................................
ARG_a_6    ............................................................
ARG_a_7    ............................................................
ARG_a_8    ............................................................
```

```
                670       680       690       700       710       720
          ....|....|....|....|....|....|....|....|....|....|....|....|
BAC       GCTGATATGAAAGGGGAGGAAGCCAATGACAGGTAATTTCATTGCTTCTGAGATGGAACC
HA_a_1    ............A................A............................
HA_a_2    ....................................................................
HA_a_3    ....................................................................
HA_a_4    ....................................................................
HA_a_5    ....................................................................
HA_a_6    ....................................................................
HA_a_7    ....................................................................
HA_a_8    ....................................................................
ARG_a_1   ..........................G.............................C...
ARG_a_3   .......A..................G.............................C...
ARG a_4   ..........................G.............................C...
ARG_a_5   ..........................G.............................C...
ARG_a_6   ..........................G.............................C...
ARG_a_7   ..........................G.............................C...
ARG_a_8   ..........................G.............................C...

                730       740       750       760       770       780
          ....|....|....|....|....|....|....|....|....|....|....|....|
BAC       CTAGATTT------------TGAGAGTTGTTTTTTTTTTCCGAATCAAAGAGCCTGTGAA
HA_a_1    ........------------........................................
HA_a_2    ........------------........................................
HA_a_3    ........------------........................................
HA_a_4    ........------------........................................
HA_a_5    ........------------........................................
HA_a_6    ........------------........................................
HA_a_7    ........------------........................................
HA_a_8    ........------------........................................
ARG_a_1   ........GGTGCATAATTT..C.....................................
ARG_a_3   ........GGTGCATAATTT..C.....................................
ARG a_4   ........GGTGCATAATTT..C.....................................
ARG_a_5   ........GGTGCATAATTT..C.....................................
ARG_a_6   ........GGTGCATAATTT..C.....................................
ARG_a_7   ........GGTGCATAATTT..C.....................................
ARG_a_8   ........GGTGCATAATTT..C............--.......................

                790       800       810       820       830       840
          ....|....|....|....|....|....|....|....|....|....|....|....|
BAC       AATGCCTTTCCATGTTCAAATTGTGATTAGAAAATATCTAAATACTCATGAATTGTTTGA
HA_a_1    ....................................................................
HA_a_2    ....................................................................
HA_a_3    ....................................................................
HA_a_4    ....................................................................
HA_a_5    ....................................................................
HA_a_6    ....................................................................
HA_a_7    ....................................................................
HA_a_8    ....................................................................
ARG_a_1   ..........................T..............--.....G.....G.T
ARG_a_3   ..........................T..............--.....G.....G.T
ARG a_4   ..........................T..............--.....G.....G.T
ARG_a_5   ..........................T..............--.....G.....G.T
ARG_a_6   ..........................T..............--.....G.....G.T
ARG_a_7   ..........................T..............--.....G.....G.T
ARG_a_8   ..........................T..............--.....G.....G.T
```

```
                 850       860       870       880       890       900
          ....|....|....|....|....|....|....|....|....|....|....|....|
BAC       TTAT-------------------GCTATG-ATGAGTTTATATTAATCACTTTACTTGTAA
HA_a_1    ....--------------------......-............................
HA_a_2    ....--------------------......-............................
HA_a_3    ....--------------------......-............................
HA_a_4    ....--------------------......-............................
HA_a_5    ....--------------------......-............................
HA_a_6    ....--------------------......-............................
HA_a_7    ....--------------------......-............................
HA_a_8    ....--------------------......-............................
ARG_a_1   .C..TTGTGAGAAAGTCAAGTAA.T..CAC....A.....T..T..G.G.......AA..
ARG_a_3   .C..TTGTGAGAGAGTCAAGTAA.T..CAC....A.....T..T..G.G.......AA..
ARG a_4   .C..TTGTGAGAAAGTCAAGTAA.T..CAC....A.....T..T..G.G.......AA..
ARG_a_5   .C..TTGTGAGAAAGTCAAGTAA.T..CAC....A.....T..T..G.G.......AA..
ARG_a_6   .C..TTGTGAGAAAGTCAAGTAA.T..CAC....A.....T..T..G.G.......AA..
ARG_a_7   .C..TTGTGAGAAAGTCAAGTAA.T..CAC....A.....T..T..G.G.......AA..
ARG_a_8   .C..TTGTGAGAAAGTCAAGTAA.T..CAC....A.....T..T..G.G.......AA..

                 910       920
          ....|....|....|....|....|..
BAC       CTTCTTGTTGTGATTGGAATCGGTTCC
HA_a_1    ...........................
HA_a_2    ...........................
HA_a_3    ...........................
HA_a_4    ...........................
HA_a_5    ...........................
HA_a_6    ...........................
HA_a_7    ...........................
HA_a_8    ...........................
ARG_a_1   ...T.......................
ARG_a_3   ...T.......................
ARG a_4   ...T.......................
ARG_a_5   ...T.......................
ARG_a_6   ...T.......................
ARG_a_7   ...T.......................
ARG_a_8   ...T.......................
```

Alignment 8-2: Nucleotide alignment (Clustal W) of RGC151 and Co3-2_M13 for the BAC sequence Co3_c2 (46,015 bp – 46,719 bp), for sequences cloned from the susceptible line HA342 (HA) and the resistant line ARG1575-2 (ARG). Predicted exons of resistance gene candidate TIR-NBS-LRR identified on HICF contig03_c2 are highlighted in grey on the BAC sequence. Dots represent conserved bases, letters and dashes indicate polymorphic sites, and deletion events, respectively. Primer sequences which were used to amplify the fragment are underlined.

```
                          10        20        30        40        50
                    ....|....|....|....|....|....|....|....|....|....|
BAC                 TGGCACCATGAGGAGTGTTTGGATGTGTTGCAAAATAGACAGGTAATCAT
HA_RGC151_1         --------------------------------------------------
HA_RGC151_2         --------------------------------------------------
HA_RGC151_7         --------------------------------------------------
HA_RGC151_3         --------------------------------------------------
HA_RGC151_4         --------------------------------------------------
HA_RGC151_5         --------------------------------------------------
HA_RGC151_6         --------------------------------------------------
HA_RGC151_8         --------------------------------------------------
ARG_RGC151_1        --------------------------------------------------
ARG_RGC151_2        --------------------------------------------------
ARG_RGC151_3        --------------------------------------------------
ARG_RGC151_4        --------------------------------------------------
ARG_RGC151_5        --------------------------------------------------
ARG_RGC151_6        --------------------------------------------------
ARG_RGC151_8        --------------------------------------------------
ARG_Co3-2_M13_3-4   ..................................................-----
ARG_Co3-2_M13_3-6   ..................................................-----
ARG_Co3-2_M13_3-7   ..................................................-----
ARG_Co3-2_M13_3-8   ..................................................-----
ARG_Co3-2_M13_4-4   .....................C............................-----
ARG_Co3-2_M13_4-8   ..................................................-----

                          60        70        80        90       100
                    ....|....|....|....|....|....|....|....|....|....|
BAC                 GTATACCTTTTCTGTGTACTTTGATTAATTTGTATCATGTATAGAC----
HA_RGC151_1         --------------------------------------------------
HA_RGC151_2         --------------------------------------------------
HA_RGC151_7         --------------------------------------------------
HA_RGC151_3         --------------------------------------------------
HA_RGC151_4         --------------------------------------------------
HA_RGC151_5         --------------------------------------------------
HA_RGC151_6         --------------------------------------------------
HA_RGC151_8         --------------------------------------------------
ARG_RGC151_1        --------------------------------------------------
ARG_RGC151_2        --------------------------------------------------
ARG_RGC151_3        --------------------------------------------------
ARG_RGC151_4        --------------------------------------------------
ARG_RGC151_5        --------------------------------------------------
ARG_RGC151_6        --------------------------------------------------
ARG_RGC151_8        --------------------------------------------------
ARG_Co3-2_M13_3-4   -----...A..AG.....A.--A.................G..T..ACAT
ARG_Co3-2_M13_3-6   -----...A..AG.....A.--A.................G..T..ACAT
ARG_Co3-2_M13_3-7   -----...A..AG.....A.--A.................G..T..ACAT
ARG_Co3-2_M13_3-8   -----...A..AG.....A.--A.................G..T..ACAT
ARG_Co3-2_M13_4-4   -----...A..AG.....A.--A.................G..T..ACAT
ARG_Co3-2_M13_4-8   -----...A..AG.....A.--A.................G..T..ACAT
```

```
                          110       120       130       140       150
                    ....|....|....|....|....|....|....|....|....|....|
BAC                 AACATTGTTAACTATTTGTTATTTCTTAGGGAACAACAACAACTCAAGGA
HA_RGC151_1         ---------------------------------------------------
HA_RGC151_2         ---------------------------------------------------
HA_RGC151_7         ---------------------------------------------------
HA_RGC151_3         ---------------------------------------------------
HA_RGC151_4         ---------------------------------------------------
HA_RGC151_5         ---------------------------------------------------
HA_RGC151_6         ---------------------------------------------------
HA_RGC151_8         ---------------------------------------------------
ARG_RGC151_1        ---------------------------------------------------
ARG_RGC151_2        ---------------------------------------------------
ARG_RGC151_3        ---------------------------------------------------
ARG_RGC151_4        ---------------------------------------------------
ARG_RGC151_5        ---------------------------------------------------
ARG_RGC151_6        ---------------------------------------------------
ARG_RGC151_8        ---------------------------------------------------
ARG_Co3-2_M13_3-4   ...T.CT....TC.G..A..-..................AT.T.....C.
ARG_Co3-2_M13_3-6   ...T.CT....TC.G..A..-..................AT.T.....C.
ARG_Co3-2_M13_3-7   ...T.CT....TC.G..A..-..................AT.T.....C.
ARG_Co3-2_M13_3-8   ...T.CT....TC.G..A..-..................AT.T.....C.
ARG_Co3-2_M13_4-4   ...T.CT....TC.G..A..-..................AT.T.....C.
ARG_Co3-2_M13_4-8   ...T.CT....TC.G..A..-..................AT.T.....C.

                          160       170       180       190       200
                    ....|....|....|....|....|....|....|....|....|....|
BAC                 CTGGCCCTTGACATGAGAACATTTGAGAATGAAATATGGAAAGAACCAAG
HA_RGC151_1         --------------------------......C.C...............
HA_RGC151_2         --------------------------......C.C...............
HA_RGC151_7         --------------------------......C.C...............
HA_RGC151_3         --------------------------......C.C...............
HA_RGC151_4         --------------------------......C.C...............
HA_RGC151_5         --------------------------......C.C...............
HA_RGC151_6         --------------------------......C.C...............
HA_RGC151_8         --------------------------......C.C...............
ARG_RGC151_1        --------------------------......C.C...............
ARG_RGC151_2        --------------------------......C.C...............
ARG_RGC151_3        --------------------------......C.C...............
ARG_RGC151_4        --------------------------......C.C...............
ARG_RGC151_5        --------------------------......C.C...............
ARG_RGC151_6        --------------------------......C.C...............
ARG_RGC151_8        --------------------------......C.C...............
ARG_Co3-2_M13_3-4   ..TA.......T..............A....G..C...............
ARG_Co3-2_M13_3-6   ..TA.......T..............A....G..C...............
ARG_Co3-2_M13_3-7   ..TA.......T..............A....G..C...............
ARG_Co3-2_M13_3-8   ..TA.......T..............A....G..C...............
ARG_Co3-2_M13_4-4   ..TA.......T..............A....G..C...............
ARG_Co3-2_M13_4-8   ..TA.......T..............A....G..C...............
```

```
                  210       220       230       240       250
                    ....|....|....|....|....|....|....|....|....|....|
BAC                 TAGTGCCAATATGCAAATGTTTGGATTTAGATCATATCCATCATTCAGTT
HA_RGC151_1         ..................................................
HA_RGC151_2         ..................................................
HA_RGC151_7         ..................................................
HA_RGC151_3         ..................................................
HA_RGC151_4         ..................................................
HA_RGC151_5         ..................................................
HA_RGC151_6         ..................................................
HA_RGC151_8         ..................................................
ARG_RGC151_1        .T............C.C.A............G.G..........A..
ARG_RGC151_2        .T............C.C.A............G.G..........A..
ARG_RGC151_3        .T............C.C.A............G.G..........A..
ARG_RGC151_4        .T............C.C.A............G.G..........A..
ARG_RGC151_5        .T............C.C.A............G.G..........A..
ARG_RGC151_6        .T............C.C.A............G.G..........A..
ARG_RGC151_8        .T............C.C.A............G.G..........A..
ARG_Co3-2_M13_3-4   .T............C.C.A............G.G..........A..
ARG_Co3-2_M13_3-6   .T............C.C.A.C..........G.G..........A..
ARG_Co3-2_M13_3-7   .T............C.C.A............G.G..........A..
ARG_Co3-2_M13_3-8   .T............C.C.A............G.G..........A..
ARG_Co3-2_M13_4-4   .T............C.C.A............G.G..........A..
ARG_Co3-2_M13_4-8   .T............C.C.A............G.G..........A..

                           260       270       280       290       300
                    ....|....|....|....|....|....|....|....|....|....|
BAC                 GTCTTTGGTGGTTATTTGGTTGGGTTTCGGGAATGTGTTCCTCATCTCGT
HA_RGC151_1         ..................................................
HA_RGC151_2         ..............G.............A................C.....
HA_RGC151_7         ..................................................
HA_RGC151_3         ..................................................
HA_RGC151_4         ..............G.............A.....................
HA_RGC151_5         ..................................................
HA_RGC151_6         ..................................................
HA_RGC151_8         ..................................................
ARG_RGC151_1        ..A..........CA...A....T..AT......TCA........T.T..
ARG_RGC151_2        ..A..........C....A....T..AT......TCA........T.T..
ARG_RGC151_3        ..A..........C....A....T..AT......CC.........T.T..
ARG_RGC151_4        ..A..........C....A....T..AT......TCA........T.T..
ARG_RGC151_5        ..A..........C....A....T..AT......TCA........T.T..
ARG_RGC151_6        ..A..........C....A....T..AT......TCA........T.T..
ARG_RGC151_8        ..A..........C....A....T..AT......TCA........T.T..
ARG_Co3-2_M13_3-4   ..A..........C....A....T..AT......TCA........T.T..
ARG_Co3-2_M13_3-6   ..A..........C....A....T..AT......TCA........A.T..
ARG_Co3-2_M13_3-7   ..A..........C....A....T..AT......TCA........T.T..
ARG_Co3-2_M13_3-8   ..A..........C....A....T..AT......TCA........T.T..
ARG_Co3-2_M13_4-4   ..A..........C....A....T..AT......TCA........T.T..
ARG_Co3-2_M13_4-8   ..A..........C....A....T..AT......TCA........T.T..
```

```
                          310       320       330       340       350
                    ....|....|....|....|....|....|....|....|....|....|
BAC                 AAAACAAAAGGTGACTTTGAGACACTTGCTTTAAGTGAAATGCGTAATTT
HA_RGC151_1         ...................................................
HA_RGC151_2         ...................................................C..
HA_RGC151_7         ...................................................
HA_RGC151_3         ...................................................
HA_RGC151_4         ...................................................C..
HA_RGC151_5         ...................................................
HA_RGC151_6         ...................................................
HA_RGC151_8         ...................................................
ARG_RGC151_1        ...........C..G.......T.T...AC....A...........C..
ARG_RGC151_2        ...........C..G.......T.T...AC....A...........C..
ARG_RGC151_3        ...........C..G.......T.T...AC....A...........C..
ARG_RGC151_4        ...........C..G.......T.T...AC....A...........C..
ARG_RGC151_5        ...........C..G.......T.T...AC....A...........C..
ARG_RGC151_6        ...........C..G.......T.T...AC....A...........C..
ARG_RGC151_8        ...........C..G.......T.T...AC....A...........C..
ARG_Co3-2_M13_3-4   ........G..C..G.......T.T...AC....A...........C..
ARG_Co3-2_M13_3-6   ...........C..G.......T.T...AC....A...........C..
ARG_Co3-2_M13_3-7   ...........C..G.......T.T...AC....A...........C..
ARG_Co3-2_M13_3-8   ...........C..G.......T-T...AC....A...........C..
ARG_Co3-2_M13_4-4   ...........C..G.......T.T...AC....A...........C..
ARG_Co3-2_M13_4-8   ...........C..G.......T.T...AC....A...........C..

                          360       370       380       390       400
                    ....|....|....|....|....|....|....|....|....|....|
BAC                 AAAACTACTCCAACTGAACTATGTACAACTCAGTGGATCCTACAAGAATT
HA_RGC151_1         ...................................................
HA_RGC151_2         ...................................................
HA_RGC151_7         ...................................................
HA_RGC151_3         ...................................................
HA_RGC151_4         ...................................................
HA_RGC151_5         ...................................................
HA_RGC151_6         ...................................................
HA_RGC151_8         ...................................................
ARG_RGC151_1        ..G......A.................................G.A....
ARG_RGC151_2        ..G......A.................................G.A....
ARG_RGC151_3        ..G......A.................................G.A....
ARG_RGC151_4        ..G......A.................T...............G.A....
ARG_RGC151_5        ..G......A.................................G.A....
ARG_RGC151_6        ..G......A......G..........................G.A....
ARG_RGC151_8        ..G......A.................................G.A....
ARG_Co3-2_M13_3-4   ..G......A.................................G.A....
ARG_Co3-2_M13_3-6   ..G.T....A.................................G.A....
ARG_Co3-2_M13_3-7   ..G......A.................................G.A....
ARG_Co3-2_M13_3-8   ..G......A.................................G.A....
ARG_Co3-2_M13_4-4   ..G......A.................................G.A....
ARG_Co3-2_M13_4-8   ..G......A.................................G.A....
```

```
                      410       420       430       440       450
                      ....|....|....|....|....|....|....|....|....|....|
BAC                   TTCCTCATGGTATAAGATGGTTATGCATGCATGGATTCCCTTTGAGTTAT
HA_RGC151_1           ..................................................
HA_RGC151_2           ...............................................N..
HA_RGC151_7           ..................................................
HA_RGC151_3           C.................................................
HA_RGC151_4           ..................................................
HA_RGC151_5           ..................................................
HA_RGC151_6           ..............................................A...
HA_RGC151_8           ..................................................
ARG_RGC151_1          .......A..............................T.........C
ARG_RGC151_2          .......A..............................T.........C
ARG_RGC151_3          .......A..............................T.........C
ARG_RGC151_4          .......A..............................T.....G....C
ARG_RGC151_5          .......A..............................T.........C
ARG_RGC151_6          .......A..............................T.........C
ARG_RGC151_8          .......A..............................T.........C
ARG_Co3-2_M13_3-4     .......A..............................T.........C
ARG_Co3-2_M13_3-6     .......A..............................T.........C
ARG_Co3-2_M13_3-7     .......A..............................T.........C
ARG_Co3-2_M13_3-8     .......A..............................T.........C
ARG_Co3-2_M13_4-4     .......A..............................T.........C
ARG_Co3-2_M13_4-8     .......A..............................T.........C

                            460       470       480       490       500
                      ....|....|....|....|....|....|....|....|....|....|
BAC                   ATTCCTTCAGACTTAGAAATGGAGAATTTGGTAGCTCTTGACTTGTCTAA
HA_RGC151_1           ..................................................
HA_RGC151_2           .....G.........C..................................
HA_RGC151_7           ..................................................
HA_RGC151_3           ..................................................
HA_RGC151_4           ...........G......................................
HA_RGC151_5           ..................................................
HA_RGC151_6           ..................................................
HA_RGC151_8           ..................................................
ARG_RGC151_1          ..A...........C........................A.......
ARG_RGC151_2          ..A...........C........................A.......
ARG_RGC151_3          ..A...........C........................A.......
ARG_RGC151_4          ..A...........C........................A.......
ARG_RGC151_5          ..A...........C........................A.......
ARG_RGC151_6          ..A...........C........................A.......
ARG_RGC151_8          ..A...........C........................A.......
ARG_Co3-2_M13_3-4     ..A...........C......................------------
ARG_Co3-2_M13_3-6     ..A...........C......................------------
ARG_Co3-2_M13_3-7     ..A...........C......................------------
ARG_Co3-2_M13_3-8     ..A..C........C......................------------
ARG_Co3-2_M13_4-4     ..A...........C......................------------
ARG_Co3-2_M13_4-8     ..A...........C......................------------
```

```
                          510       520       530       540       550
                    ....|....|....|....|....|....|....|....|....|....|
BAC                 TAGCAAACTACAACAGCTTTGGAAGAAGCCCAAGGTACTAAAATATCAAG
HA_RGC151_1         ..................................................
HA_RGC151_2         ...........................................C-.....
HA_RGC151_7         ..................................................
HA_RGC151_3         ..................................................
HA_RGC151_4         ............G.....................................
HA_RGC151_5         ..................................................
HA_RGC151_6         ................................G.................
HA_RGC151_8         ..................................................
ARG_RGC151_1        ...........................................C.....
ARG_RGC151_2        ...........................................C.....
ARG_RGC151_3        ...........................................C.....
ARG_RGC151_4        ...........................................C.....
ARG_RGC151_5        ...........................................C.....
ARG_RGC151_6        ...........................................C.....
ARG_RGC151_8        ...........................................C.....
ARG_Co3-2_M13_3-4   --------------------------------------------------
ARG_Co3-2_M13_3-6   --------------------------------------------------
ARG_Co3-2_M13_3-7   --------------------------------------------------
ARG_Co3-2_M13_3-8   --------------------------------------------------
ARG_Co3-2_M13_4-4   --------------------------------------------------
ARG_Co3-2_M13_4-8   --------------------------------------------------

                          560       570       580       590       600
                    ....|....|....|....|....|....|....|....|....|....|
BAC                 CTTCTAGTTATCTTGTTT----------AAGTATCTTTGCAACACT
HA_RGC151_1         ................----------.....................
HA_RGC151_2         ................----------.....................
HA_RGC151_7         ................----------.....................
HA_RGC151_3         ................----------.....................
HA_RGC151_4         ................----------.....................
HA_RGC151_5         ................----------.....G.....
HA_RGC151_6         ................----------.....................
HA_RGC151_8         ................----------.....................
ARG_RGC151_1        ......C....T......GATAACTCATGGTT.......A..........
ARG_RGC151_2        ......C....T......GATAACTCATGGTT.......A..........
ARG_RGC151_3        ......C....T......GATAACTCATGGTT.......A..........
ARG_RGC151_4        ......C....T......GATAACTCATGGTT.......A..........
ARG_RGC151_5        ......C....T......GATAACTCATGGTT.......A..........
ARG_RGC151_6        ......C....T......GATAACTCATGGTT.......A..........
ARG_RGC151_8        ......C....T......GATAACTCATGGTT.......A..........
ARG_Co3-2_M13_3-4   --------------------------------------------------
ARG_Co3-2_M13_3-6   --------------------------------------------------
ARG_Co3-2_M13_3-7   --------------------------------------------------
ARG_Co3-2_M13_3-8   --------------------------------------------------
ARG_Co3-2_M13_4-4   --------------------------------------------------
ARG_Co3-2_M13_4-8   --------------------------------------------------
```

```
                          610       620       630       640       650
                    ....|....|....|....|....|....|....|....|....|....|
BAC                 AACATGCATCTTAATTTT-GTTTATATGTT--AATTTAACAGCTGCTGAG
HA_RGC151_1         .................-...........--.............C...
HA_RGC151_2         ............T.....-......T.....--....A..T.......C..
HA_RGC151_7         .................-...........--..................
HA_RGC151_3         .................-...........--..................
HA_RGC151_4         .................-...........--..................
HA_RGC151_5         .................-...........--..................
HA_RGC151_6         .................-...........--..................
HA_RGC151_8         .................-...........--..................
ARG_RGC151_1        ..A.C............T......T.T..TT............A..C..
ARG_RGC151_2        ..A.C............T......T.T..TT............A..C..
ARG_RGC151_3        ..A.C............T......T.T..TT............A..C..
ARG_RGC151_4        ..A.C............T......T.T..TT............A..C..
ARG_RGC151_5        ..A.C............T......T.T..TT............A..C..
ARG_RGC151_6        ..A.C............T......T.T..TT............A..C..
ARG_RGC151_8        ..A.C............T......T.T..TT............A..C..
ARG_Co3-2_M13_3-4   --------------------------------------------------
ARG_Co3-2_M13_3-6   --------------------------------------------------
ARG_Co3-2_M13_3-7   --------------------------------------------------
ARG_Co3-2_M13_3-8   --------------------------------------------------
ARG_Co3-2_M13_4-4   --------------------------------------------------
ARG_Co3-2_M13_4-8   --------------------------------------------------

                          660       670       680       690       700
                    ....|....|....|....|....|....|....|....|....|....|
BAC                 GTCACTGAAGTTTCTTAATCTGAGTAGTTGTCATGAGCTTCTTAGAGTGG
HA_RGC151_1         ................G.................................
HA_RGC151_2         ........................A.....G.......G....G....
HA_RGC151_7         ..................................................
HA_RGC151_3         ..................................................
HA_RGC151_4         ..................................................
HA_RGC151_5         ..................................................
HA_RGC151_6         ..................................................
HA_RGC151_8         ..................................................
ARG_RGC151_1        ..................C....CG.....A......G.........
ARG_RGC151_2        ..................C....CG.....A......G.........
ARG_RGC151_3        ..................C....CG.....A....C.G.........
ARG_RGC151_4        ..................C....CG.....A......G.........
ARG_RGC151_5        ..................C....CG.....A......G.........
ARG_RGC151_6        ..................C....CG.....A......G.........
ARG_RGC151_8        ..................C....CG.....A......G.........
ARG_Co3-2_M13_3-4   --------------------------------------------------
ARG_Co3-2_M13_3-6   --------------------------------------------------
ARG_Co3-2_M13_3-7   --------------------------------------------------
ARG_Co3-2_M13_3-8   --------------------------------------------------
ARG_Co3-2_M13_4-4   --------------------------------------------------
ARG_Co3-2_M13_4-8   --------------------------------------------------
```

```
                        710       720
                    ....|....|....|....|....|.
BAC                 GTCACTTCTCTGGGCTCCCTCTACTT
HA_RGC151_1         ..........................
HA_RGC151_2         ..........................
HA_RGC151_7         ..........................
HA_RGC151_3         ..........................
HA_RGC151_4         ..........................
HA_RGC151_5         ..........................
HA_RGC151_6         ..........................
HA_RGC151_8         ..........................
ARG_RGC151_1        ..........................
ARG_RGC151_2        ..........................
ARG_RGC151_3        ..........................
ARG_RGC151_4        ..........................
ARG_RGC151_5        ..........................
ARG_RGC151_6        ..........................
ARG_RGC151_8        ..........................
ARG_Co3-2_M13_3-4   --------------------------
ARG_Co3-2_M13_3-6   --------------------------
ARG_Co3-2_M13_3-7   --------------------------
ARG_Co3-2_M13_3-8   --------------------------
ARG_Co3-2_M13_4-4   --------------------------
ARG_Co3-2_M13_4-8   --------------------------
```

Alignment 8-3: Nucleotide alignment (Clustal W) of CG1_e for the BAC sequence Co3_c2 (47,065 bp – 48,102 bp), for sequences cloned from the susceptible line HA342 (HA) and the resistant line ARG1575-2 (ARG) and for BRBScontig30846 identified in the BRBS assembly of the cDNA bulks . Predicted exons of resistance gene candidate TIR-NBS-LRR identified on HICF contig03_c2 are highlighted in grey on the BAC sequence. Dots represent conserved bases, letters and dashes indicate polymorphic sites, and deletion events, respectively. Primer sequences which were used to amplify the fragment are underlined.

```
                    10        20        30        40        50
                    |         |         |         |         |
BAC                 AGTAACCTTCCAATGTTGAAGAATTTGTTTTTGGATGGAAATCCAATGGA
HA_e_2              --------------------------------------------------
HA_e_4              --------------------------------------------------
HA_e_6              --------------------------------------------------
HA_e_7              --------------------------------------------------
HA_e_8              --------------------------------------------------
ARG_e_1             --------------------------------------------------
ARG_e_2             --------------------------------------------------
ARG_e_3             --------------------------------------------------
ARG_e_6             --------------------------------------------------
ARG_e_10            --------------------------------------------------
BRBSCo30846         ......................G....A......................

                    60        70        80        90        100
                    |         |         |         |         |
BAC                 TTCCATGCCCGATTGTGTGAGAAGCCTTAGTAGGCTGGAGACACTCAGTT
HA_e_2              --------------------------------------------------
HA_e_4              --------------------------------------------------
HA_e_6              --------------------------------------------------
HA_e_7              --------------------------------------------------
HA_e_8              --------------------------------------------------
ARG_e_1             --------------------------------------------------
ARG_e_2             --------------------------------------------------
ARG_e_3             --------------------------------------------------
ARG_e_6             --------------------------------------------------
ARG_e_10            --------------------------------------------------
BRBSCo30846         ......................A.........................T....

                   110       120       130       140       150
                    |         |         |         |         |
BAC                 TTACTTGGTGTTGGAAGCTGAAAACAGTCTTGTGTGCTCCAATCCAATTA
HA_e_2              ----........................N.......................
HA_e_4              ----........................N.......................
HA_e_6              ----........................N.......................
HA_e_7              ----........................N.......................
HA_e_8              ----........................N.......................
ARG_e_1             ----...................................................
ARG_e_2             ----...................................................
ARG_e_3             ----...................................................
ARG_e_6             ----...................................................
ARG_e_10            ----...................................................
BRBSCo30846         ..TG.......C....C.....................................
```

```
                  160       170       180       190       200
                  ....|....|....|....|....|....|....|....|....|....|
BAC               AAGCGGTTGGATGTATACTTTTGCAACTCATTGGAGAAAATAAAATTTCA
HA_e_2            ..................................................
HA_e_4            ..................................................
HA_e_6            ..................................................
HA_e_7            ..................................................
HA_e_8            ..................................................
ARG_e_1           ....A.......A..CT.......G..............C..C......
ARG_e_2           ....A.......A..CT.......G..............C..C......
ARG_e_3           ....A.......A..CT.......G..............C..C......
ARG_e_6           ....A.......A..CT.......G..............C..C......
ARG_e_10          ....A.......A..CT.......G..............C..C......
BRBSCo30846       ....A.......A..CT.......G..............C..C......

                  210       220       230       240       250
                  ....|....|....|....|....|....|....|....|....|....|
BAC               TCCAGAAAAGTCAGCTGTACCTAGAGTTTTTTATCAGACAAGTTTTAACC
HA_e_2            ..................................................
HA_e_4            ..................................................
HA_e_6            ..................................................
HA_e_7            ..................................................
HA_e_8            ..............................C...................
ARG_e_1           ...T............A...........C...G.A.C.G....G...C..
ARG_e_2           ...T............A...........C...G.A.C.G....G...C..
ARG_e_3           ...T............A...........C...G.A.C.G....G...C..
ARG_e_6           ...T............A...........C...G.A.C.G....G...C..
ARG_e_10          ...T............A...........C...G.A.C.G....G...C..
BRBSCo30846       ...T............A...........C...G.A.C.G....G...C..

                  260       270       280       290       300
                  ....|....|....|....|....|....|....|....|....|....|
BAC               TGACTGAGATTCAGCACGTATTCAAAATACAAGCTTTATCAGAAATTGAT
HA_e_2            ..................................................
HA_e_4            ..................................................
HA_e_6            ..................................................
HA_e_7            ..................................................
HA_e_8            ..................................................
ARG_e_1           ............A...A..C..............................
ARG_e_2           ............A...A..C............C.................
ARG_e_3           ............A...A..C..............................
ARG_e_6           ............A...A..C..............................
ARG_e_10          ............A...A..C..............................
BRBSCo30846       ............A...A..C..............................

                  310       320       330       340       350
                  ....|....|....|....|....|....|....|....|....|....|
BAC               GAAGAGGTATTATGCAGTTTGGGTTGGATTAATATAGCATATTTGAACCC
HA_e_2            ..............................C...................
HA_e_4            ..................................................
HA_e_6            ..................................................
HA_e_7            ..................................................
HA_e_8            ..................................................
ARG_e_1           ..............................................A....A
ARG_e_2           ..............................................A....A
ARG_e_3           ..............................................A....A
ARG_e_6           ..............................................A....A
ARG_e_10          ..............................................A....A
BRBSCo30846       ..............................................A....A
```

```
                360       370       380       390       400
           ....|....|....|....|....|....|....|....|....|....|
BAC        CTGCTGGTTGTCGAAGGCCAATGATATGGGATTGTACTATTTGGAACAAA
HA_e_2     ..................................................
HA_e_4     ..................................................
HA_e_6     ..................................................
HA_e_7     ..................................................
HA_e_8     ..................................................
ARG_e_1    ....C....T.........T.....T.T...A.A...C......C.GG..
ARG_e_2    ....C....T.........T.....T.T...A.A...C......C.GG..
ARG_e_3    ....C....T.........T.....T.T...A.A...C......C.GG..
ARG_e_6    ....C....T.........T.....T.T...A.A...C......C.GG..
ARG_e_10   ....C....T.........T.....T.T...A.A...C......C.GG..
BRBSCo30846 ....C....T.........T.....T.T...A.A...C......C.GG..

                410       420       430       440       450
           ....|....|....|....|....|....|....|....|....|....|
BAC        GGATACTTCCAGCACAGGTGGATGTCTCTTTATATATCTTTATATTTCAC
HA_e_2     ..................................................
HA_e_4     ..................................................
HA_e_6     ..................................................
HA_e_7     ...........A......................................
HA_e_8     ..................................................
ARG_e_1    .A.........................................AT...A..
ARG_e_2    .A.........................................AT...A..
ARG_e_3    .A.........................................AT...A..
ARG_e_6    .A.........................................AT...A..
ARG_e_10   .A.........................................AT...A..
BRBSCo30846 .A..............---------------------------------

                460       470       480       490       500
           ....|....|....|....|....|....|....|....|....|....|
BAC        CTTTTAATATGCATTTAAAAGCTAATAAAACTTTGATGTGACACTCAGAT
HA_e_2     ..................................................
HA_e_4     ..................................................
HA_e_6     ..................................................
HA_e_7     ..................................................
HA_e_8     ..................................................
ARG_e_1    ..................................................
ARG_e_2    ..................................................
ARG_e_3    ..................................................
ARG_e_6    ..................................................
ARG_e_10   ..................................................
BRBSCo30846 ------------------------------------------------..

                510       520       530       540       550
           ....|....|....|....|....|....|....|....|....|....|
BAC        GCTTTATGAACATGGAATATTCAGCACATATGTTCAAGGGCAAGAGGTTC
HA_e_2     ..................................................
HA_e_4     ..................................................
HA_e_6     ..............G...................................
HA_e_7     ..................................................
HA_e_8     ..................................................
ARG_e_1    ................................C.................
ARG_e_2    ................................C.................
ARG_e_3    ................................C.................
ARG_e_6    ................................C.................
ARG_e_10   ................................C.................
BRBSCo30846 ................................C.................
```

```
                560       570       580       590       600
           ....|....|....|....|....|....|....|....|....|....|
BAC        CAAAGTGGTTCACACATAGAAGCAGTGGGTCATTATTTACCTTGCAATCA
HA_e_2     ..................................................
HA_e_4     ..................................................
HA_e_6     ..................................................
HA_e_7     ..................................................
HA_e_8     ..................................................
ARG_e_1    ..G.........T..A.....G.A..........................
ARG_e_2    ..G.........T..A.....G.A..........................
ARG_e_3    ..G.........T..A.....G.A..........................
ARG_e_6    ..G.........T..A.....G.A..........................
ARG_e_10   ..G.........T..A.....G.A..........................
BRBSCo30846 ..G.......G.T..A......AC........C..............

                610       620       630       640       650
           ....|....|....|....|....|....|....|....|....|....|
BAC        TCTCCTGAAAAAGGTAAGATTAAAGGGCTAAATGTGTGCATTGTGCATAC
HA_e_2     ..................................................
HA_e_4     ..................................................
HA_e_6     ..................................................
HA_e_7     ..................................................
HA_e_8     ..................................................
ARG_e_1    .....A............................................
ARG_e_2    .....A............................................
ARG_e_3    .....A............................................
ARG_e_6    .....A............................................
ARG_e_10   .....A............................................
BRBSCo30846 ....................................T..C...A....

                660       670       680       690       700
           ....|....|....|....|....|....|....|....|....|....|
BAC        GATTTCAGGCA---AGGAAGTTGG--TCCTTCAAGAATCAAAATAACGAA
HA_e_2     ..........---.........--..........................
HA_e_4     ..........---.........--..........................
HA_e_6     ..........---.........--..........................
HA_e_7     ..........---.........--..........................
HA_e_8     ..........---.........--..........................
ARG_e_1    .......A...TGA.....A....--...A................GT..
ARG_e_2    .......A...TGA.....A....--...A................GT..
ARG_e_3    .......A...TGA.....A....--...A................GT..
ARG_e_6    .......A...TGA.....A....--...A................GT..
ARG_e_10   .......A...TGA.....A....--...A................GT..
BRBSCo30846 AG.....A...TGA..........CG.TT..A......T.......G...

                710       720       730       740       750
           ....|....|....|....|....|....|....|....|....|....|
BAC        TCTCACAAAGAAGTCCTCCTGGACATACCAGCCTATGATGTATCTTTTTC
HA_e_2     ..................................................
HA_e_4     ..................................................
HA_e_6     ..................................................
HA_e_7     ..................................................
HA_e_8     ....................................C.............
ARG_e_1    ...G.....TG.C.................................G...
ARG_e_2    ...G.....TG.C......C..........................G...
ARG_e_3    ...G.....TG.C.................................G...
ARG_e_6    ...G.....TG.C.................................G...
ARG_e_10   ...G.....TG.C.................................G...
BRBSCo30846 ...G........C..............G.......T.........G...
```

```
                760       770       780       790       800
               ....|....|....|....|....|....|....|....|....|....|
BAC            CATATAATAGCGCGTTTGGATTTGGTGATAAAGC---GGTAGTATGGTTA
HA_e_2         ....................................---.............
HA_e_4         ....................................---.............
HA_e_6         ....................................---.............
HA_e_7         ....................................---.............
HA_e_8         ....................................---.............
ARG_e_1        ..G.AG..GAT..A.......A.....G.....TTGT..............
ARG_e_2        ..G.AG..GAT..A.......A.....G.....TTGT..............
ARG_e_3        ..G.AG..GAT..A.......A.....G.....TTGT..............
ARG_e_6        ..G.AG..GAT..A.......A.....G.....TTGT..............
ARG_e_10       ..G.AG..GAT..A.......A.....G.....TTGT..............
BRBSCo30846    ..G.AG..GAT.......A.GA.......G...TAGTT.C..........

                810       820       830       840       850
               ....|....|....|....|....|....|....|....|....|....|
BAC            AGCCATTGGATGTTTGGGAAAAATGAGTTTGAAGATGGTGATAAAGTTCG
HA_e_2         ..................................................
HA_e_4         ..................................................
HA_e_6         ..................................................
HA_e_7         ..................................................
HA_e_8         ..................................................
ARG_e_1        ..................G......C.............CG.....A.
ARG_e_2        ..................G......C.............CG.....A.
ARG_e_3        ..................G......C.............CG.....A.
ARG_e_6        ..................G......C.............CG.....A.
ARG_e_10       ..................G......C.............CG.....A.
BRBSCo30846    ............................C.............G.....A.

                860       870       880       890       900
               ....|....|....|....|....|....|....|....|....|....|
BAC            CATTGACTTTAGTGTAAAACATTATAT------GAGT---TTGTCAGGTC
HA_e_2         ............................------...---..........
HA_e_4         ............................------...---..........
HA_e_6         ............................------...---.........C.
HA_e_7         ............................------...---..........
HA_e_8         ............................------...---..........
ARG_e_1        ....C.T..C.........T...T..ATTTTAGA..CCGA...ATTA..A
ARG_e_2        ....C.T..C.........T...T..ATTTTAGA..CCGA...ATTA..A
ARG_e_3        ....C.T..C.........T...T..ATTTTAGA..CCGA...ATTA..A
ARG_e_6        ....C.T..C.........T...T..ATTTTAGA..CCGA...ATTA..A
ARG_e_10       ....C.T..C.........T...T..ATTTTAGA..CCGA...ATTA..A
BRBSCo30846    ....C.A.....---------------------------------------

                910       920       930       940       950
               ....|....|....|....|....|....|....|....|....|....|
BAC            TTAGGTATGGTGGTGAAGGACCAGAATATGCAAATGTAAGGGAGTATGGT
HA_e_2         ..................................................
HA_e_4         ..................................................
HA_e_6         ..................................................
HA_e_7         ..................................................
HA_e_8         ..................................................
ARG_e_1        ..T.CC........A....................................
ARG_e_2        ..T.CC........A.....................R..........
ARG_e_3        ..T.CC........A.....................R..........
ARG_e_6        ..T.CC........A.....................R..........
ARG_e_10       ..T.CC........A.....................R..........
BRBSCo30846    ---------------------------------------------------
```

```
                        960       970       980       990       1000
                   ....|....|....|....|....|....|....|....|....|....|
BAC                ATTAGCCCCGTGTATGATGATGATGATGGTGGTGGTGGTGGAAAGCAGAA
HA_e_2             ..................................................
HA_e_4             ..................................................
HA_e_6             ..................................................
HA_e_7             ..................................................
HA_e_8             .C................................................
ARG_e_1            ........T............A.............AA-----........
ARG_e_2            ........T............A.............A.-----........
ARG_e_3            ........T............A............AA------........
ARG_e_6            ........T............A............AA------........
ARG_e_10           ........T............A............AA------........
BRBSCo30846        --------------------------------------------------

                        1010      1020      1030      1040      1050
                   ....|....|....|....|....|....|....|....|....|....|
BAC                AGAGGATCCATTGGGTTATTACAAGTCATGGAAGCACATCATTGGTGGGG
HA_e_2             ..................................................
HA_e_4             ..................................................
HA_e_6             ..................................................
HA_e_7             ..................................................
HA_e_8             ..................................................
ARG_e_1            ...T.....T.....................T..................
ARG_e_2            ...T.....T.....................T..................
ARG_e_3            ...T.....T.....................T..................
ARG_e_6            ...T.....T.....................T..................
ARG_e_10           ...T.....T.....................T..................
BRBSCo30846        --------------------------------------------------


                   ....|
BAC                ATCTC
HA_e_2             .....
HA_e_4             .....
HA_e_6             ....T
HA_e_7             .....
HA_e_8             .....
ARG_e_1            .....
ARG_e_2            .....
ARG_e_3            .....
ARG_e_6            .....
ARG_e_10           .....
BRBSCo30846        -----
```

9 Acknowledgements

I am very grateful to my academic supervisors Prof. Dr. Chris-Carolin Schön and Dr. Eva Bauer for their advice, suggestions, and support during this thesis work. Thanks to Prof. Dr. R. Hückelhoven, and Prof. Dr. C. Schwechheimer for being part of the committee.

My extra special thanks go to Dr. Volker Hahn and his colleagues Anna Elhabachi, and Sabine Schillinger at the State Plant Breeding Institute for providing me with plant material. Sunflower seeds of cmsHA342 x ARG1575-2 were kindly provided by Prof. Dr. W. Friedt, Justus-Liebig University, Gießen. Thanks to Prof. Dr. O. Spring, University of Hohenheim, for providing me with the *P. halstedii* races 100, 330, and 710.

I would like to thank my colleague and friend Grit Haseneyer for long and fruitful discussions about my thesis, science in general and about all other questions of life. Many thanks to Stefan Schwertfirm and Barbara Renz for technical assistance and Dr. V. Kuryshev for bioinformatics support. I would like to express my gratitude to the Plant Breeding group at the Technische Universität München and at the State Plant Breeding Institute, University of Hohenheim, for creating a pleasant work environment.

Thank you to Prof. Dr. S.J. Knapp, Dr. E. Bachlava and all the other members of the “Institute of Plant Breeding, Genetics and Genomics” for the possibility to visit them at the University of Georgia and for cooperation in the field of sunflower research. I would like to thank Dr. C. Saski from the Clemson University Genomics Institute for screening the HA_HBa BAC library. My gratitude is extended to Dr. S. Taudien and his colleagues from the Fritz-Lipmann-Institut, Jena, where the 454 sequencing of BAC clones and cDNA bulks was done as well as the assembly of the BAC clones. Many thanks go to the group of Dr. K.F.X. Mayer from the Helmholtz Zentrum München especially to Michael Seidel for helping with cDNA bulk analysis and BAC clone annotation.

Warm thanks are devoted to my family and Tobias who supported and believed in me throughout my life.

The financial support from Deutsche Forschungsgemeinschaft (DFG) and the scholarship from the Universität Bayern e.V. are gratefully acknowledged.

Parts of this thesis were published in the following article and are reprinting in this thesis with kind permission from Springer Science and Business Media:

Wieckhorst S, Bachlava E, Dußle C, Tang S, Gao W, Saski C, Knapp S, Schön CC, Hahn V, Bauer E (2010) Fine mapping of the sunflower resistance locus Pl_{ARG} introduced from the wild species *Helianthus argophyllus*. Theor Appl Genet 121:1633-1644

10 Curriculum vitae

Personal data

Name: Silke Wieckhorst

Date of birth: 3rd June, 1979

Place of birth: Lüneburg, Germany

Education

2007 – 2010	PhD, Technische Universität München, Freising, Germany
2006 – 2007	PhD, Universität Hohenheim, Stuttgart, Germany
2001 – 2006	Universität Hohenheim, Stuttgart, Germany
	Field of study: Agricultural Biology
	Degree: Diploma
1999 – 2001	Apprenticeship as a technical assistant in plant breeding, Lochow Petkus GmbH, Bergen-Wohlde, Germany
1999	Abitur, Gymnasium Oedeme, Lüneburg, Germany

Professional experience

2010 – present	Research scientist at KWS SAAT AG, Einbeck, Germany