CHAPTER II
LITERATURE REVIEW

Jerusalem artichoke (*Helianthus tuberosus* L.) is the member of family Asteraceae, have originated in North America (Cosgrove et al., 1991) and can be propagated using tuber. The tubers can be used as a fresh or cooked vegetable for humans and used as ingredients substituted for antibiotic in animal feed (Denoroy, 1996). The tuber accumulates carbohydrate in a form of oligofructan, inulin, whereas starch is a storage form in most plant. Inulin provides several health issues such as dietary fiber property which benefits for obesity, diabetes, and heart disease (Orafti, 2005). The inulin from Jerusalem artichoke tubers also has been proposed for many years as a possible substrate for the production of ethanol with possible use for biofuels (Denoroy, 1996). With its excellent potential, Jerusalem artichoke is an underutilized resource that possesses the potential to meet major health and energy challenges.

2.1 Molecular marker and plant genetics research

The evaluation of genetic diversity would promote the efficient use of genetic variations in the breeding program application (Paterson et al., 1991). In particular, the use of molecular markers have been used to monitor the DNA sequence variation in and among species and create new source of species by introducing favorable traits from landrace and related species (Korzun, n.d.).

Molecular genetic markers have been developed into a powerful tool to analyze genetic relationships and genetic diversity. With the use of molecular techniques it would be possible to hasten the transfer of desirable genes among varieties and to introgress novel genes from related wild species. Polygenic characters which were previously very difficult to analyse using traditional plant breeding methods would now be easily tagged using molecular markers. It would also be possible to establish genetic relationships between sexually incompatible crop plants. However, no molecular marker technique is covering all circumstances. A large

number of polymorphic markers are required to measure genetic relationships and genetic diversity in a reliable manner. This limits the use of morphological characters and isozymes, which are few or lack adequate levels and tend to be influenced by environmental factors (FAO, 2003). Genetic markers fall into one of the three broad classes: first based on visually assessable traits (morphological and agronomic traits), second based on gene product (biochemical markers), and those relying on a DNA assay (molecular markers). The idea of using genetic markers appeared very early in literatures (Sax, 1932; Wexelsen, 1933) but the development of electrophoretic assays of isozymes (Markert and Moller, 1959) and molecular markers have greatly improved our understanding in biological sciences. Various types of molecular markers are utilized to evaluate DNA polymorphism Genetic or DNA based marker techniques such as Restriction Fragment Length Polymorphism (RFLP), Random Amplified Polymorphic DNA (RAPD), Simple Sequence Repeats (SSR) and Amplified Fragment Length Polymorphism (AFLP) are now in common use for ecological, evolutionary, taxonomical, phylogenic and genetic studies of plant sciences. These techniques are well established and their advantages and limitations have been documented (Ayad et al., 1995; Agarwal et al., 2008 and Primer, 2009).

2.1.1 Sequence-Related Amplified Polymorphism (SRAP)

Recently, a new class of advanced techniques has emerged, primarily derived from combination of the earlier, more basic techniques. These advanced marker techniques combine advantageous aspects of several basic techniques. Sequence-related amplified polymorphism (SRAP; Li and Quiros, 2001) and Target Region Amplification Polymorphism (TRAP; Hu and Vick, 2003) are two kinds of newly developed molecular marker systems with the advantages of simplicity, high throughput, numerous co-dominant makers, highly reproducibility and ready to sequence, especially preferential targeting ORFs. The principle and protocol of SRAP and TRAP, which were first developing in *Brassica* and *Helianthus* crops (Li and Quiros, 2001 and Hu & Vick, 200). Primer design is a key step for SRAP and TRAP. The aim of SRAP technique (Li and Quiros, 2001) is the amplification of open reading frames (ORFs). It is based on two-primer amplification. The technique uses primers of arbitrary sequence, which are 17–21 nucleotides in length. It uses pairs of

primers with AT- or GC-rich cores to amplify intragenic fragments for polymorphism detection. The primers consist of the following elements: 1. Core sequences, which are 13-14 bases long, where the first 10 or 11 bases starting at the 5′-end, are sequences of no specific constitution (''filler'' sequences), followed by the sequence CCGG in the forward primer and AATT in the reverse primer. 2. The core is followed by three selective nucleotides at the 3′end. The filler sequences of the forward and reverse primers must be different from each other and can be 10 or 11 bases long. For the first five cycles the annealing temperature is set at 35 °C. The following 35 cycles are run at 50 °C. The amplified DNA fragments are fractionated by denaturing acrylamide gels and detected by autoradiography (Fig. 2.1).

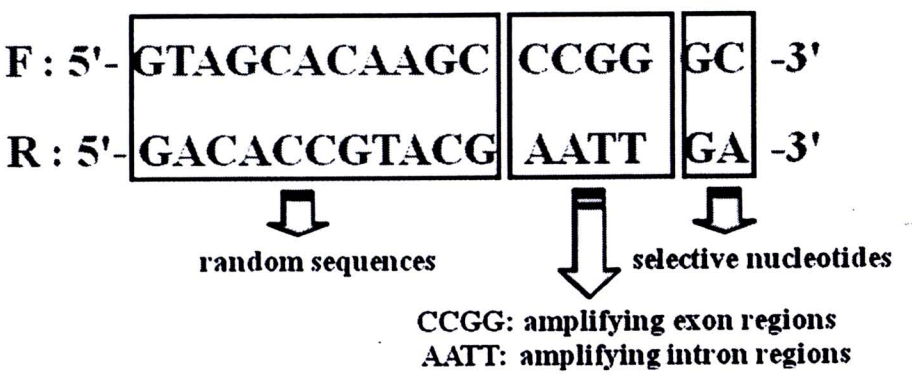

Figure 2.1 Diagrammatic representation of the SRAP primer sequences

SRAP combines simplicity, reliability, moderate throughput ratio and facile sequencing of selected bands (Li and Quiros, 2001). It targets coding sequences in the genome and results in a moderate number of co-dominant markers. Sequencing demonstrated that SRAP polymorphism results from two events, fragment size changes due to insertions and deletions, which could lead to codominant markers, and nucleotide changes leading to dominant markers. The SRAP marker system has been adapted for a variety of purposes in different crops, including map construction, gene tagging and genetic diversity studies useful molecular marker system for mapping and gene tagging (Table 2.1). Unlike SRAP, the TRAP requires cDNA or EST sequence information for fixed primer development. The annealing temperature is 35 °C. in the first 5 cycles and 50 °C. in the subsequent 30~35 cycles. The amplicons can be

separated by polyacrylamide or agarose gel, and detected by autoradiography, silver or ethidium bromide staining. SRAP markers are more consistent and repeatable than RAPD, and are less labor-intensive and timeconsuming to produce than AFLP and SSR (Li and Quiros, 2001; Santos et al., 2003; Ferriol et al., 2003).

Table 2.1 The application of SRAP markers in plant research.

Name	year	Species	Application	Primers pairs	Application from
Li & Quiros	2006	*Brassica* spp.	Diversity determination	130	Li and Quiros, 2001
Zaefizadeh and Goliev	2009	*Triticum durum*	Diversity determination compared to phenotypic trait	39	Li and Quiros, 2001
Gulsen et al.	2007	*Abelmoschus esculentus*	Diversity determination compared to phenotypic trait	39	Li and Quiros, 2001
Budak et al.	2004	*Bubalus bubalis*	Diversity determination and molecular marker methods compare	56	Li and Quiros, 2001
Comlekcioglu et al.	2010	*Solanum lycopersicon*	Genetic characterization of heat tolerant tomato	11	Li and Quiros, 2001
Youssef et al.	2011	Musa	genetic diversity	80	Li and Quiros, 2001
Xue et al.	2010	*Dendrobium*	The linkage maps of Dendrobium species	11	Li and Quiros, 2001
Castonguay et al.	2010	*Medicago sativa spp. sativa*	SRAP polymorphisms associated with superior freezing tolerance	42	Li and Quiros, 2001
Tang et al.	2010	Chinese Auricularia auricula	genetic diversity	11	Li and Quiros, 2001
Xie, et al.	2009	*MT-1 elephant grass*	Genetic variability and relationship	62	Li and Quiros, 2001
Song et al.	2009	*Salvia miltiorrhiza Bge*	Genetic diversity and population structure	6	Li and Quiros, 2001
Feng et al.	2009	*Celosia argentea*	Genetic diversity and population structure	36	Li and Quiros, 2001
Hao et al.	2008	Paeonia cultivars	identification	25	Li and Quiros, 2001
Han et al.	2008	tree peony	Molecular characterization	23	Li and Quiros, 2001

2.1.2 Expressed sequence tag (EST)

Expressed sequence tag (EST) is a short sub-sequence of a transcribed complementary DNA (cDNA) sequences provide direct evidence for all the sampled transcripts and they are currently the most important resources for transcriptome exploration. An EST is produced by one-shot sequencing of a cloned mRNA (i.e. sequencing several hundred base pairs from an end of a cDNA clone taken from a cDNA library). The resulting sequence is a relatively low quality fragment whose length is limited by current technology to approximately 500 to 800 nucleotides. (http://en.wikipedia.org/wiki/Expressed_sequence_tag). High-throughput ESTs can be generated at a reasonably low cost from either the 5' or 3' end of a cDNA clone to be able to gene discovery, complement genome annotation, aid gene structure identification, establish the viability of alternative transcripts, guide single nucleotide polymorphism (SNP) characterization and get an insight into transcriptionally active regions in any organism (Jongeneel, 2000; Dong et al., 2005; Rudd et al., 2003; Nagaraj et al., 2006).

In 1991, ESTs were used as a primary resource for human gene discovery (Adams et al., 1991). Thereafter, there has been an exponential growth in the generation and accumulation of EST data in public databases for myriad organisms. At present, the identification of ESTs has proceeded rapidly, with approximately 65.9 million ESTs now available in public databases (e.g. GenBank 2010, all species). In this respect, EST sequences are extremely valuable for identification of new genes in several aspects. The gene sequences obtained can be efficiently used for physical mapping by determining their chromosomal position. Moreover, they also contribute to the understanding of intron and exon boundaries, which will help to predict the transcribed regions of genomic sequences. EST sequencing can be used to reveal single base variations in genes for the discovery and characterization of candidate SNPs (Picoult-Newberg L, 1999, Useche FJ, Gao G, 2001).

2.1.3 Microsatellites or Simple Sequence Repeats (SSRs)

Microsatellites, also referred to as simple sequence repeats (SSRs), are the stretches of DNA consisting of exact simple tandemly repeated short motifs of 1–5 bp in length, which occur as interspersed repetitive elements in all eukaryotic genomes (Tautz and Renz, 1984) and present in both protein coding and non-coding regions of DNA sequences (Gupta et al., 1996; Toth et al., 2000; Katti et al., 2001). Variation in SSR length occurs primarily due to slipped-strand mispairing during DNA replication (Toth et al., 2000; Katti et al., 2001; Li et al., 2002), and mutations of this sort occur at a much higher frequency than do point mutations and insertions/deletions (Rossetto, 2001). Thus, SSRs reveal much higher levels of polymorphism than do most other marker systems. Compared to other molecular markers, SSRs are uniquely characterized by their simplicity, abundance, ubiquity, variation, co-dominance, reproducibility and multi-alleles among genomes (Powell et al., 1996). The polymorphism, mainly resulted from the number of repeat units, and can easily be detected by PCR using primers flanking the SSR motif. Therefore, SSRs have become a common tool broadly used in aspects of genetic mapping, molecular evolution and systematic taxonomy in most genomes.

In recent years, the popularity of SSR-based markers in plant has increased considerably (Rakoczy-Trojanowska and Bobilok, 2004), the applications including, linkage mapping, variability analysis, phytogenetic and finger printing analyses and gene tagging (table X). Because of the possibility to detect several alleles at a high frequency, SSRs is an ideal tool for identifying individual and establishing genetic diversity between them. It was well demonstrated by Hao et al. (2011), who examined 250 wheat lines with 512 primers and can be distinct the landraces from modern varieties. From the analysis of Abdurakhmonov et al. (2007), the correlation of SSR of molecular with the agronomic traits of fiber are closely linked to genes that control the qualitative characteristics of fiber and affect its growth.

However, the development of SSR markers from genomic libraries is expensive, labor intensive and time consuming (Varshney et al., 2002). Recently, genomics technology has provided researchers with access to vast amounts of

sequence information through public and private databases (Varshney et al., 2005; Bouck and Vision, 2007; Ellis and Burke, 2007 and Kong et al., 2007). The generation of microsatellite markers using EST sequences has become an attractive alternative to complement existing SSR collections and has also been successfully used in several plant species.

2.1.4 EST-SSR

Microsatellites developed from ESTs, known as EST–SSRs or genic SSRs, represent functional molecular markers as a 'putative function' for a majority of such markers. With recent increasing emphasis on functional genomics, large datasets of ESTs are being developed and provided through public and private databases (Varshney et al., 2005; Bouck and Vision, 2007; Ellis and Burke, 2007; Kong et al., 2007). More than 140 biologically and economically important plant species have been developed for EST databases (http://www.ncbi.nih.gov; http://www.tigr.org), including sunflower (*Helianthus annuus* L.), lettuce (*Lactuca sativa* L.), and safflower (*Carthamus tinctorius* L.) (Kozik et al., 2002; http://cgpdb. ucdavis.edu/cgpdb2/; Fernandez et al., 2003; Tamborindeguy et al., 2004; Ben et al., 2005; Lai et al., 2005a). As an alternative to the conventional strategy of probing repeat-enriched genomic libraries for EST-SSRs, and with evolving bioinformatic tools it is now possible to identify and develop EST-SSR markers at a large scale in a time and cost-effective manner (Scott et al., 2000; Kantety et al., 2002; Varshney et al., 2002). SSR patterns in the sequence files can be recognized by several search modules. Some programs are available in public domain such as MISA (http://pgrc.Ipkga tersleben.de/misa/), Sputnik (http://abajian.net/sputnik/index.html), SSRSEARCH (ftp://ftp.gramene.org/pub/gramene/software/scripts/ssr.pl), ESTSCA N2 (Iseli and others 1999; Lottaz et al., 2003; http://www.ch.embnet.org/software /ESTScan2.html) and WebSat (http://wsma-rtins.net/websat/). The primary limitation of using EST databases as a source of molecular markers is that this approach relies on existing genomic resources, and suitable databases are often only available to researchers who are studying economically important species. Adding to this difficulty is the fact that only a fraction of all ESTs contain an SSR. For example, it has been suggested that the frequency of SSR-containing sequences in plant-derived

EST databases is typically on the order of 2-5% (Kantety et al., 2002). However, the frequency, distribution and abundance of SSRs can be highly variable depending on the SSR search criteria, the size of the dataset, and the database-mining tools (Varshney et al., 2005). Because they are derived from genic regions, EST-SSR markers are one class of marker that can contribute to 'direct allele selection, if they are shown to be completely associated or even responsible for a targeted trait (Sorrells and Wilson, 1997). Evaluation of germplasm with SSRs derived from genes or ESTs might enhance the role of genetic markers by assaying the variation in transcribed and known-function genes. Expansion and contraction of SSR repeats in genes of known function can be tested for association with phenotypic variation or, more desirably, biological function (Ayers, et al., 1997).

2.2 Diseases on Jerusalem artichoke

Jerusalem artichoke has a few serious pest and disease problems. High yield losses can occur, especially when fungal or bacterial infestations affect the tubers. The primary disease is Sclerotinia (white mold), which can cause early wilt, stalk rot and degradation of the tubers. This pathogen also can cause severe yield reductions in dry edible bean, sunflower, and soybean (Cosgrove et al., 1991).

2.2.1 Sclerotinia wilt/rot

Sclerotinia sclerotiorum (Lib.) de Bary, an ascomycete, is a widespread pathogen that infects the stems, roots, and tubers of Jerusalem artichoke (Bisby, 1924). It is prevalent in temperate, subtropics, and tropical regions of the world. It is considered the most important sunflower disease in North America (Gulya, 1996). Infestation causes early wilt, stalk rot, and degradation of the tubers (sclerotinia wilt, white mold or cottony soft rot). *S. sclerotiorum* is the most common fungal pathogen and the most significant disease of Jerusalem artichoke globally in economic terms. Nevertheless, epidemics are rare, with single plants or small patches affected.

S. minor Jagger also causes similar symptoms of root and tuber rot and wilt. There is no apparent physiological specialization of both species, so isolates from one crop can infect another, although the degree of virulence varies. For

example, isolates of *S. sclerotiorum* from hop (*Humulus lupulus* L.) and swede (*Brassica napus* L.) and *S. minor* were lethal to Jerusalem artichoke (Keay, 1939). Sclerotinia wilt is due to the infection of the roots by mycelia derived from germinating sclerotia in the soil. Infection is dependent upon the actively growing roots coming in contact with the mycelia and can occur anytime during the plant development. The organism can spread to adjacent plants, a condition that appears to be favored by high plant densities.

2.2.2 Southern wilt/blicht/coilar rot

The disease is prevalent in the southern parts of the northern temperate zone and in subtropical and tropical regions of the world. The disease has also been reported on Jerusalem artichokes in several southern U.S. states (Farr et al., 1989, McCarter and Kays, 1984), Ghana and Malaysia (Thompson, 1928). Substantial yield losses of Jerusalem artichoke caused by *Sclerotinia sclerotiorum* have been reported mostly in the temperate regions (Cosgrove et al., 1991; Cassells and Walsh, 1995). However, pathogenicity of different isolates *S. rolfsii* in Jerusalem artichoke has been reported in the semi-arid tropics (Sennoi et al., 2010).

Sclcrotium rotfsii Sacc. can result in serious losses, especially when grown on land previously planted with Jerusalem artichokes (McCarter and Kays, 1984) or other susceptible crops. *S. rolfsii* is a soil born fungus, and it is able to persist for years in the soil (Punji, 1985). The pathogen of sclerotial disease causes damping-off of seedlings, stem canker, crown blight and rots on root, crown, bulb, tuber and fruit of various plant groups (Kwon et al., 2008). Initial symptoms of Jerusalem artichoke consisted of wilting of new shoots and leaves followed by browning and collapse of all foliage. Crown and lower stem tissues were colonized internally and externally by white, cottony mycelium. Tan, spherical sclerotia that measured approximately 1 mm in diameter formed on surfaces of the affected crowns and stems (Koike, 2004). Small, residual tubers in the soil that are not collected by mechanical or hand harvest appear to serve as a food base for the organism during the subsequent cropping season (McCarter and Kays, 1984). In addition, mechanical harvesting appears to spread the sclerotic in the field, increasing the distribution of the disease. *S.*

rolfsii also causes storage rots in tubers that appear sound at harvest (Thompson, 1928).

2.2.3 Powdery mildew

An obligate ascomycetes, *Erysiphe cichoracearum* DC., powdery mildew, can be moderately severe on some Jerusalem artichoke lines in the southern U.S., while others remain essentially disease free (McCarter, 1993). *E. cichoracearum* infects a wide range of species, especially so in the Compositae family. It is found worldwide and is especially critical in the hot, humid tropics and subtropics. Damage in temperate region trend to be limited. The disease appears as white to gray areas on the leaves and occasionally on the stems, which is caused by superficial mycelium. Typically, first symptoms are on the upper surfaces of older leaves, and the infected leaves eventually turn yellow and abscise. The white surface coloration is due to the conidia, which gave a powdery appearance. Conidia are born in long chains that are elliptical to barrel shaped. Control is possible through the development of resistant cultivars, the use of foliar fungicides, and cultural practices. There is considerable genetic variation in the Jerusalem artichoke gene pool for resistance. For example, McCarter (1993) found 3 of 36 lines displaying high levels of resistance. However, in temperate regions, powdery mildew tends to occur only late in the season, if present, and does not result in significant losses (McCarter and Kays, 1984), As a consequence, little emphasis has been placed on breeding for resistance.

2.2.4 Rust

Puccinia heliqnthi Schw., a basidiomycete, causes serious losses in Jerusalem artichoke (McCaner and Kays. 1984) and sunflower (Zimmer and Zimmerman, 1972). The organism is found only on the genus *Helianthus*, though more than 35 species are affected (Arthur, 1905; Arthur and Cummins, 1962; Hennessy and Sackston, 1972; Zimmer and Rehder, 1976). The disease can be lethal to highly susceptible genotypes. The severity of the infestation varies with the level of resistance of the host, age of the plant, and environmental conditions. Though always present, some years are much more conducive for development of the organism (McCarter and Kays, 1984). The organism was especially prevalent on the abaxial

surfaces of the leaves but can also be found on the stems. If the young foliage becomes infected, further development can be inhibited. Symptoms start on the older leaves at the base of the plant and progress upward toward the growing point. The pustules can became so abundant that the foliage appears blighted. With severe rust infections, most of the foliage is killed by the end of the growing season.

2.2.5 Apical chlorosis

A bacterial foliage disease caused by *Pseudomonas syringae* (Ps) results in significant Jerusalem artichoke losses (Shane and Baumer, 1984), primarily reported in the northern U.S. and Canada. In Minnesota in 1983, approximately 15 of 25,000 ha of Jerusalem artichokes was affected. Infected plants display extreme apical chlorosis, small dark necrotic leaf spots (1 to 2 mm in diameter) with faint chlorotic halos, and large chlorotic spots, sometimes with small gray necrotic centers. The effect of the disease on stand establishment is pronounced. Sprouts emerging from the soil can be nearly or completely chlorotic and often do not survive. Stand reductions in Minnesota were as high as 50%, with some fields plowed under. In Canada, however, mortality was limited (Laberge and Sackston, 1986). Apical chlorosis can be observed at all stages of plant development, and the leaves are uniformly chlorotic, including the veins. Some plants recover, with subsequent leaves being normal in coloration, though the original chlorotic leaves remain yellow. The bacteria are soil and tuber born and are further disseminated via rain and wind. High temperature (e.g., 28 to 30°C) and relative humidity favor the development of the pathogen (Stapp, 1961).

2.3 Plant disease resistance genes

Plant diseases can dramatically reduce crop yield and the impact of disease outbreaks is particularly acute in developing nations. Chemical controls are often beyond the means of farmers in developing nations. For these reasons, much effort has been invested towards understanding innate resistance mechanisms in plants. Resistance to disease is a major goal of all crop breeding programs. Many diseases are difficult to select in the field, therefore identification of markers associated with

disease resistance and the isolation of disease-resistance genes are important research goals.

Plant disease resistance genes (R-genes) are important components of the genetic resistance mechanism in plants (Flor, 1956). They are effective against diverse pathogens and pests, including bacteria, fungi, viruses, nematodes, and insects, and have been studied in detail in many plants like Arabidopsis, tobacco, tomato, flax, rice and maize (Grant et al., 1995; Gassmann et al., 1999; Parker et al., 1997; Jones et al., 1994; Dixon et al., 1998; Whitham et al., 1994; Lawrence et al., 1995; Wang et al., 1999: Johal and Briggs, 1992). During the past 15 years, over 70 *R* genes have been cloned and some of them have been well characterized. Within diverse *R* genes, most of them share a striking degree of homology on conserved motifs. They mainly include a nucleotide blinding site (NBS), leucine-rich repeat (LRR), a motif with homology to the cytoplasmic domains of the *Drosophila* Toll protein and the mammalian interleukin-1 receptor (TIR), a coiled-coil (CC) or leucine zipper (LZ) structure, transmembrane domain (TM), and protein kinase domain (PK) (Liu et al. 2007). Recent results indicate that these domains play significant roles in R protein interactions with effector proteins from pathogens and in activating signal transduction pathways involved in innate immunity. These *R* genes can be divided into at least four classes based on the structure of their encoded proteins (Liu et al., 2007): NBS-LRR, Receptor-like kinase (RLK), LRR-TM and TM-CC. The NBS-LRR genes represent the largest class of *R* genes (Martin et al., 2003, Belkhadir et al., 2004) and encode proteins with a variable N-terminal domain of approximately 200 amino acids, connected by a predicted NBS domain of approximately 300 amino acids (Traut, 1994) and a more variable tandem array of approximately 10 to 40 short LRR motifs (Jones et al., 1997). Furthermore, the NBS-LRR genes can be divided into three subgroups based on the motif within their N-terminus: TIR group (TIR-NBS-LRR; e.g. *L6*, *N* and *RPP5*), CC (CC-NBSLRR; e.g. *RPS2* and *RPM1*) and non-motif group (Pan et al., 2000, Steven et al., 2002). In the Arabidopsis Col-0 genome 149 NBS-LRR-encoding genes, which are 55 CC-NBS-LRR and 94 TIR-NBS-LRR genes, were discovered (Tan et al., 2007).

The NBS domain, which is also called the NB-ARC (nucleotide binding adaptor shared by NOD-LRR proteins, APAF-1, R proteins and CED4 domain) (McHale et al., 2006), contains three ATP/GTP-binding motifs known as P-loop or kinase 1a, kinase 2 and kinase 3a (Traut, 1994). The P-loop motif with the consensus sequence GXXXXGK (T/S) is involved in binding with phosphates and Mg21 ions (Saraste et al., 1990). The kinase 2, which functions in a phospho-transfer reaction, contains four consecutive hydrophobic amino acids followed by a conserved aspartate (D) residue, which coordinates the divalent metal ion on Mg-ATP. The kinase 3a motif is involved in binding purine or ribose and contains a conserved tyrosine (Y) or arginine (R) residue (Traut, 1994). The NBS domains of two NBS-LRR proteins, tomato I2 and Mi-1, have been demonstrated to be able to bind and hydrolyze ATP (Tameling et al., 2002). The ATP binding form is the active configuration of the I2 protein (Tameling et al., 2006) suggesting that the NBS domain functions as a molecular switch in signal transduction eliciting the defense response.

Usually, the LRR motif contains 20–29 residues in length with a conserved 11-residue segment LxxLxLxxN/cxL (x can be any amino acids and L can also substitute by valine, isoleucine and phenylalanine) (Bostjan and Andrey, 2001). The LRR domains are involved in protein-protein interactions in many important biological processes, such as hormone-receptor interactions, enzyme inhibition, cell adhesion and cellular trafficking (Bostjan and Andrey, 2001, Torii, 2004) and at least partly determine resistance specificity (Parniske et al., 1997, Thomas et al., 1997, Botella et al., 1998, McDowell et al., 1998, Meyers et al., 1998, Wang et al., 1999, Ellis et al., 1999, Noel et al., 1999, Dodds et al., 2001, Wulff et al., 2001, Hwang and Williamson, 2003 and Hulbert et al., 2001). Related evidence showed that the LRR domain is involved in determining the recognition specificity of several R proteins (Ellis et al., 1999, Wulff et al., 2001, Dodds et al., 2001, Hwang and Williamson, 2003). For example, the variation of LRR copy number in tomato LRR-TM genes *Cf-2*, *Cf-4*, *Cf-5 and Cf-9*, which confer resistance against the biotrophic leaf mold pathogen *Cladosporium fulvum*, determines their resistance specificity (Wulff et al., 2001). Furthermore, LRR domain exchange experiments between Cf-4/Cf-9 forcefully support that the LRR domain is critical in distinguishing resistance specificity. In addition to their role in determining the recognition specificity, the

LRR domains may also participate in defense signaling through both intra- and intermolecular interactions (Hwang and Williamson, 2003, Warren et al., 1998, Moffett et al., 2002, Belkhadir et al., 2004, Jones and Takemoto, 2004). However, whether this observation can be extended to other NBS-LRR genes requires further investigations.

The sizes of the core TIR domains vary considerably between 135 and 160 residues among the sequences reported so far (Xu et al., 2000). TIR domain interactions between receptors and adaptors play a critical role in activating conserved cellular signal transduction pathways in response to bacterial lipopolysaccharide (LPS), microbial and viral pathogens, cytokines and growth factors (Xu et al., 2000). The CC (also called LZ) serves as oligomerization domain for a wide variety of proteins including structural proteins, motor proteins and transcription factors (Nooren et al., 2001). The CC structure is conserved from viruses to plants and mammals and it has been predicted that approximately 5% of proteins encoded in sequenced genomes contain CCs (Nooren et al., 2001). CCs have been demonstrated to be important in protein-protein interactions (Nooren et al., 2001, Burkhard et al., 2001).

The TIR and CC domains mainly occur in the N-terminus of NBS-LRR genes. However, interestingly, the TIR type extensively occurs in dicots but is rare or absent in monocots, while the CC type is present in dicots and monocots (Pan et al., 2000, Bai et al., 2000, Meyers et al., 2003). In rice, there are 4 times more NBS-LRR genes in rice than in *Arabidopsis*, but no TIR type NBS-LRR genes have been identified in the rice genome (Bai et al., 2000, Monosi et al., 2004). The TIR and CC domains play an important role in R-Avr specific recognition to activate the downstream defense signaling in which two independent resistance signaling pathways, mediated through EDS1 and NDR1, respectively, have been identified (Shirasu et al., 2003). In the tobacco *N* gene, the TIR domain is essential for resistance to tobacco mosaic virus (TMV) (Mestre and Baul, 2006, Dinesh et al., 2000). Although TIR and CC domain proteins confer R-Avr proteins specific recognition, how the TIR and CC type proteins physically interact with the Avr proteins or other host proteins remains unclear.

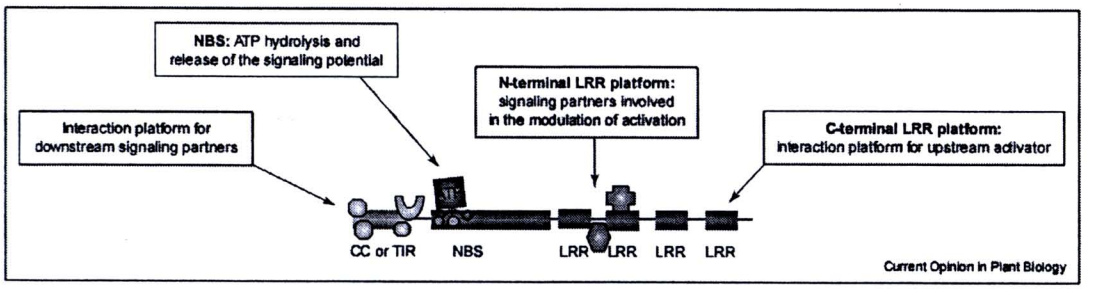

Figure 2.2 Domain structures of NBS–LRR proteins. A schematic representation of NBS–LRR proteins shows a domain-based platform for the assembly of various putative regulatory factors (Youssef et al., 2004).

Table 2.2 Conserved amino acids on disease resistance protein (Collins et al., 1998)

Conserved amino acid motif	Motif name
GVGKTT	P-loop
L(I/V/L)VLDDV	Kinase-2
GLPLAL	GLPL
KCFAFCSI	CFA
WMAG(F/I)V	WMA
MHD	MHD

2.4 Isolation and molecular cloning of plant disease resistance genes

The pace of cloning R genes increased beginning in 1994 with the cloning and characterization of R genes for resistance to several classes of pathogens, including viral, bacterial, and fungal pathogens (Staskawicz, 2001). The development of powerful techniques, specifically insertional mutagenesis and positional cloning techniques, has led to the cloning of several resistance genes from different plant species within the last 2-3 years. R genes from maize, tobacco, tomato and flax have been cloned using transposon-based gene tagging techniques. In addition R genes

from tomato and Arabidopsis have been cloned using the map-based positional cloning technique (Basim, 1997).

Table 2.3 First cloned plant disease resistance genes

Gene	Host	Pathogen	Methods	References
Hml	Maize	*Cochliobolus carbonum*	transposon-based gene tagging	Johal and Briggs, 1992
Pto	Tomato	*P. syringae* pv.tomato	map-based positional cloning	Martin et al., 1993
RPS2	Arabidopsis	*P. syringae* pv. maculicola	map-based positional cloning	Bent et. al., 1994
Prf	Tomato	*P. syringae* pv. tomato	map-based positional cloning	Salmeron et al., 1996
N	Tobacco	Tobacco Mosaic Virus	transposon-based gene tagging	Whitham et al., 1994
L6	Flax	Melamsporalini	transposon-based gene tagging	Ellis et al., 1995
Cf9	Tomato	*Clarosporium fulvum.*	transposon-based gene tagging	Jones et al., 1994
RPM1	Arabidopsis	*P. syringae* pv. maculicola	map-based positional cloning	Grant et al., 1995

All seem to encode components of signal transduction pathways. Many contain similar sequence motifs, even though they determine resistance to very different pathogens, including a virus, fungi, and bacteria. All, except *Pto*, encode

leucine-rich repeats (LRRs). *RPS2*, *RPM1*, *I2*, *N*, *RPP5*, and *L6*, but not *Cf9*, *Cf2*, or *Xa21*, also share sequences characteristic of the ATP/GTP binding sites (nucleotide binding site, NBS) of P-loop proteins (reviewed in Hammond-Kosack and Jones, 1997). The presence of these motifs and their similar organization among resistance gene products from plants as diverse as tobacco, tomato, rice, flax, and Arabidopsis suggest a common mechanism underlying disease resistance signal transduction throughout the plant kingdom. Two of the major challenges now are to piece together the remainder of the signaling pathways in which these resistance gene products act and to identify the determinants of specificity among proteins recognizing different pathogen species and strains (Salmeron et al., 1996).

However, isolation of R genes has historically involved map based cloning or transposon tagging, both of which are very labor-intensive and expensive strategies. (Graham et al., 2000). The presence of conserved domains in resistance genes presents the opportunity to clone numerous additional resistance genes from diverse species by polymerase chain reaction (PCR) with degenerate oligonucleotide primers to the conserved motifs. These predictions have been validated, as at 481 sequences have been identified that are putative R genes containing NBS/LRR domains from three sources: cloned R genes, homologous sequences in public databases, and degenerate PCR cloning (Meyers et al., 1999). Degenerate oligonucleotide primers have been used to isolate NBS-containing sequences from soybean, common bean, lentil, cowpea, pea, chickpea, alfalfa (reviewed in Yaish et al., 2004), solanaceous (potato: Leister et al., 1996; coffee: Noir et al., 2001; pepper: Pflieger et al., 1999; tomato: Seah et al., 2007; *Solanum caripense*: Trognitz and Trognitz, 2005). The identification of RGAs (resistance gene analogs) is potentially a powerful strategy for both the discovery of DNA markers closely linked to the disease-resistance loci for marker assisted selection (MAS) and the map-based cloning of these disease-resistance gene (Naik et al., 2006).

Control of organism would be best facilitated through the development of resistant cultivars. Unfortunately, there are few breeding program for Jerusalem artichoke and the resistant commercial cultivars are not currently available. As a consequence, little emphasis has been placed on breeding programs. Many diseases are difficult to select for in the field, therefore identification of markers associated

with disease resistance and the isolation of disease-resistance genes are important research goals. Furthermore, the characterization of genetic diversity among accessions in germplasm collections is extremely important for breeding programs. Selecting and assessing genotypes in germplasm molecular markers may optimize and facilitate breeding processes by separating closely related genotypes, thus increasing the efficiency and orientation of future crossings and genetic studies.

Because of the above knowledge on R genes in plant and the advantages of EST SSR markers and they are quickly obtained by electronic sorting, present in expressed regions of the genome, and relatively easy accessibility of large EST resources, increasing numbers of EST-SSR markers are now being identified and used for a variety of applications in a number of plant species like, grapes (Scott et al., 2000), sugarcane (Cordeiro et al., 2001), and cereals such as wheat, barley, rye, rice (Varshney et al., 2005). Nowadays, Jerusalem artichoke genome has not been sequenced; however 21,994 accessions of ESTs which are unigenes have been produced from Compositae Genome Project Database (retrieved May 2008, from: http://cgpdb. ucdavis.edu/cgpdb2/). We identified several RGCs by mining the Jerusalem artichoke-EST database. The analyses described here build a more complete picture of the diversity and distribution of repeat motifs in the ESTs. The preliminary investigation on genetic diversity of *H. tuberosus* genotypes obtainted from Plant Gene Resources of Canada.