UNIVERSITA' DEGLI STUDI DI NAPOLI "FEDERICO II"

DIPARTIMENTO DI AGRARIA

Sezione di Genetica e Biotecnologie Vegetali

Genomic approaches to trace the diversification history of important agronomic traits in plant

Candidato: Dott.re Antimo Di Donato

Supervisor: Prof.re Luigi Frusciante

Co-supervisor: Prof.ssa Maria Raffaella Ercolano

Genomic approaches to trace the diversification history of important agronomic traits in plant

Antimo Di Donato

University of Naples "Federico II", Department of Agricultural Sciences - Division of Plant genetics and biotechnology, Italy.

ABSTRACT

In order to investigate the diversification of important agronomic traits in plants, a conservation and evolution study of nucleotide binding genes from bacteria to plant kingdom was performed. The pathogen recognition genes were detected and classified in 102 organisms. In particular, the expansion and/or conservation of R-gene subgroups among organisms was investigated. Several large of NLR groups were found involved in important clustering events. A focus on orthologous pathogen recognition gene-rich regions in solanaceous species regions was also provided. A complete catalogue of eggplant (*Solanum melongena*) and pepper (*Capsicum annuum*) nucleotide-binding site (NBS), receptor-like protein (RLP) and receptor-like kinase (RLK) genes was generated and compared with tomato (*Solanum lycopersicum*) genomic repertoire. Orthologous relationships among clustering loci were found, and interesting reshuffling within given loci was observed for each analyzed species. The information obtained were integrated in a comparative map to highlight the evolutionary dynamics in which the PRG loci were involved. Diversification of 14 selected PRG-rich regions was also explored using a DNA target-enrichment approach. A large number of gene variants was found as well as rearrangements of single protein domain encoding sequences and changes in chromosome gene order among species. Lastly, whole-genome sequences of herbarium samples were compared to the genomes of modern tomato accessions to investigate the improvement history of the tomato crop in Italy and in Campania region in the last centuries. An aDNA extraction from herbarium tomato leaves was set up and successively used to perform aDNA sequencing sequenced. Several structural variants were detected in important genes of the ancient genomes. A comparison with a panel of wild and cultivated tomato was performed to shed light on genome pedigree history of European tomato. The findings of this thesis contribute to addressing several biological questions concerning the history of plant genome evolution and diversification.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 New challenges for crop breeding

Plant breeding efforts, from the domestication of wild plant species to the present, have played a significant role in providing the food, feed, fuel, and fiber for the development of human society that currently sustains more than 6 billion individuals living in the world (Hallauer 2011).

In last 50 years the traditional crop improvement allowed to increase yield and quality traits using massive agrochemical inputs in many species (Prohens 2011). Today, the changing climate and the growing global of population requires new solutions in development of supply and agricultural production. The food demand is estimated to increase at a rate of 100–110% between 2005 and 2050 and the agricultural production cannot be implemented by increasing the cultivated area, since it would have a strong environmental impact (Tilman et al. 2011). New varieties able to efficiently use resources in changing climate should be developed.

Recent advances in genomics field made available to the scientist and breeder several tools to study the genome and its relations with phenotype, giving the opportunity to repeat the revolution triggered by plant breeding in the 20th century. Standard genetic and breeding approach permits to study only few genes, mutations or agronomic traits at one time. The availability of huge omics data source and recent sequencing technologies may improve the discovery of genetic mutations in plant disease resistance genes and other important agronomical traits. Genomic approaches can elucidate the influence of genes or genomic regions on phenotype variations and evolution, giving the access to essential information for genetic improvement. In addition, omics data sources and NGS (Next Generation Sequencing) technologies could also accelerate the cloning and the editing of genes (Kim et al. 2014a; Steuernagel et al. 2016).

1.2 Sequencing technologies

The first sequencing methods, developed and spread in the seventies, were the Maxam and Gilbert method (Maxam & Gilbert 1977) and the Sanger method. The Sanger sequencing, based on chain-terminating dideoxynucleoside analogs that caused base-specific termination of primed DNA synthesis, had been the most widely used sequencing

method approach for at least 30 years and it remains in wide use for validation of newest techniques. The first genome sequence obtained from a eukaryotic organism was the mitochondrial human genome, published in 1981 using Sanger method (Anderson et al. 1981). The great advances in automation of DNA sequencing and the development of computer programs for the analysis of sequence data made possible the sequencing of eukaryotic genomes in the mid-80s. Chain termination sequencing of bacterial artificial chromosome (BAC)-based physical maps was the main used to perform genome sequences until first decade of this century (Bevan & Uauy 2013). In the last 10 years Next Generation Sequencing platforms, in particular the 454 (http://www.454.com) and Illumina (http://www.illumina.com), had a substantial reduction in cost per base pair and times.

NGS technologies allowed to complete several important sequencing projects of crops which were begin using old sequencing technology many years before (Garcia-Mas et al. 2012; Tomato & Consortium 2012). Therefore, numerous crop sequencing projects, which integrated different NGS technologies to exploit the advantages of each method, were launched (Xu et al. 2011; Tomato & Consortium 2012; Moghe et al. 2014). In recent years, due to the higher availability of genomic data from most important crops, it was also increased the sequencing and the re-sequencing of wild and cultivated plant genomes to improve the knowledge on crop traits. Data from plant genome sequences that can be used to develop markers, to improve the genetic mapping of agronomic traits, to detect of the genetic basis of interesting phenotypes, to reconstruction evolution or domestication of plants.

Targeted sequencing

The high automation of sequencing techniques has decreased the research costs, however, analyzing an entire genome is still challenging for little research projects (Clark et al. 2011). Genomic studies often require the analysis of dozens or hundreds samples, increasing costs further. For this reason, an alternative NGS approaches called target sequencing is quickly spreading. The term Targeted Sequencing refers to a set of techniques designed to isolate and to sequence a specific fraction of a genome. These techniques are well suited to the study of plant genomes for several reasons, primarily

fewer bases to be sequenced for a sample which means lower costs. Furthermore, the plant genomes due to high repetitive sequences tend to be very large, and often few genomic regions are associated with biological functions or agronomical traits (Kiialainen et al. 2011). There are different target sequencing techniques commercially available, among these the most popular are the hybridization-based sequence capture and the PCR amplification-based methods. In the first technologies, synthetic oligonucleotides are hybridized to regions of interest; in the second method, the region of interest are amplificated using PCR. The amplification in PCR-based method is very difficult for large genomic regions because the multiple primer pairs or probes required to cover several megabases of nucleotides. An additional problem is the allele drop-out, which occurs when a variant is located in a primer binding site hindering hybridization and stopping the amplification (Neves et al. 2013). Instead, hybridization-based method has no problems with long sequences. The hybridization-based approaches has been successfully applied to identification of mutations involved in human diseases, also it has been useful to link genetic variants to agricultural phenotypic traits of interest (Gasc et al. 2016). Other potential applications of this technique include population genomics, ancient genomics, non-model organism (Gasc et al. 2016) and isolation of new genes (Witek et al. 2016).

Ancient DNA sequencing

The remarkable progress in genetics and genomics lead to the creation new and fascinating fields of study, such as the analysis of ancient DNA. Ancient DNA (aDNA) can be extracted from biological archaeological and historical material, archival collections of herbarium or medical specimens, older than 75 years (Graham 2007). The field of ancient DNA studies was probably born in 1985 with the study of DNA material from the quagga, an extinct subspecies of plain zebra that lived in South Africa until the 19th century (Higuchi et al. 1984). This work had stimulated the study of DNA of all the oldest and best-preserved samples extracted from amber or sediments.

The nucleic acids extracted from ancient samples, unlike DNA of modern samples, had a low quality, which limits the achievable information. A number of factors promote the degradation of such genetic material, such as temperature, presence of water or air, high

pressure, exposure to light, biotic and abiotic contamination. In addition, old nucleic acids may contain a large number of post-mortem mutations as the deamination of cytosine, which increase with time and of genomic structure more susceptible to miscoding lesions, potentially leading to sequence errors, or physical destruction of the DNA molecule, thus increasing the risk for preferential amplification of exogenous contaminant sequences. Furthermore, the cytoplasmic DNA concentration is usually a thousand times higher than that of nuclear in an ancient sample (Rizzi et al. 2012). Lastly, modern human DNA and microbial DNA (ancient or modern) can contaminate aDNA samples. The described issues can influence the quality and quantity of ancient sample, DNA extraction, amplification and sequencing of aDNA. The problems that plague this field of investigation require, therefore, specific technical solutions.

Many aDNA studies on different organisms have elucidated important archaeological and evolutionary questions, showing patterns of crop domestication and migration (Der Sarkissian et al. 2015). In the last few years, the advent of new sequencing technologies have considerably increased the availability of aDNA data, thus could greatly improving our knowledge on crop evolution, adaptation and domestication. An additional fascinating aspect of aDNA investigation is the discovery of lost useful mutations that could be reintroduced in modern crops. There are different sources from which obtain plant aDNA, among these herbarium collections can be an excellent font of information. The ancient collections, preserving the ancient structure of the plant, can be used to correlate genomic data with observed phenotype. Several ancient plant genomes studies could be performed in the next future in order to elucidate the patterns of plant diversification and divergence. Last year, two studies on ancient barley (Mascher et al. 2016) and maize (Ramos-Madrigal et al. 2016) genomes provided significant insights related to domestication and origin of these modern crops.

1.3 Web platforms and bioinformatic tools for crop improvement

Basic informatic systems can provide information for facilitating many aspects of crop improvement. Several organizations share with scientists and breeders information regarding crops and their relative genomes on websites.

Nowadays, data from many plant sequencing project are available completely free on different web portals. NCBI database (http://www.ncbi.nlm.nih.gov/) is the most important for the content of omics data volumes. Other databases including plant genome sequences are Plant GDB (http://www.plantgdb.org) and Phytozome (http://www.phytozome.net). Databases are often created by the same organizations that guide the sequencing projects of a certain species or botanical family. The Arabidopsis Information Resource (TAIR) maintains a database that includes the complete genome sequence along with gene structure, gene product information, gene expression, DNA and information about the Arabidopsis research community. The Sol Genomics Network (SGN) is a family-oriented database dedicated to the Solanaceae family, the portal includes genetics and omics information about important crops such as tomato, potato, pepper and tobacco (Fernandez-Pozo et al. 2015). Some databases contain information about gene family correlated with specific agronomical traits such as PRGdb (Plant Resistance Genes database), which includes data about plant resistance genes, related pathogens and diseases (Sanseverino et al. 2009). The huge amount of data produced by omic and genetic studies, requires the development informatics tools (algorithms and software), capable of analyzing large volumes of data and simplify the study of complex biological traits.

In genetics and genomics, many bioinformatics tools were develop to browse genome sequences, analyze proteins or nucleotides, assembly or mapping reads, predict and annotate genes, perform comparative and evolutionary studies. Standard NGS technology produces short sequences typically called reads. They can be assembled using two approaches: de novo or mapping. The de novo method consist in assembling overlapped reads to create longer sequences (contigs, scaffolds or pseudomolecules). De-novo assemblies are slower and more memory demanding than mapping assemblies, but they are more much precise and exhaustive. Reads mapping allows to align sequences against an existing reference genome, building a sequence that is similar but not identical to the reference. Mapping approaches are faster than de novo assemblies, it allows to detect easily new structural variation, such as deletions, insertions and rearrangements (Li & Durbin 2009). After the mapping is possible to identify single nucleotide polymorphism (SNPs) and small InDel (insertion or the deletion of bases)Classification of proteins and extraction of motifs can be performed through a variety of tools such as Pfam (Bateman

et al. 2002), InterProScan (Jones et al. 2014) or SMART (Schultz et al. 1998). Alignment of proteins and genes is important to show similarities and differences in homolog sequences. The evolutionary history of individual gene families or plant species can be followed performing a comparative analysis.

1.4 Comparative and evolution analysis

Comparative analysis uses natural variations to understand the patterns of life at all levels - from genes to communities - and the historical relationships of individuals or higher taxa and the mechanisms and patterns that drives it (Hardison 2003). Natural variants in crop plants resulted mainly from spontaneous mutations in their wild progenitors. Crop domestication and breeding have a profound influence on the genetic diversity present in modern crops. Understanding the genetic basis of phenotypic variation and the domestication processes in crops can help us efficiently utilize these diverse genetic resources for crop improvement. The use of naturally occurring alleles has greatly increased agricultural production. Through the use of germplasm resources and genetic tools such as genome sequences, genetic populations and genome-wide association studies, crop researchers are now able to extensively and rapidly mine natural variation and associate phenotypic variation with the underlying sequence variants (Bevan & Uauy 2013). Recently, the advent of second-generation sequencing has facilitated the discovery and use of natural variation in crop design and genome-wide selection. The nearly completed sequences of plant species shed light on the history of genome evolution, and provide a foundation for advancing knowledge in many agronomically important plant species.

1.5 Genomic analysis of target traits

Crop breeders explore and use the variability of the germplasm collections to improve plant characteristics. Whether these traits are associated with yield, disease and insect resistance or quality traits they are all subjected to selection pressure. Like evolution this selection process is very slow for some traits or dramatically quick for other. Many favourable traits have been introgressed in the last years using empirical methods. The next step in genetic research would be the development of a theoretical framework that

allows reliable predictions of the phenotypic consequences when making alterations to the genome make-up of a plant (Hammer et al. 2006). Genomic information has increased exponentially during the past two decades and will enhance selection process. Optimistically, it seems further genetic progress can be sustained because as greater genetic information at the molecular level is understood and integrated with phenotypic selection (Hallauer 2011). Genomic methodologies showed to be useful to elucidate the basis of genetic traits/characteristics, to understand the phenotypic of important loci throughout the in crops belonging to Poaceae and Solanaceae species (Takeda & Matsuoka 2008). In terms of developmental aspects, terminal-branching pattern and fruit-size control seem to be the predominant determinants for the yield improvement of fruits and grains (Peng et al. 1999). They display a decrease in nucleotide diversity and increased LD after strong selection, such as during domestication and subsequent crop improvement. Recent screening showed that loci that loci controlling fruit size in tomato have been important selection targets (Chakrabarti et al. 2013). Domestication genes can identified by comparing nucleotide sequence diversity between a crop species and extant populations of wild relatives as a proxy of the ancestor species.

Plant disease resistance traits

Probably the most desired crop trait is the resistance to plant pathogens. Plant disease resistance is fundamental to obtain reliable production of food, and it provides significant reductions in agricultural use of land, water, fuel and other inputs. Plants defend them self from pathogens thorough a sophisticated defense system based on the ability of plants to distinguish the phytopathogen life-styles. The circular model describes the plant–pathogen interaction in three distinct phases: (1) interaction, (2) activation, and modulation (3) effective resistance. This model schematically showed the crucial points of two components (activation and modulation) of innate plant immunity and the resultant of their combination (Andolfo & Ercolano 2015). The activation component is essentially based on the presence at the cellular levels of specific pathogen receptors called R proteins. These proteins encoded by the pathogen recognition genes (PRGs), are characterized by some common domains such as CC (coiled-coil), NB (Nucleotide binding region), TIR (Toll-interleukin region), LRR (Leucine rich region) and K (Kinase

domain). The structures that have NB-LRR domains are divided into two classes: TNL (TIR-NB-LRR) and CNL (CC-NB-LRR) which possess, respectively, either the TIR or CC domains. TNL and CNL are usually present in the cytoplasm. The CNL group includes very important genes involved in crop disease resistance. Natural and cultivated plant populations carry inherent disease resistance. Monogenic or major gene (R gene) resistance, has been widely studied at genomic level (Sekhwal et al. 2015) and employed by breeders. New approaches for exploring resistance genes dataset could be useful for shed light in molecular and evolutionary mechanisms of this gene family and for facilitating the design of diagnostic tests, comparative analysis and new breeding program.

1.6 Scientific aims

Main goal of this thesis was to study the diversification of crop agronomical traits using genomic approaches. The first section is dedicated at the study of conservation and evolution of nucleotide binding genes from bacteria to plant kingdom. The second part reports a pilot comparison of orthologous pathogen recognition gene-rich regions in solanaceous species. In the third part ancient DNA extracted from two tomato herbarium samples was sequenced and analyzed to understand the selection routes followed by tomato growers in Campania region, with a focus on variation of candidate genes involved in determination of fruit quality traits.

2. RECONSTRUCTION OF EVOLUTIONARY HISTORY OF NLR-LIKE GENE FAMILY IN METAPHYTA KINGDOM

2.1 Introduction

Most intracellular immune receptors in plants are characterized by the presence of a nucleotide-binding site and leucine-rich repeats (NLRs, also known as, NB-LRRs or NBS-LRRs), these domains are present in the majority of cloned resistance genes (R-genes) (McHale et al. 2006). NLR protein families are divided into two classes based on the presence or absence of a toll-interleukin-1 receptor (TIR) domain in TIR-NLR (TNL) or non-TIR-NLR (n-TNL). Plant NB-LRR proteins detects the presence of fungal, nematode, bacterial, or viral pathogens elicitors and trigger the immune response. Both the NB and TIR domains have a prokaryotic origin but their fusion was observed only in plant lineage. Recent studies suggest the eukaryote innate immunity originated from their endosymbionts (Dunin-Horkawicz et al. 2014). Indeed, plant and animals system had independent origin shaped after by convergent evolution (Yue et al. 2012).

A large variation in NLR complement among and within plant species both in the sequence composition of orthologs and in the number of paralogs was observed (Y. Zhang et al. 2016). The number of NLR *genes* can vary in plant genomes from <100 to >1,000 (Yue et al. 2012; Sarris et al. 2016; Shao, Wang, et al. 2016) and some gene families are more conserved in dicots and lost or modified in monocots (Tarr & Alexander 2009; Collier et al. 2011). Although the structure and function of NLR proteins have been extensively studied, the involvement of single domain to disease resistance process in plants is still not well understood. Proteins can expand their functional repertoire in a number of ways, including residue mutations, gain and loss of domain, motif arrangement (Sarris et al. 2016). The domain arrangement is important since its modification mostly promote interactions with novel substrates or new protein partners on different pathways and processes and have specific functional and spatial relationship (Lees et al. 2016). A number of alterations that can have a considerable effect were already found (Sanseverino & Ercolano 2012).

On the following pages, it will be shown a study on genes that encode nucleotide-binding and/or leucine-rich repeat domains using data from genome sequencing of bacteria, algae and plants. A comprehensive study of genes encoding NLRs and NLR-like genes across bacterial and plant species can provide insights into the presumed history of plant NLR evolution and it can lead the discovery the means of NB protein diversification. The

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II
DIPARTIMENTO DI AGRARIA
DEPARTMENT OF AGRICULTURAL SCIENCES
Via Università, 100 – 80055 PORTICI (NA) Italia

Collegio dei Docenti del Corso di Dottorato di Ricerca in

SCIENZE AGRARIE E AGROALIMENTARI

XXIX Ciclo – XXX Ciclo

Verbale n.11

Il Collegio dei Docenti del Corso di Dottorato di Ricerca in Scienze Agrarie e Agroalimentari si è riunito in data 30 marzo 2017, alle ore 14.30 per l'esame del seguente

Ordine del giorno

1) Ammissione Esame Finale Dottorandi XXIX ciclo
2) Richieste di riservatezza per parti di tesi dottorato
3) Commissioni e date esami finali XXIX ciclo
4) Varie ed eventuali.

[omissis]

2) Richieste di riservatezza per parti di tesi dottorato

Il Dottorando Antimo Di Donato chiede con l'avvallo del tutor che – ai sensi dell'art.23 del vigente Regolamento di Dottorato, non vengano rese pubbliche per un periodo di 18 mesi le seguenti parti della tesi di dottorato sul sito FEDOA:

- da pagina 14 a pagina 29;
- da pagina 32 a pagina 53;
- da pagina 57 a pagina 73

in quanto sono presenti dati elaborati in collaborazione con aziende esterne soggetti a verifica di anteriorità a scopo di brevetto. Il Collegio approva.

[omissis]

Portici, 30 marzo 2017

Il Coordinatore
Prof. Guido D'Urso

3. COMPARISON OF SOLANACEAE ORTHOLOGOUS PATHOGEN RECOGNITION GENE-RICH REGIONS

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II
DIPARTIMENTO DI AGRARIA
DEPARTMENT OF AGRICULTURAL SCIENCES
Via Università, 100 – 80055 PORTICI (NA) Italia

Collegio dei Docenti del Corso di Dottorato di Ricerca in

SCIENZE AGRARIE E AGROALIMENTARI

XXIX Ciclo – XXX Ciclo

Verbale n.11

Il Collegio dei Docenti del Corso di Dottorato di Ricerca in Scienze Agrarie e Agroalimentari si è riunito in data 30 marzo 2017, alle ore 14.30 per l'esame del seguente

Ordine del giorno

1) Ammissione Esame Finale Dottorandi XXIX ciclo
2) Richieste di riservatezza per parti di tesi dottorato
3) Commissioni e date esami finali XXIX ciclo
4) Varie ed eventuali.

[omissis]

2) Richieste di riservatezza per parti di tesi dottorato

Il Dottorando Antimo Di Donato chiede con l'avvallo del tutor che – ai sensi dell'art.23 del vigente Regolamento di Dottorato, non vengano rese pubbliche per un periodo di 18 mesi le seguenti parti della tesi di dottorato sul sito FEDOA:

- da pagina 14 a pagina 29;
- da pagina 32 a pagina 53;
- da pagina 57 a pagina 73

in quanto sono presenti dati elaborati in collaborazione con aziende esterne soggetti a verifica di anteriorità a scopo di brevetto. Il Collegio approva.

[omissis]

Portici, 30 marzo 2017

Il Coordinatore
Prof. Guido D'Urso

4. INVESTIGATION OF EUROPEAN TOMATO IMPROVEMENT HISTORY THROUGH aDNA SEQUENCING

4.1 Introduction

Genetic analyses of ancient DNA have been used to dissect the genetic basis of traits underlying domestication in a wide range of organisms (Mascher et al. 2016). Current knowledge of plant domestication is largely derived from morphological analysis of archaeological and herbarium remains and/or population genetic analysis of present-day samples. Trace the selection history of a species can provide insights into the selection of important traits, facilitating both the management germplasm repository and the use of genetic resources (Blanca et al. 2015).

The evolutionary history of tomato (*Solanum lycopersicum*) has been clarified comparing genomes of cultivated varieties and wild species (Aflitos et al. 2014; Lin et al. 2014). Tomato domestication probably occurred in the Andean region of Ecuador and Peru and was completed in Mesoamerica (Blanca et al. 2012). Subsequently, a rapid evolution of populations under human selection led to conspicuous phenotypic transformations, as well as adaptations to varied environments (Bai & Lindhout 2007). Extensive breeding activities have modified tomato over the last centuries. Breeding was mainly focused on improving yield production, fruit quality and disease resistance traits. These efforts resulted in the introduction of many introgressions from tomato relatives and more distant wild species (Sim et al. 2011). Selection sweeps promoted the diversification and genetic differentiation in fresh and processing tomato market classes (Lin et al. 2014). The traits that most likely have been selected during the domestication of tomato were fruit morphological traits.

However, many questions about the events occurred during the domestication process remain unanswered. Notably, some changes in fruit shape became in 'modern' cultivars may originated after the tomato was brought to Europe about 500 years ago, albeit is not well understood when and where these alleles arose and how they spread through the germplasm. Multiple evolutionary processes in small cherry fruit, round large fruit, and elongated fruit have been postulated. For example, elongated accessions are evolutionary intermediates between large round and small size accessions (Lin et al. 2014). In recent years, several genes affecting these traits have been identified (Liu et al. 2002; Frary 2000; Xiao et al. 2008). Xiao et al asserted that elongated variants derived by Sun gene duplication (Xiao et al. 2008). However, other authors hypothesized that elongated

tomato fruits originated as hybrids between large round and small size tomato, and based on their distribution, they originate in Europe (Rodriguez et al. 2011). Furthermore, although several hypotheses have been proposed, the exact geographical origin of the elongated groups has not been established (Rodriguez et al. 2011). Small-scale aDNA studies can help to reveal patterns of crops adaptation and migration, however, they can't investigated the impact of these events on whole crop genomes. For this reason, whole genome scale studies on ancient genomes have been conducted in recent years, paving the way for many future studies in this fascinating field of research. Here it is reported the genome sequences of two tomato herbarium samples, which are part of the *Herbarium Porticense* collection (http://www.herbariumporticense.unina.it/it/). Whole-genome sequences of herbarium samples were compared to modern tomato accessions to reveal the relationship with wild and cultivated landraces and to investigate the improvement history of the tomato crop in Italy and in Campania region in the last centuries.

4.2 Materials and methods

Collection of Samples

The samples were taken from the *Herbarium Porticense* collection in MUSA Museum (http://www.centromusa.it/it/), University of Naples Federico II. The older samples were called SET17. According to the label, reporting information related to the identity of the species, the identity of the collector, the oldest herbarium material is 250 years old since it was collect in the eighteenth century in the historical herbaria of Neapolitan botanist Domenico Cirillo (Ricciardi & Castellano 2014a), at the time it was catalogued as "*Solanum (Lycopersicon)*". The second called LEO90 is part of the personal collection of botanist Orazio Comes (Ricciardi & Castellano 2014b), dated in 1890 and catalogued as "*Lycopersicum esculentum var. oblungum*".

aDNA extraction and PCR amplification

Total genomic DNA was isolated from herbarium leaves dated between 1700 and 1890. Approximately 0.005 g of tissue was ground in sterile 1.5 ml tubes using sterilized

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II
DIPARTIMENTO DI AGRARIA
DEPARTMENT OF AGRICULTURAL SCIENCES
Via Università, 100 – 80055 PORTICI (NA) Italia

Collegio dei Docenti del Corso di Dottorato di Ricerca in

SCIENZE AGRARIE E AGROALIMENTARI

XXIX Ciclo – XXX Ciclo

Verbale n.11

Il Collegio dei Docenti del Corso di Dottorato di Ricerca in Scienze Agrarie e Agroalimentari si è riunito in data 30 marzo 2017, alle ore 14.30 per l'esame del seguente

Ordine del giorno

1) Ammissione Esame Finale Dottorandi XXIX ciclo
2) Richieste di riservatezza per parti di tesi dottorato
3) Commissioni e date esami finali XXIX ciclo
4) Varie ed eventuali.

[omissis]

2) Richieste di riservatezza per parti di tesi dottorato

Il Dottorando Antimo Di Donato chiede con l'avvallo del tutor che – ai sensi dell'art.23 del vigente Regolamento di Dottorato, non vengano rese pubbliche per un periodo di 18 mesi le seguenti parti della tesi di dottorato sul sito FEDOA:

- da pagina 14 a pagina 29;
- da pagina 32 a pagina 53;
- da pagina 57 a pagina 73

in quanto sono presenti dati elaborati in collaborazione con aziende esterne soggetti a verifica di anteriorità a scopo di brevetto. Il Collegio approva.

[omissis]

Portici, 30 marzo 2017

Il Coordinatore
Prof. Guido D'Urso

6. REFERENCES

Aflitos, S. et al., 2014. Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing. *The Plant journal : for cell and molecular biology*, (January), pp.136–148. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25039268 [Accessed August 28, 2014].

Alexeyenko, A. et al., 2006. Automatic clustering of orthologs and inparalogs shared by multiple proteomes. In *Bioinformatics*.

Alvarez, M.E., Nota, F. & Cambiagno, D. a., 2010. Epigenetic control of plant immunity. *Molecular Plant Pathology*, 11(4), pp.563–576.

Ames, M. & Spooner, D.M., 2008. DNA from herbarium specimens settles a controversy about origins of the European potato. *American Journal of Botany*, 95(2), pp.252–257. Available at: http://www.amjbot.org/cgi/doi/10.3732/ajb.95.2.252.

Anderson, S. et al., 1981. Sequence and organization of the human mitochondrial genome. *Nature*, 290(5806), pp.457–465. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7219534%5Cnhttp://www.nature.com.libproxy.ucl.ac.uk/nature/journal/v290/n5806/pdf/290457a0.pdf.

Andolfo, G. et al., 2014. Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. *BMC plant biology*, 14(1), p.120. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4036795&tool=pmcentrez&rendertype=abstract [Accessed July 11, 2014].

Andolfo, G. et al., 2013. Genome-wide identification and analysis of candidate genes for disease resistance in tomato. *Molecular Breeding*, 33(1), pp.227–233. Available at: http://link.springer.com/10.1007/s11032-013-9928-7 [Accessed July 11, 2014].

Andolfo, G. et al., 2013. Overview of tomato (Solanum lycopersicum) candidate pathogen recognition genes reveals important Solanum R locus dynamics. *The New phytologist*, 197(1), pp.223–37. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/23163550.

Andolfo, G. & Ercolano, M.R., 2015. Plant Innate Immunity Multicomponent Model. *Frontiers in Plant Science*, 6(November), p.987. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4643146&tool=pmcentrez&rendertype=abstract.

Aversano, R. et al., 2015. *The* Solanum commersonii *Genome Sequence Provides Insights into Adaptation to Stress Conditions and Genome Evolution of Wild Potato Relatives*, Available at: http://www.plantcell.org/lookup/doi/10.1105/tpc.114.135954.

Bai, Y. & Lindhout, P., 2007. Domestication and Breeding of Tomatoes: What have We Gained and What Can We Gain in the Future? *Annals of Botany*, 100(5), pp.1085–1094. Available at: https://academic.oup.com/aob/article-lookup/doi/10.1093/aob/mcm150.

Bar-Or, C., Czosnek, H. & Koltai, H., 2007. Cross-species microarray hybridizations: a developing tool for studying species diversity. *Trends in Genetics*, 23(4), pp.200–207.

Bateman, A. et al., 2002. The Pfam Protein Families Database. , 30(1), pp.276–280.

Baumgarten, A. et al., 2003. Genome-level evolution of resistance genes in Arabidopsis thaliana. *Genetics*, 165(1), pp.309–319.

Bevan, M.W. & Uauy, C., 2013. Genomics reveals new landscapes for crop improvement. *Genome biology*, 14, p.206. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3706852&tool=pmcentrez&rendertype=abstract.

Blanca, J. et al., 2015. Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC genomics*, 16(1), p.257. Available at: http://www.biomedcentral.com/1471-2164/16/257%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4404671&tool=pmcentrez&rendertype=abstract.

Blanca, J. et al., 2012. Variation Revealed by SNP Genotyping and Morphology Provides Insight into the Origin of the Tomato W. Yan, ed. *PLoS ONE*, 7(10),

p.e48198. Available at: http://dx.plos.org/10.1371/journal.pone.0048198.

Bolger, A. et al., 2014. The genome of the stress-tolerant wild tomato species Solanum pennellii. *Nature Genetics*, 46(9), pp.1034–1038. Available at: http://www.nature.com/doifinder/10.1038/ng.3046 [Accessed July 28, 2014].

Boyko, A. et al., 2007. Transgenerational changes in the genome stability and methylation in pathogen-infected plants: (Virus-induced plant genome instability). *Nucleic Acids Research*, 35(5), pp.1714–1725.

Camacho, C. et al., 2009. BLAST plus: architecture and applications. *BMC Bioinformatics*, 10(421), p.1.

Caranta, C., Lefebvre, V. & Palloix, A., 1997. Polygenic Resistance of Pepper to Potyviruses Consists of a Combination of Isolate-Specific and Broad-Spectrum Quantitative Trait Loci. *Molecular Plant-Microbe Interactions*, 10(7), pp.872–878.

Chakrabarti, M. et al., 2013. A cytochrome P450 regulates a domestication trait in cultivated tomato. *Proceedings of the National Academy of Sciences of the United States of America*, 110(42), pp.17125–17130. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3801035&tool=pmcentrez&rendertype=abstract.

Chase, M.W. et al., 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, 181(1), pp.1–20.

Cingolani, P. et al., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w 1118; iso-2; iso-3. *Fly*, 6(2), pp.80–92.

Clark, M.J. et al., 2011. Performance comparison of exome DNA sequencing technologies. *Nature Biotechnology*, 29(10), pp.908–914. Available at: http://dx.doi.org/10.1038/nbt.1975.

Collier, S.M., Hamel, L.-P. & Moffett, P., 2011. Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. *Molecular plant-microbe interactions*, 24(8), pp.918–931.

Coppe, A., Danieli, G.A. & Bortoluzzi, S., 2006. REEF: searching REgionally Enriched Features in genomes. *BMC bioinformatics*, 7(1), p.453. Available at: http://www.biomedcentral.com/1471-2105/7/453 [Accessed July 16, 2014].

Dangl, J.L. & Jones, J.D.G., 2001. Plant pathogens and integrated defence responses to infection. *Nature*, 411(6839), pp.826–833. Available at: http://www.nature.com/doifinder/10.1038/35081161.

Destefanis, M. et al., 2015. A disease resistance locus on potato and tomato chromosome 4 exhibits a conserved multipartite structure displaying different rates of evolution in different lineages. *BMC plant biology*, 15, p.255. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4619397&tool=pmcentrez&rendertype=abstract.

Djian-Caporalino, C. et al., 2007. Root-knot nematode (Meloidogyne spp.) Me resistance genes in pepper (Capsicum annuum L.) are clustered on the P9 chromosome. *Theoretical and Applied Genetics*, 114(3), pp.473–486.

Dunin-Horkawicz, S., Kopec, K.O. & Lupas, A.N., 2014. Prokaryotic ancestry of eukaryotic protein networks mediating innate immunity and apoptosis. *Journal of Molecular Biology*, 426(7), pp.1568–1582.

Ercolano, M.R. et al., 2012. Genetic and genomic approaches for R-gene mediated disease resistance in tomato: Retrospects and prospects. *Plant Cell Reports*, 31, pp.973–985.

Fernandez-Pozo, N. et al., 2015. The Sol Genomics Network (SGN)-from genotype to phenotype to breeding. *Nucleic Acids Research*, 43(D1), pp.D1036–D1041.

Finn, R.D., Clements, J. & Eddy, S.R., 2011. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research*, 39(SUPPL. 2).

Frary, A., 2000. fw2.2: A Quantitative Trait Locus Key to the Evolution of Tomato Fruit Size. *Science*, 289(5476), pp.85–88. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10884229.

Gabriëls, S.H.E.J. et al., 2007. An NB-LRR protein required for HR signalling mediated by both extra- and intracellular resistance proteins. *Plant Journal*, 50(1), pp.14–28.

Garcia-Mas, J. et al., 2012. The genome of melon (Cucumis melo L.). *Proceedings of the National Academy of Sciences*, 109(29), pp.11872–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3406823&tool=pmcentrez&rendertype=abstract.

Gasc, C., Peyretaillade, E. & Peyret, P., 2016. Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. *Nucleic acids research*, (8), p.gkw309-. Available at: http://nar.oxfordjournals.org/content/early/2016/04/21/nar.gkw309.full.

González, V.M. et al., 2013. High presence/absence gene variability in defense-related gene clusters of Cucumis melo. *BMC genomics*, 14, p.782. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3845527&tool=pmcentrez&rendertype=abstract.

Graham, E.A.M., 2007. DNA reviews: Ancient DNA. *Forensic Science, Medicine, and Pathology*, 3(3), pp.221–225.

Grover, C.E., Salmon, A. & Wendel, J.F., 2012. Targeted sequence capture as a powerful tool for evolutionary analysis. *American Journal of Botany*, 99(2), pp.312–319.

Grube, R.C., Radwanski, E.R. & Jahn, M., 2000. Comparative genetics of disease resistance within the solanaceae. *Genetics*, 155(2), pp.873–887.

Gugerli, F., Parducci, L. & Petit, R.J., 2005. Ancient plant DNA: Review and prospects. *New Phytologist*, 166(2), pp.409–418.

Hallauer, A.R., 2011. Evolution of plant breeding. *Crop Breeding and Applied Biotechnology*, 11(3), pp.197–206. Available at: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1984-70332011000300001&lng=en&nrm=iso&tlng=en.

Hammer, G. et al., 2006. Models for navigating biological complexity in breeding improved crop plants. *Trends in Plant Science*, 11(12), pp.587–593.

Hardison, R.C., 2003. Comparative genomics. *PLoS Biology*, 1(2).

Hayashi, N. et al., 2010. Durable panicle blast-resistance gene Pb1 encodes an atypical

CC-NBS-LRR protein and was generated by acquiring a promoter through local genome duplication. *Plant Journal*, 64(3), pp.498–510.

Higuchi, R. et al., 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature*, 312(5991), pp.282–284.

Hirakawa, H. et al., 2014. Draft Genome Sequence of Eggplant (Solanum melongena L.): the Representative Solanum Species Indigenous to the Old World. *DNA research : an international journal for rapid publication of reports on genes and genomes*, pp.1–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25233906.

Hoberman, R. & Durand, D., 2005. The Incompatible Desiderata of Gene Cluster Properties. *Proc. of RECOMB 2005 International Workshop on Comparative Genomics, RCG 2005*, 3678, pp.73–87.

Hoberman, R., Sankoff, D. & Durand, D., 2005. The statistical significance of max-gap clusters. *Comparative Genomics*, pp.55–71. Available at: http://www.springerlink.com/index/VQQN6C6ECN9T0NM0.pdf.

Holub, E.B., 2001. The arms race is ancient history in Arabidopsis, the wildflower. *Nature reviews. Genetics*, 2(7), pp.516–527.

Jacob, F., Vernaldi, S. & Maekawa, T., 2013. Evolution and conservation of plant NLR functions. *Frontiers in Immunology*, 4(SEP).

Jia, Y. et al., 2015. Extreme expansion of NBS-encoding genes in Rosaceae. *BMC genetics*, 16, p.48. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4417205&tool=pmcentrez&rendertype=abstract.

Jones, P. et al., 2014. InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30(9), pp.1236–1240.

Joshi, R.K. & Nayak, S., 2013. Perspectives of genomic diversification and molecular recombination towards R-gene evolution in plants. *Physiology and Molecular Biology of Plants*, 19(1), pp.1–9.

Jupe, F. et al., 2012. Identification and localisation of the NB-LRR gene family within the potato genome. *BMC genomics*, 13, p.75. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3297505&tool=pmcentrez&rendertype=abstract.

Katoh, K. & Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), pp.772–780.

Kawchuk, L.M. et al., 2001. Tomato Ve disease resistance genes encode cell surface-like receptors. *Proceedings of the National Academy of Sciences of the United States of America*, 98(11), pp.6511–5. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=33499&tool=pmcentrez&rendertype=abstract.

Kearse, M. et al., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), pp.1647–1649.

Kiialainen, A. et al., 2011. Performance of microarray and liquid based capture methods for target enrichment for massively parallel sequencing and SNP discovery. *PloS one*, 6(2), p.e16486. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3036585&tool=pmcentrez&rendertype=abstract [Accessed November 24, 2014].

Kim, S. et al., 2014a. Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. *Nature genetics*, 46(3), pp.270–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24441736 [Accessed July 21, 2014].

Kim, S. et al., 2014b. Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. *Nature genetics*, 46(3), pp.270–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24441736.

Krzywinski, M. et al., 2009. Circos: An information aesthetic for comparative genomics. *Genome Research*, 19(9), pp.1639–1645.

Lees, J.G. et al., 2016. Functional innovation from changes in protein domains and their combinations. *Current Opinion in Structural Biology*, 38, pp.44–52.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools.

Bioinformatics, 25(16), pp.2078–2079.

Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), pp.1754–1760.

Li, J. et al., 2010. Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. *Molecular Genetics and Genomics*, 283(5), pp.427–438.

Lin, T. et al., 2014. Genomic analyses provide insights into the history of tomato breeding. *Nature genetics*, (October). Available at: http://www.ncbi.nlm.nih.gov/pubmed/25305757 [Accessed October 13, 2014].

Liu, J. et al., 2002. A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proceedings of the National Academy of Sciences*, 99(20), pp.13302–13306. Available at: http://www.pnas.org/content/99/20/13302.short.

Lu, Y., Huggins, P. & Bar-Joseph, Z., 2009. Cross species analysis of microarray expression data. *Bioinformatics*, 25(12), pp.1476–1483. Available at: http://bioinformatics.oxfordjournals.org/content/25/12/1476.full.

Luo, S. et al., 2012. Dynamic Nucleotide-Binding Site and Leucine-Rich Repeat-Encoding Genes in the Grass Family. *Plant Physiology*, 159(1), pp.197–210.

Marone, D. et al., 2013. Plant Nucleotide Binding Site–Leucine-Rich Repeat (NBS-LRR) Genes: Active Guardians in Host Defense Responses. *International Journal of Molecular Sciences*, 14(4), pp.7302–7326. Available at: http://www.mdpi.com/1422-0067/14/4/7302/.

Mascher, M. et al., 2016. Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nature Genetics*, (July). Available at: http://www.nature.com/doifinder/10.1038/ng.3611.

Maxam, a M. & Gilbert, W., 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74(2), pp.560–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/265521%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC392330.

McHale, L. et al., 2006. Plant NBS-LRR proteins: adaptable guards. *Genome biology*, 7,

p.212.

McHale, L.K. et al., 2012. Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant physiology*, 159(4), pp.1295–308. Available at: http://www.plantphysiol.org/content/159/4/1295.

Meyers, B.C. et al., 2003. Genome-wide analysis of NBS-LRR – encoding genes in Arabidopsis. *The Plant Cell*, 15(April), pp.809–834. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=152331&tool=pmcentrez&rendertype=abstract.

Meyers, B.C. et al., 1998. The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. *The Plant cell*, 10(11), pp.1817–1832.

Moghe, G.D. et al., 2014. Consequences of Whole-Genome Triplication as Revealed by Comparative Genomic Analyses of the Wild Radish Raphanus raphanistrum and Three Other Brassicaceae Species. *The Plant cell*, 26(5), pp.1925–1937. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4079359&tool=pmcentrez&rendertype=abstract [Accessed July 10, 2014].

Molinier, J. et al., 2006. Transgeneration memory of stress in plants. *Nature*, 442(7106), pp.1046–1049.

Nandety, R.S. et al., 2013. The role of TIR-NBS and TIR-X proteins in plant basal defense responses. *Plant physiology*, 162(3), pp.1459–72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3707564&tool=pmcentrez&rendertype=abstract.

De Natale, A., 2007. Herbarium Porticense. Available at: http://www.herbariumporticense.unina.it/doc/pdf/Erbario/Herbarium_Porticense.pdf.

de Natale, A. & Cellinese, N., 2009. Imperato, Cirillo, and a series of unfortunate events: A novel approach to assess the unknown provenance of historical herbarium specimens. *Taxon*, 58(3), pp.963–970.

Nazar, R.N. et al., 2010. DNA chip analysis in diverse organisms with unsequenced genomes. *Molecular Biotechnology*, 44(1), pp.8–13.

Neves, L.G. et al., 2013. Whole-exome targeted sequencing of the uncharacterized pine genome. *The Plant journal : for cell and molecular biology*, 75(1), pp.146–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23551702 [Accessed November 14, 2014].

Olsen, J.L. et al., 2016. The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea. *Nature*, 530(7590), pp.331–335. Available at: http://dx.doi.org/10.1038/nature16548%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/26814964%5Cnhttp://www.nature.com/doifinder/10.1038/nature16548.

Paradis, E., Claude, J. & Strimmer, K., 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2), pp.289–290.

Peart, J.R. et al., 2005. NRG1, a CC-NB-LRR protein, together with N, a TIR-NB-LRR protein, mediates resistance against tobacco mosaic virus. *Current Biology*, 15(10), pp.968–973.

Pease, J.B. et al., 2016. Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. *PLoS Biology*, 14(2), pp.1–24. Available at: http://dx.doi.org/10.1371/journal.pbio.1002379.

Peele, H.M. et al., 2014. Loss and retention of resistance genes in five species of the Brassicaceae family. *BMC plant biology*, 14(Cc), p.298. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4232680&tool=pmcentrez&rendertype=abstract.

Peng, J. et al., 1999. “Green revolution” genes encode mutant gibberellin response modulators. *Nature*, 400(July), pp.256–261.

Prohens, J., 2011. Plant Breeding: A Success Story to be Continued Thanks to the Advances in Genomics. *Frontiers in Plant Science*, 2(September), pp.1–3.

Qin, C. et al., 2014. Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. *Proceedings of the National Academy of Sciences of the United States of America*, 111(14), pp.5135–40. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3986200&tool=pmcentrez&rendertype=abstract.

Quinlan, A.R. & Hall, I.M., 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), pp.841–842.

Ramos-Madrigal, J. et al., 2016. Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into the Early Stages of Maize Domestication. *Current Biology*, pp.1–7. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0960982216311204.

Rehrig, W.Z. et al., 2014. *CaDMR1* Cosegregates with QTL *Pc5.1* for Resistance to in Pepper (*Capsicum annuum*). *The Plant Genome*, 7(2), pp.1–12. Available at: https://www.crops.org/publications/tpg/abstracts/7/2/plantgenome2014.03.0011.

Remm, M., Storm, C.E. & Sonnhammer, E.L., 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *Journal of molecular biology*, 314(5), pp.1041–1052.

Ricciardi, M. & Castellano, M.L., 2014a. Domenico Cirillo's Collections. *Nuncius*, 29(2), pp.499–530. Available at: http://booksandjournals.brillonline.com/content/journals/10.1163/18253911-02902008.

Ricciardi, M. & Castellano, M.L., 2014b. Domenico Cirillo's Collections. *Nuncius*, 29(2), pp.499–530.

Richly, E., Kurth, J. & Leister, D., 2002. Mode of amplification and reorganization of resistance genes during recent Arabidopsis thaliana evolution. *Molecular biology and evolution*, 19(1), pp.76–84.

Rizzi, E. et al., 2012. Ancient DNA studies: new perspectives on old samples. *Genetics, selection, evolution : GSE*, 44(1), p.21. Available at: http://www.gsejournal.org/content/44/1/21.

Rodriguez, G.R. et al., 2011. Distribution of SUN, OVATE, LC, and FAS in the Tomato Germplasm and the Relationship to Fruit Shape Diversity. *PLANT PHYSIOLOGY*, 156(1), pp.275–285. Available at: http://www.plantphysiol.org/cgi/doi/10.1104/pp.110.167577.

Sacco, A. et al., 2015. Exploring a tomato landraces collection for fruit-related traits by the aid of a high-throughput genomic platform. *PLoS ONE*, 10(9), pp.1–20.

Sanseverino, W. et al., 2009. PRGdb: A bioinformatics platform for plant resistance gene analysis. *Nucleic Acids Research*, 38(November 2009), pp.814–821.

Sanseverino, W. & Ercolano, M.R., 2012. In silico approach to predict candidate R proteins and to define their domain architecture. *BMC Research Notes*, 5(1), p.678. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3532234&tool=pmcentrez&rendertype=abstract.

Der Sarkissian, C. et al., 2015. Ancient genomics. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370(1660), p.20130387. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4275894&tool=pmcentrez&rendertype=abstract.

Sarris, P.F. et al., 2016. Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biology*, 14(1), p.8. Available at: http://www.biomedcentral.com/1741-7007/14/8.

Schultz, J. et al., 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proceedings of the National Academy of Sciences of the United States of America*, 95(11), pp.5857–64. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=34487&tool=pmcentrez&rendertype=abstract.

Sekhwal, M. et al., 2015. Disease Resistance Gene Analogs (RGAs) in Plants. *International Journal of Molecular Sciences*, 16(8), pp.19248–19290. Available at: http://www.mdpi.com/1422-0067/16/8/19248/.

Seo, E. et al., 2016. Genome-wide Comparative Analyses Reveal the Dynamic Evolution of Nucleotide-Binding Leucine-Rich Repeat Gene Family among Solanaceae Plants. *Frontiers in Plant Science*, 7(August), p.1205.

Shao, Z.-Q., Xue, J.-Y., et al., 2016. Large-Scale Analyses of Angiosperm Nucleotide-Binding Site-Leucine-Rich Repeat Genes Reveal Three Anciently Diverged Classes with Distinct Evolutionary Patterns. *Plant Physiology*, 170(4), pp.2095–2109. Available at: http://www.plantphysiol.org/lookup/doi/10.1104/pp.15.01487.

Shao, Z.-Q. et al., 2014. Long-Term Evolution of Nucleotide-Binding Site-Leucine-Rich Repeat (NBS-LRR) Genes: Understandings Gained From and Beyond the Legume Family. *Plant physiology*, 166(September), pp.217–234. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25052854.

Shao, Z.-Q., Wang, B. & Chen, J.-Q., 2016. Tracking ancestral lineages and recent expansions of NBS-LRR genes in angiosperms. *Plant Signaling & Behavior*, 2324(June), pp.00–00. Available at: http://www.tandfonline.com/doi/full/10.1080/15592324.2016.1197470.

Sim, S.-C. et al., 2011. Population structure and genetic differentiation associated with breeding history and selection in tomato (Solanum lycopersicum L.). *Heredity*, 106(6), pp.927–935. Available at: http://dx.doi.org/10.1038/hdy.2010.139.

Spoel, S.H. & Dong, X., 2012. How do plants achieve immunity? Defence without specialized immune cells. *Nature Reviews Immunology*, 12(2), pp.89–100. Available at: http://www.nature.com/doifinder/10.1038/nri3141.

Steuernagel, B. et al., 2016. Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nature Biotechnology*, (August 2015). Available at: http://www.nature.com/doifinder/10.1038/nbt.3543.

Stevens, M.A. & Rick, C.M., 1986. Genetics and breeding. In *The tomato crop*. Springer, pp. 35–109.

Takeda, S. & Matsuoka, M., 2008. Genetic approaches to crop improvement: responding to environmental and population changes. *Nature reviews. Genetics*, 9(6), pp.444–57. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18475268.

Tamura, K. et al., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12), pp.2725–2729.

Tarr, D.E.K. & Alexander, H.M., 2009. TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. *BMC research notes*, 2, p.197.

Tian, D. et al., 2003. Fitness costs of R-gene-mediated resistance in Arabidopsis thaliana. *Nature*, 423(6935), pp.74–7. Available at: http://dx.doi.org/10.1038/nature01588.

Tilman, D. et al., 2011. Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of Americaof the United States of America*, 108(50), pp.20260–20264. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3250154&tool=pmcentrez&rendertype=abstract.

Tomato, T. & Consortium, G., 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, 485(7400), pp.635–41. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3378239&tool=pmcentrez&rendertype=abstract.

Vossen, J.H., Jo, K.R. & Vosman, B., 2014. Mining the genus solanum for increasing disease resistance. In *Genomics of Plant Genetic Resources: Volume 2. Crop Productivity, Food Security and Nutritional Quality*. Springer, pp. 27–46.

Wang, G.-L. et al., 1998. Xa21D Encodes a Receptor-Like Molecule with a Leucine-Rich Repeat Domain That Determines Race-Specific Recognition and Is Subject to Adaptive Evolution. *The Plant Cell*, 10(5), p.765. Available at: http://www.jstor.org/stable/10.2307/3870663?origin=crossref.

Wang, Y. et al., 2008. Sequencing and comparative analysis of a conserved syntenic segment in the solanaceae. *Genetics*, 180(1), pp.391–408.

Wei, C., Chen, J. & Kuang, H., 2016. Dramatic Number Variation of R Genes in Solanaceae Species Accounted for by a Few R Gene Subfamilies. *Plos One*, 11(2), p.e0148708. Available at: http://dx.plos.org/10.1371/journal.pone.0148708.

Whitham, S. et al., 1994. The product of the tobacco mosaic virus resistance gene N: similarity to toll and the interleukin-1 receptor [published erratum appears in Cell 1995 May 5;81(3):466]. *Cell*, 78(Department of Plant Pathology, University of California, Berkeley 94720.), pp.1101–1115.

Wickham, H., 2011. ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(2), pp.180–185. Available at: http://doi.wiley.com/10.1002/wics.147.

Witek, K. et al., 2016. Accelerated cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing. *Nature Biotechnology*, (August 2015), pp.1–8.

Available at: http://www.nature.com/doifinder/10.1038/nbt.3540.

Wu, C.-H. et al., 2016. The NLR helper protein NRC3 but not NRC1 is required for Pto-mediated cell death in Nicotiana benthamiana. *New Phytologist*, 209, pp.1344–1352.

Xiao, H. et al., 2008. A Retrotransposon-Mediated Gene Duplication Underlies Morphological Variation of Tomato Fruit. *Science*, 319(2008), pp.1527–1530.

Xu, X. et al., 2011. Genome sequence and analysis of the tuber crop potato. *Nature*, 475(7355), pp.189–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21743474%5Cnhttp://dx.doi.org/10.1038/nature10158.

Yeaman, S., 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences of the United States of America*, 110(19), pp.E1743-51. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3651494&tool=pmcentrez&rendertype=abstract.

Yoshida, K. et al., 2013. The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine. *eLife*, 2(2), pp.1–25. Available at: http://elifesciences.org/lookup/doi/10.7554/eLife.00731.

Yue, J.-X. et al., 2012. Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. *New Phytologist*, 193(4), pp.1049–1063. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22212278.

Zhang, R. et al., 2014. Paleo-evolutionary plasticity of plant disease resistance genes. *BMC genomics*, 15(1), p.187. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24617999 [Accessed July 11, 2014].

Zhang, Y. et al., 2004. Expression of RPS4 in tobacco induces an AvrRps4-independent HR that requires EDS1, SGT1 and HSP90. *The Plant journal : for cell and molecular biology*, 40(2), pp.213–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15447648.

Zhang, Y. et al., 2016. The Diversification of Plant NBS-LRR Defense Genes Directs the Evolution of MicroRNAs That Target Them. *Molecular Biology and Evolution*,

p.msw154. Available at: http://mbe.oxfordjournals.org/lookup/doi/10.1093/molbev/msw154.

Zhang, Y.-M. et al., 2016. Uncovering the dynamic evolution of nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes in Brassicaceae. *Journal of Integrative Plant Biology*, 58(2), pp.165–177. Available at: http://doi.wiley.com/10.1111/jipb.12365.

Zhong, Y. & Cheng, Z.-M. (Max), 2016. A unique RPW8-encoding class of genes that originated in early land plants and evolved through domain fission, fusion, and duplication. *Scientific Reports*, 6(August), p.32923. Available at: http://www.nature.com/articles/srep32923.

SUPPLEMENTARY DATA

Organism	**Family**	**Other Taxonomic info**	**Source of data**
Ananas cosmus	Bromeliaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Aquilegia coerulea	Ranunculaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Arabidopsis halleri	Brasicaceaee	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Arabidopsis lyrata	Brasicaceaee	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Arabidopsis thaliana	Brasicaceaee	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Beta vulgaris	Chenopodiaceae	Viridiplantae	http://bvseq.molgen.mpg.de/Genome/Download/RefBeet-1.2/
Boechera stricta	Brasicaceaee	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Beta vulgaris	Chenopodiaceae	Viridiplantae	http://bvseq.molgen.mpg.de/Genome/Download/RefBeet-1.2/
Boechera stricta	Brasicaceaee	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Beta vulgaris	Chenopodiaceae	Viridiplantae	http://bvseq.molgen.mpg.de/Genome/Download/RefBeet-1.2/
Boechera stricta	Brasicaceaee	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Brachypodium distachyon	Poaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Brachypodium stacie	Poaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Bradyrhizobium diazoefficiens	Bradyrhizobiaceae	Bacteria	https://phytozome.jgi.doe.gov/pz/portal.html
Brassica rapa	Brasicaceaee	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Capsella grandiflora	Brasicaceaee	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Capsella rubella	Brasicaceaee	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Capsicum annuum Zunla	Solanaceae	Viridiplantae	ftp://ftp.solgenomics.net
Capsicum anuum CM334	Solanaceae	Viridiplantae	ftp://ftp.solgenomics.net

Carica papaya	Caricaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Chlamydomonas reinhardtii	Chlamydomonadaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Chloroflexus aurantiacus	Chloroflexaceae	Bacteria	ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/Chloroflexus_aurantiacus/
Chloroherpeton thalassium	Ignavibacteriaceae	Bacteria	ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/Chloroherpeton_thalassium
Chondrus crispus	Gigartinaceae	Red alga	http://ftp.gramene.org/CURRENT_RELEASE/data/fasta/Viridiplantaes_rhodophyta1_collection/
Citrullus lanatus	Cucurbitaceae	Viridiplantae	ftp://www.icugi.org/pub/genome/watermelon/97103/v1/
Citrus clementina	Rutaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Citrus sinensis	Rutaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Clostridium cellulovorans	Clostridiaceae	Bacteria	ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/Clostridium_cellulovorans/
Coccomyxa subellipsoidea	Coccomyxaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Coffea canephora	Rubiaceae	Viridiplantae	http://coffee-genome.org/download
Cucumis melo	Cucurbitaceae	Viridiplantae	https://melonomics.net/files/Genome/Melon_genome_v3.5_Garcia-Mas_et_al_2012/
Cucumis sativus	Cucurbitaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Cyanophora paradoxa	Glaucocystaceae	//	http://cyanophora.rutgers.edu/cyanophora/Cyanophora_CLC_112010.fasta
Dacus carota	Apiaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Dunaliella salina	Dunaliellaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Ectocarpus siliculosus	Ectocarpaceae	Brown Alga	https://bioinformatics.psb.ugent.be/gdb/ectocarpus/
Eragrostis tef	Poaceae	Viridiplantae	http://130.92.252.158/tef/version1/
Eucalyptus grandis	Myrtaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Eutrema salsugineum	Brassicaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Fragaria vesca	Rosaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Galdieria sulphuraria	Galdieriaceae	Red alga	http://ftp.gramene.org/CURRENT_RELEASE/data/fasta/Viridiplantaes_rhodophyta1_collection/
Gloeobacter violaceus	Gviolaceus	Bacteria	http://www.uniprot.org/taxonomy/251221
Glycine max	Fabaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Gossypium raimondii	Malvaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Hordeum vulgare	Poaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Kadua laxiflora	Rubiaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Kalanchoe marnieriana	Crassulaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Lactuca sativa	Asteraceae	Viridiplantae	http://gviewer.gc.ucdavis.edu/fgb2/gbrowse/lechuga_version_1_2/
Linum usitatissimum	Linaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Lotus japonicus	Fabaceae	Viridiplantae	ftp://ftp.kazusa.or.jp/pub/lotus/lotus_r3.0/
Malus domestica	Rosaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Manihot esculenta	Euphorbiaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Marchantia polymorpha	Marchantiaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Medicago truncatula	Fabaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Methanosarcina mazei go1	Methanosarcinaceae	Bacteria	ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/Methanosarcina_mazei

Micromonas pusilla	Mamiellaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Micromonas sp.RCC299	Mamiellaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Mimulus guttatus	Phrymaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Musa acuminata	Musaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Nicotiana benthamiana	Solanaceae	Viridiplantae	ftp://ftp.solgenomics.net
Nicotiana sylvestris	Solanaceae	Viridiplantae	ftp://ftp.solgenomics.net
Nicotiana tabacum	Solanaceae	Viridiplantae	ftp://ftp.solgenomics.net
Nostoc punctiforme PCC 73102	Nostocaceae	Bacteria	ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/Nostoc punctiforme
Olea europea	Oleaceae	Viridiplantae	ftp://climb.genomics.cn/pub/10.5524/100001_101000/100201/
Oropetium thomaeum	Poaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Oryza sativa	Poaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Ostreococcus lucimarinus	Bathycoccaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Panicum hallii	Poaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Panicum virgatum	Poaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Petunia axilaris	Solanaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Petunia infilata	Solanaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Phaseolus vulgaris	Fabaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Phoenix dactylifera	Arecaceae	Viridiplantae	ftp://ftp.ncbi.nlm.nih.gov/genomes/Phoenix_dactylifera/GFF
Phyllostachys heterocycla	Poaceae	Viridiplantae	http://202.127.18.221/bamboo/down.php
Physcomitrella patens	Funariaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Picea abies	Pinaceae	Viridiplantae	http://congenie.org/start
Pinus taeda	Pinaceae	Viridiplantae	http://dendrome.ucdavis.edu/ftp/Genome_Data/genome/pinerefseq/Pita/v1.01/gene_models/
Populus trichocarpa	Salicaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Prunus Persica	Rosaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Rhodopirellula baltica	Planctomycetaceae	Bacteria	http://www.genome.jp/dbget-bin/get_linkdb?uniprot:Q7UEH8_RHOBA
Ricinus communis	Euphorbiaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Salix purpurea	Salicaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Selaginella moellendorffii	Selaginellaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Sesamum indicum	Pedaliaceae	Viridiplantae	http://ocri-genomics.org/Sinbase/login.htm
Setaria italica	Poaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Setaria viridis	Poaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Solanum lycopersicum	Solanaceae	Viridiplantae	ftp://ftp.solgenomics.net
Solanum melongena	Solanaceae	Viridiplantae	ftp://ftp.solgenomics.net
Solanum pennellii	Solanaceae	Viridiplantae	ftp://ftp.solgenomics.net
Solanum pimpinellifolium	Solanaceae	Viridiplantae	ftp://ftp.solgenomics.net
Solanum tuberosum phureja	Solanaceae	Viridiplantae	ftp://ftp.solgenomics.net
Sorghum bicolor	Poaceae	Viridiplantae	ftp://ftp.solgenomics.net

Spirodela polyrhiza	Aracaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Thauera aminoaromatica S2	Rhodocyclacea e	Bacteria	ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/Thauera_aminoaromatica
Theobroma cacao	Malvaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Trifolium pratense	Fabaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Triticum aestivum	Poaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Triticum urartu	Poaceae	Viridiplantae	ftp://ftp.ensemblgenomes.org/pub/Viridiplantaes/release-33/fasta/triticum_urartu/
Vitis vinifera	Vitaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Volvox carteri	Volvocaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Zea mays	Poaceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html
Zostera marina	Zosteraceae	Viridiplantae	https://phytozome.jgi.doe.gov/pz/portal.html

Tables S1. The 102 sequenced genomes used for identification of NLR-like genes and their download sources.