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** ................................................................. 3
   1.1 New challenges for crop breeding .................................. 4
   1.2 Sequencing technologies ............................................. 4
   1.3 Web platforms and bioinformatic tools for crop improvement .......................... 7
   1.4 Comparative and evolution analysis ................................. 9
   1.5 Genomic analysis of target traits ................................... 9
   1.6 Scientific aims ......................................................... 11

2. **RECONSTRUCTION OF EVOLUTIONARY HISTORY OF NLR-LIKE GENE FAMILY IN METAPHYTA KINGDOM** ......................................................... 12
   2.1 Introduction .............................................................. 13
   2.2 Materials and methods ............................................... 14
   2.3 Results ................................................................. 15
   2.4 Discussions ............................................................ 27

3. **COMPARISON OF SOLANACEAE ORTHOLOGOUS PATHOGEN RECOGNITION GENE-RICH REGIONS** ................................................................. 30
   3.1 Introduction .............................................................. 31
   3.2 Materials and methods ............................................... 32
   3.3 Results ................................................................. 36
   3.4 Discussions ............................................................ 50

4. **INVESTIGATION OF EUROPEAN TOMATO IMPROVEMENT HISTORY THROUGH aDNA SEQUENCING** ................................................................. 54
   4.1 Introduction .............................................................. 55
   4.2 Materials and methods ............................................... 56
   4.3 Results ................................................................. 58
   4.4 Discussions ............................................................ 69

5. **CONCLUSIONS AND PERSPECTIVES** ..................................... 72
6. **REFERENCES** ...................................................................... 75

**SUPPLEMENTARY DATA** .......................................................... 90
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 (Kiiialainen 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 amplified 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 Phytosome (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
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
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
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
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


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:


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.


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.


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.


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:


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


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.


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


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.


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


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,


### SUPPLEMENTARY DATA

<table>
<thead>
<tr>
<th>Organism</th>
<th>Family</th>
<th>Other Taxonomic info</th>
<th>Source of data</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ananas comosus</em></td>
<td>Bromeliaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Aquilegia coerulea</em></td>
<td>Ranunculaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Arabidopsis halleri</em></td>
<td>Brassicaceee</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Arabidopsis lyrata</em></td>
<td>Brassicaceee</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>Brassicaceee</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Beta vulgaris</em></td>
<td>Chenopodiaceae</td>
<td>Viridiplantae</td>
<td><a href="http://bvseq.molgen.mpg.de/Genome/Download/RefBeet-1.2/">http://bvseq.molgen.mpg.de/Genome/Download/RefBeet-1.2/</a></td>
</tr>
<tr>
<td><em>Boechera stricta</em></td>
<td>Brassicaceee</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Beta vulgaris</em></td>
<td>Chenopodiaceae</td>
<td>Viridiplantae</td>
<td><a href="http://bvseq.molgen.mpg.de/Genome/Download/RefBeet-1.2/">http://bvseq.molgen.mpg.de/Genome/Download/RefBeet-1.2/</a></td>
</tr>
<tr>
<td><em>Brachypodium distachyon</em></td>
<td>Poaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Brachypodium stacie</em></td>
<td>Poaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Bradyrhizobium diazoefficiens</em></td>
<td>Bradyrhizobium diaeae</td>
<td>Bacteria</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Brassica rapa</em></td>
<td>Brassicaceee</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Capsella grandiflora</em></td>
<td>Brassicaceee</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Capsella rubella</em></td>
<td>Brassicaceee</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td>Taxonomy</td>
<td>Class</td>
<td>Kingdom</td>
<td>URL</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>---------------------</td>
<td>------------------</td>
<td>---------------------------------------------------------------------</td>
</tr>
<tr>
<td>Carica papaya</td>
<td>Caricaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td>Chlamydomonas reinhardtii</td>
<td>Chlamydomonadae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td>Citrullus lanatus</td>
<td>Cucurbitaceae</td>
<td>Viridiplantae</td>
<td>ftp://www.icugi.org/pub/Genome/Watermelon/97103/v1/</td>
</tr>
<tr>
<td>Citrus clementina</td>
<td>Rutaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Citrus sinensis</td>
<td>Rutaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Coccomyxa subellipsoidea</td>
<td>Coccomyxaaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Coffea canephora</td>
<td>Rubiaceae</td>
<td>Viridiplantae</td>
<td><a href="http://coffee-genome.org/download">http://coffee-genome.org/download</a></td>
</tr>
<tr>
<td>Cucumis melo</td>
<td>Cucurbitaceae</td>
<td>Viridiplantae</td>
<td><a href="https://melonomics.net/files/Genome/Melon_genome_v3.5_Garcia-Mas_et_al_2012/">https://melonomics.net/files/Genome/Melon_genome_v3.5_Garcia-Mas_et_al_2012/</a></td>
</tr>
<tr>
<td>Cucumis sativus</td>
<td>Cucurbitaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Cyanophora paradoxa</td>
<td>Glaucocystaceae</td>
<td>//</td>
<td><a href="http://cyanophora.rutgers.edu/cyanophora/Cyanophora_CLC_112010.fasta">http://cyanophora.rutgers.edu/cyanophora/Cyanophora_CLC_112010.fasta</a></td>
</tr>
<tr>
<td>Dacus carota</td>
<td>Apiaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Dunaliella salina</td>
<td>Dunaliellaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Eragrostis tef</td>
<td>Poaceae</td>
<td>Viridiplantae</td>
<td><a href="http://130.92.252.158/tef/version1/">http://130.92.252.158/tef/version1/</a></td>
</tr>
<tr>
<td>Eucalyptus grandis</td>
<td>Myrtaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Extrema salguineum</td>
<td>Brassicaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Fragaria vesca</td>
<td>Rosaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Gloeobacter violaceus</td>
<td>Giviolaceae</td>
<td>Bacteria</td>
<td><a href="http://www.uniprot.org/taxonomy/251221">http://www.uniprot.org/taxonomy/251221</a></td>
</tr>
<tr>
<td>Glycine max</td>
<td>Fabaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Gossypium raimondii</td>
<td>Malvaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Hordeum vulgare</td>
<td>Poaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Kadua laxiflora</td>
<td>Rubiaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Kalanchoe marnieriana</td>
<td>Crassulaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Lactuca sativa</td>
<td>Asteraceae</td>
<td>Viridiplantae</td>
<td><a href="http://gviewer.gc.ucdavis.edu/fgb2/gbrowse/lechuga_version_1_2/">http://gviewer.gc.ucdavis.edu/fgb2/gbrowse/lechuga_version_1_2/</a></td>
</tr>
<tr>
<td>Linum usitatissimum</td>
<td>Linaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Lotus japonicus</td>
<td>Fabaceae</td>
<td>Viridiplantae</td>
<td>ftp://ftp.kazusa.or.jp/pub/lotus/lotus_r3.0/</td>
</tr>
<tr>
<td>Malus domestica</td>
<td>Rosaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Manihot esculenta</td>
<td>Euphorbiaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Marchantia polymorpha</td>
<td>Marchantiaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td>Medicago truncatula</td>
<td>Fabaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/page.html">https://phytozome.jgi.doe.gov/pz/page.html</a></td>
</tr>
<tr>
<td><strong>Micromonas pusilla</strong></td>
<td>Mamiellaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td><strong>Micromonas sp.</strong> RCC299</td>
<td>Mamiellaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Minimus guttatus</strong></td>
<td>Phrymaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Musa acuminate</strong></td>
<td>Musaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Nicotiana benthamiana</strong></td>
<td>Solanaceae</td>
<td>Viridiplantae</td>
<td>ftp://ftp.solgenomics.net</td>
</tr>
<tr>
<td><strong>Nicotiana sylvestris</strong></td>
<td>Solanaceae</td>
<td>Viridiplantae</td>
<td>ftp://ftp.solgenomics.net</td>
</tr>
<tr>
<td><strong>Nicotiana tabacum</strong></td>
<td>Solanaceae</td>
<td>Viridiplantae</td>
<td>ftp://ftp.solgenomics.net</td>
</tr>
<tr>
<td><strong>Orthepium thomaeum</strong></td>
<td>Poaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Oryza sativa</strong></td>
<td>Poaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Ostreococcus lucimarinus</strong></td>
<td>Bathycoccaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Panicum hallii</strong></td>
<td>Poaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Panicum virgatum</strong></td>
<td>Poaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Petonia asilari</strong></td>
<td>Solanaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Petonia inflata</strong></td>
<td>Solanaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Phaseolus vulgaris</strong></td>
<td>Fabaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Phyllostachys heterocycla</strong></td>
<td>Poaceae</td>
<td>Viridiplantae</td>
<td><a href="http://202.127.18.221/bamboo/down.php">http://202.127.18.221/bamboo/down.php</a></td>
</tr>
<tr>
<td><strong>Physcomitrella patens</strong></td>
<td>Funariaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Picea abies</strong></td>
<td>Pinaceae</td>
<td>Viridiplantae</td>
<td><a href="http://congenie.org/start">http://congenie.org/start</a></td>
</tr>
<tr>
<td><strong>Pinus taeda</strong></td>
<td>Pinaceae</td>
<td>Viridiplantae</td>
<td><a href="http://dendrome.ucdavis.edu/ftp/Genome_Data/genome/pine_refseq/Pita/v1.01/gene_models/">http://dendrome.ucdavis.edu/ftp/Genome_Data/genome/pine_refseq/Pita/v1.01/gene_models/</a></td>
</tr>
<tr>
<td><strong>Populus trichocarpa</strong></td>
<td>Salicaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Prunus Persica</strong></td>
<td>Rosaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Ricinus communis</strong></td>
<td>Euphorbiaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Salix purpurea</strong></td>
<td>Salicaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Selaginella moellendorfii</strong></td>
<td>Selaginellaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Sesamum indicum</strong></td>
<td>Pedaliaceae</td>
<td>Viridiplantae</td>
<td><a href="http://oci-genomics.org/Sinbase/login.htm">http://oci-genomics.org/Sinbase/login.htm</a></td>
</tr>
<tr>
<td><strong>Setaria italica</strong></td>
<td>Poaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Setaria viridis</strong></td>
<td>Poaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><strong>Solanum lycopersicum</strong></td>
<td>Solanaceae</td>
<td>Viridiplantae</td>
<td>ftp://ftp.solanomics.net</td>
</tr>
<tr>
<td><strong>Solanum melongena</strong></td>
<td>Solanaceae</td>
<td>Viridiplantae</td>
<td>ftp://ftp.solanomics.net</td>
</tr>
<tr>
<td><strong>Solanum pennellii</strong></td>
<td>Solanaceae</td>
<td>Viridiplantae</td>
<td>ftp://ftp.solanomics.net</td>
</tr>
<tr>
<td><strong>Solanum pimpinellifolium</strong></td>
<td>Solanaceae</td>
<td>Viridiplantae</td>
<td>ftp://ftp.solanomics.net</td>
</tr>
<tr>
<td><strong>Solanum tuberosum</strong></td>
<td>Solanaceae</td>
<td>Viridiplantae</td>
<td>ftp://ftp.solanomics.net</td>
</tr>
<tr>
<td><strong>Sorghum bicolor</strong></td>
<td>Poaceae</td>
<td>Viridiplantae</td>
<td>ftp://ftp.solanomics.net</td>
</tr>
<tr>
<td>Species</td>
<td>Family</td>
<td>Division</td>
<td>Website</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td><em>Spirodela polyrhiza</em></td>
<td>Aracaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Theobroma cacao</em></td>
<td>Malvaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Trifolium pratense</em></td>
<td>Fabaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Triticum aestivum</em></td>
<td>Poaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Vitis vinifera</em></td>
<td>Vitaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Volvox carteri</em></td>
<td>Volvocaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Zea mays</em></td>
<td>Poaceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
<tr>
<td><em>Zostera marina</em></td>
<td>Zosteraceae</td>
<td>Viridiplantae</td>
<td><a href="https://phytozome.jgi.doe.gov/pz/portal.html">https://phytozome.jgi.doe.gov/pz/portal.html</a></td>
</tr>
</tbody>
</table>

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