Characterization of Resistance Gene Analogues (RGAs) in Apple (*Malus × domestica* Borkh.) and Their Evolutionary History of the Rosaceae Family

Michele Perazzoli1,3, Giulia Malacarne1,3, Angela Baldo2, Laura Righetti1, Aubrey Bailey2, Paolo Fontana1, Riccardo Velasco1, Mickael Malnoy1*

1 Research and Innovation Centre, Fondazione Edmund Mach, San Michele all’Adige, Italy, 2 United States Department of Agriculture-Agricultural Research Service Plant Genetic Resources Unit, Geneva, New York, United States of America

**Abstract**

The family of disease resistance gene analogues (RGAs) with a nucleotide-binding site (NBS) domain accounts for the largest number of disease resistance genes and is one of the largest gene families in plants. We have identified 868 RGAs in the genome of the apple (*Malus × domestica* Borkh.) cultivar ‘Golden Delicious’. This represents 1.51% of the total number of predicted genes for this cultivar. Several evolutionary features are pronounced in *M. domestica*, including a high fraction (80%) of RGAs occurring in clusters. This suggests frequent tandem duplication and ectopic translocation events. Of the identified RGAs, 56% are located preferentially on six chromosomes (Chr 2, 7, 8, 10, 11, and 15), and 25% are located on Chr 2. TIR-NBS and non-TIR-NBS classes of RGAs are primarily exclusive of different chromosomes, and 99% of non-TIR-NBS RGAs are located on Chr 11. A phylogenetic reconstruction was conducted to study the evolution of RGAs in the Rosaceae family. More than 1400 RGAs were identified in six species based on their NBS domain, and a neighbor-joining analysis was used to reconstruct the phylogenetic relationships among the protein sequences. Specific phylogenetic clades were found for RGAs of *Malus*, *Fragaria*, and *Rosa*, indicating genus-specific evolution of resistance genes. However, strikingly similar RGAs were shared in *Malus*, *Pyrus*, and *Prunus*, indicating high conservation of specific RGAs and suggesting a monophyletic origin of these three genera.

> Citation: Perazzoli M, Malacarne G, Baldo A, Righetti L, Bailey A, et al. (2014) Characterization of Resistance Gene Analogues (RGAs) in Apple (*Malus × domestica* Borkh.) and Their Evolutionary History of the Rosaceae Family. PLoS ONE 9(2): e83844. doi:10.1371/journal.pone.0083844

**Editor:** Leandro Peña, Instituto Valenciano de Investigaciones Agrarias, Spain

**Received:** June 26, 2013; **Accepted:** November 17, 2013; **Published:** February 5, 2014

**Copyright:** © 2014 Perazzoli et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The research was supported by the Provincia Autonoma di Trento. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: mickael.malnoy@fmach.it

---

**Introduction**

When a genome sequence is available, the analysis of large gene families can contribute to the understanding of major events responsible for molecular evolution. This is the case for resistance gene analogues (RGAs) with a nucleotide-binding site (NBS) domain [1–5]. The NBS domain is part of the larger NB-ARC domain that hydrolyses ATP and GTP and functions as a molecular switch for signal transduction after pathogen recognition [6]. Many resistance proteins encoded by RGAs contain a leucine-rich repeat (LRR) domain [7,8], involved in protein–protein interactions and in pathogen recognitions [9]. Proteins codified by RGAs can be further classified according to the presence of the toll/interleukin-1 receptor (TIR) or other N-terminal features, such as coiled-coil (CC) and BED finger (Bed) [3,10,11]. The N-terminal features are involved in downstream specificity and signaling regulation [12]. RGAs evolved for pathogen recognition and frequently matched with specific pathogen avirulence factors to trigger signal transduction cascades and defense responses [9].

The genome sequencing of model plants has enabled the study of RGAs families in monocots and dicots, including *Arabidopsis thaliana* [11,13], *Brassica rapa* [14], *Carica papaya* [15,16], *Cucumis sativus* [17], *Glycine max* [18,19], *Zea mays* [20,21], *Medicago truncatula* [22], *Oryza sativa* [23–25], *Populus trichocarpa* [26], *Sorghum bicolor* [27], *Vitis vinifera* [2,28,29], *Brachypodium distachyon* [30,31], *Solanum tuberosum* [32], and *Solanum lycopersicum* [33]. According to these studies, approximately 0.2–1.3% of genes predicted in plant genomes corresponds to RGAs, which occur at a density of 0.3–1.6 per mega base (Mb). The genome of apple (*Malus × domestica* Borkh.) also contains a large number of RGAs [34]. Apple is characterized by recent whole genome duplication (WGD) [34]. The role and relevance of such radical genomic changes in plant evolution was largely demonstrated, but the number and timing of WGDs in the different plant species was only partially understood [35,36]. Polyploidy is common in angiosperms [22,37], and most if not all extant species are thought to be ancient polyploids [38]. However, ancestral genomes are in most cases dispersed on multiply rearranged chromosomes, having also suffered wholesale gene losses [5,39]. Given that synonymous substitutions are immune to selection pressure [40], the per-site synonymous substitution rate (Ks) is widely used to infer the time of WGD and to describe the relationships among chromosomes [2,34].
In this study, cluster organization of RGAs and their distribution across chromosomes were analyzed in terms of recent duplication of the apple genome. In addition, the phylogenesis of RGAs from the domesticated and wild *Malus* species, including also other Rosaceae, *P. trichocarpa*, and *V. vinifera* RGAs, was considered to clarify the evolutionary history of apple and its related species.

**Results**

**Classes of RGAs in *Malus × domestica***

Based on the presence of the NBS domain, 868 RGAs were identified in the genome of the *M. domestica* cultivar ‘Golden Delicious’, and all of them showed a significant (more than 90%) protein similarity with RGAs of *A. thaliana*, *P. trichocarpa*, and *V. vinifera*. In addition, 124 putative RGAs were found, and they were not further analyzed. By domain analysis, RGAs were assigned to TIR-NBS-LRR (TNL) and CC-NBS-LRR (CNL) classes. In particular, 505 RGAs were classified as NBS-LRR (NL), including CNL subclass, and 231 RGAs were classified as TIR-NBS (TN), including TIR-NBS-LRR (TNL), NBS-LRR-TIR (NLT), TIR-CC-NBS-LRR (TCNL), TIR-CC-NBS (TCN), and TIR-NBS (TN) subclasses (Table 1). In addition, 132 RGAs were characterized only by the presence of the NBS (N) or CC-NBS (CN) domains. The 868 RGAs accounted for 1.51% of *M. domestica* predicted genes, a percentage slightly higher than that in other plant genomes (Table 1). The density of RGAs per Mb was similar for *M. domestica* and other genomes with the exception of *Z. mays*, *C. papaya*, *C. sativus*, and *S. bicolor*.

The mean exon number detected in apple RGAs was 4.51, and the number of exons of CNL class (3.46) was lower than the number of TNL class (6.41; P<0.001). Thus, the number of exons in RGAs of *M. domestica* was consistent with the number in *A. thaliana* and *B. rapa* but higher than the number in *V. vinifera*, *P. trichocarpa*, and *O. sativa* (Table 1). Moreover, 23% of CNL RGAs are encoded by a single exon, while all TNL have at least three exons.

**Table 1.** Classification and organization of resistance gene analogues (RGAs) with a nucleotide-binding site (NBS) domain in different plant genomes.

| Characteristic | *Malus × domestica* | Arabidopsis thaliana | Populus trichocarpa | Vitis vinifera | Oryza sativa | Cucumis sativus | Sorghum bicolor | Brassica rapa | Brachypodium distachyon | Glycine max | Zea mays |
|---------------|----------------------|----------------------|--------------------|---------------|--------------|----------------|----------------|---------------|------------------------|------------|----------|
| Number of total predicted genes | 57,524 | 27,228 | 45,654 | 33,514 | 41,911 | 26,682 | 28,591 | 27,640 | nd | 25,532 | 46,430 | 32,540 |
| Genome size (Mb) | 750 | 125 | 485 | 487 | 487 | 389 | 243 | 730 | 529 | 272 | 1,115 | 2,500 |
| N<sup>a</sup> of RGAs | 868 | 178 | 402 | 391 | 535 | 61 | 54 | 211 | 92 | 178 | 429 | 129 |
| NBS-LRR class<sup>a</sup> | (58%) | (32%) | (59%) | (57%) | (89%) | (57%) | (74%) | (30%) | (18%) | (89%) | 55% | (74%) |
| (27%) | 57 | 236 | 194 | 480 | nd | 31 | 184 | 17 | 212 | 236 | 95 |
| TIR-NBS class<sup>b</sup> | (27%) | (64%) | (23%) | (13%) | (1%) | (13%) | (74%) | (30%) | (18%) | (89%) | 55% | [30] |
| Other RGAs<sup>c</sup> | (15%) | (4%) | (18%) | (30%) | (10%) | (30%) | (25%) | (30%) | (36%) | (11%) | 9% | (26%) |
| RGAs/total genes (%) | 1.51 | 0.65 | 0.88 | 1.01 | 1.27 | 0.23 | 0.19 | 0.76 | 0.92 | 0.28 | 0.65 | 0.038 |
| RGAs per Mb | 1.16 | 1.42 | 0.82 | 0.7 | 1.5 | 0.25 | 0.15 | 0.28 | 0.92 | 0.28 | 0.056 |
| Average number of exons in RGAs | 4.51 | 4.19 | 2.35 | 3.96 | 3.72 | nd | nd | nd | nd | 4.2 | nd | nd |
| Number of Single RGAs | 156 | 46 | 135 | 55 | 104 | 20 | 12 | 18 | nd | 4.2 | nd | nd |
| Number of Clusters | 152 | 39 | 75 | 52 | 157 | 11 | 13 | 24 | nd | nd | nd |
| Maximum Number of RGAs in clusters | 21 | 11 | 19 | 26 | 11 | 9 | 7 | nd | nd | 5 | nd |
| Average Number of RGAs in clusters | 4.11 | 3.21 | 3.75 | 5.78 | 3.48 | 3.72 | 2.92 | nd | 2.54 | nd | nd |

<sup>a</sup>NBS-LRR class includes: NBS-LRR (NL) and CC-NBS-LRR (CNL). Percentage (%) of this class relative to the total number of RGAs is reported in brackets.

<sup>b</sup>TIR-NBS class includes: TIR-NBS-LRR (TNL), NBS-LRR-TIR (NLT), TIR-CC-NBS-LRR (TCNL), TIR-CC-NBS (TCN), and TIR-NBS (TN). Percentage (%) of this class relative to the total number of RGAs is reported in brackets.

<sup>c</sup>Class of other RGAs includes: NBS (N) and CC-NBS (CN). Percentage (%) of this class relative to the total number of RGAs is reported in brackets.

nd: not declared by the authors.

doi:10.1371/journal.pone.0083844.t001
Figure 1. Chromosomal organization of RGAs in *Malus × domestica*. A: Phylogenetic analysis of NBS domain was carried out by neighbor-joining method [65] on RGAs protein sequences from *M. domestica* cultivar 'Golden delicious'. Major phylogenetic clades (from CN1 to CN5 and from
were present in a cluster, and the largest cluster contained 21 RGA
(located on Chr 2). Several clusters of RGA can be associated with
QTLs affecting disease resistance of Malus genotypes (Figure 1).

As previously shown in Arabidopsis [6,11], RGA of TIR-NBS and
non-TIR-NBS classes had different topologies in the phylogenetic
analysis (Figure 1A). In particular, six major TIR-NBS clades (numbered from TN1 to TN6) and five non-TIR-NBS major clades (numbered from CN1 to CN5) were identified in apple. RGA of TIR-NBS class were mainly located on Chr 2, 5, 9, 12, 15, 16, and 17, with Chr 16 hosting the TIR-NBS class almost exclusively (Figure S1A and Table S1). Chr 3, 4, 8, 11, 13, and 14 were mainly characterized by non-TIR-NBS class, and Chr 11 had almost exclusively RGA of non-TIR-NBS class. Considering TIR-NBS and non-TIR-NBS phylogenetic clades, the major clade TN6 represented more than one-third of the RGA on Chr 1 and 6, while the major clade CN4 included more than half of the RGA on Chr 11 and 14 (Figure S1A). Moreover, the major clade TN4 was located preferentially (63%) on Chr 2 (Figure S1B).

Phylogeny of RGA in Domesticated and Wild Malus Species

Twenty-four wild Malus species (Table S2) were considered, and
PCR fragments were amplified from germplasm. After sequence
comparison, unique fragments were translated in amino acid
sequences (Table S1), and 115 of them matched NBS sequences of
known resistance proteins with an E-value lower than 1E−10.
Phylogenetic analysis indicated that RGA of wild Malus species
mostly grouped in clades that included sequences of the
domesticated apple (Figure 2). A significant fraction of phyloge
netic clades contained only a few RGA, probably due to the short
sequence of the NBS domain used for this analysis. Some clades
consisted mainly of sequences from wild species and contained
only few RGA of the domesticated apple.

Phylogeny of RGA among Rosaceae Species

A total of 693 Rosaceae RGA sequences at NCBI were
downloaded (75 from Rubus, 293 from Prunus, 16 from Fragaria,
125 from Rosa, 34 from Pyrus, and 150 public sequences from
Malus species) and compared to the 868 RGA of M. domestica and
the 210 sequences obtained from wild Malus species (Table S1). In

Table 2. Organization and distribution of resistance gene analogues (RGA) with a nucleotide-binding site (NBS) domain in the apple (Malus × domestica) chromosomes.

| Chromosome | Number of RGA | Genome organization of RGA |
|------------|---------------|---------------------------|
| Number of single RGA | Number of Clusters | Average Number of RGA/cluster |
| 1 | 43 | 10 | 7 | 4.7 |
| 2 | 109 | 14 | 15 | 6.3 |
| 3 | 47 | 12 | 11 | 3.2 |
| 4 | 27 | 12 | 6 | 2.5 |
| 5 | 48 | 11 | 11 | 3.4 |
| 6 | 9 | 2 | 2 | 3.5 |
| 7 | 57 | 4 | 11 | 4.8 |
| 8 | 76 | 11 | 14 | 4.6 |
| 9 | 40 | 7 | 10 | 3.3 |
| 10 | 56 | 14 | 14 | 3.0 |
| 11 | 79 | 7 | 10 | 7.2 |
| 12 | 37 | 11 | 6 | 4.3 |
| 13 | 17 | 9 | 4 | 2.0 |
| 14 | 22 | 6 | 4 | 4.0 |
| 15 | 58 | 14 | 14 | 3.1 |
| 16 | 14 | 3 | 4 | 2.8 |
| 17 | 39 | 9 | 8 | 3.8 |
| Not anchored RGA | 90 | – | – | – |
| Total | 868 | 156 | 152 | 4.1 |

doi:10.1371/journal.pone.0083844.t002
the phylogenetic tree of Rosaceae species (Figure 3), 49 clades were specific to the genus *Malus*, and included sequences from two or more *Malus* species. Most of the remaining clades were represented by *RGAs* from two or more Rosaceae genera. In particular, three clades comprised *RGAs* of *Malus*, *Pyrus*, and *Prunus*, indicating a monophyletic origin of the three genera and strong conservation of some *RGA* sequences in these plants. Few clades were represented by non-apple *RGAs*, and clades specific to *Fragaria* or *Rosa* were also present.

Comparison of *RGAs* among *Malus × domestica*, *Populus trichocarpa*, and *Vitis vinifera*

*RGAs* sequences can also be compared across different plant families, and a phylogenetic tree of *RGAs* from *M. domestica*, wild *Malus* species, *V. vinifera*, and *P. trichocarpa* (Table S1) was obtained (Figure 4). Several clades comprised sequences from two or three species, and two major clades, named Md1 and Md2, comprised only sequences of *M. domestica* (Figure 4). However, sequences of the Md1 clade were grouped in three subclades in the phylogenetic tree of *RGAs* from Rosaceae species (Figure S2). *RGAs* of subclades Md1 sc2 and Md1 sc3 did not show similarity with any Rosaceae *RGAs*, whereas sequences of Md1 subclade 1 (Md1 sc1) shared significant similarity with four *RGAs* of *Pyrus* (Figure S2). Clade Md2 included one and two *RGAs* from *Rubus* and *Rosa*, respectively. Most of the *RGAs* of the clade Md2 are located on Chr 2, 3, 7, 11, 12, and 15.

Duplication of *RGAs in Malus × Domestica*

To study the recent duplication of *RGAs* in the *M. domestica* genome, *Ks* values were determined, and results from recent gene duplications were highlighted (Figure S3). Links among different *RGAs* helped to describe the relationships among the duplicated apple chromosomes [34]. Homologous apple chromosomes had more than 10 links, except for Chr 15 and 16, which hosted only a low number of *RGAs*. Chr 6 was not included in this analysis because it contains only nine *RGAs*, six of them derived from the recent WGD. Moreover, the duplicated chromosomes had *RGAs* belonging to the same phylogenetic clades (Figure S4).
Discussion

To counteract pathogens, plants rely on the innate immunity of their cells and on systemic signals emanating from infection sites [9,41]. Pathogen effectors from very diverse organisms are recognized by resistance proteins encoded by RGAs and activate plant defense responses [6,9]. NBS-mediated disease resistance is effective against obligate biotrophic and hemibiotrophic pathogens [42].

In apple, the abundance of RGAs is only partly related to genome size (750 Mb), which is much smaller than in maize (2300 Mb; [21]) or soybean (1115 Mb; [19]). The TIR-NBS class accounts for the largest group of RGAs in A. thaliana (64%; [11]) and B. rapa (64%; [14]). In P. trichocarpa [26], V. vinifera [2,5,28,29], and C. papaya [16,30], the percentage of TIR-NBS class is much lower than in the previously mentioned species. The TIR-NBS class is present at a very low frequency in O. sativa (1%; [24]) and S. bicolor (1%; [27]) and is absent in B. distachyon and Z. mays [30], supporting the conclusion that this class is specific for dicotyledons.

In apple, 231 RGAs of TIR-NBS class have been identified, and they are mainly located on Chr 2, 5, 9, 12, 15, 16, and 17. However, the number of RGAs belonging to non-TIR-NBS class in apple (505) is greater than in all other species considered, and these RGAs are mainly located on Chr 3, 4, 8, 11, 13, and 14. The existence of chromosome-specific RGAs classes suggests that groups of chromosomes evolved separately, but further analyses are required to test this hypothesis. In grapevine, the existence of two chromosome groups has been inferred based on RGAs cluster similarity, and the two groups seem to have evolved independently.

Figure 3. Phylogenesis of RGAs from Malus species (wild and domesticated apple), Pyrus communis, Prunus species, Fragaria ananassa, Rubus idaeus, and Rosa species. Phylogenetic analysis of the NBS domain was carried out by the neighbor-joining method [65] using RGA sequences of domesticated and wild Malus species (green), Pyrus spp. (yellow), Prunus spp. (purple), Fragaria spp. (red), Rosa spp. (orange), and Rubus spp. (blue). Proteins present in contiguous positions of the tree are merged (collapsed branches are indicated by the + sign). Phylogenetic tree indicates 49, three and one clades specific to Malus spp., Fragaria spp. and Rosa spp., respectively. Clades with RGAs of different genera: three clades of Malus spp. and Prunus spp.; seven clades of Malus spp. and Pyrus spp.; two clades of Malus spp. and Rubus spp.; four clades of Malus spp. and Rosa spp.; two clades of Fragaria spp. and Rosa spp.; two clades of Malus spp., Rosa spp., and Rubus spp.; three clades of Malus spp., Prunus spp., and Rubus spp.; four clades of Malus spp., Prunus spp., Rosa spp., and Rubus spp.; three clades of Malus spp., Prunus spp., Pyrus spp., Rosa spp., and Rubus spp.; two clades of Malus spp., Fragaria spp., Prunus spp., Rosa spp., and Rubus spp., one clade of Malus spp., Fragaria spp., Pyrus spp., Rosa spp., and Rubus spp.

doi:10.1371/journal.pone.0083844.g003
Moreover, the TIR-NBS class is specific for only one of the two components of *V. vinifera* genome, suggesting an independent evolution of the RGA classes [2].

In apple, 56% of RGAs (435 of 778 anchored) are located preferentially on six chromosomes, with 14% located on Chr 2. In large gene families, genes are commonly organized in clusters and superclusters [4,5,11,14,16,25,26], as demonstrated here for the apple genome. Of the RGAs clusters in apple, 71% (108 of 152) include RGAs from the same phylogenetic clade, and 29% RGAs from two to three different clades. Clusters frequently consist of tandem duplications of the same gene [5,43]. Heterogeneous clusters, in which sequences belong to different phylogenetic lineages, are also present, most probably as a result of different molecular mechanisms like ectopic recombination, chromosomal translocation, and gene transposition, as has been recently highlighted for the grapevine genome [2]. This kind of genome evolution could be explained in terms of a positive selection for cluster complexity, which could serve as the basis for the generation of new resistance specificities [4,44]. The role of tandem duplication in the apple genome is supported by low Ks values among RGAs of the same cluster, as is already known for other species [2,5,14,22,43]. Gene duplication in a position different from the original cluster has to be preceded by gene transposition, as predicted for *A. thaliana* and *V. vinifera* RGAs [1,2].

Thus, a successful transposition is the starting point for the creation of a new RGA cluster, and the selection for disease resistance could favor the process [5,45]. Moreover, analysis of RGA transposition has indicated that *V. vinifera* putative component genomes may have evolved independently and later fused and evolved together in the same nucleus [2].

Velasco et al. [34] have shown that recent WGD has increased the chromosome number in apple from nine in the putative ancestor to the current 17. The recent duplication of RGAs due to a WGD event supports the existence of i) a tetraploid state of the genome in which a pair of chromosomes exists with a second homologous pair; ii) duplications inside chromosomes, particularly for Chr 11 where recent duplications can be observed; and iii) duplications in different chromosomes, suggesting recent events of gene transposition. Eight of the 17 chromosomes (Chr 3 and 11, 5 and 10, 9 and 17, and 13 and 16) represent a direct duplication of...
four ancestral chromosomes, and each of the extant Chr 4, 6, 12, and 14 derives from translocation between two ancestral chromosomes [34]. More complex events have generated the remaining five chromosomes that are derived from starting three ancestral chromosomes. The different clades of RGAs along duplicated chromosomes indicate a similar position of orthologous RGAs along each chromosome doublets (Chr 3 and 11, 5 and 10, 9 and 17, and 13 and 16). These results strongly support the origin of the apple chromosomes as described by Velasco et al. [34] and indicate that RGA distribution might be used to dissect plant genome evolution [2]. As is the case for other species, the process of gene duplication has shaped the apple genome in different ways, including the selective retention of paralogs associated with specific biological processes, the amplification of specific gene families, and an extensive subfunctionalization of paralogs. Both the major WGD event and small-scale duplications could be responsible for the high number of the apple RGAs. A remarkable feature of gene duplication in apple is the high proportion of paralogs showing divergent expression patterns [46]. Extensive subfunctionalization could have contributed to the acquisition of new traits specific to apple or to the Pyrinae lineage [47]. Sequences of Euroidis genomes provide evidence of ancient genome duplications that occurred early in evolution, suggesting a polyploid origin for most Eudicots [28,48].

Most of the RGAs of wild Malus species are closely related RGAs of the domesticated apple. Whereas RGAs sequencing from wild Malus species was partial and could include alleles of the same gene, phylogenetic analysis revealed specific clades of wild Malus species, indicating, as expected, the potential to enlarge the gene family size between species could be attributed to gene duplication, deletion, pseudogenization, and functional diversification [56–58]. The last case is supported by attributed to gene duplication, deletion, pseudogenization, and [54,55]. The variation of gene families in the dicotyledonous plants. Similar situations were reported for other reasons could explain the variation of RGAs in domesticated apple. Moreover, the comparison of apple RGAs with those of other Rosaceae indicates the existence of specific clades for apple. In addition, several clades include a mixture of RGAs from Malus, Pyrus, and Prunus, indicating that similar resistance genes are still shared in different genera of the Rosaceae. While these results support the monophyletic origin of the three genera, clades specific for each genus were also found. The existence of genus- or species-specific clades indicates the existence of mechanisms for cluster conservation, as reported by Plocki et al. [49].

Phylogenetic relationships within the Rosaceae inferred from RGAs are consistent with phylogenies based on chloroplast and nuclear genes [50,51]. The phylogenetic analysis of the RGAs from Malus, Vitis, and Populus shows that Malus contains two large non-TIR-NBS clades that are specific to Malus. This inference should be considered with caution, because the RGA sequences used in our analysis are from only a few species. Several other reasons could explain the variation of RGAs in Rosaceae species, such as the inter-specific variation of the RGA family size observed in dicotyledonous plants. Similar situations were reported for other gene families in the Archeae [52], bacteria [52,53], and mammals [54,55]. The variation of RGA family size between species could be attributed to gene duplication, deletion, pseudogenization, and functional diversification [56–58]. The last case is supported by the necessity of a species to adapt to rapidly changing pathogen populations.

Concluding Remarks

This paper analyses the RGAs of Malus spp. and other Rosaceae species to reveal specific evolutionary features of M. domestica. RGAs of M. domestica are mainly located in clusters and are mapped preferentially on six chromosomes. TIR-NBS and non-TIR-NBS classes of RGAs are located in different chromosome groups. Phylogenetic reconstruction in the Rosaceae family revealed specific clades of RGAs for Malus spp., Fragaria spp., and Rosa spp., indicating genus-specific evolution of resistance genes. However, strikingly similar RGAs were shared in different species of Malus, Pyrus, and Prunus highlighting a monophyletic origin of these three genera and the high conservation of some RGA sequences in these plants.

Materials and Methods

Identification of RGAs in the Apple Genome

The RGA sequences were identified from the predicted proteins of M. domestica cultivar ‘Golden Delicious’ [34] based on their NB-ARC domain profile (PF00931 [59]) using HMMER [60]. Putative RGA alleles were identified as predicted genes that have more than 90% of sequence similarity and overlap with another RGA along each scaffold of the heterozygous apple genome. Apple RGAs were validated by BLAST-N analysis (more than 90% protein sequence similarity) against known A. thaliana, P. trichocarpa, and V. vinifera genes. RGAs were grouped in different classes based on the presence of the domains TIR, LRR, CC, and BED finger [43]. The motifs were derived from the domain profiles retrieved from PFAM (http://pfam.janelia.org), PANTHER (http://www.pantherbio.org/), and SMART (http://smart.embl-heidelberg.de) databases and from the COILS program; a stringent threshold of 0.9 was used so that CC domains were specifically detected [61]. Resistance-related proteins were also identified based on kinase domains (IPR000719, PF07714, PF00069). Additional putative apple resistance genes were selected using BLAST and Arabidopsis proteins as reference sequences, based on a 60% similarity threshold.

Identification of RGA Clusters in the Apple Genome

The Arabidopsis definition of RGA cluster [4] was adopted: two or more RGAs in a cluster should be located within an average of 250 Kb and should not be interrupted by more than 21 open reading frames different from RGAs, as previously adopted for grapevine RGA clusters [2].

Isolation of RGAs from Wild Species

Four pairs of degenerate primers targeting the NBS domain [62,63] were used to amplify RGA sequences from 26 different Malus accessions present in the USDA apple germplasm collection at Geneva (NY, USA) (www.ars-grin.gov/npgs/index.html; Table S2). The homologous sequences represent the following species: M. baccata, M. floribunda, M. fusca, M. halliana, M. honeanesis, M. huphemensis, M. kansuensis, M. microalbus, M. orientalis, M. pratii, M. prunifolia, M. punica, M. robusta, M. sargentii, M. sieboldii, M. sieversii, M. sikkimensis, M. sylvestris, M. sublobata, M. sylvestris, M. transitoria, and M. yunnanensis (Table S2). PCR fragments were cloned in pGEMT easy (Promega), and two clones for each fragment were sequenced. Sequences were screened, cleaned, and compared with resistance genes previously identified in Rosaceae and in other Angiosperms. BLAST DNA similarity searches were performed against the RGA sequences of the apple genome using a collection of established RGAs. The RGAs were translated using tBLAST-N. Clones were filtered based on hit quality, because most of the RGA clones encoded between 24 and 40 amino acid residues. Queries having only a single hit below 90% identity were removed, and those with multiple smaller hits were annotated manually. RGA sequences from wild Malus species were submitted to the NCBI database (www.ncbi.nlm.nih.gov) under the accession numbers reported in Table S1.
Phylogenetic Analyses

Public RGA sequences from Rosaceae, P. trichocarpa, and V. vinifera Release 2 were downloaded from GenBank (http://www.ncbi.nlm.nih.gov; Table S1). RGA sequences from wild Malus species were also included (Table S1). Protein sequences of NBS domain of RGAs from M. domestica were aligned together with NBS sequences of wild Malus species, P. trichocarpa, V. vinifera and with the other Rosaceae species using hidden Markov models with the Sequence Alignment and Modeling Software System (SAM-T2K [64]); the sequences were formatted for analysis with the Phylip phylogenetic inference package [65].

The SEQUENCING tool of the Phylip package was used to generate 500 bootstraps of the dataset, and the PROTDIST tool was used to construct 500 bootstrapping distance matrices using the Dayhoff PAM matrix [65]. These matrices were jumbled twice and processed with the FITCH tool to create a phylogenetic tree. A neighbor-joining tree of the 500 bootstraps was also constructed (jumbling the sequence input order twice), and a majority-rule consensus tree was assembled.

Determination of the Ks Value

Based on a CLUSTALW nucleotide alignment of M. domestica RGAs sequences, a total of 302,253 Ks values were obtained [66]. The connections between chromosomes were defined on the basis of the number of RGAs and Ks values. A connection between two chromosomes was accepted if at least ten RGAs had a Ks value lower than or equal to the first quartile of 0.25 [34].

Supporting Information

Figure S1 A: Distribution (percentage) of the major phylogenetic clades of apple RGAs (Figure 1A) on the 17 M. domestica chromosomes (Chr). B: Percentage of chromosome (Chr) assignment to the major phylogenetic clades. Colours of major phylogenetic clades and chromosomes are listed below each chart. (TIF)

Figure S2 Phylogenesis of RGAs from Rosaceae species. Phylogenetic analysis of the NBS domain was carried out by the neighbor-joining method [65] using RGA sequences of domesticated and wild Malus species (green), Pyrus spp. (yellow), Prunus spp. (purple), Fragaria spp. (red), Rosa spp. (orange), and Rubus spp. (blue). The composition of the phylogenetic clades (Md1 and Md2; Figure 4) and subclades (sc) of sequences mainly from M. domestica is highlighted. Proteins present in contiguous positions on the tree are merged (collapsed branches are indicated by the + sign). (TIF)

Figure S3 Connections between apple chromosomes based on Ks values from pairwise comparisons of RGAs. Joining lines represent connections between two RGAs among duplicated chromosomes [35] (blue, red, pink, green), among not duplicated chromosomes (yellow), and within the same chromosome (gray). Each line represents a connection between two RGAs with a Ks value lower than 0.25 [35]. A connection between two chromosomes was accepted if at least ten pairwise comparisons had a Ks value lower than 0.25. (TIF)

Figure S4 Distribution of RGAs among chromosome (Chr) doublets derived from the recent whole genome duplication of apple [34]. Colours of major phylogenetic clades (Figure 1A) are indicated. (TIF)

Table S1 List of accession numbers and abbreviations of resistance gene analogues (RGAs) with a nucleotide-binding site (NBS) domain from Malus × domestica, Populus trichocarpa, Vitis vinifera, wild Malus species, Fragaria spp., Prunus spp., Pyrus communis, Rubus idaeus, and Rosa roxburghii. Chromosome location, code, and class based on protein domain analysis are indicated for each of the RGAs of M. domestica. (XLS)

Table S2 List of wild Malus species accessions (USDA apple germplasm collection at Geneva, NY, USA; www.ars-grin.gov/npgs/index.html) used for the isolation of RGAs. (DOCX)

Acknowledgments

The authors would like to thank Prof. Francesco Salamini for critical discussions and reading of the manuscript. The authors would like to thanks Dr Iezzoni Dr Volk for proving the sequence of RGA from cherry.

Author Contributions

Conceived and designed the experiments: RV MM. Performed the experiments: A Baldo A Bailey PF. Analyzed the data: GM MP LR MM. Contributed reagents/materials/analysis tools: RV MM. Wrote the paper: GM MP A Baldo MM.

References

1. Freeling M, Lyons E, Pedersen B, Alam M, Ming R, et al. (2008) Many or most genes in Arabidopsis transposed after the origin of the order Brassicales. Genome Res 18: 1924–1937.
2. Malacarne G, Perazzolli M, Cestaro A, Sterck L, Fontana P, et al. (2012) Deconstruction of the paleopolyploid grapevine genome based on the analysis of transposition events involving NBS resistance genes. PLoS one 7: e29762.
3. Meyers BC, Kaushik S, Nandety RS (2005) Evolving disease resistance genes in Arabidopsis. BMC Plant Biol 7: 56.
4. Richly E, Kureth J, Leister D (2002) Mode of amplification and reorganization of resistance genes during recent Arabidopsis thaliana evolution. Mol Biol Evol 19: 76–84.
5. Yang S, Zhang X, Yue JX, Tian D, Chen JQ (2000) Recent duplications dominate NBS-encoding gene expansion in two Woody species. Mol Genet Genomics 260: 167–199.
6. Meyers BC, Dickerman AW, Michelmore RW, Sivaramanakrishnan S, Sohrad BW, et al. (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J 20: 317–332.
7. Leister D, Ballvera A, Salamini F, Gehardt C (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nat Genet 14: 421–429.
8. Yu YG, Buss GR, Marof GR, MA (1996) Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. Proc Natl Acad Sci U S A 93: 11751–11756.
9. Duong TL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. Nature 411: 826–833.
10. Bai J, Pennill LA, Ning J, Lee SW, Ramalingam J, et al. (2002) Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. Genome Res 12: 1071–1084.
11. Meyers BC, Kozik A, Greigo A, Kaang H, Michelmore RW (2005) Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell 17: 809–834.
12. DeYoung BJ, Innes RW (2006) Plant NBS-LRR proteins in pathogen sensing and host defense. Nat Immunol 7: 1213–1219.
13. Tan X, Meyers BC, Kozik A, Al West M, Morgante M, et al. (2007) Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in Arabidopsis. BMC Plant Biol 7: 56.
14. Mun J-H, Yu H-J, Park S, Park B-S (2009) Genome-wide identification of NBS-encoding resistance genes in Gossypium hirsutum. Mol Genet Genomics 282: 637–651.
15. Ming R, Hou SB, Feng Y, Yu QY, Dionne-Laporte A, et al. (2008) The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnnaeus). Nature 452: 991–996.
16. Porter BW, Paidi M, Ming R, Alam M, Nishijima WT, et al. (2009) Genome-wide analysis of *Caesia papaya* reveals a small NBS resistance gene family. Mol Genet Genomics 281: 699-626.

17. Huang S, Li R, Zhang Z, Li L, Gu X, et al. (2009) The genome of the cucumber, *Cucumis sativus*. Nat Genet 41: 1275-1281.

18. Kang JY, Kim KH, Shim S, Yoon MY, Sun S, et al. (2012) Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. BMC Plant Biol 12.

19. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, et al. (2010) Genome sequence of the palaeopolyploid soybean. Nature 463: 178-183.

20. Cheng Y, Li X, Jiang H, Ma W, Miao W, et al. (2012) Systematic analysis and comparison of nucleotide-binding site disease resistance genes in maize. Feds J 279: 283-324. doi: 10.1079/FEDS.2012.279.283.

21. Schanbacher PS, Ware D, Fulton RS, Stein JC, Wei F, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112-1115.

22. Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, et al. (2007) The grapevine genome: complexity, diversity, and dynamics. Science 326: 1112-1115.

23. Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, et al. (2008) Genome-wide identification of NBS-LRR genes and their association with disease resistance in soybean. BMC Plant Biol 12.

24. Monosi B, Wisser RJ, Pennill L, Hulbert SH (2004) Full-genome analysis of *Medicago truncatula* genome: complexity, diversity, and dynamics. Science 326: 1112-1115.

25. Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, et al. (2004) Genome-wide pattern of numbers of gramineous-rich repeat genes in the model plant *Medicago truncatula*. Plant Physiol 146: 5-21.

26. Matsushita T, Wu JZ, Kanamori H, Katayose Y, Fujimura M, et al. (2005) The map-based sequence of the rice genome. Nature 436: 793-800.

27. Monosi B, Wisser RJ, Pennill L, Hubert SR (2004) Full-genome analysis of *Medicago truncatula* genome: diversity, and dynamics. Science 326: 1112-1115.

28. Ameline-Torregrosa C, Wang B-B, O'Bleness MS, Deshpande S, Zhu H, et al. (2008) Identification and characterization of nucleotide-binding-site-Lecuine-rich repeat genes in the model plant *Medicago truncatula*. Plant Physiol 146: 5-21.

29. Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, et al. (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine phyla. Nature 449: 463-465.

30. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, et al. (2010) The genome sequence of the tomato, *Solanum lycopersicum*. Nature 465: 556.

31. Araki H, Jing Z, Zhou T, Zhou Y, Jin J, et al. (2004) Genome-wide mapping of *NBS-LRR* genes in the *Medicago truncatula* genome. Genet Genomics 281: 609-626.

32. Calenge F, Valancourt J, Tuller T, Van de Weg WE, Van der Linden CG, et al. (2004) Identification of nucleotide-binding site disease resistance genes in maize. Febs J 66: 619-636.

33. Calenge F (2006) Both stable and unstable QTLs for resistance to powdery mildew are detected in apple after four years of field assessments. Mol Genet Genomics 269: 101-108.

34. Calenge F, Faure A, Boisselier C, Soufflet-Freslon V, Gianfranceschi L, Patocchi A, et al. (2008) Inheritance of an apple powdery mildew resistance gene *Rm16*. Interact 23: 608-617.

35. Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K (2009) The *Sorghum bicolor* genome. Genome Res 19: 155-167.

36. Lee SY, Seo JS, Rodriguez-Lanetty M, Lee DH (2003) Comparative analysis of superfamilies of NBS-encoding disease resistance gene analogs in cultivated and wild apple species. Mol Genet Genomics 269: 101-108.

37. Plocik A, Layden J, Kesseli R (2004) Comparative analysis of NBS domain sequences of *NBS-LRR* disease resistance genes from sunflower, lettuce, and chicory. Mol Phylogenet Evol 31: 153-163.

38. Ding J, Zhang W, Jing Z, Chen J-Q, Tian D (2007) Unique pattern of *NBS-LRR* gene families and its genetic association with an NBS-LRR candidate resistance gene. Plant Breed 126: 476-481.
78. Cevik V, King J (2002) High-resolution genetic analysis of the Sd-1 aphid resistance locus in *Malus* spp. Theor Appl Genet 105: 346–354.

79. Bus VGM, Chagné D, Bassett HCM, Bowatte D, Calenge F, et al. (2008) Genome mapping of three major resistance genes to woolly apple aphid (*Eriosoma lanigerum* Hausm.). Tree Genet Genomes 4: 223–236.

80. Zhang X, Feng Y, Cheng H, Tian D, Yang S, et al. (2011) Relative evolutionary rates of NBS-encoding genes revealed by soybean segmental duplication. Mol Genet Genomics 285: 79–90.