

RESEARCH ARTICLE

Open Access

Paleo-evolutionary plasticity of plant disease resistance genes

Rongzhi Zhang^{1,2}, Florent Murat¹, Caroline Pont¹, Thierry Langin¹ and Jerome Salse^{1*}

Abstract

Background: The recent access to a large set of genome sequences, combined with a robust evolutionary scenario of modern monocot (*i.e.* grasses) and eudicot (*i.e.* rosids) species from their founder ancestors, offered the opportunity to gain insights into disease resistance genes (R-genes) evolutionary plasticity.

Results: We unravel in the current article (i) a R-genes repertoire consisting in 7883 for monocots and 15758 for eudicots, (ii) a contrasted R-genes conservation with 23.8% for monocots and 6.6% for dicots, (iii) a minimal ancestral founder pool of 384 R-genes for the monocots and 150 R-genes for the eudicots, (iv) a general pattern of organization in clusters accounting for more than 60% of mapped R-genes, (v) a biased deletion of ancestral duplicated R-genes between paralogous blocks possibly compensated by clusterization, (vi) a bias in R-genes clusterization where Leucine-Rich Repeats act as a 'glue' for domain association, (vii) a R-genes/miRNAs interome enriched toward duplicated R-genes.

Conclusions: Together, our data may suggest that R-genes family plasticity operated during plant evolution (i) at the structural level through massive duplicates loss counterbalanced by massive clusterization following polyploidization; as well as at (ii) the regulation level through microRNA/R-gene interactions acting as a possible source of functional diploidization of structurally retained R-genes duplicates. Such evolutionary shuffling events led to CNVs (*i.e.* Copy Number Variation) and PAVs (*i.e.* Presence Absence Variation) between related species operating in the decay of R-genes colinearity between plant species.

Keywords: R-genes, Duplication, Plasticity, Evolution

Background

Pathogen attacks from fungi [1], viruses [2], nematodes [3] or bacteria [4], compelled plants to prevent damages by engaging an "arms race" with these organisms. Therefore, plants have developed a battery of defense mechanisms involving (1) PTI (PAMP-Triggered Immunity) triggered by PAMP (Pathogen-Associated Molecular Patterns) [5-7] and (2) ETI (Effector-Triggered Immunity) triggered by effectors leading to hypersensitive response (referenced as HR [8]). Therefore, constant evolution leading to novel mechanisms is crucial for plant defense processes as well as adaptation to biotic stresses. The most studied disease resistance proteins encoding genes (hereafter R-genes) or genes involved in disease resistance pathways are Nucleotide-Binding-Sites (NBS) [9,10], Leucine-Rich

Repeats (LRR) [9,10], Toll-Interleukin1 Receptors (TIR), WRKY transcription factors [11,12], Lysine Motif (LysM) families [13,14], and Protein Kinase families (hereafter referenced as PKinase) [15,16]. R-genes can then be functionally classified into five distinct groups consisting in CNL (genes encoding proteins with coiled-coil, nucleotide binding site, leucine-rich repeat domains, *i.e.* CC-NBS-LRR), TNL (genes encoding proteins with Toll-interleukin receptor-like, nucleotide binding site, leucine-rich repeat domains, *i.e.* TIR-NBS-LRR), RLP (genes encoding proteins with receptor serine-threonine kinase like, extracellular leucine rich repeat domains, *i.e.* ser/thr-LRR), RLK (genes encoding proteins with kinase, extracellular leucine-rich repeat domains, *i.e.* Kin-LRR), and RGA (includes all other genes conferring resistance through different molecular mechanisms) classes [17]. LRR-RLK, LRR, LysM, LysM-kinase act as pattern-recognition receptors (PRR) involved in the PTI pathway, while NBS-LRR commonly responds in the frame of the

* Correspondence: jsalse@clermont.inra.fr

¹INRA/UBP UMR 1095 GDEC 'Génétique, Diversité et Ecophysiologie des Céréales', 5 chemin de Beaulieu, 63100 Clermont-Ferrand, France

Full list of author information is available at the end of the article

ETI pathway [18,19]. Finally, WRKY and protein-kinases, associated with protein domains encoded by R-genes (hereafter R-domains), can also be activated by PRRs in disease resistance pathways [18-20].

R-genes have been reported to be ancient and conserved genes that have been detected in gymnosperms, plants and animals to ensure immunity [21-23]. However, comparative genomic analyses have shown that R-genes are associated with a great structural diversity in vertebrates and plants. For example, the presence of TIR domains in conifers and mosses indicated that TIR may represent an ancestral R-gene family with shared functionality with their mammalian or insect homologues regarding innate immunity [21,22,24-26]. TIR genes typically expanded in eudicot genomes, while they have been reported to be absent (or at least rare) in grass genomes [27-31]. Moreover, tandem and segmental duplications have been reported as a source of structural plasticity of NBS-LRR genes in plant genomes [32]. Furthermore, PAV (Presence/Absence Variation) polymorphisms often exist in a population or between species [33-36]. Overall, small-scale studies (*i.e.* few R-genes families/domains and/or few plant species investigated) have suggested R-genes as one of the most plastic gene family in plants associated with intense structural shuffling in the course of evolution leading to synteny erosion or alternatively loss [37]. For example, evolutionary investigations of R-genes in *Arabidopsis* and rice have been conducted suggesting contrasted amplification of TNL and CNL families as well as clusterization of NBS-LRRs *via* segmental and tandem duplications or ectopic gene conversions [38].

Few studies have investigated the conservation of R-genes across a large set of plant species and at the whole-genome level. Genome sequences from flowering plants that are derived from a common ancestor 135 to 250 million years ago (mya) are increasingly available in the public domain for evolutionary studies. Recent paleohistorical studies demonstrated that modern grass genomes, including Panicoideae (sorghum [*Sorghum bicolor*], [39] maize [*Zea mays*], [40]), Ehrhartoideae (rice [*Oryza sativa*], [41]), and Pooideae (*Brachypodium distachyon*; [42]), were shaped from $n = 5$ to 12 ancestral grass karyotypes (AGKs) containing a minimal set of 6045 ordered protogenes with a minimum physical size of 33 Mb [43-45] through whole-genome duplication (WGD) and ancestral chromosome fusion events. Likewise, the recent comparison of numerous eudicot genomes (*i.e.* mainly eurosids), including grape (*Vitis vinifera*; [46]), poplar [47], *Arabidopsis thaliana* [48], soybean *Glycine max*; [49], and cacao (*Theobroma cacao*; [50]), revealed that modern eudicot genomes derived from an $n = 7$ ancestor that went through a paleohexaploidization event to reach a $n = 21$ intermediate followed by numerous

lineage-specific WGDs and chromosome fusion events [46,51]. During the last 135 to 250 million years of evolution, the protein-coding gene families have been then shaped by various gene duplication mechanisms, including WGDs (or polyploidization), segmental duplications, and tandem duplications. It is now well established that all modern diploid plant species are highly shuffled paleopolyploids [52-55].

Duplication (WGDs, segmental duplications and tandem duplications) were proposed as the major mechanisms driving R-genes family expansion or contraction from their traceable ancestral copies [56-58]. However, a systematic and detailed study of the paleohistorical evolution of R-genes across plant subfamilies including rosids species (*Arabidopsis thaliana* [48], Grape [59], Apple [60], Poplar [47], Soybean [49], Lotus [61], Strawberry [62], Cacao [50] and Papaya [63]), and grasses (Rice [41], Maize [40], *Sorghum bicolor* [39] and *Brachypodium distachyon* [42]) is still lacking. Particularly, how R-genes have behaved following polyploidization events is not well established. Such a precise investigation of the paleohistory of R-genes during the last 250 million years of evolution will unravel precise mechanisms that lead to the reduced conservation of R-genes observed between modern plant species.

Results

Disease resistance gene mapping, conservation and evolutionary patterns

To identify the largest set of plant R-genes, three complementary methods (see Methods section) were combined (illustrated as Additional file 1: Figure S1) consisting in (1) the detection of PFAM [64] domains, (2) the exploitation of public genome annotations, and (3) the use of the Plant Resistance Gene Database, PRGdb (<http://prgdb.cbm.fvg.it/index.php>, [65]). The integration of the three previous approaches allowed to construct a non-redundant set of putative R-genes in the plant species considered in this study (Table 1, Additional file 1: Figure S2 and Additional file 2: Dataset S1). Based on the genome annotation approach, 2697 R-genes were identified in monocots, and 3021 in eudicots, corresponding to a total set of 5718 plant R-genes (Table 1). Regarding the PFAM domain identification procedure, 8013 sequences in monocots and 14996 in eudicot species were identified, corresponding to 23009 putative R-genes sequences in total (Table 1). Finally, the PRGdb aligned on the 13 genome sequences investigated unraveled 1874 sequences in monocots and 2001 sequences in eudicot species. These three complementary methods lead us to deliver the most complete and non-redundant list in angiosperms consisting in 23641 R-genes sequences, 7883 for the monocots and 15758 for the eudicots (*cf* Table 1 and Additional file 2: Dataset S1). The identified R-genes families deriving from the annotation, PFAM and PRGdb

Table 1 R-genes catalog and conservation in plant genomes

Species	Genome data			R-genes data set				R-genes conservation		
	Nb chr.	size (Mbp)	Nb gene	Nb ortholog	Annot	Pfam	PRG db	Non-redundant R-genes*	OrthoSeq	%
Monocot										
<i>Oryza sativa (rice)</i>	12	372	41046	Ref	1180	2294	1316	2637 (889;539;10;16;100;1544;214;102)	Ref	Ref
<i>Sorghum bicolor</i>	10	659	34008	6147(14,9%)	570	1616	159	1717 (516;284;2;20;98;1158;68;39)	413 (116;16;0;7;5;347;19;3)	24,1%
<i>Zea mays (maize)</i>	10	2365	32540	4454(10,85%)	480	2630	399	1867 (558;140;13;15;106;1255;0;45)	319 (104;6;2;3;3;261;0;6)	17,1%
<i>Brachypodium distachyon</i>	5	271	25504	8533(20,78%)	467	1473	nd	1662 (435;199;2;11;80;1094;0;104)	495 (137;19;0;4;7;417;0;1)	29,8%
Monocot Total	-	-	133098	-	2697	8013	1874	7883 (2398;1162;27;62;384;5051;282;290)	1227 (357;41;2;14;15;1025;19;10)	-
Eudicot										
<i>Vitis vinifera (grape)</i>	19	302	21189	Ref	256	892	170	1078 (421;229;50;10;37;557;0;91)	Ref	Ref
<i>Arabidopsis thaliana</i>	5	119	33198	2389(11,27%)	564	1407	1105	1559 (436;168;125;12;73;1010;3;126)	74 (27;0;0;1;0;58;0;2)	4,7%
<i>Populus trichocarpa (poplar)</i>	19	294	30260	4555(21,5%)	212	1229	134	1297 (413;122;31;24;63;930;0;32)	122 (53;4;1;4;6;89;0;2)	9,4%
<i>Carica papaya</i>	9	234	19205	3199(15,1%)	113	674	nd	703 (200;50;13;11;42;515;0;0)	101 (45;1;1;1;4;74;0;0)	14,4%
<i>Glycine max (soybean)</i>	20	949	46164	4013(18,94%)	1023	3104	318	3310 (1097;411;1;166;49;179;2172;170;146)	148 (51;0;1;3;8;114;6;3)	4,5%
<i>Malus x domestica (apple)</i>	17	742	58979	3498(16,51%)	853	4135	243	4252 (1638;860;340;9;123;2292;0;117)	125 (41;0;1;1;7;94;0;2)	2,9%
<i>Lotus japonicus</i>	6	500	15470	1720(8,12%)	nd	664	26	668 (165;77;49;0;34;443;0;5)	46 (18;1;0;0;2;36;0;1)	6,9%
<i>Fragaria vesca (strawberry)</i>	7	240	34809	3289(15,52%)	nd	1452	1	1452 (483;154;148;3;52;884;0;0)	108 (44;2;1;0;1;89;0;0)	7,4%
<i>Theobroma cacao</i>	10	430	27814	4472(20,1%)	nd	1439	4	1439 (492;220;14;4;55;934;0;0)	149 (55;3;0;1;10;113;0;0)	10,4%
Eudicot Total	-	-	265899	-	3021	14996	2001	15758 (5345;2291;936;122;658;9737;173;517)	873 (334;11;5;1;1;38;667;6;10)	-

*Number of non-redundant R-genes.

Nb, number.

Nd, not detected.

Numbers in bracket refers to LRR, NBS, TIR, LysM, WRKY, Pkinase, Ser/Thr and Disease resistance related genes respectively.

In bold, total non-redundant detected R-genes.

Ref, reference.

approaches are detailed in Table 1 and Additional file 1: Figure S2, showing that the total number of non-redundant R-genes in plants is close to the dataset obtained with the PFAM detection approach, suggesting that such method is the most appropriate in delivering the largest and most complete set of R-genes in any genomic sequence of interest.

The conservation of R-genes between genomes was then investigated using rice and grape as reference genomes respectively for monocots and eudicots as they represent the most closely related modern genome structures of the reconstructed plant ancestral karyotypes [66]. The 2637 rice R-gene sequences were aligned against the three other monocot species available, *i.e.* sorghum, *Brachypodium* and maize. Similarly, the 1078 grape R-genes sequences were aligned against the eight eudicot species investigated (Table 1, 'Evolutionary Data' column). 413 orthologous R-genes were identified in sorghum corresponding to 24.1% of the reference dataset (*i.e.* rice). Similarly, 319 orthologous R-genes were found in maize (17.1% of the reference) and 495 in *Brachypodium* (29.8% of the reference). This result suggests that R-genes may appear as more conserved than the total protein-coding genes, with respectively 24.4%, 17.1%, and 29.8% of rice orthologous R-genes characterized in sorghum, maize and *Brachypodium* compared to 14.9%, 10.85%, and 20.78% of overall annotated protein-coding genes conservation observed for the same species (P -value < 5% Fisher Exact Test, Additional file 1: Table S1). However, if the protein kinase family is excluded, as it consists in the largest (14788 kinases) one compared to the six others (with an average of 1264 R-genes for these families), the synteny conservation observed for R-genes is similar to one observed for the total annotated protein-coding genes. The R-genes synteny (excluding protein kinases) at the chromosome level between grasses is illustrated as Figure 1A. In contrast to grasses, in eudicot species R-genes are significantly less conserved than the annotated protein-coding genes (except for papaya) with 4.7%, 9.4%, 4.5%, 2.9%, 6.9%, 7.4%, and 10.4% orthologous R-genes identified in *Arabidopsis*, poplar, soybean, apple, lotus, strawberry, and cacao compared to 11.27%, 21.5%, 18.94%, 16.51%, 8.12%, 15.52%, and 20.10% for the conservation of total protein-coding genes for the same species (P -value < 5% Fisher's Exact Test; Additional file 1: Table S1). This result may suggest a reduced syntenic conservation of R-genes in eudicots compared to monocots.

R-genes were classified into seven distinct groups, according to their specific encoded protein domains (R-domains). In monocots, we identified 2398 LRR, 1162 NBS, 27 TIR, 62 LysM, 384 WRKY, 5333 Protein-kinases, and 290 RG. In eudicots, 5345 LRR, 2291 NBS, 936 TIR, 122 LysM, 658 WRKY, 9910 Protein-kinases, and 517 RG were characterized. The distribution of the R-domain

repertoire excluding Pkinases in the 13 plant species investigated is illustrated as Figure 1B with a color code that illuminates the six different R-domains (*i.e.* for a total of 7743 LRR, 3453 NBS, 963 TIR, 171 LysM, 1042 WRKY and 807 RG). Regarding the six different R-domains investigated, LRR and NBS are more abundant in the investigated plant genomes (Figure 1B). LRR and NBS consist in, on average, more than 50% and 20% of detected R-genes in both eudicots and monocots respectively. Few WRKY domains were detected in plants (~3.98% and 3.33% in monocot and eudicot species respectively). However, the number of TIR domains appeared much abundant in eudicots than the monocot species (Additional file 1: Table S2) as previously reported [27-31], which may indicated a specific amplification of such domain during rosids paleohistory. The distribution of R-gene families structured into PTI (consisting in LRR-RLK, LRR, LysM, and LysM kinase), ETI (consisting in NBS-LRR), other Pattern Recognition Receptors (PRRs) divided into R-domains combinations (including NBS, TIR, RG, NBS-Pkinase, NBS-WRKY, TIR-NBS, TIR-NBS-Pkinase, TIR-Pkinase, hereafter 'R-combination') and genes involved in disease resistance pathway (including WRKYs and protein-kinases, hereafter 'R-pathway') is available as Additional file 1: Figure S3. The observed distribution of R-gene families in the investigated species from the more abundant is R-pathway (13189) > PTI (5844) > R-combination (2525) > ETI (2070).

In order to reconstruct the ancestral R-genes repertoire in plants, we used the recently reconstructed ancestral monocot (5 protochromosomes) and eudicot (7 protochromosomes) karyotypes to investigate the R-genes evolutionary dynamic. In Figure 1C, the evolutionary scenario of the modern grass genomes deriving from a $n = 5$ ancestor is illustrated [66]. Circular distributions illuminated the conservation rate of R-genes families excluding Pkinases (Figure 1C top) and their abundances within the different species (Figure 1C bottom). In the four grasses, the distribution of R-domains appeared very similar, except for RG, more abundant in *Brachypodium* compared to the other grasses investigated (Fisher Exact Test P -value = 4.10E-08, 1.46E-09 and 7.38E-08 in comparison to rice, sorghum, and maize; illustrated as red star in Figure 1C bottom). This phenomenon can simply be explained by the differences in RGA annotation and functional characterization efforts in the different species investigated. Finally, WRKY appeared less abundant in rice compared with the other grasses (Fisher Exact Test P -value = 3.41E-04, 2.70E-05, 9.47E-07 in comparison to rice *Brachypodium*, sorghum, and maize; illustrated as blue star in Figure 1C bottom). According to the reconstructed five protochromosomes, dating back to ~50-70 mya before the speciation of the four modern species investigated (containing 1622, 732, 845, and 832

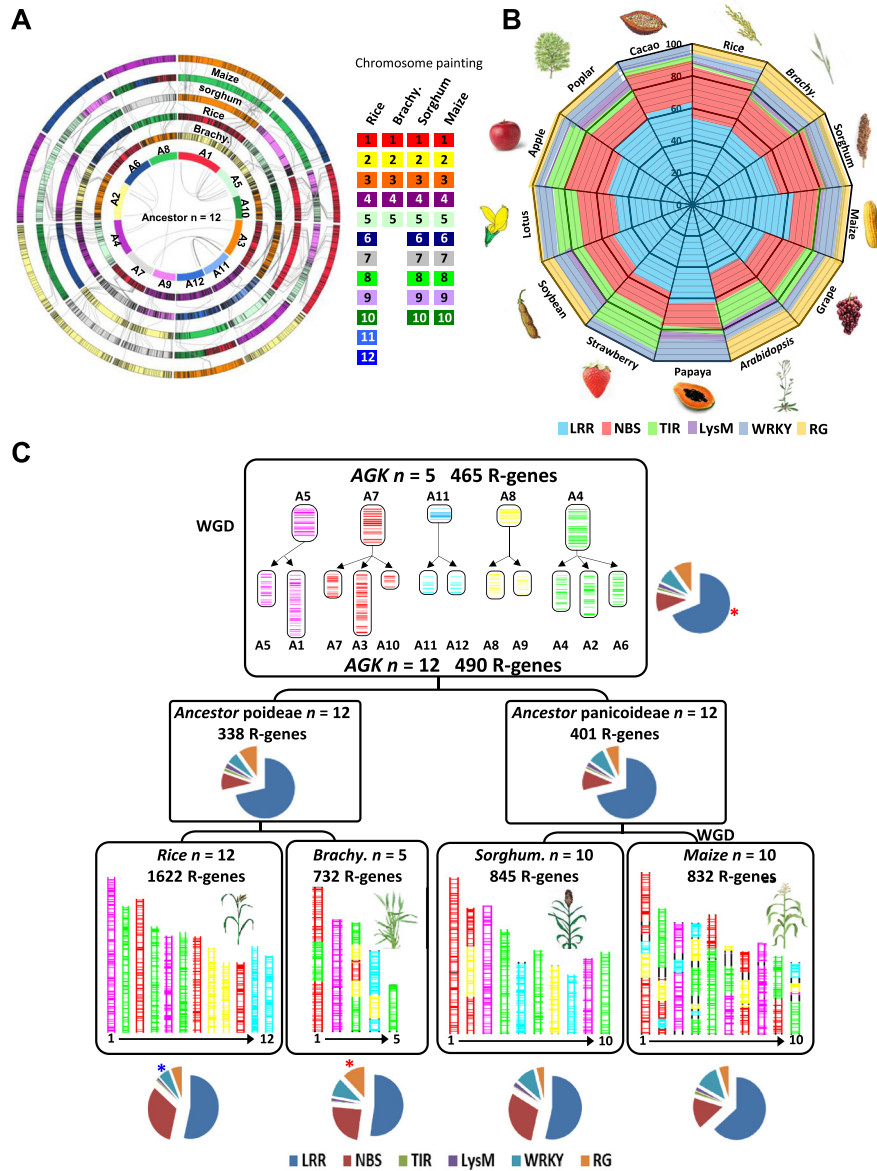


Figure 1 R-genes conservation and evolution in plants. (A) Grass genome synteny is illustrated as concentric circles. The chromosomes are highlighted with a color code (right) that illuminates the $n = 12$ monocot ancestral genome structure (inner circle A1 to A12). Any radius of the circle shows orthologous chromosomes between *Brachypodium*, rice, sorghum, and maize genomes. Maize genome is depicted as a double circle originating from the maize-specific recent WGD. Colinear R-genes are linked with black lines between circles, and ancestral duplicated R-genes are linked with black lines at the center of the circle. **(B)** R-genes content from 13 plant genomes including monocots (rice, *Brachypodium*, sorghum, and maize) and eudicots (*Arabidopsis*, Grape, Cacao, Papaya, Strawberry, Poplar, Lotus, Apple, and Soybean). The color code (bottom) highlights the R-gene classes investigated (LRR, NBS, TIR, LysM, RG). **(C)** Evolutionary scenario of R-genes in monocots. The modern grass genome structures (bottom) are depicted with a five-color code that illuminates their relationship with the $n = 5$ (A5, A7, A11, A8, A4) and $n = 12$ (A1 to A12) ancestors (top), according to Murat et al. [67]. The characterized R-genes are illustrated as vertical bars on the chromosomes of modern and ancestral genomes. The percentages of R-gene classes (LRR, NBS, TIR, LysM, RG), highlighted with the color code legend at the bottom, are shown with circular distributions for the four monocot genomes (bottom), the rice/*Brachypodium* and sorghum/maize ancestral genome intermediates (center), as well as for the ancestral karyotype (top). Statistically enriched and impoverished R-gene families are illustrated respectively with red and blue dots on the circular distributions.

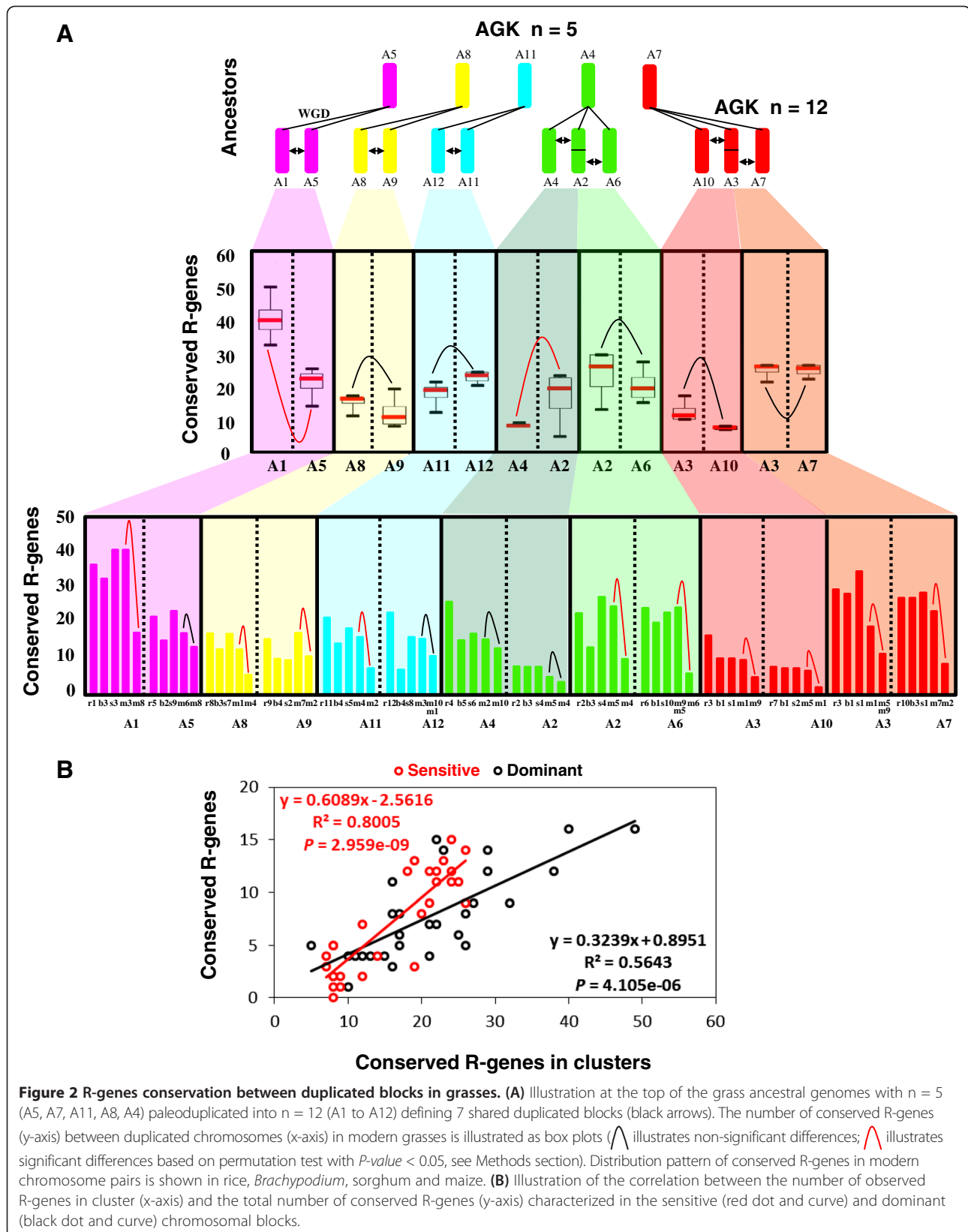
R-genes excluding Pkinase domains in rice, *Brachypodium*, sorghum, and maize respectively), we were able to reconstruct a minimal founder (conserved) pool of 465 ancestral R-domains consisting in 361 LRR, 54 NBS, 6 TIR, 11 LySM, 42 WRKY, and 52 RG (Figure 1C top, Additional file 3: Dataset S2). Based on the same strategy, the evolutionary scenario of the eudicots has been used to unravel a minimal founder pool of 150 R-genes (Additional file 1: Figure S4). In order to understand in more details the evolution of the major PTI/ETI families, we have reconstructed their ancestral pools. The results suggests that PTI genes content in the ancestors are significantly higher than observed in each modern species (P -value = 6.33E-04, 3.98E-04, 3.37E-04 in grasses ancestor, rice-*Brachypodium* ancestor and sorghum-maize ancestor respectively, Additional file 1: Figure S5), while the ETI genes content is lower in ancestors compared to modern species (P -value = 1.12E-04, 5.03E-04, 3.97E-04 in grasses ancestor, rice-*Brachypodium* ancestor and sorghum-maize ancestor respectively). This result may suggest an opposite evolutionary trend between PTI and ETI families that are respectively lost and gained in the course of evolution.

R-genes plasticity in response to duplication events

We wanted to investigate the impact of polyploidy (or whole genome duplication, hereafter WGD) in shaping the modern R-genes repertoire deriving from a founder pool of 465 (dating back from ~100 mya) and 150 (dating back from ~250 mya) R-genes for respectively the grasses and rosids. While massive duplicated gene deletion in the course of evolution following WGD has been reported in the literature [68,69], then leading to orthologous dominant (*i.e.* retention of ancestral genes) and sensitive (*i.e.* deletion of ancestral genes) blocks, the particular evolutionary fate of R-genes in response of such diploidization phenomenon is still not well established. To understand the R-genes family plasticity in response to WGDs, we used monocots as a model system to investigate the retention of R-genes (excluding Pkinases) in duplicated fragment pairs (Figure 2A). It has been shown that protein-coding genes behave differently in response to this diploidization process. Diploidization resistant genes (*i.e.* gene functions retained as duplicates following WGDs) are mainly transcription factors (TFs), transcription regulators (TRs) as well as miRNAs to a less extent, whereas the remaining gene families are considered as diploidization sensitive in returning to a singleton status after WGDs *via* selective gene deletion between dominant and sensitive chromosomal blocks [70,71]. At the whole genome level, we observed that only ~5% (25 out of 465) ancestral R-genes mapped on the grass ancestor ($n = 12$) were co-retained (*i.e.* paralogous genes observed in the ancestral duplicated chromosome pairs), a much lower

rate than the one reported for transcription factors (TF) as well as miRNA genes with up to 50% of observed co-retention of ancestral duplicates [68,72]. Therefore, the observed lower co-retention of ancestral paralogous R-genes may suggest that R-genes act as diploidization sensitive genes in returning to a singleton status after WGD. However, for 71.43% (five out of seven) duplicated chromosome pairs, the previous characterized deletion of the diploidization sensitive R-genes does not follows the subgenome dominance hypothesis (except for A1/A5 and A2/A4; highlighted with red connecting lines in Figure 2A bottom) in deriving dominant and sensitive blocks [68,73]. Instead, we observed that R-genes retention after WGD is equally distributed between ancestral chromosome pairs A8/9, A11/A12, A2/A6, A3/A7, and A3/A10 (permutation test with P -value < 5%, Figure 2A and Additional file 1: Table S3) in the four grass species. Intriguingly, after the recent WGD in maize, for 64.29% (nine out of fourteen paralogous pairs highlighted with red connecting lines in Figure 2A bottom) of the duplicated chromosomes, R-genes deletion was partitioned between the paralogous pairs (permutation test with P -value < 5%, Additional file 1: Table S4 and Additional file 1: Figure S6).

Such observed absence in R-genes deletion partitioning in grasses between ancestral duplicated chromosomes may be due to R-gene clusters identified as more abundant in ancestral sensitive chromosomes compared to dominant chromosomal compartment ($R^2 = 0.64$ with P -value = 4.73E-06 for sensitive chromosomes and $R^2 = 0.34$ with P -value = 3.70E-03 for dominant chromosomes; Figure 2B and Additional file 1: Table S5). For example, between A2 (reported as dominant) and A4 (reported as sensitive), there is no cluster located in the A2 in contrast to seven genes in three clusters on A4, while between A1 (reported as dominant) and A5 (reported as sensitive) more clusters in the A5 (16 genes in five clusters) do not reverse or reduce the reported dominance of A1 (24 genes in ten clusters). In contrast, R-genes clusters in maize did not affected the observed bias retention of duplicated R-genes between paralogous fragments excluding for A5 (m6 vs m8), A12 (m3 vs m1/10), A2 (m4 vs m5) and A4 (m2 vs m10); Figure 2A (bottom), Additional file 1: Table S6. This result may indicate that the random deletion of R-genes after WGD, not following the known subgenome dominance rule for the ancestral tetraploidization, may be a consequence of the high plasticity of such gene family evolving particularly in local tandem duplications (also referenced as clusterization in the next section) that may have compensated the ancestral biased deletion of duplicates in known sensitive subgenomes in the course of evolution. However, for the recent WGD in maize dating back to 5 mya, tandem duplications or clusters can't offset the dominance/sensitivity effect in such short period of



time. Thus, our data led to the hypothesis that R-genes, identified as diploidization sensitive genes, may have followed the subgenome dominance hypothesis that was compensated in the course of history by a reshuffling return flow consisting in local tandem duplications, enriching sensitive genomic compartments in R-genes content.

R-genes plasticity via clusterization and transposition mechanisms

While R-genes have been reported to be clustered in grass chromosomes [34] based on few species or loci investigated, a large-scale investigation of this phenomenon is still lacking in monocots and dicots. The structural definition of a R-gene cluster was considered following a previous study where linked (*i.e.* clustered) R-genes were not interrupted by more than eight non-R-genes [74]. In Additional file 1: Table S7, we reported that there are about 69% and 63% R-genes on average organized in clusters in monocots and eudicots respectively, suggesting that R-genes families expanded by lineage-specific tandem duplications leading to duplicated gene copy variants associated with high sequence similarities. Surprisingly, in poplar for example, we detected only 32% of R-genes organized in clusters using the same strategy, then unraveling possible specific patterns of R-gene clusterization between species. The typology of the R-genes clusters differs between species then reinforcing the concept of a recent clusterization process with most (58% on average) of them consisting in clusters of two locally duplicated R-genes, especially in maize (72%), while the largest and rare cluster, made of six R-genes, was only observed in rice (Additional file 1: Figure S6).

The particular evolution of R-genes *via* clusterization was highly dynamic through lineage-specific rearrangements leading to the observed conservation/erosion of R-genes colinearity between grasses, referenced as Copy Number Variation (CNV) and Presence/Absence Variation (PAV). The Figure 3A illustrates an orthologous R-gene locus in grasses involving a 178 kb rice region on chromosome 1 (containing four R-genes in clusters), a 68 kb region on the *Brachypodium* chromosome 2 (a single R-gene), a 161 kb region of the sorghum chromosome 3 (a single R-gene), and the two orthologous regions in maize on chromosomes 3 (232 kb) and 8 (113 kb) both with no R-gene annotated. This example illustrates the extreme structural variation in R-gene content between orthologous regions. On the two maize paralogous regions no R-gene were identified, while transposable elements (TEs) are found with high concentration (green rectangles as shown in Figure 3A). Their presences suggest that TEs may be involved in the loss of R-genes through illegitimate recombination. The reconstruction of the evolutionary history of this locus illustrates the plasticity of the R-gene

family where a single ancestral R-gene is retained in the modern *Brachypodium* and sorghum genomes but evolved into CNV in rice (four copies) and PAV (no R-gene) in maize.

The Figure 3B illustrates the retention of duplicated loci with 90% of sequence similarity in cacao, one locus with one R-gene (Tc06p002810 consisting TIR-NBS-LRR domains), and the duplicated locus harboring two R-genes (*i.e.* Tc03p00890 with a TIR domain and Tc03p00900 with NBS-LRR domains). We located precisely TSD (Target Site Duplication) motifs and long terminal reverse duplication, suggesting Tc06p002810 as the acceptor site (Figure 3B, Dotplot horizontal axis) and the duplicated region with two genes as the donor site (Figure 3B, Dotplot vertical axis). The acceptor site is characterized by a 7 kb fragmental deletion, which is located in the intergenic region of the donor site between Tc03p00890 and Tc03p00900 (with no transposable element or repeat detected in this particular fragment). The deletion of the intergenic fragment between the neighbor genes may have led to the read through of the ORFs leading the two neighbor genes fused into a single one (Tc06p002810) in the course of evolution. These paralogous regions from cacao may then suggest duplication as a major process resulting in domain shuffling (then reducing colinearity between species) between tandem duplicated R-genes.

Such R-gene structural plasticity may also be driven by TEs as we illustrated in the Figure 3C with two tandem duplicated R-genes with about 80% of sequence similarity from the rice genome, *i.e.* Os06g16300 (LRR domain) and Os06g16330 (LRR-Pkinase domains). Using LTR_Finder [75], we identified a 240 bp ancient LTRs flanking the Os06g16300 gene (dark blue arrow) as well as 5 bp TSD motifs associated with 1.1 kb recent LTRs (light blue arrow) flanking two transposase genes, *i.e.* Os06g16310 and Os06g16320. We then proposed an evolutionary scenario for this locus, where Os06g16300 is the ancestral gene (conserved with the modern *Brachypodium* gene, Bradi1g43690) that have been partially (illustrated as red exons) duplicated in tandem as Os06g16330 and finally physically separated by the TE-based transposition of the two transposases (Os06g16310 and Os06g16320). Overall, this example of tandem duplication followed by TE-based transposition events illustrates another source of R-gene plasticity reported in the current analysis, leading to R-gene synteny erosion between closely related species.

Taking into account the previous case examples obtained from cacao and rice genomes and in order to investigate R-gene synteny erosion at the whole genome level, we aligned the non-syntenic R-genes with the total R-genes repertoire. Using the parameters CIP \geq 70% and CALP \geq 70% [66] to identify the paired non-self matches, and according to the similarity of flanked protein-coding genes between the paired R-genes using

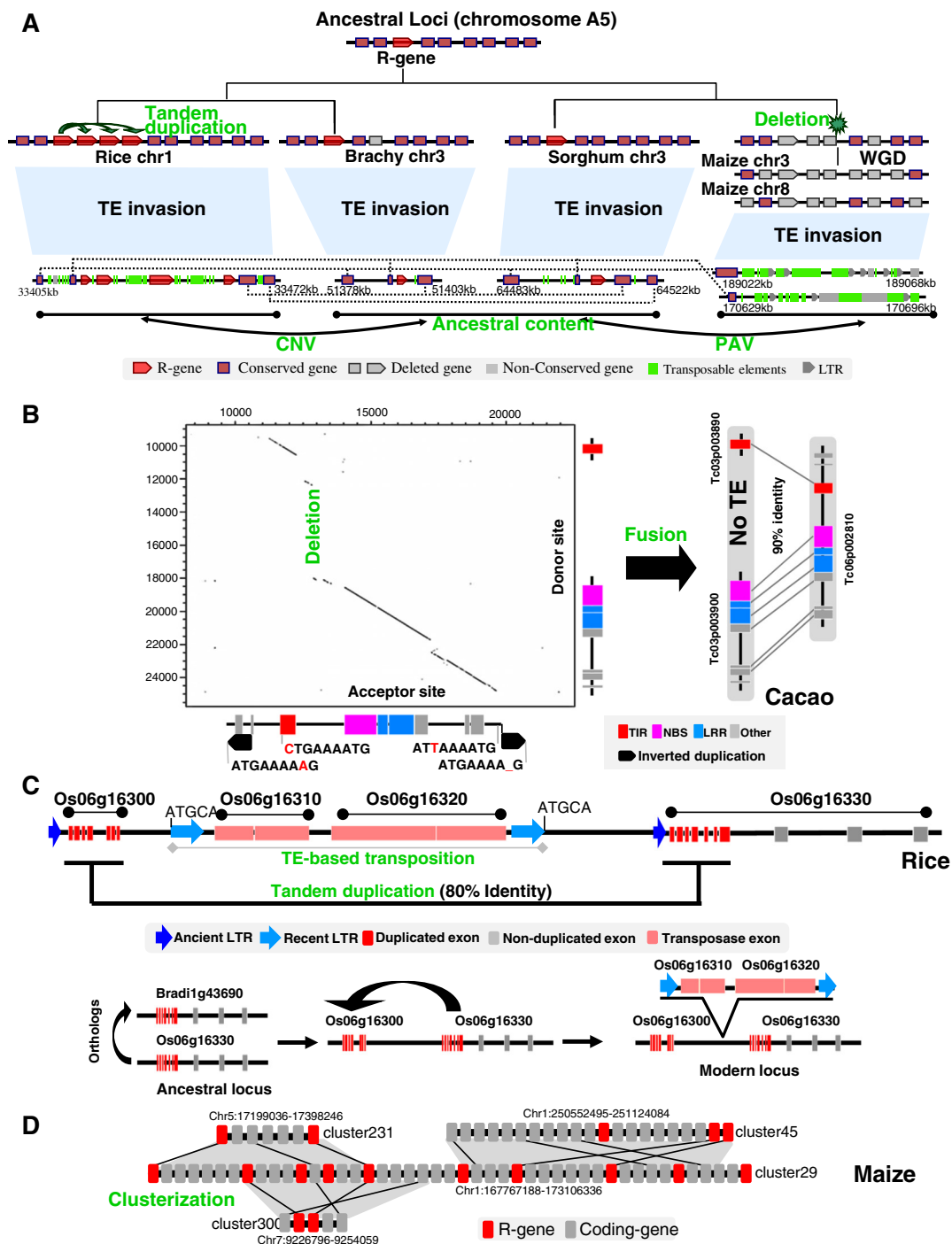


Figure 3 Different sources of R-genes plasticity in plants. **(A)** Evolutionary history of a locus located on the ancestral chromosome A5 showing R-gene conservation as well as CNV and PAV between rice, *Brachypodium*, sorghum and maize. Conserved genes, non-conserved genes, deleted genes and transposable elements are illustrated according to the legend at the bottom. Dotted black lines link orthologous genes between modern loci at the bottom. The ancestral gene content is illustrated at the top. **(B)** Illustration of an example of R-gene fusion in cacao after duplication. The dotplot illustrates the two copy-paste regions flanked by TSD (Target Site Duplication as black arrows) motifs and repeated sequence motifs (ctgaaaatg/attaaaatg). R-genes (TIR, NBS, LRR) classes are illustrated according to the color code at the bottom. **(C)** Illustration of a R-gene (Os06g16330) partially duplicated (Os06g16300) in rice separated by a transposition event (Os06g16310- Os06g16320). The reconstructed evolutionary scenario is illustrated based on a color code illuminating repeat (ancient, recent) and gene (duplicated, non-duplicated, transposase) content. **(D)** Duplicated R-gene cluster plasticity in maize. One central R-gene cluster (chromosome 1) consisting in R-genes (red) and non R-genes (grey) is duplicated on homeologous regions (chromosomes 1-5-7) with distinct R-gene contents.

E-value $< e^{-10}$ as a blast threshold, we distinguished segmental from single-gene duplications (also referenced as Small-Scale Duplication *i.e.* SSD) from these gene pairs (detailed in Methods section). In rice, *Brachypodium*, sorghum, and maize, we found 13.04% (153 out of 1173), 9.08% (74 out of 815), 12.35% (115 out of 931), and 35.63% (404 out of 1134) R-genes loci (R-genes located in the same clusters were considered as a single locus) involved in single-gene duplications (Additional file 1: Table S8 and Additional file 4: Dataset S3), which is higher than the 5% to 7% of single-gene duplication frequency reported for the total annotated protein-coding genes in grasses [76]. Among grasses, the single-gene duplication frequency in maize is significantly higher than in rice, *Brachypodium*, and sorghum respectively (P -value = $6.32E-24$, $2.28E-29$, and $1.71E-22$ in Fisher's Exact Test respectively, cf Additional file 1: Table S9). In addition, we observed hotspots of single-gene duplications where R-loci showing higher sequence similarity with at least two other non-related R-loci was considered as hotspot (Figure 3D). In maize, 51.73% (209 out of 404) of the single-gene duplications frequency was observed, a much higher rate compared to 23.53% (36 out of 153), 37.84% (28 out of 74), and 31.30% (36 out of 115) in rice, *Brachypodium*, and sorghum respectively (Additional file 1: Table S8). We can then speculate that the recent WGD in maize, dating back to 5 mya, may have promoted and accelerated R-gene singleton duplication frequency compared to the other grasses.

Homologous R-genes sequences within clusters generated by tandem duplications provided the structural template to form novel R-gene informs through domain recombinations. We characterized all the different R-domains in modern clusters and observed a specific domain affinity for clusterization (Additional file 1: Figure S7A and Additional file 1: Table S2). NBS-LRR and LRR-Pkinase combinations are observed as representing the majority of domain combinations in clusters (on average 31.78% and 42.63% out of the total R-domains in clusters for NBS-LRR, 51.08% and 46.47% for LRR-Pkinase domains, respectively in monocots and eudicots), compared to rare observed combinations in clusters for LRR-NBS-Pkinase-WRKY or LRR-TIR-WRKY (Additional file 1: Table S2). More interestingly, we observed a preference or affinity in domain combinations where more than 90% of them included LRR (Additional file 1: Figure S7B), with a preferential observed R-domain association with Pkinase (59% and 52% in monocots and eudicots respectively; Additional file 1: Figure S7C) and NBS (41% and 46% in monocots and eudicots respectively; Additional file 1: Figure S7C) domains. Therefore, our data confirm and largely refine previous conclusions suggesting LRR as a 'glue' for domain association leading to new combinations of R-gene domains observed in modern species, one major

source of R-gene plasticity. This dynamic recombination of R-domains within clusters, especially enriching NBS-LRR associations, may promote the development a novel source of disease resistance in the investigated species.

R-gene plasticity mediated by miRNA/R-gene interactome MiRNAs, as a versatile class of post-transcriptional gene regulator, are reported to be involved in a large variety of cellular processes, including development and defense responses in plants [77-79]. Small RNA cloning and high-throughput sequencing from plants infected by pathogens have shown that many microRNAs [80-82] and siRNAs [83] may be involved in biotic defense responses through up or down regulation of targeted gene expression. We wanted then to investigate whether the miRNA/R-gene interactome had an impact on the R-genes evolutionary plasticity as the role of miRNAs in plant immunity system has been largely reported in the literature [80-82]. To unveil if the plant paleoevolution has affected or even shaped the R-gene/miRNA interactome, we investigated miRNAs potentially targeting R-genes in the four monocots investigated (rice, *Brachypodium*, sorghum, and maize) as well as in nine eudicots (grape, *Arabidopsis*, strawberry, cacao, papaya, poplar, soybean, apple, and lotus).

We considered resistance genes as *in silico* targets of miRNAs based on sequence mismatch scores using Targetfinder algorithm [84] (detailed in Methods section). On average, we characterized 33.31% and 35.60% R-genes predicted as *in silico* targets of miRNA in monocots and eudicots respectively, significantly higher than for non R-genes with 11.48% and 13.06% respectively (P -value = $9.013e-05$ in paired student t-test, Additional file 1: Table S10). No highly significant differences were observed between non-conserved and conserved R-genes targeted *in silico* by miRNAs in monocot (Table 2). In eudicots, these differences (P -value = $3.48E-02$ between conserved and non-conserved miRNA-targeted resistance genes) are likely to be associated with the numerous rounds of WGD. We observed a correlation ($r = 0.7133$ with P -value = 0.03) between the number of WGD rounds and the number of *in silico* miRNA/R-gene interactions that took place in the plant paleohistory (Figure 4A). This observation may suggest that successive WGDs may have increased or putatively shaped the R-gene/miRNA *in silico* interactome. After recent WGDs, for example, in soybean, ~50% of retained R-genes (Figure 4B) are potential targeted by miRNAs with mismatch score of ≤ 4 . One explanation could be that additional species-specific R-genes copies (deriving from lineage-specific WGDs), then leading to R-gene functional redundancy, may be repressed at the expressional level through miRNAs. Such suggested impact of miRNA regulation on duplicated R-genes expression may need to be biologically and functionally validated.

Table 2 miRNA repertoire targeting conserved and lineage-specific R-genes in plants

Lineage-specific WGD rounds	Conserved ¹				Non-conserved ²		
	Species	Targets ³	Total orth ⁴	(%) ⁵	Targets ³	Total non-orth ⁴	(%) ⁵
Monocot							
0 WGD	OS	178	521	34.17	771	2116	36.44
0 WGD	BD	156	495	31.52	319	1167	27.34
0 WGD	SB	158	413	38.26	489	1304	37.50
1 WGD	ZM	105	319	32.92	448	1548	28.94
Eudicot							
0 WGD	VV	29	282	10.28	124	796	15.58
0 WGD	TC	77	149	51.68	554	1290	42.95
0 WGD	CP	22	101	21.78	116	602	19.27
0 WGD	FV	39	108	36.11	275	1344	20.46
1 WGD	MD	49	125	39.20	1516	4127	36.73
1 WGD	PT	59	122	48.36	389	1175	33.11
1 WGD	LJ	23	46	50.00	219	622	35.21
2 WGD	AT	51	74	68.92	685	1485	46.13
2 WGD	GM	72	148	48.65	1307	3162	41.33

Note: OS, BD, SB, ZM, VV, TC, CP, FV, MD, PT, LJ, AT, and GM represent Rice, *Brachypodium*, Sorghum, Maize, Grape, Cacao, Papaya, Strawberry, Poplar, Lotus, *Arabidopsis*, and Soybean, respectively; ¹Conserved R-genes are associated with orthologous genes with other monocots and eudicots species; ²Non-conserved R-genes do not have orthologous genes with other species either monocots and eudicots; ³Number of R-genes targeted by miRNAs; ⁴Number of total orthologous genes; ⁵Percentage of R-genes targeted by miRNAs.

Finally, we investigated the R-domains/miRNA affinity and observed that NBS/TIR > LRR > WRKY/PKinase domains are preferentially targeted by miRNAs (P -value = 2.17E-03 and 8.45E-03, 1.21E-03, and 3.97E-03 with paired student t-test for NBS vs LRR, TIR vs LRR, LRR vs WRKY, LRR vs Pkinase, respectively; Figure 4C & Additional file 1: Table S11). We also observed that 67% and 63% of R-genes clusters are targeted by miRNA in contrast to 33% and 37% of singleton R-genes, respectively in eudicots and monocots (Additional file 1: Figure S8 and Additional file 1: Table S12). Overall, the interaction affinity between species-specific R-genes (in clusters and in majority involving NBS/TIR domains) and miRNAs after WGD can be considered as a major source of R-gene family plasticity in plants, as part of a possible functional diploidization of structurally retained duplicated R-genes.

Discussion

Diploidization following duplication as a major source of R-genes structural plasticity

Most of the investigated rosids (grape, *Arabidopsis*, soybean, poplar, and cacao) species experienced up to three WGD events, whereas the investigated grasses (rice, maize, sorghum, and *Brachypodium*) went through one shared ancestral WGD during their evolution, except for maize which experienced a recent extra-WGD 5 mya [66]. Biased erosion of duplicated gene redundancy between sister blocks has been characterized recently in plants defining

dominant and sensitive blocks [68,69]. In our current analysis, the identification of 23641 R-genes sequences in angiosperms established a higher R-genes conservation in grasses (on average 23.8%) compared to rosids (on average 6.6%) suggesting that successive rounds of WGDs act as a decay into R-genes conservation, as a primer source of R-gene plasticity. The evolutionary investigation of the characterized R-genes repertoire allowed the reconstruction of minimal ancestral pool of 465 and 150 founder R-genes respectively for the grasses and rosids.

Tandem duplication or clusterization played an important role in R-genes plasticity leading to structural variations such as CNV/PAV between species, which are thought to contribute to the reported tremendous R-genes diversity [85]. Special expansion of tandem duplications especially in sensitive chromosomes, as a rapid counterbalance flow of the duplicates deletion phenomenon, may have compensate R-genes loss in such chromosomal fragments as an expected consequence of the known diploidization process. R-genes clusters may have been shaped by classically proposed shuffling mechanisms such as replication slippage, segmental duplication *via* homologous/non-homologous unequal crossover, transposition *via* ectopic recombination/TE capture [86]. Such clusters may have been shaped by domain shuffling events [87], domain breakage and fusion so that R-domain combinations such as LRR-NBS-PK-TIR and LRR-NBS-PK-WRKY might be the result of local shuffling and recombination

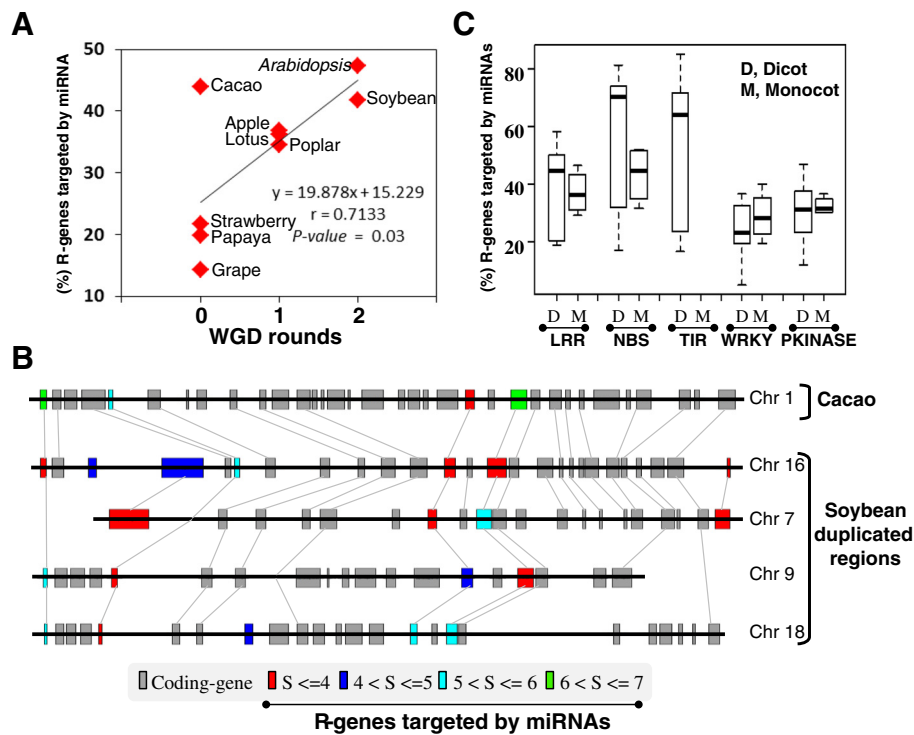


Figure 4 R-genes/miRNAs interactome in plants. (A) Illustration of the percentage of R-genes targeted by miRNA in dicots species classified according the number of experienced WGDs (x-axis). The regression curve, correlation and associated *P*-value are mentioned. **(B)** Illustration of a micro-synteny locus between cacao (one region) and soybean (four duplicated regions) harboring R-genes targeted by miRNA (according to the number of sequence mismatches between miRNA and R-genes from 4 to 7 identified as miRNA target score and highlighted with a color code at the bottom). Grey bars represent non R-genes. **(C)** Illustration of the percentage of R-genes targeted by miRNA (y-axis) in dicots (D) and monocots (M) species classified according to the investigated R-domains (LRR, NBS, TIR, WRKY, Pkinase; x-axis).

events. Altogether, R-gene clusterization, a center source of plasticity, triggered a serial of reshuffling events to make rapid copy variation leading to PAVs and CNVs between species, as a putative source for R-gene structural diversity. In the current analysis we observed up to 60% of R-genes organized in clusters in grasses and rosids. Genic and intergenic sequence repeats within R-clusters generated by duplication, transpositions and insertions provide a structural template that allows mis-pairing during recombination giving rise to unequal crossovers and interlocus gene conversions/rearrangements. The resulting R-domain combinations appeared not random with LRR as a 'glue' for domain association leading to new resistance gene isoforms in modern plant species.

MicroRNA/R-gene interactome as a major source of R-genes functional diploidization

Recently, new evidences have been proposed regarding miRNAs regulating NBS-LRR in plants such as miR2109/miR2118/miR1507 in Medicago, miR482/miR2118 in tomato, and miR6019/miR6020 in tobacco guiding the cleavage of transcript of NBS-LRRs, and then triggering the

secondary phased of siRNA production by RNA-dependent RNA polymerase [88-90]. Thus, MiRNAs may be involved in defense immunity in regulating R-genes expression level. Our *in silico* analysis suggests that miRNA may target preferentially duplicated R-genes either deriving from WGDs and more interestingly from local tandem duplications (*i.e.* clusters). This observation may suggest that the presence of redundant duplicated R-genes copies, when retained after diploidization, may require modification or specialization in expression/regulation through possibly miRNA interaction. Moreover, a specific R-domain affinity was observed for miRNA *in silico* interaction toward LRR/NBS/TIR, which may indicate to some extent a domain preference for R-gene/miRNA interaction. Overall, our data may suggest miRNAs as a dosage regulator playing a possible role in R-genes functional redundancy erosion following large or local duplication events. The preferential post-transcriptional regulation of duplicated R-genes by miRNA can be proposed as part of a functional diploidization process in response to duplications to maintain a perfect dosage balance regarding the product of R-genes duplicates.

Putative model of R-genes paleohistory in plants

Based on our analysis and previous studies, we proposed an hypothetical evolutionary model in Figure 5 illustrating the conclusion regarding the reported R-gene conservation/diversity, polyploidization, domain reshuffling as well

as microRNA/R-gene interactome. If duplicated R-genes are deleted between paralogous fragments after whole genome duplications following the general subgenome dominance hypothesis [68,69], they can be then consider as dosage-sensitive genes in returning to singleton status

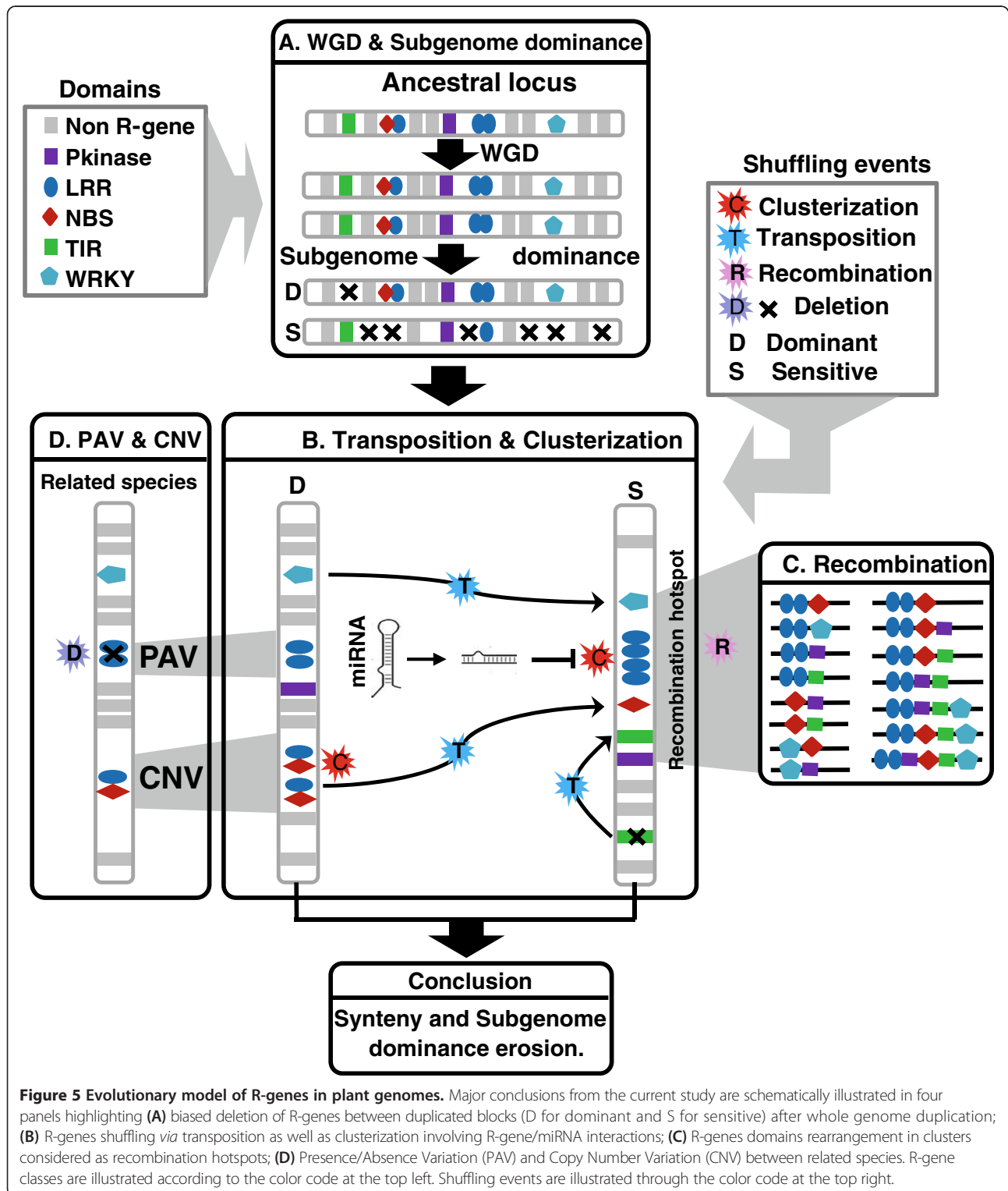


Figure 5 Evolutionary model of R-genes in plant genomes. Major conclusions from the current study are schematically illustrated in four panels highlighting (A) biased deletion of R-genes between duplicated blocks (D for dominant and S for sensitive) after whole genome duplication; (B) R-genes shuffling via transposition as well as clusterization involving R-gene/miRNA interactions; (C) R-genes domains rearrangement in clusters considered as recombination hotspots; (D) Presence/Absence Variation (PAV) and Copy Number Variation (CNV) between related species. R-gene classes are illustrated according to the color code at the top left. Shuffling events are illustrated through the color code at the top right.

after WGDs and then are biased distributed on the modern pairs of duplicated chromosomes (Figure 5A). However, during evolution, if the deletion of R-genes in sensitive chromosome is compensated by high frequency of tandem duplication (clusterization) and/or transposition events, an equivalent number of R-genes between dominant and sensitive fragments is observed in modern plant genomes (Figure 5B). Despite the previous structural R-genes plasticity scenario, R-genes functional redundancy within clusters or in general between duplicated loci may be counterbalanced through microRNAs regulation with a specific affinity for LRR/NBS/TIR domains. Therefore, R-genes colinearity as well as subgenome dominance following WGDs has been eroded in the course of evolution. Finally R-genes clusterization may be considered as R-domain recombination hotspots, potential source of new domain combinations then possibly facilitating the neoformation or neo-functionalization of R-genes isoforms (Figure 5C). Overall, when comparing modern plant species for their R-genes content, copy number variation (CNV), or Present/Absent variation (PAV) are generally observed (Figure 5D).

Conclusions

We reconstructed the R-genes paleohistory in plant unraveling duplications (either whole genome, small-scale clusters or single-gene based) as the major source of structural (CNV, PAV, domain recombination) or even potentially functional (enhanced miRNA regulation for R-gene clusters and specific R-domains) plasticity that may have promote the development a novel source of disease resistance in the course of the evolution of the different investigated species. The conserved role of similar R-gene families (especially TIR and NBS-LRR) in both plant and animal defense systems suggest a common and ancestral origin. The current reconstruction of the ancestral gene pool in angiosperm opens the perspective to determine the origin of innate immunity mechanism in eukaryotes.

Methods

R-gene identification and mapping

R-genes were selected from monocots *Oryza sativa*, *Sorghum bicolor*, *Zea mays*, *Brachypodium distachyon*, and eudicots *Arabidopsis thaliana*, *Populus trichocarpa*, *Carica papaya*, *Glycine max*, *Lotus japonicas*, *Fragaria vesca*, and *Theobroma cacao*, on the basis of functional annotations available on Phytosome (<http://www.phytosome.net/>), Plant GDB (<http://www.plantgdb.org/>). *Vitis vinifera* and *Malus x domestica* R-genes annotation were retrieved from [59] and [60] supplementary data. The R-genes identification methods is illustrated in Additional file 1: Figure S1A. *PFAM domain identification* – Putative R-genes were investigated using profile Hidden Markov Model. Several PFAM profiles [64]

were used to extract putative R-gene proteins within the 13 genomes investigated: LRR: pf00560, pf07723, pf07725, pf12799, pf01463, pf08263; NB-ARC: pf00931; TIR: pf01582; LysM: pf01476; Pkinase: pf00069; WRKY: pf03106. PFAM profiles were identified within genomes using the hmsearch algorithm (e-value cut: 1e-10) from HMMER3 (<http://hmmer.janelia.org/>; [91]). *PRGdb* – R-genes sequences from the Plant Resistance Gene database were downloaded (<http://prgdb.cbm.fvg.it/index.php>, [65]). Annotation, PFAM and PRGdb –based R-genes were aligned against Soybean (46194 protein sequences), Cacao (27814 protein sequences), Strawberry (34809 protein sequences), Lotus (15470 protein sequences), Papaya (19205 protein sequences), Poplar (30260 protein sequences), Apple (58979 protein sequences), *Brachypodium*, (32255 protein sequences) Sorghum (36338 protein sequences), and Maize (53764 protein sequences) genome data using BLASTP (PFAM and Annotation R-gene sequences) and BLASTX (PRGdb sequences). BLAST results were parsed using CIP (Cumulative Identity Percentage) and CALP (Cumulative Alignment Length Percentage) parameters (70% as minimum threshold) delivering a non-redundant list of R-genes for each species [44].

Orthologs/Paralogs identification and synteny relationships

Orthologous and paralogous R-genes were identified aligning Rice RefBank against *Brachypodium*, Sorghum and Maize using BLASTALL. BLASTP results are parsed with CIP and CALP parameters set to 60% and 70% as minimum threshold for ortholog identification and 60% and 70% as minimum threshold for paralog identification as described in [44]. Ancestral relationship between monocot species were represented as concentric circles with the visualization tool Circos [92]. Relationship between rosids species were investigated with the same protocol. Grape RefBank was aligned against Soybean, Cacao, Strawberry, Lotus, Papaya, Poplar and Apple genome data. BLAST results were parsed with the same parameters described previously. The synteny relations at the chromosome levels were considered using public synteny data available for both Monocotyledones and Eudicotyledones [66].

R-genes not located in the syntenic region were BLAST aligned against the total R-genes content. The gene pairs excluding self matches (CIP \geq 70%, CIAP \geq 70%) were considered as single-gene duplication and used to the further analysis. Then we selected 40 flanking genes windows surrounding R-genes pairs. If flanking pairs with E-value \leq e-10 are observed these paired R-genes were then considered as part of a segmental duplication, otherwise, as single-gene duplication.

Permutation-test for R-genes partitioning between duplicated blocks

In the absence of any biased retention/deletion of ancestral R-genes content (N) after WGD (null hypothesis), the post-duplication R-genes content is $2N = N + N$. After million years of evolution and associated shuffling, n_1 and n_2 R-genes are observed in modern duplicated chromosomes. We simulated the random deletion of R-genes 1000 times (number of deleted R-gene is equal with $(N - n_1)$ and $(N - n_2)$ in the two duplicated chromosomes respectively), and derived two sample datasets (Sample1: $X_1, X_2, X_3, \dots, X_{1000}$; Sample: $Y_1, Y_2, Y_3, \dots, Y_{1000}$) corresponding to the random deletion of R-genes between blocks. We performed Z-tests (see formula below) to test significant differences in the distribution of retained R-genes between the two duplicated chromosome pairs (null hypothesis rejected).

$$Z = \frac{X - u}{\frac{\sigma}{\sqrt{n}}} \left(u = \frac{\sum_{i=0}^{1000} (X_i - Y_i)}{n} \right);$$

$X = n_1 - n_2$; $n = 1000$; σ^2 , variance)

MiRNA identification associated with R-genes as targets

Mature miRNAs dataset from miRBase (<http://www.mirbase.org/>; Release 18) was used to predict R-genes as targets in the investigated plant genomes Targetfinder algorithm (<http://carringtonlab.org/resources/targetfinder/>) with score < 4 [84]. To reduce the false positive, secondary structures of the identified mature miRNA was validated using MiReNA software [93]. R-genes targeted by miRNA with validated secondary structure and with a mismatch score < 4 , are considered as *in silico* targets of miRNA. The detailed pipeline is illustrated in Additional file 1: Figure S1B.

Additional files

Additional file 1: Table S1. R-genes conservation in plants. **Table S2.** R-genes domains/family diversity in plants. **Table S3.** Number of R-genes in the ancestral duplicated chromosomes in grasses. **Table S4.** Number of R-genes in the recent duplicated maize chromosomes. **Table S5.** Number of R-genes clusters in ancient duplicated grass chromosomes. **Table S6.** Number of R-genes clusters in recent duplicated maize chromosomes. **Table S7.** R-genes clusters distribution in plants. **Table S8.** R-genes duplication frequency in maize. **Table S9.** R-genes duplication frequency in maize compared to other grasses. **Table S10.** R-genes targeted by miRNAs in plants. **Table S11.** R-domains targeted by miRNAs in eudicots. **Table S12.** R-genes cluster loci targeted by miRNAs in plants. **Figure S1.** R-genes and miRNA detection pipelines. **Figure S2.** R-domains distribution in plant genomes. **Figure S3.** R-genes family distribution in plant genomes. **Figure S4.** R-genes paleohistorical evolution in eudicots. **Figure S5.** Evolutionary scenario of R-genes families in monocots. **Figure S6.** R-genes distribution and content in clusters. **Figure S7.** R-domains combination in clusters. **Figure S8.** R-gene clusters targeted by miRNAs in plants.

Additional file 2: Dataset S1. R-gene repertoire and associated miRNAs in plants. The table provides the catalog of R-genes characterized in grasses including rice (OS), *Brachypodium* (BD), sorghum (SB), and maize (ZM), and rosids including grape (VV), *Arabidopsis* (AT), papaya (CP), Cacao (TC), Soybean (GM), Lotus (LJ), Medicago (MD), Strawberry (FV), and Poplar (PT).

Additional file 3: Dataset S2.

Ancestral R-gene content in grass ancestor. The table provides the catalog of R-genes characterized in the 5 ancestral chromosomes of grasses with associated R-domains and modern R-genes representatives characterized on rice, *Brachypodium*, sorghum and maize duplicated fragments.

Additional file 4: Dataset S3.

Duplicated R-genes repertoire in grasses. The table provides the catalog of non-syntenic duplicated R-genes characterized in rice (OS), *Brachypodium* (BD), sorghum (SB), and maize (ZM).

Competing interests

The authors (RZ, FM, CP, TL and JS) declare they have no competing interests as part of the same research unit (INRA/UBP UMR 1095 GDEC 'Génétique, Diversité et Ecophysiologie des Céréales', 5 chemin de Beaulieu, 63100 Clermont-Ferrand, France) when the study performed.

Authors' contributions

RZ, FM, and CP performed the analyses. RZ, TL and JS wrote the manuscript. All authors read and approved the final manuscripts.

Acknowledgements

This work has been supported by grants from the Agence Nationale de la Recherche (program ANR Blanc-PAGE, ref: ANR-2011-BSV6-00801).

Author details

¹INRA/UBP UMR 1095 GDEC 'Génétique, Diversité et Ecophysiologie des Céréales', 5 chemin de Beaulieu, 63100 Clermont-Ferrand, France. ²Institute of Crop Sciences, Shandong Academy of Agricultural Sciences, Jinan 250100, China.

Received: 5 September 2013 Accepted: 25 February 2014

Published: 12 March 2014

References

1. Glazebrook J: Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu Rev Phytopathol* 2005, **43**:205–227.
2. Pallas V, Garcia JA: How do plant viruses induce disease? Interactions and interference with host components. *J Gen Virol* 2011, **92**(Pt 12):2691–2705.
3. Soriano IR, Riley IT, Potter MJ, Bowers WS: Phytoecdysteroids: a novel defense against plant-parasitic nematodes. *J Chem Ecol* 2004, **30**(10):1885–1899.
4. Choy A, Roy CR: Autophagy and bacterial infection: an evolving arms race. *Trends Microbiol* 2013, **21**(9):451–456.
5. Chisholm ST, Coaker G, Day B, Staskawicz BJ: Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* 2006, **124**(4):803–814.
6. Medzhitov R, Janeway CA Jr: Innate immunity: the virtues of a nonclonal system of recognition. *Cell* 1997, **91**(3):295–298.
7. Nurnberger T, Brunner F: Innate immunity in plants and animals: emerging parallels between the recognition of general elicitors and pathogen-associated molecular patterns. *Curr Opin Plant Biol* 2002, **5**(4):318–324.
8. Mur LAJ, Kenton P, Lloyd AJ, Ougham H, Prats E: The hypersensitive response; the centenary is upon us but how much do we know? *J Exp Bot* 2008, **59**(3):20.
9. Hammond-Kosack KE, Parker JE: Deciphering plant-pathogen communication: fresh perspectives for molecular resistance breeding. *Curr Opin Biotechnol* 2003, **14**(2):177–193.
10. Glowacki S, Macioszek VK, Kononowicz AK: R proteins as fundamentals of plant innate immunity. *Cell Mol Biol Lett* 2011, **16**(1):1–24.
11. Qiu D, Xiao J, Ding X, Xiong M, Cai M, Cao Y, Li X, Xu C, Wang S: OsWRKY13 mediates rice disease resistance by regulating defense-related genes in salicylate- and jasmonate-dependent signaling. *Mol Plant Microbe Interact* 2007, **20**(5):492–499.
12. Shimono M, Koga H, Akagi A, Hayashi N, Goto S, Sawada M, Kurihara T, Matsushita A, Sugano S, Jiang CJ, Kaku H, Inoue H, Takatsuji H: Rice WRKY45 plays important roles in fungal and bacterial disease resistance. *Mol Plant Pathol* 2012, **13**(1):83–94.
13. Spaink HP: Specific recognition of bacteria by plant LysM domain receptor kinases. *Trends Microbiol* 2004, **12**(5):201–204.
14. Buist G, Steen A, Kok J, Kuipers OP: LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol Microbiol* 2008, **68**(4):838–847.
15. Witte CP, Keinath N, Dubiella U, Demouliere R, Seal A, Romeis T: Tobacco calcium-dependent protein kinases are differentially phosphorylated in vivo as part of a kinase cascade that regulates stress response. *J Biol Chem* 2010, **285**(13):9740–9748.

16. Kuruusu T, Hamada J, Nokajima H, Kitagawa Y, Kiyoduka M, Takahashi A, Hanamata S, Ohno R, Hayashi T, Okada K, Koga J, Hirochika H, Yamane H, Kuchitsu K: **Regulation of microbe-associated molecular pattern-induced hypersensitive cell death, phytoalexin production, and defense gene expression by calneurin B-like protein-interacting protein kinases, OsCIPK14/15, in rice cultured cells.** *Plant Physiol* 2010, **153**(2):678–692.
17. Bent AF: **Plant disease resistance genes: function meets structure.** *Plant Cell* 1996, **8**(10):1757–1771.
18. Wang J, Tan S, Zhang L, Li P, Tian D: **Co-variation among major classes of LRR-encoding genes in two pairs of plant species.** *J Mol Evol* 2011, **72**(5–6):498–509.
19. Panstruga R, Parker JE, Schulze-Lefert P: **SnapShot: plant immune response pathways.** *Cell* 2009, **136**(5):978. e971–973.
20. Yue JX, Meyers BC, Chen JQ, Tian D, Yang S: **Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes.** *New Phytol* 2012, **193**(4):1049–1063.
21. Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND: **Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily.** *Plant J* 1999, **20**(3):317–332.
22. Akita M, Valkonen JP: **A novel gene family in moss (*Physcomitrella patens*) shows sequence homology and a phylogenetic relationship with the TIR-NBS class of plant disease resistance genes.** *J Mol Evol* 2002, **55**(5):595–605.
23. Bertin J, Nir WJ, Fischer CM, Tayber OV, Errada PR, Grant JR, Keilty JJ, Gosselin ML, Robison KE, Wong GH, Glucksmann MA, DiStefano PS: **Human CARD4 protein is a novel CED-4/Apaf-1 cell death family member that activates NF-kappaB.** *J Biol Chem* 1999, **274**(19):12955–12958.
24. Meyers BC, Morgante M, Michelmore RW: **TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in Arabidopsis and other plant genomes.** *Plant J* 2002, **32**(1):77–92.
25. Liu JJ, Ekramoddoullah AK: **Isolation, genetic variation and expression of TIR-NBS-LRR resistance gene analogs from western white pine (*Pinus monticola* Dougl. ex. D. Don).** *Mol Genet Genomics* 2003, **270**(5):432–441.
26. Girardin SE, Sansonetti PJ, Philpott DJ: **Intracellular vs extracellular recognition of pathogens—common concepts in mammals and flies.** *Trends Microbiol* 2002, **10**(4):193–199.
27. Tarr DE, Alexander HM: **TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders.** *BMC Res Notes* 2009, **2**:197.
28. Bai J, Pennill LA, Ning J, Lee SW, Ramalingam J, Webb CA, Zhao B, Sun Q, Nelson JC, Leach JE, Hulbert SH: **Diversity in nucleotide binding site-leucine-rich repeat genes in cereals.** *Genome Res* 2002, **12**(12):1871–1884.
29. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW: **Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis.** *Plant Cell* 2003, **15**(4):809–834.
30. Yang S, Zhang X, Yue JX, Tian D, Chen JQ: **Recent duplications dominate NBS-encoding gene expansion in two woody species.** *Mol Genet Genomics* 2008, **280**(3):187–198.
31. Porter BW, Paidi M, Ming R, Alam M, Nishijima WT, Zhu YJ: **Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family.** *Mol Genet Genomics* 2009, **281**(6):609–626.
32. Leister D: **Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene.** *Trends Genet* 2004, **20**(3):116–122.
33. Luo S, Zhang Y, Hu Q, Chen J, Li K, Lu C, Liu H, Wang W, Kuang H: **Dynamic nucleotide-binding site and leucine-rich repeat-encoding genes in the grass family.** *Plant Physiol* 2012, **159**(1):197–210.
34. Guo YL, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D: **Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in Arabidopsis.** *Plant Physiol* 2011, **157**(2):757–769.
35. Shen J, Araki H, Chen L, Chen JQ, Tian D: **Unique evolutionary mechanism in R-genes under the presence/absence polymorphism in Arabidopsis thaliana.** *Genetics* 2006, **172**(2):1243–1250.
36. Yang S, Feng Z, Zhang X, Jiang K, Jin X, Hang Y, Chen JQ, Tian D: **Genome-wide investigation on the genetic variations of rice disease resistance genes.** *Plant Mol Biol* 2006, **62**(1–2):181–193.
37. Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Schulze-Lefert P: **Rapid reorganization of resistance gene homologues in cereal genomes.** *Proc Natl Acad Sci U S A* 1998, **95**(1):370–375.
38. Meyers BC, Kaushik S, Nandety RS: **Evolving disease resistance genes.** *Curr Opin Plant Biol* 2005, **8**(2):129–134.
39. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Ollilar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, et al: **The Sorghum bicolor genome and the diversification of grasses.** *Nature* 2009, **457**(7229):551–556.
40. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, et al: **The B73 maize genome: complexity, diversity, and dynamics.** *Science* 2009, **326**(5956):1112–1115.
41. International Rice Genome Sequencing P: **The map-based sequence of the rice genome.** *Nature* 2005, **436**(7052):793–800.
42. International Brachypodium I: **Genome sequencing and analysis of the model grass *Brachypodium distachyon*.** *Nature* 2010, **463**(7282):763–768.
43. Salse J, Abrouk M, Bolot S, Guilhot N, Courcelle E, Faraut T, Waugh R, Close TJ, Messing J, Feuillet C: **Reconstruction of monocotyledonous proto-chromosomes reveals faster evolution in plants than in animals.** *Proc Natl Acad Sci U S A* 2009, **106**(35):14908–14913.
44. Salse J, Abrouk M, Murat F, Quraishi UM, Feuillet C: **Improved criteria and comparative genomics tool provide new insights into grass paleogenomics.** *Brief Bioinform* 2009, **10**(6):619–630.
45. Bolot S, Abrouk M, Masood-Quraishi U, Stein N, Messing J, Feuillet C, Salse J: **The 'inner circle' of the cereal genomes.** *Curr Opin Plant Biol* 2009, **12**(2):119–125.
46. Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Huguency P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A, Segures B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gasparo G, Dumas V, et al: **The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla.** *Nature* 2007, **449**(7161):463–467.
47. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson D, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, et al: **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray).** *Science* 2006, **313**(5793):1596–1604.
48. Arabidopsis Genome I: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature* 2000, **408**(6814):796–815.
49. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, et al: **Genome sequence of the palaeopolyploid soybean.** *Nature* 2010, **463**(7278):178–183.
50. Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, Abrouk M, Murat F, Fouet O, Poulain J, Ruiz M, Roguet Y, Rodier-Goud M, Barbosa-Neto JF, Sabot F, Kudrna D, Ammiraju JS, Schuster SC, Carlson JE, Sallet E, Schiex T, Dievart A, Kramer M, Gelley L, Shi Z, Bérard A, et al: **The genome of *Theobroma cacao*.** *Nat Genet* 2011, **43**(2):101–108.
51. Abrouk M, Murat F, Pont C, Messing J, Jackson S, Faraut T, Tannier E, Plomion C, Cooke R, Feuillet C, Salse J: **Palaeogenomics of plants: synteny-based modelling of extinct ancestors.** *Trends Plant Sci* 2010, **15**(9):479–487.
52. Paterson AH, Bowers JE, Chapman BA: **Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics.** *Proc Natl Acad Sci U S A* 2004, **101**(26):9903–9908.
53. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH: **Synteny and collinearity in plant genomes.** *Science* 2008, **320**(5875):486–488.
54. Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH: **Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps.** *Genome Res* 2008, **18**(12):1944–1954.
55. Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K: **The flowering world: a tale of duplications.** *Trends Plant Sci* 2009, **14**(12):680–688.
56. Bowers JE, Chapman BA, Rong J, Paterson AH: **Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events.** *Nature* 2003, **422**(6930):433–438.
57. Lawton-Rauh A: **Evolutionary dynamics of duplicated genes in plants.** *Mol Phylogenet Evol* 2003, **29**(3):396–409.
58. Blanc G, Wolfe KH: **Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes.** *Plant Cell* 2004, **16**(7):1667–1678.

59. Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, Malacarne G, Iliev D, Coppola G, Wardell B, Micheletti D, Macalma T, Facci M, Mitchell JT, Perazzolli M, Eldredge G, Gatto P, Oyzerski R, Moretto M, Gutin N, Stefanini M, Chen Y, Segala C, Davenport C, Demattè L, Mraz A, et al: **A high quality draft consensus sequence of the genome of a heterozygous grapevine variety.** *PLoS One* 2007, **2**(12):e1326.
60. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, et al: **The genome of the domesticated apple (*Malus x domestica* Borkh).** *Nat Genet* 2010, **42**(10):833–839.
61. Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, Fujishiro T, Katoh M, Kohara M, Kishida Y, Minami C, Nakayama S, Nakazaki N, Shimizu Y, Shinpo S, Takahashi C, Wada T, Yamada M, Ohmido N, Hayashi M, Fukui K, Baba T, Nakamichi T, Mori H, Tabata S: **Genome structure of the legume, *Lotus japonicus*.** *DNA Res* 2008, **15**(4):227–239.
62. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, Burns P, Davis TM, Slovin JP, Bassil N, Hellens RP, Evans C, Harkins T, Kodira C, Desany B, Crasta OR, Jensen RV, Allan AC, Michael TP, Setubal JC, Celton JM, Rees DJ, Williams KP, Holt SH, Ruiz Rojas JJ, Chatterjee M, et al: **The genome of woodland strawberry (*Fragaria vesca*).** *Nat Genet* 2011, **43**(2):109–116.
63. Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, Salzberg SL, Feng L, Jones MR, Skelton RL, Murray JE, Chen C, Qian W, Shen J, Du P, Eustice M, Tong E, Tang H, Lyons E, Paull RE, Michael TP, Wall K, Rice DW, Albert H, Wang ML, Zhu YJ: **The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus).** *Nature* 2008, **452**(7190):991–996.
64. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M: **The Pfam protein families database.** *Nucleic Acids Res* 2014, **42**(Database issue):D222–D230.
65. Sanseverino W, Roma G, De Simone M, Faino L, Melito S, Stupka E, Frusciante L, Ercolano MR: **PRGdb: a bioinformatics platform for plant resistance gene analysis.** *Nucleic Acids Res* 2010, **38**(Database issue):D814–D821.
66. Salse J: **In silico archeogenomics unveils modern plant genome organisation, regulation and evolution.** *Curr Opin Plant Biol* 2012, **15**(2):122–130.
67. Murat F, Xu JH, Tannier E, Abrouk M, Guilhot N, Pont C, Messing J, Salse J: **Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution.** *Genome Res* 2010, **20**(11):1545–1557.
68. Abrouk M, Zhang R, Murat F, Li A, Pont C, Mao L, Salse J: **Grass microRNA gene paleohistory unveils new insights into gene dosage balance in subgenome partitioning after whole-genome duplication.** *Plant Cell* 2012, **24**(5):1776–1792.
69. Schnable JC, Freeling M, Lyons E: **Genome-wide analysis of syntenic gene deletion in the grasses.** *Genome Biol Evol* 2012, **4**(3):265–277.
70. Thomas BC, Pedersen B, Freeling M: **Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes.** *Genome Res* 2006, **16**(7):934–946.
71. Sankoff D, Zheng C, Zhu Q: **The collapse of gene complement following whole genome duplication.** *BMC Genomics* 2010, **11**:313.
72. Xiong Y, Liu T, Tian C, Sun S, Li J, Chen M: **Transcription factors in rice: a genome-wide comparative analysis between monocots and eudicots.** *Plant Mol Biol* 2005, **59**(1):191–203.
73. Freeling M: **Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition.** *Annu Rev Plant Biol* 2009, **60**:433–453.
74. Richly E, Kurth J, Leister D: **Mode of amplification and reorganization of resistance genes during recent Arabidopsis thaliana evolution.** *Mol Biol Evol* 2002, **19**(1):76–84.
75. Xu Z, Wang H: **LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons.** *Nucleic Acids Res* 2007, **35**(Web Server issue):W265–W268.
76. Wicker T, Buchmann JP, Keller B: **Patching gaps in plant genomes results in gene movement and erosion of colinearity.** *Genome Res* 2010, **20**(9):1229–1237.
77. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP: **MicroRNAs in plants.** *Genes Dev* 2002, **16**(13):1616–1626.
78. Lim LP, Glasner ME, Yekta S, Burge CB, Bartel DP: **Vertebrate microRNA genes.** *Science* 2003, **299**(5612):1540.
79. Bagga S, Pasquinelli AE: **Identification and analysis of microRNAs.** *Genet Eng* 2006, **27**:1–20.
80. Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, Estelle M, Voinnet O, Jones JD: **A plant miRNA contributes to antibacterial resistance by repressing auxin signaling.** *Science* 2006, **312**(5772):436–439.
81. Li Y, Zhang Q, Zhang J, Wu L, Qi Y, Zhou JM: **Identification of microRNAs involved in pathogen-associated molecular pattern-triggered plant innate immunity.** *Plant Physiol* 2010, **152**(4):2222–2231.
82. Du P, Wu J, Zhang J, Zhao S, Zheng H, Gao G, Wei L, Li Y: **Viral infection induces expression of novel phased microRNAs from conserved cellular microRNA precursors.** *PLoS Pathog* 2011, **7**(8):e1002176.
83. Katiyar-Agarwal S, Morgan R, Dahlbeck D, Borsani O, Villegas A Jr, Zhu JK, Staskawicz BJ, Jin H: **A pathogen-inducible endogenous siRNA in plant immunity.** *Proc Natl Acad Sci U S A* 2006, **103**(47):18002–18007.
84. Allen E, Xie Z, Gustafson AM, Carrington JC: **microRNA-directed phasing during trans-acting siRNA biogenesis in plants.** *Cell* 2005, **121**(2):207–221.
85. McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, Hyten DL, Gerhardt DJ, Jeddeloh JA, Stupar RM: **Structural variants in the soybean genome localize to clusters of biotic stress-response genes.** *Plant Physiol* 2012, **159**(4):1295–1308.
86. Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdil P, Joudrier P, Gautier MF, Cattolico L, Beckert M, Aubourg S, Weissenbach J, Caboche M, Bernard M, Leroy P, Chalhou B: **Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*).** *Plant Cell* 2005, **17**(4):1033–1045.
87. Sun X, Wang GL: **Genome-wide identification, characterization and phylogenetic analysis of the rice LRR-kinases.** *PLoS One* 2011, **6**(3):e16079.
88. Zhai J, Jeong DH, De Paoli E, Park S, Rosen BD, Li Y, González AJ, Yan Z, Kitto SL, Grusak MA, Jackson SA, Stacey G, Cook DR, Green PJ, Sherrier DJ, Meyers BC: **MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs.** *Genes Dev* 2011, **25**(23):2540–2553.
89. Li F, Pignatta D, Bendix C, Brunkard JO, Cohn MM, Tung J, Sun H, Kumar P, Baker B: **MicroRNA regulation of plant innate immune receptors.** *Proc Natl Acad Sci U S A* 2012, **109**(5):1790–1795.
90. Eckardt NA: **A microRNA cascade in plant defense.** *Plant Cell* 2012, **24**(3):840.
91. Finn RD, Clements J, Eddy SR: **HMMER web server: interactive sequence similarity searching.** *Nucleic Acids Res* 2011, **39**(Web Server issue):W29–W37.
92. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.** *Genome Res* 2009, **19**(9):1639–1645.
93. Mathelier A, Carbone A: **MiRENA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data.** *Bioinformatics* 2010, **26**(18):2226–2234.

doi:10.1186/1471-2164-15-187

Cite this article as: Zhang et al.: Paleo-evolutionary plasticity of plant disease resistance genes. *BMC Genomics* 2014 **15**:187.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

