3 Biotech (2018) 8:453
https://doi.org/10.1007/s13205-018-1478-6

ORIGINAL ARTICLE

Genome-wide identification, characterization, and evolutionary analysis of NBS-encoding resistance genes in barley

Yosra Habachi-Houimli[1] · Yosra Khalfallah[1] · Maha Mezghani-Khemakhem[1] · Hanem Makni[1,2] · Mohamed Makni[1] · Dhia Bouktila[1,3]

Received: 20 July 2018 / Accepted: 11 October 2018 / Published online: 19 October 2018

Abstract

In this study, a systematic analysis of Nucleotide-Binding Site (NBS) disease resistance (*R*) gene family in the barley, *Hordeum vulgare* L. cv. Bowman, genome was performed. Using multiple computational analyses, we could identify 96 regular NBS-encoding genes and characterize them on the bases of structural diversity, conserved protein signatures, genomic distribution, gene duplications, differential expression, selection pressure, codon usage, regulation by microRNAs and phylogenetic relationships. Depending on the presence or absence of CC and LRR domains; the identified NBS genes were assigned to four distinct groups; NBS–LRR (53.1%), CC-NBS–LRR (14.6%), NBS (26%), and CC-NBS (6.3%). NBS-associated domain analysis revealed the presence of signal peptides, zinc fingers, diverse kinases, and other structural features. Eighty-five of the identified NBS-encoding genes were mapped onto the seven barley chromosomes, revealing that 50% of them were located on chromosomes 7H, 2H, and 3H, with a tendency of NBS genes to be clustered in the distal telomeric regions of the barley chromosomes. Nine gene clusters, representing 22.35% of total mapped barley NBS-encoding genes, were found, suggesting that tandem duplication stands for an important mechanism in the expansion of this gene family in barley. Phylogenetic analysis determined 31 HvNBS orthologs from rice and *Brachypodium*. 87 out of 96 HvNBSs were supported by expression evidence, exhibiting various and quantitatively uneven expression patterns across distinct tissues, organs, and development stages. Fourteen potential miRNA-*R* gene target pairs were further identified, providing insight into the regulation of NBS genes expression. These findings offer candidate target genes to engineer disease-resistant barley genotypes, and promote our understanding of the evolution of NBS-encoding genes in Poaceae crops.

Keywords *Hordeum vulgare* · Nucleotide-binding site · Disease-resistance genes · Genome analysis

Abbreviations

CC	Coiled-coil (domain)
IBSC	The International Barley Genome Sequencing Consortium
LRR	Leucine-rich repeat (domain)
NBS	Nucleotide-binding site (domain)
NLR	NOD-like receptors
miRNA	MicroRNAs
TIR	Toll-interleukin-1 receptor (domain)

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13205-018-1478-6) contains supplementary material, which is available to authorized users.

✉ Dhia Bouktila
dhia_bouktila2000@yahoo.fr

[1] Université de Tunis El Manar, Faculté des Sciences de Tunis, Unité de Recherche Génomique des Insectes Ravageurs des Cultures d'intérêt agronomique (GIRC, UR11ES10), El Manar, 2092 Tunis, Tunisia

[2] Université de Tunis, Institut Supérieur de l'Animation pour la Jeunesse et la Culture (ISAJC), Bir El Bey, Tunisia

[3] Université de Jendouba, Institut Supérieur de Biotechnologie de Béja (ISBB), 9000 Béja, Tunisia

Introduction

Within the family of Poaceae, the tribe of the Triticeae contains two of the world's most important food crops, namely wheat and barley. Cultivated barley, *Hordeum vulgare* L. ssp. *vulgare*, is the fourth most important cereal crop worldwide, after wheat, maize, and rice and is among the top ten crop plants in the world (https://www.croptrust.org/crop/barley/). It is a true diploid ($2n = 2x = 14$, HH) species that was first domesticated about 10,000 years ago, from its wild relative *H. vulgare* L. ssp. *spontaneum* in the area of the Middle East,

known as the Fertile Crescent (Badr et al. 2000). Barley is widely adapted to diverse environmental conditions and is more stress tolerant than its close relative, wheat (Nevo et al. 2012). Until the late nineteenth century, barley genotypes existed as highly heterogeneous landraces adapted to different environments. Over the past 100 years, the landraces have mostly been replaced in agriculture by pureline varieties with reduced genetic diversity (Nevo 1992). Extensive cultivation, intensive breeding and selection have resulted in thousands of modern varieties of barley, distinguishable by a number of key traits, e.g., feed or malting, winter or spring growth habit, hulled or hull-less, and six-, four- or two-row varieties. Although less productive, local primitive landraces of barley and wild barley populations are of great importance to broaden the narrow genetic base of new agricultural cultivars (Abdel-Ghani et al. 2008), especially with regard to resistance to biotic stresses (Abdel-Ghani et al. 2008; Zeng et al. 2014).

The barley genome has been proposed as a model for genomic research in the Triticeae (Schulte et al. 2009). The International Barley Genome Sequencing Consortium (IBSC) has recently published a gene-space assembly (IBSC 2012) of cultivated barley as an enabling platform for genome-assisted basic research and crop improvement. This barley genome has a size of 5.1 Giga-base pairs, with ~98% (4.98 Gb) physically mapped, and more than 3.90 Gb anchored to a high-resolution genetic map (IBSC 2012). The barley gene-set has been estimated to approximately 30,400 genes based on exon detection, gene modelling and gene-family-directed comparison with the genomes of sorghum, rice, *Brachypodium*, and *Arabidopsis*. Among these genes, 26,159 (86%) were defined as High-Confidence (HC) genes, based on comparison with reference plant genomes. An additional characteristic of the barley genome, as previously reported by Wicker et al. (2009) and further confirmed by the IBSC survey, was the abundance of repetitive DNA, with approximately 84% of the genome made of mobile elements or other repeat structures. The IBSC survey has provided a physical and functional resource that describes the barley gene-space in a structured whole-genome context. Even so, annotation of eukaryotic genomes remains a long-winded task. Indeed, protein databases grow constantly, leading to better annotations. For crop genomes, a major challenge in the post-genome sequencing era is to obtain high-quality annotations (Yandell and Ence 2012) so that the full potential of genome sequences can be realized for functional studies and comparative analysis (Loveland et al. 2012). In this regard, annotating single protein families or sets of related protein families (making up a single biological system) may lead to improvements when compared to the whole-genome systematic annotation.

Plant diseases are among the many major factors that limit the yield of crops. Plants need to defend themselves against attack from viruses, microbes, and invertebrates. There are two branches of plant immunity; one uses transmembrane pattern recognition receptors (PRRs) that respond to slowly evolving PAMPs (Chisholm et al. 2006); the second acts largely inside the cell, using protein products encoded by most resistance (*R*) genes (Dangl and Jones 2001; Zipfel 2008). In plant genomes, *R* genes convey disease resistance against pathogens by producing *R* proteins. These receptors are mainly intracellular, and they can specifically interact with pathogen effectors coded by the avirulence (*Avr*) genes, either directly, following the "gene for gene" model (Flor 1971), or indirectly, following the guard or decoy models. In these models, the interaction between receptor and effector can occur through so-called "decoy domains" fused to NOD-like receptors (NLRs), which can be the virulence target ("guard model") or a structural mimic of such a target ("decoy model") (van der Hoorn and Kamoun 2008). Many *R* genes are single genotype locus genes that confer resistance against one pathogen (Timmerman-Vaughan et al. 2000). Of these, more than 100 have been cloned from various plant species, revealing that most of them contain a nucleotide-binding site (NBS) usually attached to a leucine-rich repeat (LRR) domain (Borrelli et al. 2018). These plant NBS–LRR proteins (also called NLR, NB–LRR or NB–ARC–LRR proteins) are typically categorized into the TIR or non-TIR class, based on the identity of the sequences that precedes the NB domain (Meyers et al. 2005). The TIR class of plant NBS–LRR proteins contains an amino-terminal domain with homology to the Toll and interleukin 1 receptors. The non-TIR class is less well defined, but most NBS–LRR proteins of this class contain α-helical-coiled coil-like sequences in their amino-terminal domain (Pan et al. 2000). The NBS domain was proposed to be crucial for adenosine triphosphate (ATP)-binding and hydrolysis, and the presence of bound ATP or ADP may determine whether the *R* protein is in an active or inactive signaling state (Lukasik and Takken 2009). The LRR domain is responsible mainly for the specificity of the interaction between the *R* gene product and the avirulence factors (Yoshimura et al. 1998). The NBS domains of all characterized *R* genes contain several highly conserved motifs: a P-loop, kinase-2, kinase-3a, and Gly-Leu-Pro-Leu (GLPL) domains (Tan and Wu 2012). The P-loop and kinase-2 motifs are thought to be ATP- and guanosine triphosphate (GTP)-binding sites (Meyers et al. 1999).

Genome-wide and comparative genomic analyses have indicated that plant genomes can encode several hundreds of NBS-type proteins, and that there is a great diversity in the number and distribution of the subclasses of these genes (reviewed in Marone et al. 2013). The extreme numbers are ~50 in papaya (Porter et al. 2009) and *Cucumis sativus* (Wan et al. 2013), to ~1200 in *Aegilops tauschii* (Jia et al. 2013). Proliferation of *R* genes often involves

tandem or segmental duplication, and hence the genes tend to be clustered. It has been postulated that the rapid copy number evolution of NBS genes is driven by gene loss or expansion within a species through repeated cycles of duplication, divergence, and eventual loss by pseudogenization or deletion in response to diverse pathogens (Li et al. 2010; Gu et al. 2015). The rapid gene copy number evolution (expansion or contraction) is a fundamentally important strategy for a species to adapt to its changing environment. As NBS genes are expected to be under continuous selection pressure for alleles that allow the plant to defend itself against quickly changing species-specific pathogen spectrum, the expansion and/or contraction of this family has been documented as faster than other gene groups (Li et al. 2010).

The recent availability of the whole-genome sequences of a number of crop and model species, such as *Arabidopsis thaliana* (The Arabidopsis Genome Initiative 2000), rice (International Rice Genome Sequencing Project 2005) and hexaploid wheat (The International Wheat Genome Sequencing Consortium 2014), has given rise to a suite of new studies assessing the NBS genes genomic dynamics. The results of these studies showed that the NBS-encoding gene family contains a high proportion of duplicated genes, including tandem and segmental duplication events (Meyers et al. 2003; Yang et al. 2008). Tandem duplication may play a major role in the expansion of NBS-encoding genes (Mun et al. 2009), while segmental duplication has only been found to have contributed to a substantial increase in NBS gene number in *Glycine max* (Schmutz et al. 2010). However, genomic analysis of *Oryza sativa* (Yu et al. 2002), *A. thaliana* (Nobuta et al. 2005), *Vitis vinifera*, and *Populus trichocarpa* (Yang et al. 2008) indicates that most duplicated NBS-encoding genes are lost after whole-genome duplication (WGD) events.

Applying genomic tools for a better knowledge of the NBS-encoding gene family in barley could have a great applicability for fundamental studies, such as phylogenetic analyses (Lozano et al. 2015), population genetics (Joshi and Nayak 2013), evolutionary research (Gu et al. 2015), as well as for pathogen and pest resistance (Martin et al. 2011; Biruma et al. 2012). Modern plant breeding is witnessing an important paradigm shift, from Marker-Assisted Selection (MAS) to a more acute large-scale Gene-Assisted Selection (GAS) (Ruane and Sonnino 2007). In this paper, the barley genome was investigated to identify, de novo, members of the NBS-encoding disease resistance gene family. The obtained candidate genes were submitted to a characterization package, including their structural diversity, conserved protein signatures, gene duplications, differential expression, selection pressure, synonymous codon usage bias, regulation by microRNAs (miRNAs) and phylogenetic relationships. This genome-scale characterization provides a genomic resource that will act as a basis for developing disease-resistant barley cultivars.

Materials and methods

Retrieval, identification, and classification of candidate barley NBS-encoding *R* genes

The Whole Genome Shotgun Sequence assembly of barley cv. Bowman (2,077,901 contigs; 1,779,486,241 bp) was downloaded from the National Center for Biotechnology Information NCBI database (http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=CAJX01). Method used to identify the NBS-encoding genes in *H. vulgare* consisted of the three following steps:

(a) HMM-based sequence retrieval: all contigs of the barley genome were translated in six reading frames, using transeq algorithm from the software package of the European Molecular Biology Open Software Suite (EMBOSS version 6.5.0.0, ftp://emboss.open-bio.org/pub/EMBOSS/) and the resulting files were used to construct a local database in the subsequent search. We used a hidden Markov model (HMM) for the NBS domain from the Pfam database (PF00931; http://pfam.sanger.ac.uk/search). The threshold expectation value was set to 10^{-10} (Tan and Wu 2012), a value determined empirically to filter out most of the spurious hits. This HMM profile was applied to the six-frame translation of barley contigs, using hmmsearch tool (HMMER 3.0, http://hmmer.janelia.org/) and all nucleotide contigs corresponding to NBS domains were retrieved using seqret tool (EMBOSS version 6.5.0.0, ftp://emboss.open-bio.org/pub/EMBOSS/), merged into a unique multiFasta file using DNA Baser sequence assembler v. 3.5.4.2 (http://www.dnabaser.com/), and checked for any redundant contigs, using cdhit-est (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi?cmd=cd-hit-est) and manual removal. The obtained non-redundant contigs were split into individual Fasta files using seqretsplit command of EMBOSS.

(b) Gene identification: the non-redundant contigs were submitted individually to ab initio gene prediction, using FGENESH (http://www.softberry.com) program. At this stage, three hundred and eight (308) gene predictions showing a complete ORF structure extending from ATG to stop codon, were selected to further analyze their potential protein products.

(c) Gene classification: Meyers et al. (1999, 2003) defined NBS domain as a region ranging up to ~300 amino acids that is composed of eight well-known characteristic motifs: P-loop (Kinase-1a), Kinase-2, RNBS-A,

RNBS-B, RNBS-C, RNBS-D, GLPL, and MHDV. Therefore, we performed a thorough investigation of barley NBS putative proteins with MEME package (Bailey and Elkan 1994) to check for the conservation of the NBS domain. Analysis of the 308 gene models with a complete open reading frame (based on gene prediction) and an NBS structure (based on the Pfam database) and identified 96 genes with highly conserved NBS region, which were considered as regular NBS genes, while the length of non-regular NBS domain was shorter than that of normal NBS domain and lacked one or more major motifs (Cheng et al. 2010). Subsequently, all candidate genes were evaluated to further verify whether they encoded TIR, CC, NBS, or LRR motifs using Geneious R8 (http://www.geneious.com) with added plugins (coiled-coil and transmembrane prediction tools), which enabled submitting all candidate genes for protein signature analysis via internet connection to InterProScan with its 13 integrated databases (Gene3D, HAMAP, PANTHER, Pfam, PIR super family, PRINTS, PROFILE, PROSITE, SignalP, SMART, Pro-Dom, SUPERFAMILY, and TIGRFAMs), predicting coiled-coils (CC), predicting cellular topology (i.e., protein cytoplasmic, transmembrane, and extracellular regions). The detailed protein domain information was used to classify the NBS-encoding genes into subgroups.

Chromosomal locations of NBS-encoding genes

Nucleotide sequences of the 96 regular barley NBS genes identified in our study were used to retrieve their physical positions on barley chromosomes. For this purpose, we used the 'Barleymap' server (http://floresta.eead.csic.es/barleymap/align/) (Cantalapiedra et al. 2015), to perform a GMAP search of the 96 HvNBS *R* genes determined in this study, against the barley physical map (IBSC 2012), with highly stringent parameters applied: minimum sequence identity = 98% and minimum query coverage = 95%. Results were obtained as a CSV-formatted file and a graphical plot of the mapped *H. vulgare* NBS genes on the barley seven chromosomes.

Comparative phylogeny

Ninety-six (96) regular barley NBS proteins (HvNBS) obtained after checking for the conservation of the NBS domain motifs by MEME, were used. For *O. sativa* and *Brachypodium distachyon* sequences, we retrieved the predicted whole proteomes and domain annotations from ftp servers of the Rice Genome Annotation Project (RGAP 7.0, http://rice.plantbiology.msu.edu/), and the Munich Information Center for Protein Sequences (MIPS, ftp://ftpmips.helmholtz-muenchen.de/plants/brachypodium/v1.2), respectively. Using the domain annotations of both proteomes, all predicted proteins containing an NB-ARC domain (Pfam PF00931) were retrieved using seqret (EMBOSS) and placed in separate files. MEME package (Bailey and Elkan 1994) was used to crop NBS domains, from both species, with a conserved distribution of the major motifs. Multiple alignments of the amino acid sequences were performed by MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/). The phylogenetic tree was constructed using the NJ method (Saitou and Nei 1987) implemented in the Molecular Evolutionary Genetics Analysis software version 5.0 (MEGA 5.0) (Tamura et al. 2011). Bootstrapping (1000 replicates) was used to evaluate the degree of support for a particular node in the phylogenetic tree. Branch lengths were assigned by pairwise calculations of the genetic distances.

Gene duplication and selection pressure

To detect potential gene duplication, we aligned and calculated all of the relevant genes identified in *H. vulgare*. We considered the following parameters as evidence of recent gene duplication (1) the alignment covered > 70% of the longer gene; (2) the aligned region had an identity > 70%; (3) only one duplication event was counted for tightly linked genes (Gu et al. 2002; Tan and Wu 2012). A block of duplications (gene family) was defined if more than one gene was involved in the duplication.

The nonsynonymous/synonymous rate ratio is an important measure of the mode and strength of natural selection acting on nonsynonymous mutations in protein-coding genes. Generally, Ka/Ks = 1, > 1 and < 1 indicate neutral, positive, and purifying selection, respectively. The ratios of nonsynonymous substitution (Ka) to synonymous substitution (Ks) were computed in the gene families. The nucleotide coding sequences (CDSs) in each gene family were aligned by ClustalW and the values of Ka, Ks, and Ka/Ks were calculated by MEGA 5.0.

Synonymous codon usage bias

Using the algorithm "cusp" in EMBOSS explorer, we calculated the following parameters: (1) over-all-codons percentage of $G + C$ across all combined CDSs; (2) For each CDS solely, we determined the first, second, and third codon positions (GC1, GC2, and GC3), as well as the GC12 value that was calculated as the mean of GC1 and GC2. A neutrality plot (GC12 vs. GC3s) (Kawabe and Miyashita 2003) was established; (3) For each CDS, we calculated the following parameters: A3 (frequencies of adenine at the third positions of codons), T3 (frequencies of thymine at the third positions of codons), G3 (frequencies of guanine at the third positions of codons), and C3

(frequencies of cytosine at the third positions of codons). These statistics (A3, T3, G3 and C3) were used to examine the influence of GC content on codon usage, by constructing a bias plot (Sueoka 1999a).

Digital expression analysis

To gain insight into the expression profiles of NBS genes in *H. vulgare* in different tissues and development stages, we compared the 96 HvNBS genes identified, against the latest barley annotation data accessed through the Plant Genome and Systems Biology (PGSB) group database (barley project) (http://pgsb.helmholtz-muenchen.de/plant/barley/index.jsp). This PGSB database provides a BLAST server (IPK server, Leibniz Institute of Plant Genetics and Crop Plant Research, Germany), where sequences can be searched against the barley assemblies of the IBSC (http://webblast.ipk-gatersleben.de/barley/). All regular barley NBS genes, identified in the present study, were used as queries by BLASTP (Altschul et al. 1997), against all genes of the IBSC's survey, with default parameters. BLASTP hits were subsequently filtered, based on at least 90% of sequence similarity and 90% of sequence coverage.

Regulation by microRNAs

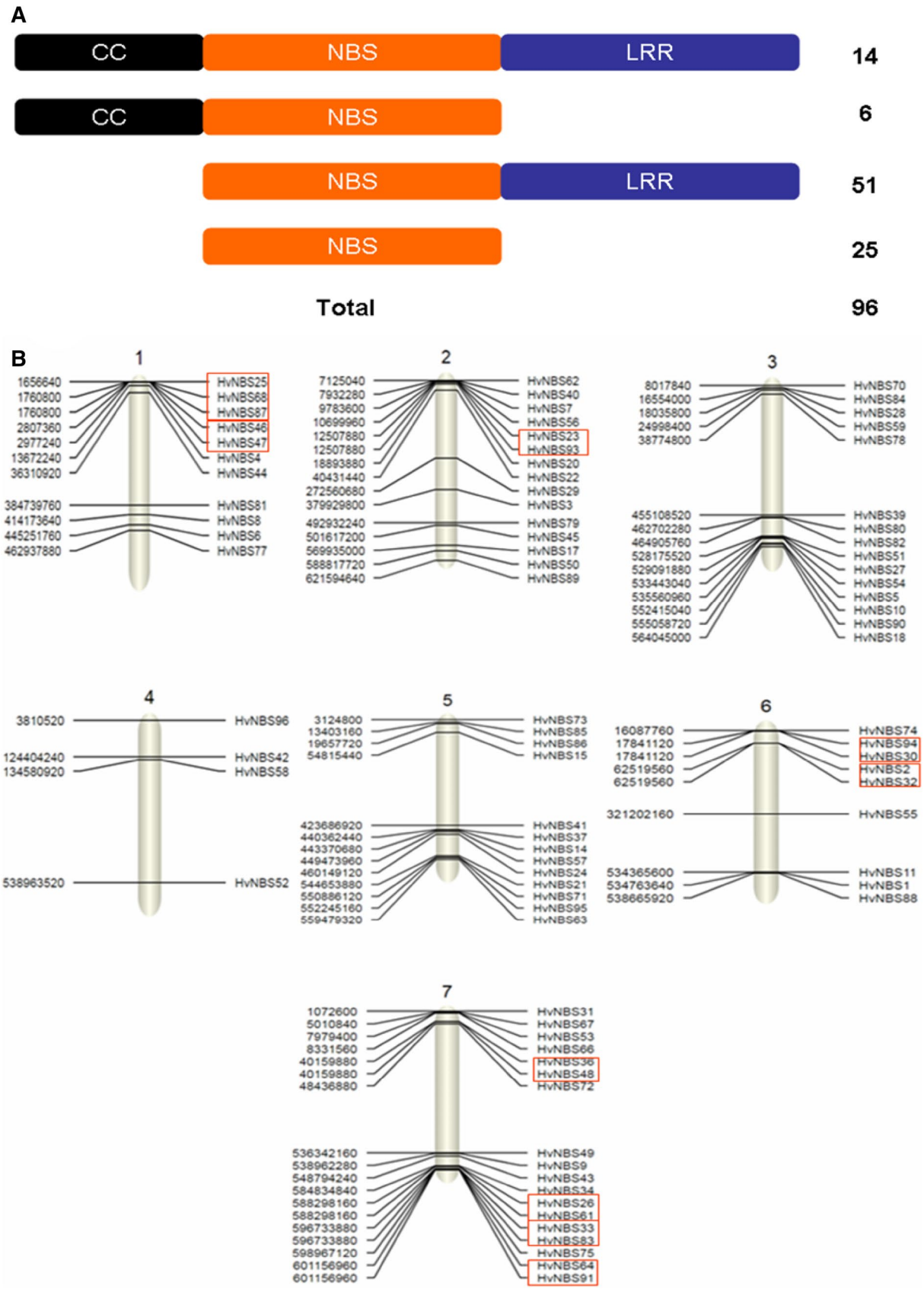
All available mature miRNAs of *H. vulgare* (Hvu-miRs) were collected from the Sanger's miRBase (http://www.mirbase.org, Release 21, 2014), for all tissues in *H. vulgare*. These Hvu-miRs were downloaded as three experimental sets, namely ER0000000318 (7-days leaf: 36 Hvu-miR species), ER0000000319 (inflorescence: 42 Hvu-miR species) and ER0000000320 (7-days leaf inoculated with the fungus *Blumeria*, agent of barley powdery mildew: 39 Hvu-miR species). The number of Hvu-miRs was later reduced to 43, by arranging miRNAs according to the sequence of the mature form and checking for any redundancies (Supplementary data Table S1). In addition to these Hvu-miRs, we used members of the miR482/2118 superfamily, including "miR482 type" and "miR2118 type" (Shivaprasad et al. 2012). One hundred and sixty (160) and 89 sequences of miR482 and miR2118, respectively, were obtained from miRBase (http://www.mirbase.org/). Sequence redundancies in the mature form were eliminated leading to the retention of 103 miR482/2118 sequences from 25 plant species (Supplementary data Table S1). Targets of miRNA were identified using psRNATarget software (http://plantgrn.noble.org/psRNATarget/) (Dai and Zhao 2011). The analysis was performed using default search parameters (Naqvi et al. 2011), after selecting the option "user-submitted small RNAs/user-submitted transcripts".

Results

Identification and structural features of NBS proteins in barley

Availability of the complete *H. vulgare* genome sequences has made it possible to identify all the NBS-encoding genes in this plant species. Initially, three hundred and eight (308) gene analogs displaying a complete ORF structure extending from ATG to stop codon were obtained by gene prediction from barley contigs whose translations were identified by HMM search. Analysis of the potential protein products from these 308 gene analogs identified 96 genes with highly conserved NBS regions that were considered as regular NBS genes, while the remaining 212 genes were characterized by a non-regular length of NBS domain (shorter than 2/3 of normal NBS domain) and/or lacked one or more of the conserved motifs characteristic of the NBS domain (Supplementary data Figure S1).

For the 96 regular NBS genes, the number of introns, exons and their lengths were determined from gene prediction output. Exon–intron analysis shows that barley-NBS genes are composed of an average of 3.34 exons, ranging from one exon to 14 exons in HvNBS96 (Supplementary Data Table S2). Of the 96 genes, 19, 13, 14, and 18 had one, two, three, and four exons, respectively; thus, two-thirds (66.66%) of the genes were found to contain one to four exons. Among the 96 regular NBS protein products, the shortest was HvNBS1 (378 aa), while the longest (HvNB96) was of 2,153 aa (Supplementary Data Table S2). Two cell arrangements were identified with respect to cell topology: proteins with only a cytoplasmic domain (90.62%), and those with cytoplasmic, transmembrane and extracellular domains (9.37%) (Supplementary Data Table S2). Depending on the presence or absence of CC and LRR domains; we could assign the 96 gene models identified to four distinct groups; NBS–LRR (53.1%), CC-NBS–LRR (14.6%), NBS (26%) and CC-NBS (6.3%) (Fig. 1a and Supplementary data Table S2). The total number of NBS-type genes identified in barley (96) was close to that of *Brachypodium* and maize, but was twofold lower than that of sorghum, and fivefold lower than that of rice and *Triticum urartu* (Table 1). The superiority of NBS–LRR subgroup within NBS-type genes was a common feature shared by barley, rice, and *T. urartu*. Unlike these species, in *Brachypodium*, maize, and sorghum, NBS-type genes are mostly CC-NBS–LRR genes (Table 1). In addition to the CC and LRR domains, additional structural features were present in the identified models. These characteristics include the presence of signal peptides, zinc fingers, diverse types of kinases, transcription factors and others. The structural diversity that


A
CC
NBS
LRR
14
CC
NBS
6
NBS
LRR
51
NBS
25
Total
96
B
1
1656640 HvNBS25
1760800 HvNBS68
1760800 HvNBS87
2807360 HvNBS46
2977240 HvNBS47
13672240 HvNBS4
36310920 HvNBS44
384739760 HvNBS81
414173640 HvNBS8
445251760 HvNBS6
462937880 HvNBS77
2
7125040 HvNBS62
7932280 HvNBS40
9783600 HvNBS7
10699960 HvNBS56
12507880 HvNBS23
12507880 HvNBS93
18893880 HvNBS20
40431440 HvNBS22
272560680 HvNBS29
379929800 HvNBS3
492932240 HvNBS79
501617200 HvNBS45
569935000 HvNBS17
588817720 HvNBS50
621594640 HvNBS89
3
8017840 HvNBS70
16554000 HvNBS84
18035800 HvNBS28
24998400 HvNBS59
38774800 HvNBS78
455108520 HvNBS39
462702280 HvNBS80
464905760 HvNBS82
528175520 HvNBS51
529091880 HvNBS27
533443040 HvNBS54
535560960 HvNBS5
552415040 HvNBS10
555058720 HvNBS90
564045000 HvNBS18
4
3810520 HvNBS96
124404240 HvNBS42
134580920 HvNBS58
538963520 HvNBS52
5
3124800 HvNBS73
13403160 HvNBS85
19657720 HvNBS86
54815440 HvNBS15
423686920 HvNBS41
440362440 HvNBS37
443370680 HvNBS14
449473960 HvNBS57
460149120 HvNBS24
544653880 HvNBS21
550886120 HvNBS71
552245160 HvNBS95
559479320 HvNBS63
6
16087760 HvNBS74
17841120 HvNBS94
17841120 HvNBS30
62519560 HvNBS2
62519560 HvNBS32
321202160 HvNBS55
534365600 HvNBS11
534763640 HvNBS1
538665920 HvNBS88
7
1072600 HvNBS31
5010840 HvNBS67
7979400 HvNBS53
8331560 HvNBS66
40159880 HvNBS36
40159880 HvNBS48
48436880 HvNBS72
536342160 HvNBS49
538962280 HvNBS9
548794240 HvNBS43
584834840 HvNBS34
588298160 HvNBS26
588298160 HvNBS61
596733880 HvNBS33
596733880 HvNBS83
598967120 HvNBS75
601156960 HvNBS64
601156960 HvNBS91

◂**Fig. 1** Overview of the structure and genomic distribution of barley NBS-encoding genes. **a** The distribution of NBS gene structures within the barley genome (CNL, CN, NL, and N). **b** Physical location and corresponding clusters distribution of NBS-encoding genes. Chromosome numbers are indicated at the top end of the chromosome. Grey bars represent chromosomes. Codes on the right side stand for HvNBS genes characterized in the present study, while numbers on the left of each chromosome stand for their corresponding physical locations (bp). Colorful boxes stand for clusters of NBS-encoding genes

exists among NBS proteins in barley is remarkable, suggesting that diversifying selection has played an important role in the evolution of *R* genes in this agronomically important species. Description and representation of each feature is illustrated in Supplementary data Table S3.

Chromosomal distribution of barley NBS-encoding genes

Among 96 HvNBS genes reported in this study, we could assign 85 to their chromosomal locations, based on their comparison with the barley physical map (Fig. 1b, Supplementary Data Table S4). The remaining genes could not be mapped because they did not result in gene counterparts with enough significance. There were 18 genes that were localized on chromosome 7H, 15 each on chromosomes 2H and 3H, whereas only four genes were located on chromosome 4H. The remaining 33 NBS-encoding genes were located on chromosomes 1H (11), 5H (13), and 6H (9).

Holub (2001) defined a gene cluster as a chromosome region that contains two or more genes within 200 kb. Using this standard, we found that 19 NBS genes, representing 22.35% of total mapped barley NBS-encoding genes, were located in 9 clusters and the remaining 66 genes were singletons (Fig. 1b). Among the 9 clusters, the highest number (4) was located on chromosome 7H, two were located on chromosome 1H, two on chromosome 6H, one on chromosome 2H, while clusters were absent from chromosomes 3H, 4H, and 5H. Most gene clusters contained two genes, although a cluster of 3 genes was identified as chromosome 1H. As previously reported by The International Barley Genome Sequencing Consortium (2012), there is a clear tendency of NBS genes to be clustered in the distal telomeric regions of the barley chromosomes. The chromosomal distribution of barley NBS-type genes suggests that tandem duplication stands for an important mechanism in the expansion of NBS-encoding genes in barley, as previously established in *Brassica rapa* (Mun et al. 2009), grapevine and poplar (Yang et al. 2008).

Comparative phylogeny

Phylogenetic analysis by the NJ method was conducted to determine the relationships among NBS-encoding genes in barley and its two close relatives, rice and *Brachypodium*. We constructed a composite phylogenetic tree for candidate genes from these three species. A primary dataset including 507 proteins from three species (96 of barley, 292 of rice and 119 of *Brachypodium*) was submitted to phylogeny reconstruction. For the obtained phylogeny size was too large, we selected 71 OsNBS and 42 BdNBS sequences, on the basis of the clades containing one or more barley genes with at least a counterpart from rice or *Brachypodium*, for a better readability of HvNBS sequences and their evolutionary relationships with rice or *Brachypodium* homologs. Similar to rice (Zhou et al. 2004) and maize (Cheng et al. 2012), the phylogenetic tree for the barley NBS-encoding genes (Fig. 2) produced many radiating branches that lacked resolution and did not indicate relationships or classification schemes among the NBS-encoding genes. Based on relationships in the tree, we detected 10 HvNBS paralogous clusters and 31 sequences orthologous to HvNBS ones: 18 from *B. distachyon* (BdNBS) and 13 from *O. sativa* (OsNBS) (Fig. 2). The identified orthologous genes may be useful for future transgenic research in Poaceae, given that orthologous genes tend to have similar structures and functions.

Gene duplication and selection pressure

During evolution, gene duplications have greatly expanded the NBS gene family in both monocot and eudicot lineages (Tan and Wu 2012). In this study, we confirmed these genome duplications. A total of 12 NBS genes with duplication events were identified in barley, and subsequently divided into six multigene families. The maximal number of members of a multigene family was two (Supplementary Data Table S5). The percentage of the multigene families in NBS genes of *H. vulgare* genome (12.5%) was clearly lower than in the four Poaceae species used for comparison (maize, sorghum, rice, and *Brachypodium*) (Table 2).

Positive selection drives the host–pathogen coevolution and selection for new resistance specificities. To detect and measure the direction and intensity of selection, we estimated the ratios of the nonsynonymous substitution to the synonymous substitution (Ka/Ks or d*N*/d*S*) using MEGA 5.0 in all families. The ratio of Ka/Ks in each duplicated gene pair was less than one, revealing that all duplicated genes underwent purifying selection (Table 3). The average value of Ka/Ks (the ratio of non-synonymous substitutions to synonymous substitutions) in HvNBS sequences was 0.385. This value is larger than the average (genomic) Ka/Ks value of barley of 0.2214 (Jia et al. 2013), suggesting a purifying

Table 1 Numbers and architectures of regular NBS-encoding genes, in barley and five Poaceae relatives

Letter code	Architecture, at protein level, with respect to the four major component domains in monocots					
	Barley[a]	*Triticum urartu*[b]	*Brachypodium*[c]	Maize[d]	Rice[d]	Sorghum[d]
CC–NBS–LRR	14	161	102	58	160	119
NBS–LRR	51	202	12	31	304	54
CC–NBS	6	61	11	11	7	19
NBS	25	87	1	7	45	44
Total	96	511	126	107	519	236

[a]Data from the current study

[b]Data from Ling et al. (2013)

[c]Data from Tan and Wu (2012)

[d]Data from Cheng et al. (2012)

selection that is weaker in NBS sequences, in comparison with the whole-barley genome.

Synonymous codon usage bias

The GC content in all three codon positions across HvNBS genes was found to be equal to 47.4%, indicating that the AT content in HvNBS genes is higher than GC content. A neutrality plot (GC12 vs. GC3s) revealed the relationship between mutation and selection bias (Fig. 3a). Points were distributed along a nearly-parallel line to abscissa, indicating that genes were completely non-neutral. The regression coefficient (slope) was clearly smaller than unity ($R^2 = 0.3913$), indicating that GC mutational bias doesn't lead to similar GC content in all codon positions (Sueoka 1999b). It seems, therefore, that the synonymous codons bias is more shaped by natural selection than by neutral mutation. To confirm this hypothesis, we constructed a bias plot of the third codon position between G and C content and between A and T content (Fig. 3b). Natural selection can be detected if GC or AT are disproportionately found in the third codon letter (Wright 1990). The results showed that GC or AT were used disproportionately, with G and T used more frequently than C and A in the third position of most HvNBS genes.

Expression profiles and interaction with microRNAs

Gene expression profiles were determined for 87 of 96 barley NBS-containing genes identified in the present study, based on their close homologs hosted in the barley genome explorer, Barlex (Colmsee et al. 2015). Expression data of nine genes could not be obtained, most probably because of presence/absence sequence variation between cv. Bowman and cv. Morex contigs. For the remaining genes, expression has been detected in multiple tissues/development stages simultaneously, namely in: "developing tillers at six-leaf stage, third internode" (65/96), "roots from seedlings, 10 cm shoot stage" (67/96), "shoots from seedlings, 10 cm shoot stage" (64/96), "developing inflorescences, 1–1.5 cm" (55/96), "developing grain, bracts removed 15 DPA" (62/96), "developing grain, bracts removed 5 DPA" (57/96), "young developing inflorescences, 5mm" (2/96), "4-day embryos" (64/96), "Etiolated seedling, dark cond. (10 DAP)" (60/96), "Inflorescences, lemma (42 DAP)" (58/96),"Inflorescences, lodicule (42 DAP)" (54/96),"Epidermal strips (28 DAP)" (65/96),"Roots (28 DAP)" (67/96),"Senescing leaves (56 DAP)" (63/96),"Inflorescences, rachis (35 DAP)" (62/96), and "Dissected inflorescences, palea (42 DAP)" (59/96) (Supplementary data Table S6). However, expression was quantitatively uneven across these tissues/development stages, as illustrated in Fig. 4, with maximum expression detected in "Senescing leaves (56 DAP)" in 16 of these genes and in "developing inflorescences, 1–1.5 cm" in 13 of them.

The psRNATarget algorithm (Dai and Zhao 2011) that predicts targets of plant miRNAs identified the mRNAs of barley candidate-disease resistance proteins as having binding sites for *H. vulgare* miRNAs, as well as miR482/miR2118 superfamily species. Of the 96 predicted NBS-containing proteins encoded in the barley genome, there were 11 and 3 with target motifs of miR482/2118 and HvmiRs, respectively (Supplementary data Table S7). The number of miRNA species potentially involved in post-transcriptional regulation of barley NBS genes was as low as eight, among which three were identified from *H. vulgare* miRs (hvu-mi5048a, hvu-miR5048b and hvu-miR6205), and five from miR482/miR2118 superfamily (aly-miR472-3p, aqc-miR482c, ata-miR2118a-3p, ata-miR2118a-5p, bdi-miR2118a) (Supplementary data Table S7).

Discussion

Barley is one of the world's four major cereal crops, alongside bread wheat, rice, and maize. Diseases of barley, mostly caused by fungal pathogens and, secondly, by bacteria,

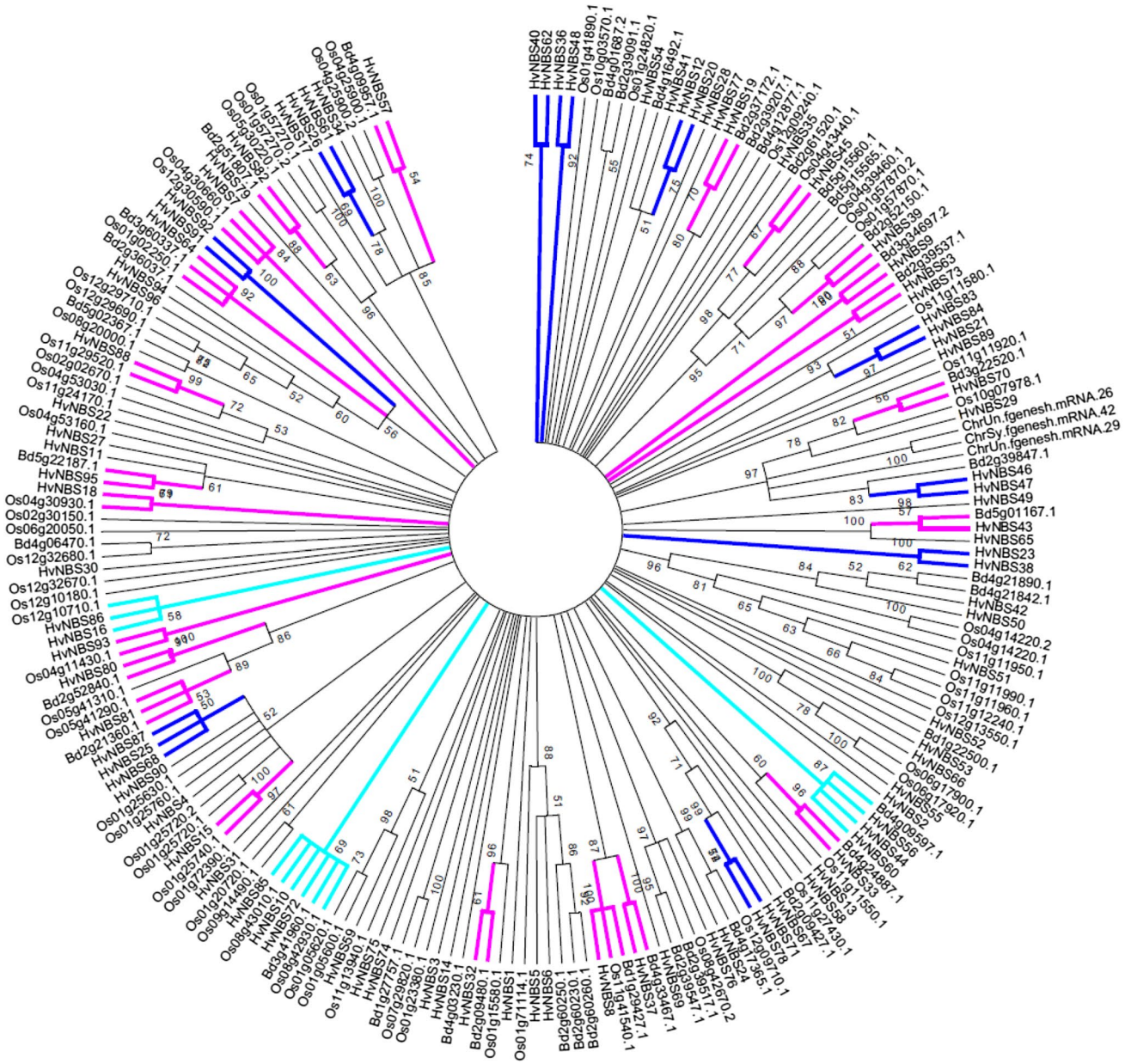


Fig. 2 Neighbor-Joining-based phylogeny reconstructed using the MEGA 5.0 program (Tamura et al. 2011). An alignment of 209 core NBS sequences was used: 96 of barley (HvNBS), 71 of rice (OsNBS) and 42 of Brachypodium (BdNBS). Blue clades contain only HvNBS paralogs, cyan clades contain at least two HvNBS paralogs with ortholog Bd/OsNBS sequences; magenta clades contain a unique HvNBS sequence with at least one Bd/OsNBS ortholog. HvNBS, BdNBS, and OsNBS core NBS sequences were aligned and used to generate a 1000 bootstrap Neighbor-Joining phylogenetic tree

viruses, and nematodes, are important constraints of its production and quality. A significant challenge in barley breeding is the introduction of durable disease-resistance traits into agronomically important varieties. However, only about five resistance genes have been identified in barley, which confer resistance to powdery mildew (Büschges et al. 1997; Halterman and Wise 2004; Seeholzer et al. 2010) and stem rust (Mirlohi et al. 2008). The identification and annotation of barley *R* genes, in particular those of the NBS family, will shed light on breeding strategies for disease control in crops. This aim is currently greatly facilitated due to advances in genomic data and computational ability. Also, genome-wide cultivar-specific analyses are useful to determine variation across cultivars, which are related to crop yield.

Our present study has identified 96 regular NBS-encoding genes in the genome of *H. vulgare* cv. Bowman. These genes were classified into four gene subgroups, namely NL, CNL, N, and CN. NL genes were found in greatest abundance

Table 2 Comparison of duplications, in the NBS-encoding *R* genes, between barley and four plant genomes

Organization	Barley[a]	Maize[b]	*Sorghum*[b]	Rice[b]	*Brachypodium*[c]
Unique genes	84	68	136	240	77
Multigenes	12	41	102	279	49
Gene family number	6	17	20	93	20
Maximal family members	2	6	7	10	7
Average members per family	2	2.41	6.10	3.00	2.45
Multigenes/single-gene families	0.14	0.60	2.83	1.16	0.64
Percentage of multigenes	12.5	37.6	47.3	53.7	34.9

[a]Data from the current study

[b]Data from Cheng et al. (2012)

[c]Data from Tan and Wu (2012)

(53.1%). Many comparative analyses have suggested that the NBS gene family in each genome is characterized by variable copy numbers and different distributions among structural subclasses (Gu et al. 2015). For example, some genomes, such as the apple genome, contain over 1000 NBS genes, although the genome size is relatively small (about 740 Mb) (Velasco et al. 2010). In contrast, in the maize genome, which exceeds 2 Gb, only 109 NBS–LRR genes were identified (Cheng et al. 2012). In the grass species, a great variation has been documented in the total number of NBS–LRR genes, from about 100 in maize to 500 in rice, testifying to the variable evolutionary patterns of NBS genes in different plant genomes. Such patterns may include specific expansion events in some species. For example, in bread wheat, the NBS genes have expanded greatly due to the polyploidy (Gu et al. 2015). Additionally, there are specific preferences in gene functions between species with the same ploidy level. For example *Aegilops tauschii*, the D-genome donor to modern wheat, contains about threefold more NBS genes than the A-genome progenitor, *T. urartu*, because the A genome tends to control morphological traits, while the D genomes prefers controlling the reaction to biotic and abiotic stresses (Gu et al. 2015). In some other cases, drastic number variations in NBS genes are due to the evolutionary speed of this family in each species, as a strategy of natural selection. In plants with rapidly evolving pathogens, disease-resistance genes must also evolve quickly; therefore the pseudogenization processes may be slowed down.

Table 3 Selective pressures of NBS-encoding genes in the barley genome

Cluster no.	HvNBS genes	Ka	Ks	Ka/Ks	Selection type
1	HvNBS71, HvNBS78	0.142	0.367	0.387	Purifying
2	HvNBS43, HvNBS65	0.162	0.468	0.346	Purifying
3	HvNBS36, HvNBS48	0.081	0.224	0.361	Purifying
4	HvNBS46, HvNBS47	0.104	0.274	0.379	Purifying
5	HvNBS23, HvNBS38	0.088	0.182	0.483	Purifying
6	HvNBS12, HvNBS20	0.143	0.404	0.354	Purifying

We have demonstrated the possible presence of some atypical domains such as zinc fingers (Zf-BED), AvrRpt-cleavage and DNA-Binding domains, together with the possible presence of three types of kinases (serine/threonine protein kinase, SNF1-serine/threonine Pkinase and Pkinase). The inventory and analysis of these additional domains should have an importance into refining the classification of NBS-type proteins into subclasses and, especially, a better knowledge of their involvement in signal transduction during the plant resistance response to biotic stresses. For example, HvNBS83 was characterized by an AvrRpt-cleavage domain. This domain was reported in Solanaceae, where it was shown that it allows the *R* protein to respond to more than one *Avr* protein, as demonstrated with *Pto* and *Rpm1* genes (Luo et al. 2009). In HvNBS89, we have identified sucrose nonfermenting-1 serine/threonine protein kinase (SNF1-Ser/Thr PKinase) domain. The yeast SNF1 is a Ser/Thr kinase family implicated in cell adaptation to glucose deprivation (Carlson 1999). In Planta, it has been suggested the implication of SNF1-related protein kinase (SnRK1) in sugar signaling and in the control of carbohydrate and starch metabolism (Thelander et al. 2004). More recently, Gissot et al. (2006) demonstrated the existence of interactions in the cytosol between the catalytic subunit AKIN$\beta\gamma$ (restricted to the plant kingdom) of SNF1-related protein kinases and two leucine-rich repeats related to pathogen resistance proteins. To the best of our knowledge, there is no previous report of SNF1-Ser/Thr PKinase, as integral part of an NBS-type *R* protein. Therefore, we hypothesize that NBS proteins containing this domain play a role in metabolic responses to nutritional and environmental stresses, in addition to plant–pathogen interactions. We have found that HvNBS89, possessing SNF1-Ser/Thr Protein Kinase

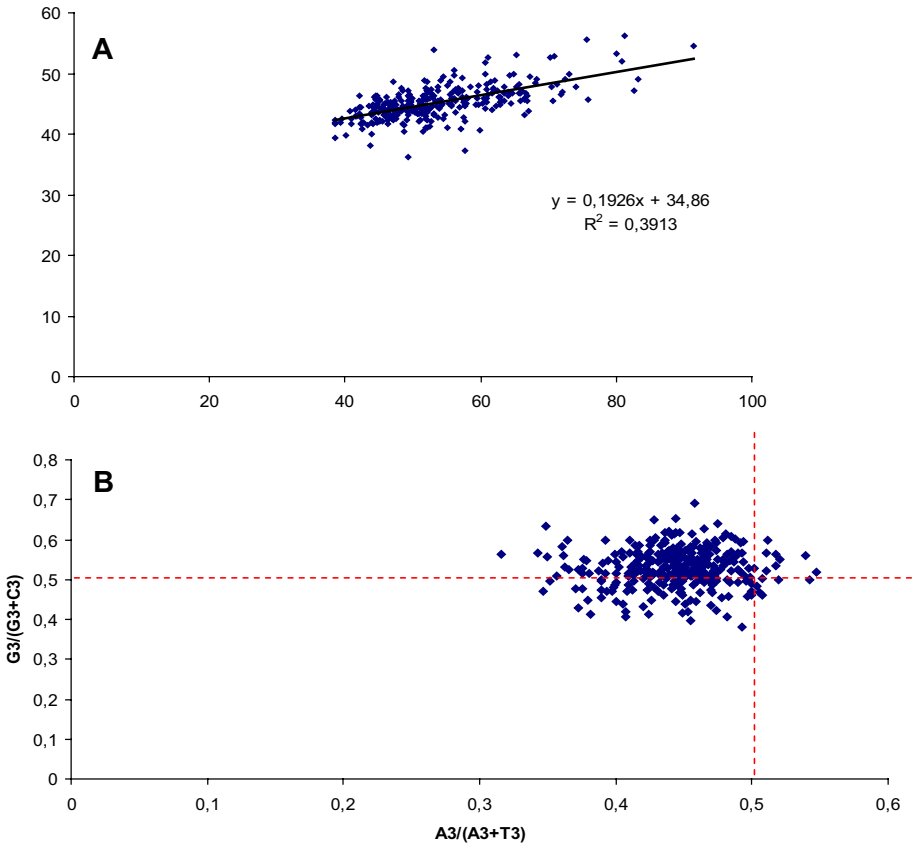


Fig. 3 Impact of mutation and natural selection on codon usage preference. In the neutrality plot (**a**), the weak correlation between GC12 and GC3 denotes the weak impact of mutational bias in favour of selective bias. In the bias plot of the third codon position (**b**), the disproportionate use of GC or AT, confirms the major role played by natural selection in shaping the synonymous codons bias

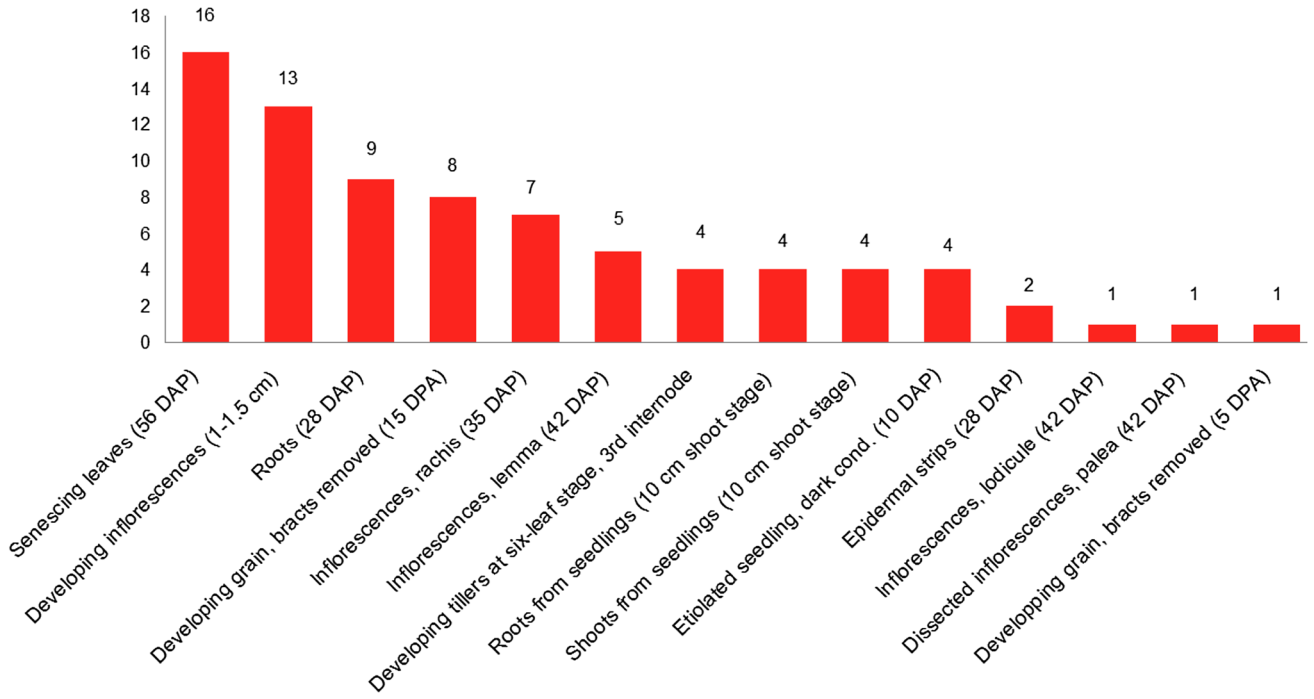


Fig. 4 Distribution of "maximum expression" among barley tissues and development stages, for 87 barley NBS candidate genes identified in this study, showing sufficient homology with the barley genes in the barley genome explorer (Barlex) (http://apex.ipk-gatersleben.de/apex/f?p=284:10:4747726930277)

might be expressed in any tissue of plant. The existence of these domains in association with the NBS one strongly suggests a great functional divergence and versatility of this gene family in barley. In fact, a number of recent studies (Ellis 2016; Krattinger and Keller 2016) demonstrated that NBS proteins (NLRs) with non-canonical domain architectures provide new molecular leads for improving disease resistance in plants. These composite immune receptors are thought to arise from fusions between NLRs and additional domains that serve as "baits" for the pathogen-derived effector proteins, thus enabling pathogen recognition (Sarris et al. 2016). Several names have been proposed to describe these proteins, including "integrated decoys", "integrated sensors" and NLR-IDs.

Eighty-five (85) NBS-encoding genes were mapped onto barley chromosomes (88.5%), showing a relatively even distribution among the chromosomes (with the exception of chromosome 4H), but clearly contrasted within each chromosome. Nearly the quarter (22.35%) of these genes reside in a clustered manner, which is thought to facilitate the evolution of *R* genes through tandem duplication and gene recombination, while genes that are not clustered together could be more stable and may play the role of pioneers, establishing new locations for future clusters (Friedman and Baker 2007). These results are similar to those reported by Andersen et al. (2016) on *H. vulgare* cv. Morex. In both studies, the maximal number of genes was mapped on chromosome 7H, genes tended to be located in the extra-pericentromeric regions of chromosomes, and most clusters were composed of only two genes, with none detected on chromosome 4H.

Unlike dicots where previous phylogenetic analyses of NBS domain sequences revealed two distinct clades: TNLs and non-TNLs, similar phylogenetic analyses conducted in monocots often give rise to star-like topologies. Similar to rice (Zhou et al. 2004) and maize (Cheng et al. 2012), the phylogenetic tree for the barley NBS-encoding genes produced many radiating branches that lacked resolution and did not indicate relationships or classification schemes among the NBS-encoding genes.

In genetics, the Ka/Ks ratio is used as an indicator of selective pressure acting on a protein-coding gene. In our study, all Ka/Ks ratios were less than one, indicating a prevalence of purifying selection. This result is the same as that previously reported by Anderson et al. (2016) in NBS genes of barley cv. Morex. Even though purifying selection was detected, a rapid evolutionary pattern of barley NBS genes could be evidenced from their average Ka/Ks ratio (0.385) that was larger than the genome-wide Ka/Ks value, suggesting a purifying selection that is weaker in NBS sequences, in comparison with the whole barley genome. This fact could be explained as one of the strategies used by the plant species to protect itself against rapidly evolving pathogens. Ka/Ks values are widely variable among species. In NBS genes that possess an LRR C-terminal domain, the LRR region is assumed to mainly recognize the pathogen effectors (Yoshimura et al. 1998). For this reason, it is often believed to undergo positive selection. Here, the Ka/Ks ratios of full CDS were evaluated, because nearly third of the studied genes (31/96) are lacking the LRR domain. Different plants, with different life histories (e.g. ploidy levels) and different affected pathogens, might be driven by specific ecological environments into diverse evolution of NBS genes. For example, in cultivated bread wheat, Gu et al. (2015) have reported an average Ka/Ks ratio, for the full NBS genes CDS, of 0.532, which is higher than the value we report for barley. The higher value in wheat, together with the presence of several genes with $Ks = 0$, and high proportion of present/absent NBS-encoding genes between bread wheat and its diploid progenitors indicate that the NBS–LRR genes evolved very rapidly after natural polyploidization, and that their expansion was influenced by artificial selection. In sorghum, the reported ratio of Ka/Ks for complete CDS domain (0.39) was close to barley (Yang and Wang 2016).

A number of factors have been proposed to explain the mechanism of codon usage bias. Natural selection and mutation are two typical and recognized hypotheses (Duret 2002). Natural selection occurs in highly expressed genes, such as translation factors and ribosomal proteins, to ensure efficient and/or accurate translation. Further, selection-mutation-drift theory has been proposed as the model under which codons are used in the genome (Duret 2002). This model proposes that selection favours the optimal codon over minor codons, while mutational pressure and genetic drift allow the minor codons to persist. Codon usage bias has been documented in a wide variety of species, such as *B. distachyon* (Liu et al. 2012), *Populus Tremula* (Ingvarsson 2010), rice (Liu et al. 2004) and maize (Liu et al. 2010). In these species, the relative contributions of mutation and various forms of natural selection, on codon usage bias, were found greatly different. In barley, our findings revealed a weak impact of mutational bias in favour of selective bias, which confirmed the tendency to expansion evidenced from Ka/Ks.

Eighty-seven (87) out of 96 HvNBSs were supported by expression evidence, and exhibited distinct expression patterns in different tissues and organs. One explanation is that barley NBS-encoding genes may have temporal and spatial expression patterns, which vary by tissue type, developmental stage or genotype. For example, HvNBS68 exhibited a tissue-specific expression pattern in shoots from seedlings, which suggests that it may function only in resistance to a disease of the roots. Conversely, some genes (e.g. HvNBS7, HvNBS20, HvNBS37 and HvNBS43) lacked any tissue-specificity (Supplementary data Table S6), which may suggest that these proteins act somehow as "gatekeeper" *R* proteins conferring pleiotropic effects. Could such genes

act as "helper NBS–LRRs", mediating signal transduction downstream of various different NBS–LRR receptors for activation during effector-triggered immunity, as previously made evident for ADR1 family in *A. thaliana* (Bonardi et al. 2011)?

Furthermore, the expression data revealed that the most duplicated NBS-encoding gene groups exhibited diverse expression patterns among members. For example, HvNBS71 and HvNBS78 belong to the same duplicated gene family, but are characterized by different expression profiles (Supplementary data Table S6). These observations strongly suggest that the functional diversification of the retained duplicated genes is a major feature of the long-term evolution of this gene family in barley, as previously hypothesized in *A. thaliana* (Blanc and Wolfe 2004).

In plants, miRNAs regulate a number of fundamental functions, such as organogenesis, meristem development, seed development and germination, leaf and flower morphogenesis, signal transduction, hormone interaction and response to environmental stresses (Guo et al. 2005; Nikovics et al. 2006; Das et al. 2015; Bai et al. 2017). Disease resistance NBS gene family has been shown to be targeted by multiple, independent miRNA families (Zhai et al. 2011; Zhu et al. 2013; Li et al. 2012; Habachi-Houimli et al. 2016), and there is increasing evidence that small RNAs are involved in regulating plant immunity (Fei et al. 2016). Because the expression profiles of plant *R* genes are often mediated by factors other than pathogen infection, such as tissue type, developmental stage, or environmental conditions (Collins et al. 1999), miRNA-mediated spatiotemporal regulation of *R* gene expression greatly enhances the optimization of plant defense responses in terms of reducing fitness costs of defense signaling. For this reason, it was very important to explore, in our study, the influence of microRNAs on barley NBS genes. Bioinformatics analyses conducted in the present study, using psRNATarget algorithm, showed that 14.58% (14 out of 96) mRNAs of barley NBS genes represented potential targets of miR482/2118 and HvmiRs. This number is lower than that reported in sorghum (37.5%) and rice (36.4%) (Zhang et al. 2014). There is a tight association between the diversity and evolution of plant NBS–LRRs and miRNAs (Zhang et al. 2016), suggesting that such a lower proportion of miRNAs targeting disease-resistance genes in barley may be associated with a higher amount of gene retention and neo-functionalization following gene duplication, particularly through escaping expression repression exerted by miRNAs. Recently, Bai et al. (2017), using small RNA sequencing, have identified a large number of miRNA-target pairs in barley seeds. Of those, 16 known miRNAs with 40 target genes, and 3 novel miRNAs with 4 target genes were confirmed based on degradome sequencing data. Two miRNA-target pairs, namely miR7757/MLOC_17471 and miR9863/MLOC_24045, corresponded to microRNAs targeting barley NBS–LRR putative disease-resistance genes, suggesting that miRNAs contribute to the control of plant defense, notably through the regulation of *R* gene expression. Our present study identified 14 miRNA-*R* gene target pairs, providing a useful source of information that is likely to be further confirmed through real-time quantitative reverse transcription (qRT) PCR analysis of miRNAs and their target levels under specific biotic stresses.

Conclusion

The results of this study provide a genomic framework for the exploitation of candidate NBS-encoding genes in barley, and contribute to a deeper understanding of their evolutionary mode in the contexts of Poaceae, monocotyledons and Planta. The observed diversity of barley NBS proteins indicates the variety of recognition molecules available to detect the diverse biotic challenges to this crop species. By in-depth experimental validation of the expression of priorized candidates, we expect convergent evidence for the discovery of genes, which will be relevant to a more durable pathogen resistance in barley.

Acknowledgements We gratefully acknowledge Abdennour Sébéi (Centre Régional de Recherches sur les Grandes Cultures, CRRGC, Béja, Tunisia), for his permanent logistic help. This study was financially supported by the Tunisian Ministry of Higher Education and Scientific Research.

Author contributions YHH analyzed data using bioinformatics tools, discussed the results and co-wrote the manuscript, under the guidance of DB. YK helped in performing bioinformatics analyses. MMK, HM, and MM helped to perform this study and contributed to discussions. DB designed this study, guided YHH in all analyses and discussions, and co-wrote the manuscript. All authors have read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Abdel-Ghani AH, Al-Ameiri NS, Karajeh MR (2008) Resistance of barley landraces and wild barley populations to powdery mildew in Jordan. Phytopathol Mediterr 47:92–97

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Andersen EJ, Ali S, Reese RN, Yen Y, Neupane S, Nepal MP (2016) Diversity and evolution of disease resistance genes in Barley (*Hordeum vulgare* L.). Evol Bioinform 12:99–108. https://doi.org/10.4137/EBO.S38085

Badr A, Müller K, Schäfer-Pregl R, El Rabey H, Effgen S, Ibrahim HH, Pozzi C, Rohde W, Salamini F (2000) On the origin and domestication history of Barley (*Hordeum vulgare*). Mol Biol Evol 17:499–510

Bai B, Shi B, Hou N, Cao Y, Meng Y, Bian H, Zhu M, Han N (2017) microRNAs participate in gene expression regulation and phytohormone cross-talk in barley embryo during seed development and germination. BMC Plant Biol 17:150. https://doi.org/10.1186/s12870-017-1095-2

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: Proceedings of the second international conference on intelligent systems for molecular biology, AAAI Press, Menlo Park, pp 28–36

Biruma M, Martin T, Fridborg I, Okori P, Dixelius C (2012) Two loci in sorghum with NB-LRR encoding genes confer resistance to *Colletotrichum sublineolum*. Theor Appl Genet 124:1005–1015

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell 16:1679–1691

Bonardi V, Tang S, Stallmann A, Roberts M, Cherkis K, Dangl JL (2011) Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. Proc Natl Acad Sci USA 108:16463–16468

Borrelli GM, Mazzucotelli E, Marone D, Crosatti C, Michelotti V, Valè G, Mastrangelo AM (2018) Regulation and evolution of NLR genes: a close interconnection for plant immunity. Int J Mol Sci 19:1662. https://doi.org/10.3390/ijms19061662

Buschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J, Topsch S, Vos P, Salamini F, Schulze-Lefert P (1997) The barley Mlo gene: a novel control element of plant pathogen resistance. Cell 88:695–705

Cantalapiedra CP, Boudiar R, Casas AM, Igartua E, Contreras-Moreira B (2015) BARLEYMAP: physical and genetic mapping of nucleotide sequences and annotation of surrounding loci in barley. Mol Breed 35:13. https://doi.org/10.1007/s11032-015-0253-1

Carlson M (1999) Glucose repression in yeast. Curr Opin Microbiol 2:202–207

Cheng X, Jiang H, Zhao Y, Qian Y, Zhu S, Cheng B (2010) A genomic analysis of disease-resistance genes encoding nucleotide binding sites in *Sorghum bicolor*. Genet Mol Biol 33:292–297

Cheng Y, Li X, Jiang H, Ma W, Miao W, Yamada T, Zhang M (2012) Systematic analysis and comparison of nucleotide-binding site disease resistance genes in maize. FEBS J 279:2431–2443

Chisholm ST, Coaker G, Day B, Staskawicz BJ (2006) Host–microbe interactions: shaping the evolution of the plant immune response. Cell 124:803–814

Colmsee C, Beier S, Himmelbach A, Schmutzer T, Stein N, Scholz U, Mascher M (2015) BARLEX—the Barley draft genome explorer. Mol Plant 8:964–966. https://doi.org/10.1016/j.molp.2015.03.009

Collins N, Drake J, Ayliffe M, Sun Q, Ellis J, Hulbert S, Pryor T (1999) Molecular characterization of the maize Rp1-D rust resistance haplotype and its mutants. Plant Cell 11:1365–1376

Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. https://doi.org/10.1093/nar/gkr319

Dangl JL, Jones JDG (2001) Plant pathogens and integrated defense responses to infection. Nature 411:826–833

Das SS, Karmakar P, Nandi AK, Sanan-Mishra N (2015) Small RNA mediated regulation of seed germination. Front Plant Sci 6:828. https://doi.org/10.3389/fpls.2015.00828.eCollection2015

Duret L (2002) Evolution of synonymous codon usage in metazoans. Curr Opin Genet Dev 12:640–649

Ellis JG (2016) Integrated decoys and effector traps: how to catch a plant pathogen. BMC Biol 14:13. https://doi.org/10.1186/s12915-016-0235-8

Fei Q, Zhang Y, Xia R, Meyers BC (2016) Small RNAs add zing to the zig-zag-zig model of plant defenses. Mol Plant Microbe Interact 29:165–169

Flor HH (1971) Current status of the gene-for-gene concept. Ann Rev Phytopathol 9:275–296

Friedman AR, Baker BJ (2007) The evolution of resistance genes in multiprotein plant resistance systems. Curr Opin Genet Dev 17:493–499

Gissot L, Polge C, Jossier M, Girin T, Bouly J-P, Kreis M, Thomas M (2006) AKINβγ contributes to SnRK1 heterotrimeric complexes and interacts with two proteins implicated in plant pathogen resistance through its KIS/GBD sequence. Plant Physiol 142:931–944

Gu Z, Cavalcanti A, Chen FC, Bouman P, Li WH (2002) Extent of gene duplication in the genomes of *Drosophila*, nematode and yeast. Mol Biol Evol 19:256–262

Gu L, Si W, Zhao L, Yang S, Zhang X (2015) Dynamic evolution of NBS–LRR genes in bread wheat and its progenitors. Mol Genet Genom 290:727–738

Guo HS, Xie Q, Fei JF, Chua NH (2005) MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for *Arabidopsis* lateral root development. Plant Cell 17:1376–1386

Halterman DA, Wise RP (2004) A single-amino acid substitution in the sixth leucine-rich repeat of barley MLA6 and MLA13 alleviates dependence on RAR1 for disease resistance signalling. Plant J 38:215–226

Holub EB (2001) The arms race is ancient history in *Arabidopsis*, the wildflower. Nat Rev Genet 2:516–527

Habachi-Houimli Y, Khalfallah Y, Makni H, Makni M, Bouktila D (2016) Large-scale bioinformatic analysis of the regulation of the disease resistance NBS gene family by microRNAs in Poaceae. C R Biol 339:347–356

Ingvarsson PK (2010) Natural selection on synonymous and nonsynonymous mutations shapes patterns of polymorphism in *Populus tremula*. Mol Biol Evol 27:650–660

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800. https://doi.org/10.1038/nature03895

Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, Appels R, Pfeifer M, Tao Y, Zhang X, Jing R, Zhang C, Ma Y, Gao L, Gao C, Spannagl M, Mayer KFX, Li D, Pan S, Zheng F, Hu Q, Xia X, Li J, Liang Q, Chen J, Wicker T, Gou C, Kuang H, He G, Luo Y, Keller B, Xia Q, Lu P, Wang J, Zou H, Zhang R, Xu J, Gao J, Middleton C, Quan Z, Liu G, Wang J, International Wheat Genome Sequencing Consortium, Yang H, Liu X, He Z, Mao L, Wang J (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91–95

Joshi RJ, Nayak S (2013) Perspectives of genomic diversification and molecular recombination towards *R*-gene evolution in plants. Physiol Mol Biol Plants 19:1–9. https://doi.org/10.1007/s12298-012-0138-2

Kawabe A, Miyashita NT (2003) Patterns of codon usage bias in three dicot and four monocot plant species. Genes Genet Syst 78:343–352

Krattinger SG, Keller B (2016) Trapping the intruder—immune receptor domain fusions provide new molecular leads for improving disease resistance in plants. Genome Biol 17:23. https://doi.org/10.1186/s13059-016-0891-6

Li J, Ding J, Zhang W, Zhang Y, Tang P, Chen J-Q, Tian D, Yang S (2010) Unique evolutionary pattern of numbers of gramineous NBS–LRR genes. Mol Genet Genom 283:427–438

Li F, Pignatta D, Bendix C, Brunkard JO, Cohn MM, Tung J, Sun H, Kumar P, Bakker B (2012) MicroRNA regulation of plant innate immune receptors. Proc Natl Acad Sci USA 109:1790–1795

Ling HQ, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y, Gao C, Wu H, Li Y, Cui Y, Guo X, Zheng S, Wang B,

Yu K, Liang Q, Yang W, Lou X, Chen J, Feng M, Jian J, Zhang X, Luo G, Jiang Y, Liu J, Wang Z, Sha Y, Zhang B, Wu H, Tang D, Shen Q, Xue P, Zou S, Wang X, Liu X, Wang F, Yang Y, An X, Dong Z, Zhang K, Zhang X, Luo MC, Dvorak J, Tong Y, Wang J, Yang H, Li Z, Wang D, Zhang A, Wang J (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496:87–90

Liu Q, Feng Y, Zhao X, Dong H, Xue Q (2004) Synonymous codon usage bias in *Oryza sativa*. Plant Sci 167:101–105

Liu H, He R, Zhang H, Huang Y, Tian M, Zhang J (2010) Analysis of synonymous codon usage in *Zea mays*. Mol Biol Rep 37:677–684

Liu H, Huang Y, Du X, Chen Z, Zeng X, Chen Y, Zhang H (2012) Patterns of synonymous codon usage bias in the model grass *Brachypodium distachyon*. Genet Mol Res 11:4695–4706

Loveland JE, Gilbert JG, Griffiths E, Harrow JL (2012) Community gene annotation in practice. J Biol Databases Curation 2012:bas009. https://doi.org/10.1093/database/bas009

Lozano R, Hamblin MT, Prochnik S, Jannink JL (2015) Identification and distribution of the NBS–LRR gene family in the *Cassava* genome. BMC Genom 16:360. https://doi.org/10.1186/s12864-015-1554-9

Lukasik E, Takken FL (2009) STANDing strong, resistance proteins instigators of plant defence. Curr Opin Plant Biol 12:427–436

Luo Y, Caldwell KS, Wroblewski T, Wright ME, Michelmore RW (2009) Proteolysis of a negative regulator of innate immunity is dependent on resistance genes in tomato and *Nicotiana benthamiana* and induced by multiple bacterial effectors. Plant Cell 21:2458–2472

Marone D, Russo MA, Laidò G, De Leonardis AM, Mastrangelo AM (2013) Plant nucleotide binding site-leucine-rich repeat (NBS–LRR) genes: active guardians in host defense responses. Int J Mol Sci 14:7302–7326

Martin T, Biruma M, Fridborg I, Okori P, Dixelius C (2011) A highly conserved NB-LRR encoding gene cluster effective against *Setosphaeria turcica* in sorghum. BMC Plant Biol 11:151. https://doi.org/10.1186/1471-2229-11-151

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Yong ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within nucleotide-binding superfamily. Plant J 20:317–332

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS–LRR-encoding genes in *Arabidopsis*. Plant Cell 15:809–834

Meyers BC, Kaushik S, Nandety RS (2005) Evolving disease resistance genes. Curr Opin Plant Biol 8:129–134

Mirlohi A, Brueggeman R, Drader T, Nirmala J, Steffenson BJ, Kleinhofs A (2008) Allele sequencing of the barley stem rust resistance gene Rpg1 identifies regions relevant to disease resistance. Phytopathology 98:910–918

Mun JH, Yu HJ, Park S, Park BS (2009) Genome-wide identification of NBS-encoding resistance genes in *Brassica rapa*. Mol Genet Genom 282:617–631

Naqvi AR, Choudhury NR, Mukherjee SK, Haq QM (2011) In silico analysis reveals that several tomato microRNA/microRNA* sequences exhibit propensity to bind to tomato leaf curl virus (ToLCV) associated genomes and most of their encoded open reading frames (ORFs). Plant Physiol Biochem 49:13–17

Nevo E (1992) Origin, evolution, population genetics, and resources for breeding of wild barley, *Hordeum spontaneum*, in the fertile crescent. In: Showrey PR (ed) Barley, genetics, biochemistry, molecular biology and biotechnology. C.A.B. International, Wallingford, pp 19–43

Nevo E, Fu Y-B, Pavlicek T, Khalifa S, Tavasi M, Beiles A (2012) Evolution of wild cereals during 28 years of global warming in Israel. Proc Natl Acad Sci USA 109:3412–3415

Nikovics K, Blein T, Peaucelle A, Ishida T, Morin H, Aida M, Laufs P (2006) The balance between the MIR164A and CUC2 genes controls leaf margin serration in *Arabidopsis*. Plant Cell 18:2929–2945

Nobuta K, Ashfield T, Kim S, Innes RW (2005) Diversification of non-TIR class NB-LRR genes in relation to whole-genome duplication events in *Arabidopsis*. Mol Plant Microbe Interact 18:103–109

Pan Q, Wendel J, Fluhr R (2000) Divergent evolution of plant NBS–LRR resistance gene homologues in dicot and cereal genomes. J Mol Evol 50:203–213

Porter BW, Paidi M, Ming R, Alam M, Nishijima WT, Zhu YJ (2009) Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. Mol Genet Genom 281:609–626

Ruane J, Sonnino A (2007) Marker-assisted selection as a tool for genetic improvement of crops, livestock, forestry and fish in developing countries: an overview of the issues. In: Guimaraes E, Ruane J, Sonnino A, Scherf B, Dargie JD (eds) Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish. Food and Agriculture Organization of the United Nations, Rome, ISBN: 978-92-5-105717-9

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sarris PF, Cevik V, Dagdas G, Jones JDG, Krasileva KV (2016) Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. BMC Biol 14:8. https://doi.org/10.1186/s12915-016-0228-7

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Schulte D, Close TJ, Graner A, Langridge P, Matsumoto T, Muehlbauer G, Sato K, Schulman AH, Waugh R, Wise RP, Stein N (2009) The International Barley Sequencing Consortium—at the threshold of efficient access to the barley genome. Plant Physiol 149:142–147

Seeholzer S, Tsuchimatsu T, Jordan T, Bieri S, Pajonk S, Yang W, Jahoor A, Shimizu KK, Keller B, Schulze-Lefert P (2010) Diversity at the Mla powdery mildew resistance locus from cultivated barley reveals sites of positive selection. Mol Plant Microbe Interact 23:497–509

Shivaprasad PV, Chen HM, Patel K, Bond DM, Santos BA, Baulcombe DC (2012) A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. Plant Cell 24:859–874

Sueoka N (1999a) Translation-coupled violation of Parity Rule 2 in human genes is not the cause of heterogeneity of the DNA G + C content of third codon position. Gene 238:53–58

Sueoka N (1999b) Two aspects of DNA base composition: G + C content and translation-coupled deviation from intra-strand rule of A = T and G = C. J Mol Evol 49:49–62

Tamura K, Peterson D, Peterson N, Steecher G, NeiMand Kumar S (2011) MEGA: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Tan S, Wu S (2012) Genome wide analysis of nucleotide-binding site disease resistance genes in *Brachypodium distachyon*. Comp Funct Genom. https://doi.org/10.1155/2012/418208 **(Article ID 418208)**

The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

The International Barley Genome Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716

The International Wheat Genome Sequencing Consortium (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788. https://doi.org/10.1126/science.1251788

Thelander M, Olson MA, Ronne H (2004) Snf1-related protein kinase 1 is needed for growth in a normal day–night light cycle. EMBO J 23:1900–1910

Timmerman-Vaughan GM, Frew TJ, Weeden NF (2000) Characterization and linkage mapping of R-gene analogous DNA sequences in pea (*Pisum sativum* L.). Theor Appl Genet 101:241–247

van der Hoorn RAL, Kamoun S (2008) From guard to decoy: a new model for perception of plant pathogen effectors. Plant Cell 20:2009–2017

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, Zini E, Eldredge G, Fitzgerald LM, Gutin N, Lanchbury J, Macalma T, Mitchell JT, Reid J, Wardell B, Kodira C, Chen Z, Desany B, Niazi F, Palmer M, Koepke T, Jiwan D, Schaeffer S, Krishnan V, Wu C, Chu VT, King ST, Vick J, Tao Q, Mraz A, Stormo A, Stormo K, Bogden R, Ederle D, Stella A, Vecchietti A, Kater MM, Masiero S, Lasserre P, Lespinasse Y, Allan AC, Bus V, Chagne D, Crowhurst RN, Gleave AP, Lavezzo E, Fawcett JA, Proost S, Rouze P, Sterck L, Toppo S, Lazzari B, Hellens RP, Durel CE, Gutin A, Bumgarner RE, Gardiner SE, Skolnick M, Egholm M, Van de Peer Y, Salamini F, Viola R (2010) The genome of the domesticated apple (Malus × domestica Borkh). Nat Genet 42:833–839

Wan H, Yuan W, Bo K, Shen J, Pang X, Chen J (2013)) Genome-wide analysis of NBS-encoding disease resistance in *Cucumis sativus* and phylogenetic study of NBS-encoding genes in Cucurbitaceae crops. BMC Genom 14:109. https://doi.org/10.1186/1471-2164-14-109

Wicker T, Taudien S, Houben A, Keller B, Graner A, Platzer M, Stein N (2009) A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. Plant J 59:712–722

Wright F (1990) The effective number of codons used in a gene. Gene 87:23–29

Yandell M, Ence D (2012) A beginner's guide to eukaryotic genome annotation. Nat Rev Genet 13:329–342

Yang X, Wang J (2016) Genome-wide analysis of NBS–LRR genes in sorghum genome revealed several events contributing to NBS–LRR gene evolution in grass species. Evol Bioinform Online 12:9–21

Yang SH, Zhang XH, Yue JX, Tian DC, Chen JQ (2008) Recent duplications dominate NBS-encoding gene expansion in two woody species. Mol Genet Genom 280:187–198

Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang ZX, Kono I, Kurata N, Yano M, Iwata N, Sasaki T (1998) Expression of Xa1 a bacterial blight-resistance gene in rice is induced by bacterial inoculation. Proc Natl Acad Sci USA 95:1663–1668

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science 296:79–92

Zeng XQ, Luo XM, Wang YL, Xu QJ, Bai LJ, Yuan HJ, Tashi N (2014) Transcriptome sequencing in a Tibetan barley landrace with high resistance to powdery mildew. Sci World J. https://doi.org/10.1155/2014/594579 **(ID 594579)**

Zhai J, Jeong DH, De Paoli E, Park S, Rosen BD et al (2011) MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. Genes Dev 25:2540–2553

Zhang R, Murat F, Pont C, Langin T, Salse J (2014) Paleo-evolutionary plasticity of plant disease resistance genes. BMC Genom 15:187

Zhang Y, Xia R, Kuang H, Meyers BC (2016) The diversification of plant defense genes directs the evolution of microRNAs that target them. Mol Biol Evol 33:2692–2705

Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D (2004) Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS–LRR genes. Mol Genet Genom 271:402–415

Zhu Q-H, Fan L, Liu Y, Xu H, Llewellyn D, Wilson I (2013) miR482 regulation of *NBS–LRR* defense genes during fungal pathogen infection in cotton. PLoS One 8:e84390. https://doi.org/10.1371/journal.pone.0084390

Zipfel C (2008) Pattern-recognition receptors in plant innate immunity. Curr Opin Immunol 20:10–16