AJ ACADEMIC JOURNALS
expand your knowledge

African Journal of Agricultural Research

Full Length Research Paper

A clue for generating a new leucine-rich repeat gene in maize

Zheng Yang[1] and Kui Xiang[2*]

[1]College of Agronomy, Sichuan Agricultural University, Chengdu, Sichuan, 611130, China.
[2]Maize research institute, Sichuan Agricultural University, Chengdu, Sichuan, 611130, China.

Plant leucine-rich repeat (LRR) proteins play an important role in cell adhesion and signaling, neuronal development, disease resistance response and pathogen recognition. Therefore, the origin and evolution of the LRR genes had been studied deeply. However, there were no evidence of generating new LRR genes. In this study, the genomic and amino acid sequences of the LRR genes or proteins were aligned based on Maize GDB and NCBI databases. The result showed that the part sequences of GRMZM5G851515 and part sequences of GRMZM2G167560 were consistent with parts of GRMZM2G343449. It was indicated that GRMZM2G343449 generated from GRMZM5G851515, GRMZM2G167560 and other genes. Evolution analysis also supported that GRMZM2G343449 and homologous genes were newborn. Meanwhile, they existed in only angiosperm that is closely related to human life, suggesting that they might be retained through the artificial selection. GRMZM2G343449 and its homologous genes in Poales were an independent branch of evolution which might be related with environment adaptability because there were more diseases and insect pests in the growth and development of crops than those of other species. The results indicated that GRMZM2G343449 and its homologous genes generated from stress resistance. This study provided key information for finding new generated LRR, which might be a clue for searching newborn genes from maize and other species in the big data era currently.

Key words: Leucine-rich repeat, evolution, generation, maize

INTRODUCTION

Protein containing leucine-rich repeat (LRR) domain (PLRR) existed in all life forms, from viruses to eukaryotes (Bella et al., 2008). Cell adhesion and signaling (Hohenester et al., 2006; Kresse and Schonherr, 2001), neuronal development (Chen et al., 2006), disease resistance response and pathogen recognition in plants (Gay and Gangloff, 2007; Martinon and Tschopp, 2005; West et al., 2006) are very important for the plant life. Currently, three typical PLRRs are plant receptors, FLS2 (Gomez-Gomez and Boller, 2000), EFR (Zipfel et al.,

2006) and Xa21 (Song et al., 1995) which showed that PLRR is important for plant life, especially for disease resistance.

PLRR had different types based on the other different conversed domains (Bella et al., 2008). PLRRs contained tandems of two or more LRRs forming the continuously expanding LRR superfamily (Buchanan and Gay, 1996). For example, there were six PLRR types, such as LRR-Transmembrane (T)-Protein kinase (PK) (LRR-T-PK), LRR-PK, NB-LRR, LRR-T, LRR, and ATPases (A)-LRR (A-LRR) in maize (Li et al., 2016). More than 200 PLRR were distributed on the whole chromosomes of maize (Li et al., 2016). LRR domain was an essential part of PLRR. It was one of the six prolific types of protein which was a widespread structural motif (Andrade et al., 2001). The solvent-exposed amino acid residues of LRR in some PLRRs were involved in recognizing pathogens (Bergelson et al., 2001). Therefore, the study of evolutionary dynamics of PLRR is important for plant disease resistance.

In the evolution of LRR, The LRR existed in Eubacteria, Archaebacteria, Protista, Fungi, Plantae and Animalia (Yue et al., 2012). The number of LRR in Eubacteria, Archaebacteria and Fungi was far less than that in Plantae and Aminalia (Yue et al., 2012). It indicated that the evolution tendency of the LRR was similar to the tendency of evolving organisms and the evolution analysis of the LRR was important for the whole kingdom.

Previous studies showed that LRR and NBS domains existed before the split of prokaryotes and eukaryotes and the fusion of LRR and NBS domain was observed only in land plant lineages (Yue et al., 2012). However, there were no evidences to explain how to generate a new PLRR currently. In this study, we used three maize proteins for explaining the generation of a new PLRR gene, although the evidences were not integrated. Meanwhile, we analyzed the elevation of the new generated PLRR. This study aims to provide a clue for other research on generating new genes in the era of big data currently.

METHODOLOGY

Obtaining data

The genomic sequences and amino acid sequences were downloaded from the MaizeGDB (http://www.maizegdb.org/) (Lawrence et al., 2004) and NCBI database (http://www.ncbi.nlm.nih.gov/guide/) (Gene-duplication events of LRR genes (Li et al., 2016)).

Data analysis

SMART protein motif analysis (http://smart.embl-heidelberg.de/) (Letunic et al., 2012; Schultz et al., 1998) was used. The sequence alignment was carried out using DNAman Version 6.0. (Higgins and Sharp, 1989).

Data sampling

For each species, protein entries matching the LRR domain of GRMZM2G343449 in the NCBI database (http://www.ncbi.nlm.nih.gov/guide/) were identified as LRR-encoding protein using the blastp search with an *E*-value cut-off of 10^{-4} (Eddy, 1998; Yue et al., 2012). For all LRR-encoding proteins in our data set, the amino acid sequences were aligned and constructed as a phylogenetic trees using MEGA6 with the bootstrap of 1000 replicates (Tamura et al., 2013).

RESULTS

New born for one PLRR based on the DNA and protein sequences

On the basis of gene-duplication events of LRR genes (Li et al., 2016), we only found that GRMZM2G343449 on chromosome 4 and GRMZM5G851515 on chromosome 2, their amino acid sequences of N-terminals were identical among the duplication LRR genes. The identical region consisted of 295 amino acids (Figure 1). The percentages of identical continuous amino acids were 89.9 and 74.1% in GRMZM2G343449 and GRMZM5G851515, respectively. GRMZM2G343449 possessing 328 amino acids was on bin4.06, whereas GRMZM5G851515 possessing 398 amino acids was on bin 2.04 (Table 1). Both of them had four introns and belonged to the LRR-type (Figure 2A and 2B). Identical amino acids were not equal to identical DNA sequence. Therefore, the nucleic acid sequences of GRMZM2G343449 and GRMZM5G851515 were subsequently aligned. The result showed that there were 1298bp identical nucleic acids between GRMZM2G343449 and GRMZM5G851515 at the 3'-terminal except for six SNPs. Meanwhile, most different regions of the two proteins were in the first exon (Figure 3), suggesting that the first exon of GRMZM2G343449 or GRMZM5G851515 might derived from another encoded protein or or other unknown fragment on Chromosomes.

Looking for origins of generating one PLRR

To detect which gene was the origin, we blasted the DNA sequences of the two genes in NBCI and MaizeGDB databases, respectively. The references were the default values. On the basis of the about 13Gb genome from MaizeGDB that consist of four types of bases (A, T, C and G), more than 17 continuous DNAs might be a new gene using the power method of math. To a certain extent, continuous 170bp DNAs stand for a new gene in maize. There were no genes that had more than 170 bp identical DNAs to GRMZM5G851515. However, GRMZM2G167560 (NCBI accession number: EU972243) had 200 bp identical DNAs to 851th bp-1050th bp of GRMZM2G343449 (Figure 4), suggesting that GRMZM2G343449 might consist of GRMZM2G167560,

Figure 1. The amino acid sequence alignment of GRMZM2G343449 and GRMZM5G851515. The amino acids in Black background mean same amino acids between the two proteins.

Table 1. The details of GRMZM2G343449_T01 and GRMZM5G851515_T01 on chromosomes.

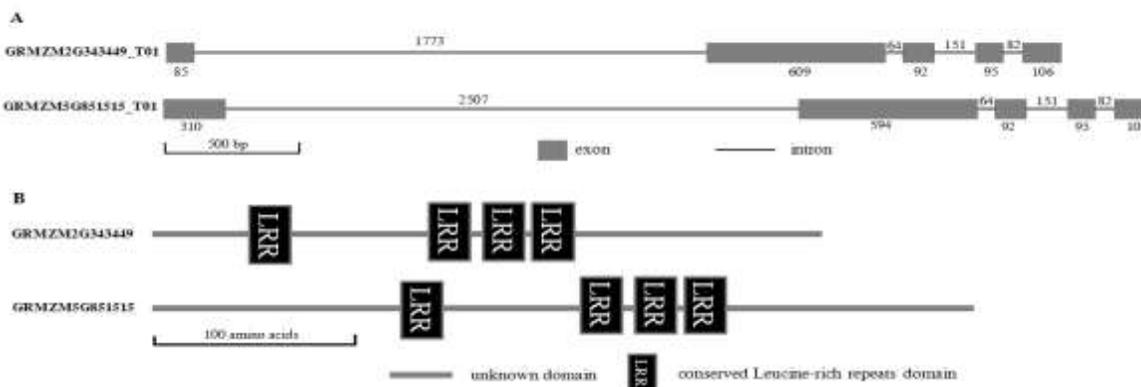| Maize DGB ID (Ensembl)	Chromosome	Bin	Number of amino acid
GRMZM2G343449_T01	Chr4:162859562..162862618	4.06	328
GRMZM5G851515_T01	Chr2:65851212..65855212	2.04	398

Figure 2. The structures of genes and proteins of GRMZM2G343449 and GRMZM5G851515. A, The structure of genes GRMZM2G343449 and GRMZM5G851515. B, The structure of proteins GRMZM2G343449 and GRMZM5G851515.

parts of GRMZM5G851515 and other DNA sequences with unknown origins (Figure 5). It is indicated that GRMZM2G167560 and GRMZM5G851515 existed before generating GRMZM2G343449. It was a clue for analyzing how new LRR or other genes were created in the period of big data.

The evolution analysis of the new generated PLRR

On the basis of an *E*-value cut-off of 10^{-4}, 23 homologous genes of GRMZM2G343449 were from 22 species in angiosperm except for GRMZM2G343449 in maize, whereas were not from gymnosperm, bryophyte, fungi and animals (Figure 6). It indicated the GRMZM2G343449 and its homologous were new-born and angiosperm-specific. The 23 species were closely related to human life. Among these species, most of them were edible. Though monocotyledon was not separated from dicotyledon completely based on clustering result, the species in Poales were clustered into together, which were distinguish from other orders (Figure 6). It indicated that the GRMZM2G343449 and its homologous genes in Poales were relatively independent in the process of

Figure 3. The genomic alignment of GRMZM2G343449 and GRMZM5G851515. Black bases mean same bases between the two genes.

```
Genomic_GRMZM2G343449        TGTCTGAATCCAAGCTCGCGATGAAAACAGCAGCCAGAAAGAGCGCTCACCGACCGTGTCGCGGCGGCGCTTCTGCACAGAGCTCGGATCCAACCAACCC    900
Genomic_GRMZM2G167560-1-200bp ...............................................CGACCGTGTCGCGGCGGCGCTTCTGCACAGAGCTCGGATCCAACCAACCC     50
Consensus                                                                   cgaccgtgtcgcggcggcgcttctgcacagagctcggatccaaccaaccc

Genomic_GRMZM2G343449        GCACGGGCGCGTCCGAGGCGCGGAACCACCGTACCCAGCGGCGGAGCAGCGGGGAGAGAGATCGAGAAGCTTTCGCTGTTGCCGCGCCGGTAAAGAAAAC   1000
Genomic_GRMZM2G167560-1-200bp GCACGGGCGCGTCCGAGGCGCGGAACCACCGTACCCAGCGGCGGAGCAGCGGGGAGAGAGATCGAGAAGCTTTCGCTGTTGCCGCGCCGGTAAAGAAAAC    150
Consensus                    gcacgggcgcgtccgaggcgcggaaccaccgtacccagcggcggagcagcggggagagagatcgagaagctttcgctgttgccgcgccggtaaagaaaac

Genomic_GRMZM2G343449        AAGCTCCTTCTGCGAGAGGACTAGGAGGAGGTTTTGCCAACACGTCGAGCAGAGCAGGGGTCGGCCGGGTCCGGGGGATAAGCGAGGCGAGGGGGACGGG   1100
Genomic_GRMZM2G167560-1-200bp AAGCTCCTTCTGCGAGAGGACTAGGAGGAGGTTTTGCCAACACGTCGAGC.................................................    200
Consensus                    aagctccttctgcgagaggactaggaggaggttttgccaacacgtcgagc
```

Figure 4. The genomic alignment of GRMZM2G343449 and GRMZM2G167560 at the 5' terminal. Black bases mean same bases between the two genes.
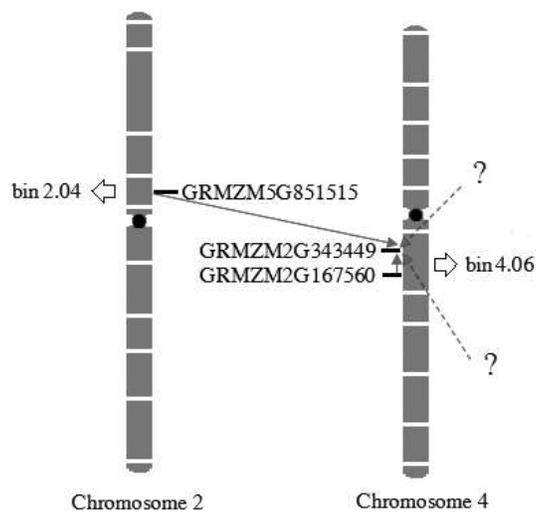

Figure 5. The mimetic diagram of generating GRMZM2G343449. ? Means unknown genes or fragments on chromosomes.

evolution.

DISCUSSION

LRR domain exists before the split of prokaryotes and eukaryotes (Yue et al., 2012). However, the LRR protein GRMZM2G343449 might be born in modern times which might be born after GRMZM2G167560, part of GRMZM5G851515. It might result from the help of transposons because transposons usually contributed into the generation of new genes. There was a large number of transposons in maize, such as Helitron transposons (Li and Dooner, 2009), Ac/Ds (Lazarow et al., 2013) and Mutator (Walbot et al., 1988). They might result into the generation of GRMZM2G343449.

In plant, LRR domains of several R-proteins is the major determinants of recognizing the specificity of Avr factors (Jones and Jones, 1997; Ellis et al., 2000; Leister and Katagiri, 2000). Amino acids in the LRR might also influence the interaction with host factors (Banerjee et al., 2001). Adaptive divergence among LRR proteins had been investigated in tomato (Parniske et al., 1997), rice (Wang et al., 1998) and Arabidopsis (Botella et al., 1998; McDowell et al., 1998; Noel et al., 1999). The LRR region often evolved at fast rates unusually (Bergelson et al., 2001). However, there were no direct evidences of evolution for LRR protein. In this study, the generation of GRMZM2G343449 provided the clue of evolution for LRR protein. It is a clue on how to born new LRR or other genes based on current big data.

Adaptive divergence and allelic polymorphism were the two types of evolutionary dynamics (Bergelson et al., 2001). The adaptive variants coexist with other alleles (Bergelson et al., 2001). R-proteins possessing LRR domains were usually associated with recognize pathogens (Bergelson et al., 2001). The generation of GRMZM2G343449 might result from the interaction of
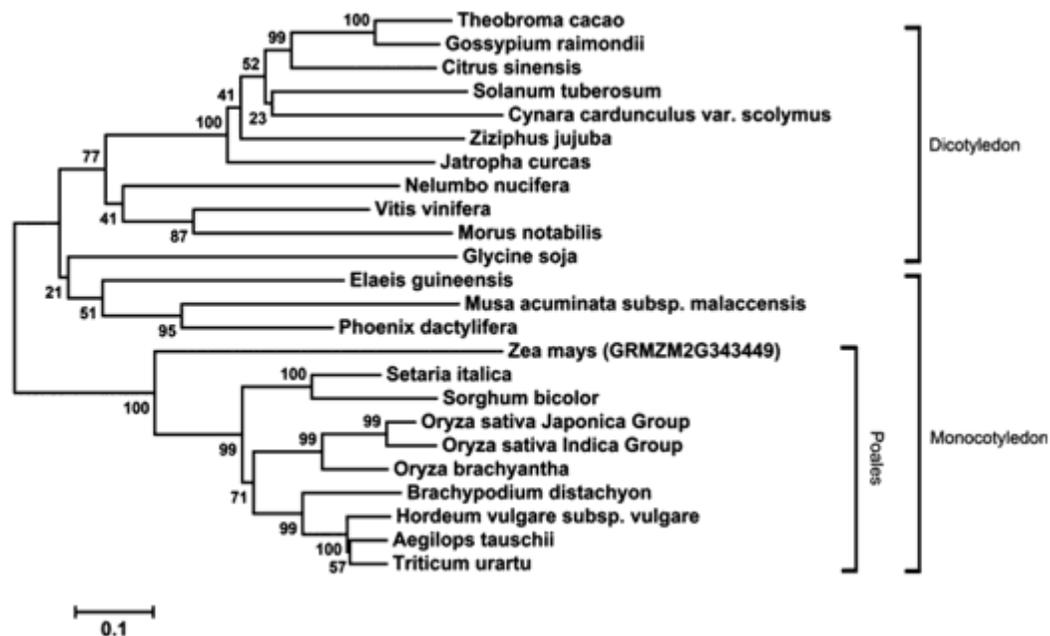
Figure 6. The phylogenetic tree of GRMZM2G343449 and its homologous genes. The NCBI ID of homologous gene from *Setaria italic* is XP_004982915.1, from *Sorghum bicolor* is KXG38364.1, from *Oryza brachyantha* is XP_006662430.2, from *Oryza sativa Japonica* Group is XP_015612820.1, from *Oryza sativa Indica* Group is EAY78879.1, from *Aegilops tauschii* is EMT14287.1, from *Hordeum vulgare* subsp. *Vulgare* is BAJ94896.1, from *Triticum urartu* is EMS46943.1, from *Brachypodium distachyon* is XP_003574034.1, from *Aegilops tauschii* is EMT33856.1, from *Musa acuminata* subsp. *malaccensis* is XP_009390803.1, from *Phoenix dactylifera* is XP_008801287.1, from *Theobroma cacao* is XP_007029633.1, from *Ziziphus jujube* is XP_015895303.1, from *Nelumbo nucifera* is XP_010249615.1, from *Solanum tuberosum* is XP_015163330.1, from *Setaria italic* is XP_004961982.1, from *Nelumbo nucifera* is XP_010241269.1, from *Phoenix dactylifera* is XP_008801288.1, from *Vitis vinifera* is XP_010646797.1, from *Morus notabilis* is XP_010106115.1, from *Glycine soja* is KHN31501.1, from *Jatropha curcas* is KDP35941.1, from *Elaeis guineensis* is XP_010910731.1, from *Gossypium raimondii* is XP_012470043.1, from *Cynara cardunculus* var. *scolymus* is KVI00715.1, from *Elaeis guineensis* is XP_010908122.1, and from *Citrus sinensis* is XP_015387143.1.

maize and pathogen. It indicated that GRMZM2G343449 might be associated with pathogen resistance of maize.

GRMZM2G343449 and its homologous genes existed in only 23 species of angiosperm, which does not exis extensively in angiosperm. These species were closely related to human life. It suggests that not only GRMZM2G343449 and its homologous genes might be new-born, but also they might be retained through the artificial selection. The homologous genes of GRMZM2G343449 in Poales were independent branch of evolution, it might be related with environment adaptability because there are more diseases and insect pests in the production of crops than in other species. It further indicated that GRMZM2G343449 and its homologous genes might be associated with stress resistance.

Conclusion

The current study revealed that GRMZM2G343449 might

be generated from GRMZM5G851515, GRMZM2G167560 and other genes. Evolution analysis also supported that GRMZM2G343449 as an homologous genes might be newborn and retained through the artificial selection. GRMZM2G343449 and its homologous genes in Poales being an independent branch of evolution indicated that GRMZM2G343449 and its homologous genes generated from stress resistance. This study provided a clue for searching newborn genes from maize and other species

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGEMENTS

REFERENCES

Andrade A, Perez-Iratxeta C, Ponting P (2001). Protein repeats: Structures, functions, and evolution. Journal of Structural Biology 134:117-131.

Bella J, Hindle L, McEwan A, Lovell C (2008). The leucine-rich repeat structure. Cellular and Molecular Life Sciences 65:2307-2333.

Bergelson J, Kreitman M, Stahl E, Tian D (2001). Evolutionary dynamics of plant R-genes. Science 292:2281-2285.

Botella MA, Parker JE, Frost LN, Bittner-Eddy PD, Beynon JL, Daniels MJ, Holub EB, Jones JD (1998). Three genes of the Arabidopsis RPP1 complex resistance locus recognize distinct Peronospora parasitica avirulence determinants. Plant Cell 10:1847-1860.

Buchanan SG, Gay NJ (1996). Structural and functional diversity in the leucine-rich repeat family of proteins. Progress in Biophysics & Molecular Biology 65: 1-44.

Chen Y, Aulia S, Li L, Tang BL (2006). AMIGO and friends: an emerging family of brain-enriched, neuronal growth modulating, type I transmembrane proteins with leucine-rich repeats (LRR) and cell adhesion molecule motifs. Brain Research Reviews 51: 265-274.

Eddy SR (1998). Profile hidden Markov models. Bioinformatics 14: 755-763.

Ellis J, Dodds P, Pryor T (2000). Structure, function and evolution of plant disease resistance genes. Current Opinion in Plant Biology 3:278-284.

Gay NJ, Gangloff M (2007). Structure and function of Toll receptors and their ligands. Annual Review of Biochemistry 76:141-165.

Gomez-Gomez L, Boller T (2000). FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Molecular Cell 5:1003-1011.

Higgins DG, Sharp PM (1989). Fast and Sensitive Multiple Sequence Alignments on a Microcomputer. Computer Applications in the Biosciences 5:151.

Hohenester E, Hussain S, Howitt JA (2006). Interaction of the guidance molecule Slit with cellular receptors. Biochemical Society Transactions 34:418-421.

Jones DA, Jones JDG (1997). The Role of Leucine-Rich Repeat Proteins in Plant Defences. Advances in Botanical Research 24(8):89-167.

Kresse H, Schonherr E (2001). Proteoglycans of the extracellular matrix and growth control. Journal of Cellular Physiology 189: 266-274.

Lawrence CJ, Dong Q, Polacco ML, Seigfried TE, Brendel V (2004). MaizeGDB, the community database for maize genetics and genomics. Nucleic Acids Research 32:D393-397.

Lazarow K, Doll ML, Kunze R (2013). Molecular biology of maize Ac/Ds Elements: An Overview. Methods in Molecular Biology 1057:59-82.

Leister RT, Katagiri F (2000). A resistance gene product of the nucleotide binding site -leucine rich repeats class can form a complex with bacterial avirulence proteins in vivo. Plant Journal 22: 345-354.

Letunic I, Doerks T, Bork P (2012). SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Research 40:D302-D305.

Li WT, Xiang K, Zhang ZM, Yuan GS, Lin HJ, Pan GT (2016). Systematic analysis of leucine-rich repeat disease resistance genes in maize. Maydica 61(1):10.

Li Y, Dooner HK (2009). Excision of Helitron transposons in maize. Genetics 182(1):399-402.

Martinon F, Tschopp J (2005). NLRs join TLRs as innate sensors of pathogens. Trends in Immunology 26:447-454.

McDowell JM, Dhandaydham M, Long TA, Aarts MG, Goff S, Holub EB, Dangl JL (1998). Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of Arabidopsis. Plant Cell 10:1861-1874.

Noel L, Moores TL, van Der Biezen EA, Parniske M, Daniels MJ, Parker JE, Jones JD (1999). Pronounced intraspecific haplotype divergence at the RPP5 complex disease resistance locus of Arabidopsis. Plant Cell 11:2099-2112.

Parniske M, Hammond-Kosack KE, Golstein C, Thomas CM, Jones DA, Harrison K, Wulff BB, Jones JD (1997). Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato. Cell 91:821-832.

Schultz J, Milpetz F, Bork P, Ponting CP (1998). SMART, a simple modular architecture research tool: identification of signaling domains. Proceedings of the National Academy of Sciences of the United States of America 95:5857-5864.

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. Science 270:1804-1806.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution 30:2725-2729.

Walbot V, Britt AB, Luehrsen K, McLaughlin M, Warren C (1988). Regulation of mutator activities in maize. Basic Life Sciences 47:121-135.

Wang GL, Ruan DL, Song WY, Sideris S, Chen L, Pi LY, Zhang S, Zhang Z, Fauquet C, Gaut BS, Whalen MC, Ronald PC (1998). Xa21D encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. Plant Cell 10:765-779.

West AP, Koblansky AA, Ghosh S (2006). Recognition and signaling by toll-like receptors. Annual Review of Cell and Developmental Biology 22:409-437.

Yue JX, Meyers BC, Chen JQ, Tian D, Yang S (2012). Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. New Phytologist 193:1049-1063.

Zipfel C, Kunze G, Chinchilla D, Caniard A, Jones JD, Boller T, Felix G (2006). Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. Cell 125:749-760.