

Bioinformatic analysis of the nucleotide binding site-encoding disease-resistance genes in foxtail millet (*Setaria italica* (L.) Beauv.)

Y.B. Zhu^{1,2*}, X.Q. Xie^{3*}, Z.Y. Li², H. Bai², L. Dong², Z.P. Dong² and J.G. Dong¹

¹Molecular Plant Pathology Laboratory, Agricultural University of Hebei, Baoding, China

²Millet Institute, Hebei Academy of Agricultural and Forestry Sciences, Shijiazhuang, China

³School of Life Science, Peking University, Beijing, China

*These authors contributed equally to this study.

Corresponding authors: Z.P. Dong / J.G. Dong

E-mail: zhibaoshi001@163.com / shmdjg@hebau.edu.cn

Genet. Mol. Res. 13 (3): 6602-6609 (2014)

Received January 30, 2013

Accepted August 1, 2013

Published August 28, 2014

DOI <http://dx.doi.org/10.4238/2014.August.28.5>

ABSTRACT. The nucleotide-binding site (NBS) disease-resistance genes are the largest category of plant disease-resistance gene analogs. The complete set of disease-resistant candidate genes, which encode the NBS sequence, was filtered in the genomes of two varieties of foxtail millet (Yugu1 and 'Zhang gu'). This study investigated a number of characteristics of the putative NBS genes, such as structural diversity and phylogenetic relationships. A total of 269 and 281 NBS-coding sequences were identified in Yugu1 and 'Zhang gu', respectively. When the two databases were compared, 72 genes were found to be identical and 164 genes showed more than 90% similarity. Physical positioning and gene family analysis of the NBS disease-resistance genes in the genome revealed that the number of genes on each chromosome was similar in both varieties. The eighth chromosome contained the largest

number of genes and the ninth chromosome contained the lowest number of genes. Exactly 34 gene clusters containing the 161 genes were found in the Yugu1 genome, with each cluster containing 4.7 genes on average. In comparison, the 'Zhang gu' genome possessed 28 gene clusters, which had 151 genes, with an average of 5.4 genes in each cluster. The largest gene cluster, located on the eighth chromosome, contained 12 genes in the Yugu1 database, whereas it contained 16 genes in the 'Zhang gu' database. The classification results showed that the CC-NBS-LRR gene made up the largest part of each chromosome in the two databases. Two TIR-NBS genes were also found in the Yugu1 genome.

Key words: Foxtail millet; Gene clusters; Phylogenetic tree; Database; Nucleotide-binding site

INTRODUCTION

A resistance gene (R), known as an encoding receptor gene, is a type of gene whose product can directly or indirectly identify the virulence gene product of pathogens and then express a defense response. Previous studies have shown that a resistance gene-encoded product was the signal that switched on the plant disease-resistance response, and thus played a crucial role in the plant disease-resistance process (Somerville and Somerville, 1999; Kohler et al., 2008). Therefore, cloning and studying disease-resistance genes can help researchers understand the plant disease-resistance mechanism. According to an analysis of the protein structure encoded by the currently cloned disease-resistance genes, the nucleotide-binding site (NBS) disease-resistance genes represent the largest category of plant disease-resistance genes. NBS domains have been identified in many prokaryotic and eukaryotic proteins, and their ability to bind ATP or GTP is essential for their biological activities, such as growth and differentiation of cells and resistance to disease (Pan et al., 2000; Shirano et al., 2002). The NBS consists of four conservative regional structures: the P-loop region, the kinase 2a region, the GLPL region, and the MHDV region. The NBS-leucine-rich repeat (LRR) class of R proteins, which contain an LRR domain and an NBS, can be further subdivided into two classes based on the secondary structure of the N terminus. The first class is CC-NBS-LRR (CNL), which contains CN, a putative coiled-coil region, at the N terminus. The second class is TIR-NBS-LRR (TNL), which contains an N-terminal TIR region that shares homology with the Toll protein of *Drosophila* and the interleukin-1 receptor (IL-1R) of mammals.

Foxtail millet (*Setaria italica* (L.) Beauv.) is one of China's traditional crops. It has many advantageous characteristics, such as drought resistance, forage use, and a high nutrient content (Jusuf and Pernes, 1985; Zohary and Hopf, 2000; Bettinger et al., 2010). However, diseases remain a serious threat to the quality and yield production of foxtail millet, especially foxtail millet rust and foxtail millet blast, which can lead to severe plant lodging and crop failure. Therefore, attempts to clarify the interaction between pathogens and their host plants has become an important research subject in order to reveal the plant disease-resistance mechanism to enable breeding-resistant lines (Baker et al., 1997; Hammond-Kosack and Jones, 1997). This study investigated the NBS-type resistance genes at the genome-wide scale using

bioinformatic methods that utilized recently published foxtail millet genomic data (Jeffrey et al., 2012; Zhang et al., 2012). Several characteristics, such as classification, chromosomal localization, and phylogenetic evolution, were also analyzed in order to provide a theoretical basis for foxtail millet-resistance gene cloning and to increase understanding of the resistance mechanism to aid in the cultivation of resistant varieties.

MATERIAL AND METHODS

Plant materials and genomic DNA information

Whole genome data for foxtail millet Yugu1 and 'Zhang gu' were selected for the study, and complete genetic information was obtained from the following website: ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v8.0/Sitalica/ and <http://gigadb.org/foxtail-millet/>.

NBS disease-resistance gene databases

The whole genome sequences for foxtail millet varieties Yugu1 and 'Zhang gu' were downloaded from the website and then the BioEdit software (Hall, 1999) was used to construct the local database. The homologous sequences with P values above 10^{-6} were screened based on the conserved domain of the NBS sequences. Then, all of the screened sequences were determined using protein-based sequencing and regional and family classification information from the Pfam database (<http://pfam.sanger.ac.uk/search>), CILS (http://ch.embnet.org/software/COILS_form.html). Finally, the ClustalX software (Thompson et al., 1997) was used to create multiple sequences for the selected nucleotide sequences and to remove the repetitive sequences in the candidate genes.

NBS-type gene classification

The NBS region ran from the pre-P-loop structure to the MHDV structure. However, before the pre-P-loop structure and after the MHDV structure, there were some other specific structures, which were classified according to their differences at the N-terminal and C-terminal.

Physical location of the NBS-type genes

By using BLAST to compare the candidate genes with the whole genome alignment published on various websites, the specific location of each NBS class gene found on a chromosome was confirmed, and then the project mapping software, AutoCAD 2010, was used to locate all of the candidate genes in the chromosomes.

Construction of NBS-type gene phylogenetic trees

ClustalX was used to perform multiple sequence alignments for the standard NBS-type disease-resistance genes for the whole genomes of the foxtail millet varieties Yugu1 and 'Zhang gu'. MEGA5.1 (Tamura et al., 2011) was then used to construct the neighbor-joining phylogenetic trees in order to analyze evolutionary relationships between NBS class disease-resistance genes in the two varieties of foxtail millet.

RESULTS

Construction of a candidate gene database for the NBS class

The candidate genes without the NBS structure region were removed from the whole genome sequences of Yugu1 by BLAST, and 289 candidate gene sequences were obtained. Twenty redundant sequences were further removed by Pfam analysis and multiple sequencing, revealing a total of 269 genes containing the NBS structure. The ‘Zhang gu’ database contained 281 NBS genes. A total of 72 identical genes and 164 genes with more than 90% similarity were identified between the two NBS databases. Yugu1_025869 and Yugu1_025868 from the Yugu1 database were classified as overlapping sequences, and they showed 90% similarity to Zhanggu_10009415. Yugu1_025910 and Zhanggu_10035205 had identical genes. Zhanggu_10020497 was present only in the ‘Zhang gu’ database (Figure 1).

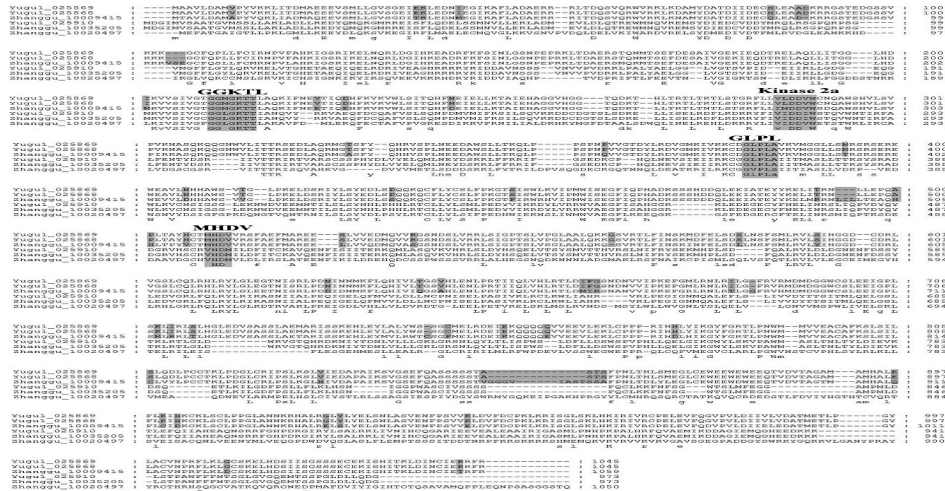


Figure 1. Part of the match results for the two databases (Yugu1 database and ‘Zhang gu’ database).

Classification of NBS-type genes

Among the 269 NBS-type-resistance genes, 249 gene motifs from the NBS region were highly conserved, with a complete open reading frame (ORF) called a standard NBS-type gene (Wang et al., 2009b). The other genes were non-standard NBS-type genes (non-regular NBS genes). The standard NBS genes are divided into six major categories (Li et al., 2010): CC-NBS, CC-NBS-LRR, NBS, NBS-LRR, and TIR-NBS, and others (Figure 2). The CC-NBS-LRR type was found to be the most common type in both databases. The ‘Zhang gu’ database contained NBS genes of all categories except the TIR-NBS type.

Physical map of the NBS class

The chromosome distributions (Figure 3) of the identified NBS genes in the Yugu1 database showed that chromosome 8 contained the largest number of NBS genes, followed

by chromosome 2 and chromosome 3. With respect to NBS-type gene classes, chromosome 4 contained three types: NL, CNL, and others, which were the least frequent. The CC-NBS-LRR class on each chromosome was more common than the other classes. When localizing the genes in the ‘Zhang gu’ database on each chromosome, five genes were found with no position on the chromosomes. When comparing these chromosomes in the two databases, most genes were in the same position and the positions of a few genes were end to end (Figure 4). The differences may have been caused by several reasons. For example, the genes might have come from different varieties or errors might have occurred during the sequencing assembly process. Additionally, it was also interesting that chromosome 9 was the longest chromosome in both the ‘Zhang gu’ and Yugu1 databases, but it contained the lowest number of NBS-coding sequences.

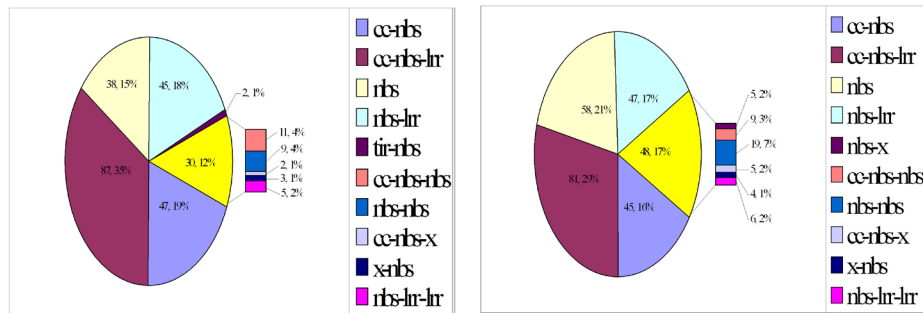


Figure 2. Classification of NBS-resistance genes in foxtail millet (left, Yugu1; right, ‘Zhang gu’).

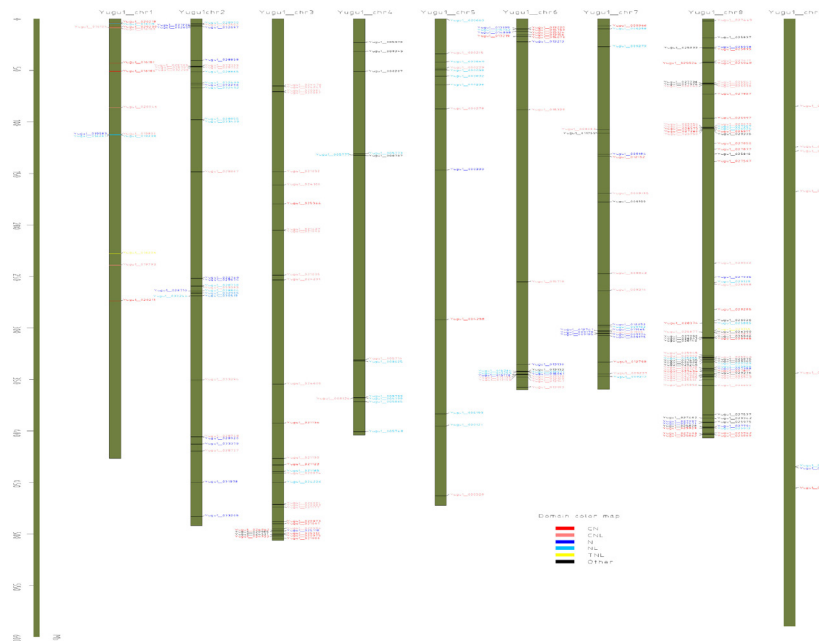


Figure 3. Chromosomal distribution of the NBS database for Yugu1.

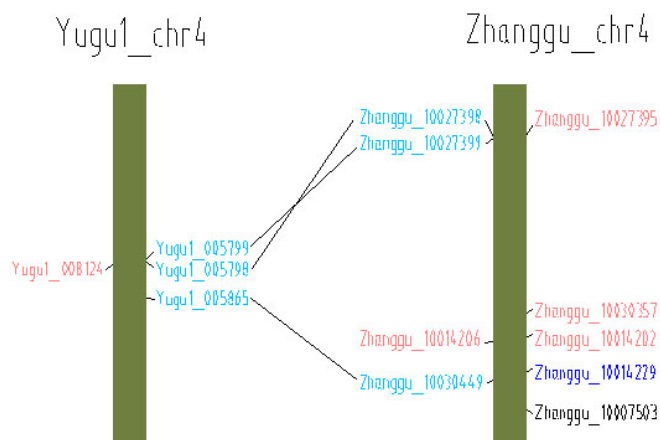


Figure 4. Comparison between parts of chromosome 4 found in the two databases ((left, Yugu1; right, ‘Zhang gu’).

Based on the definition of the gene cluster in Holub (2001), the gene clusters contained more than three gene flocks in 200-kb nucleotide units. In the 249 NBS-type genes, 34 gene clusters contained a total number of 161 genes, with each cluster containing an average of 4.7 genes. The largest gene cluster contained 12 genes located on chromosome 8. Gene clusters are formed by gene duplication. Most of them belonged to the same gene family and showed a high homology, which indicated that replication on chromosome 8 may have occurred at a large-scale, resulting in a high gene density in the chromosomal distribution.

Contribution analysis of the NBS class

Five distinct branches were obtained among the phylogenetic trees for the 249 NBS-type genes from Yugu1, but the number of resistance genes within the different branches was different (Figure 5A). This was particularly apparent in the ‘Zhang gu’ phylogenetic trees (Figure 5B). The classification is not completely clear, as only the gene clusters combined together (Figure 5). This suggested that these genes did not have a separate mode of evolution and that evolution of the resistance genes was more complex than expected.

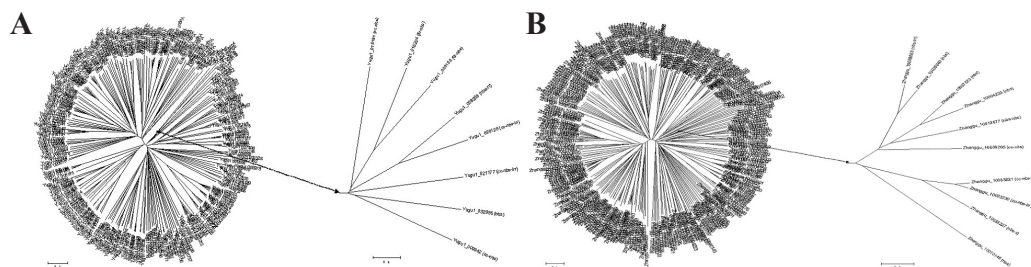


Figure 5. Phylogenetic tree (N-J type) for the two varieties of foxtail millet genome NBS class genes. **A.** Yugu1 database; **B.** ‘Zhang gu’ database.

DISCUSSION

The 269 genes containing the NBS structure accounted for 0.67% of the genes in the Yugu1 genome and 0.72% of the whole genome in the 'Zhang gu' database. The results showed that foxtail millet contained fewer NBSs than rice (1%). More than half of the rice NBS-type genes were shown to have generated by duplication, and the homology among these genes was high (Zhou et al., 2004). This may be the reason for the relatively low number of foxtail millet-resistance genes. Chromosome 8 in the foxtail millet genes was found in collinear blocks with chromosome 11 of rice, which contains the largest number of NBSs and the two largest clusters (Zhang et al., 2012). This is consistent with our results, which indicated that there are likely to be differences and similarities in disease-resistance between the two cultivars of foxtail millet. By comparing the differences and similarities between the two databases, it should be possible to narrow down the search for disease-resistance genes or defense-responsive genes.

Gene cluster analysis can explain the reasons behind gene duplication, which is one of the most obvious characteristics of resistance genes. Holub (2001) defined gene duplication as an alignable nucleotide sequence that covers >80% of the longer gene, or the region of identity between them encompasses >75% of the alignable region, and only one duplication is recorded for tightly linked genes. Gene duplication frequently leads to an increase in the number of gene clusters and members as well as the number of homologous genes. Due to the similarity between the homologous gene sequences, gene and genetic recombination can be greatly facilitated. The few or no replicated genes on chromosome 9 may explain why it had the lowest number of NBS genes.

Phylogenetic analysis of the Yugu1 genome NBS-type-resistance genes showed five branches that diverged from the star distribution of rice (Yang et al., 2008; Zhang et al., 2010), corn (Wang et al., 2009a), and other crops (Radwan et al., 2008; Ameline-Torregrosa et al., 2008). *Arabidopsis* had two distinct branches, but this result was mainly due to differences between the TIR structure and the CC structure (Meyers et al., 2003; Wang et al., 2009b). However, in this study, the two TIR structures were not separated on a distinct branch, but they were rather twigged onto a big branch. This may be because the TIR-NBS structure only contained two genes and the blast, whereas the other sites had high homology.

TIR-NBS disease-resistance genes have been widely reported in dicotyledons. However, TIR-type disease-resistance genes were not found in monocots until 2004 when Zhou et al. undertook a BLAST search using the improved Hidden Markov Model (HMM). TIR-type genes were first found in rice chromosome 1. However, the reason for a lack of a typical LRR domain is unclear. In the present study, two TIR-NBS disease-resistance genes were found in foxtail millet that were similar to genes containing the TIR structure found in rice and sorghum (Paterson et al., 2009), both of which do not have the LRR domain. There are receptors that are similar to the IL-1 receptor of the *Drosophila* Toll protein, which is also found in mammals and in plants. Increasing research in this field is accumulating knowledge as to whether the plant immune system is similar to that of *Drosophila* or mammals. Therefore, there is a need to conduct further research to discover the specific functions of these two genes.

ACKNOWLEDGMENTS

We are grateful to Zhu Yantao (Hebei University of Engineering) and Jiang Hongxin (Nankai University) for assistance with software and programming language, respectively. Research supported by the Ministry of Science and Technology of China (Grant#2011BAD06B01),

by the National Natural Science Foundation of China (Grants #31271787 and #31101163), the Natural Science Foundation of China of Hebei (C2013301037, C2014301028), and Genetically Modified Organisms to Cultivate New Varieties of Major Projects (#2014ZX0800909B).

REFERENCES

- Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, et al. (2008). Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol.* 146: 5-21.
- Baker B, Zambryski P, Staskawicz B and Dinesh-Kumar SP (1997). Signaling in plant-microbe interactions. *Science* 276: 726-733.
- Bettinger RL, Barton L and Morgan C (2010). The origins of food production in north China: A different kind of agricultural revolution. *Evol. Anthropol.* 19: 9-21.
- Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41: 95-98.
- Hammond-Kosack KE and Jones JD (1997). Plant disease resistance genes. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 48: 575-607.
- Holub EB (2001). The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat Rev. Genet.* 2: 516-527.
- Jeffrey LB, Jeremy S, Wang H and Percifield R (2012). Reference genome sequence of the model plant *Setaria*. *Nat. Biotechnol.* 30: 555-561.
- Jusuf M and Pernes J (1985). Genetic variability of foxtail millet (*Setaria italica* P. Beauv.). *Theor. Appl. Genet.* 71: 385-391.
- Kohler A, Rinaldi C, Duplessis S, Baucher M, et al. (2008). Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant Mol. Biol.* 66: 619-636.
- Li J, Ding J, Zhang W, Zhang Y, et al. (2010). Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. *Mol. Genet. Genomics* 283: 427-438.
- Meyers BC, Kozik A, Griego A, Kuang H, et al. (2003). Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 15: 809-834.
- Pan Q, Wendel J and Fluhr R (2000). Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J. Mol. Evol.* 50: 203-213.
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, et al. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457: 551-556.
- Radwan O, Gandhi S, Heesacker A, Whitaker B, et al. (2008). Genetic diversity and genomic distribution of homologs encoding NBS-LRR disease resistance proteins in sunflower. *Mol. Genet. Genomics* 280: 111-125.
- Shirano Y, Kachroo P, Shah J and Klessig DF (2002). A gain-of-function mutation in an *Arabidopsis* Toll Interleukin1 receptor-nucleotide binding site-leucine-rich repeat type R gene triggers defense responses and results in enhanced disease resistance. *Plant Cell* 14: 3149-3162.
- Somerville C and Somerville S (1999). Plant functional genomics. *Science* 285: 380-383.
- Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, et al. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876-4882.
- Wang JM, Jing HY, Zhao Y and Xiang Y (2009a). Genome-wide analysis of NBS-encoding disease resistance genes in Maize inbred line B73. *Acta Agronom. Sin.* 35: 566-570.
- Wang Y, Li ZY, Tang XL and Lu S (2009b). Bioinformatic analysis of the NBS-LRR gene family in *Arabidopsis thaliana*. *Chin. Agr. Sci. Bull.* 25: 40-45.
- Yang S, Gu T, Pan C, Feng Z, et al. (2008). Genetic variation of NBS-LRR class resistance genes in rice lines. *Theor. Appl. Genet.* 116: 165-177.
- Zhang G, Liu X, Quan ZW and Cheng SF (2012). Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat. Biotechnol.* 30: 549-554.
- Zhang XM, Wei JL, Zhang D and Wang CT (2010). Isolation and sequence analysis of NBS resistance gene analogues in Wild Rice. *J. Plant Genet. Resour.* 11: 741-748.
- Zhou T, Wang Y, Chen JQ, Araki H, et al. (2004). Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol. Genet. Genomics* 271: 402-415.
- Zohary D and Hopf M (2000). Domestication of Plants in the Old World: The Origin and Spread of Cultivated Plants in West Asia, Europe, and the Nile Valley (3rd edn.). Oxford University Press, New York.