Long noncoding RNAs in the model species *Brachypodium distachyon*

Concetta De Quattro¹, Mario Enrico Pè² & Edoardo Bertolini¹,²

Eukaryotic genomes are pervasively transcribed and only a small portion of the transcribed sequences belongs to protein coding genes. High-throughput sequencing technology contributed to consolidate this perspective, allowing the identification of numerous noncoding RNAs with key roles in biological processes. Long noncoding RNAs (lncRNAs) are transcripts longer than 200 nt with limited phylogenetic conservation, expressed at low levels and characterized by tissue/organ specific expression profiles. Although a large set of lncRNAs has been identified, the functional roles of lncRNAs are only beginning to be recognized and the molecular mechanism of lncRNA-mediated gene regulation remains largely unexplored, particularly in plants where their annotation and characterization are still incomplete.

Using public and proprietary poly-(A)+ RNA-seq data as well as a collection of full length ESTs from several organs, developmental stages and stress conditions in three *Brachypodium distachyon* inbred lines, we describe the identification and the main features of thousands IncRNAs. Here we provide a genome-wide characterization of IncRNAs, highlighting their intraspecies conservation and describing their expression patterns among several organs/tissues and stress conditions. This work represents a fundamental resource to deepen our knowledge on long noncoding RNAs in C₃ cereals, allowing the *Brachypodium* community to exploit these results in future research programs.

In the past decade the complexity of eukaryotic transcriptomes, by which genomic regions are largely transcribed into RNAs and give rise to processed and regulated coding and noncoding transcripts, has been revealed¹. The advent of high-throughput sequencing technologies and computational methods played a crucial role in boosting the annotation of a large number of transcripts, including those rare sequences that could not be detected using other methods, including array-based approaches².

Noncoding RNAs (ncRNAs) are a broad class of molecules, accounting for 90% of the genome, and are ubiquitous components of the transcriptomes³. Based on their biological roles, ncRNAs have been divided into: (i) structural ncRNAs (ribosomal RNAs, transfer RNAs, small nuclear RNAs and small nucleolar RNAs) and (ii) regulatory ncRNAs, better known as small ncRNAs (microRNA and small interfering RNAs) and long ncRNAs (lncRNAs)⁵. Although small regulatory RNAs have received large attention in the past decade (miRBase version 21 contains more than 28,000 entries)⁶, lncRNAs have been poorly studied, especially in plants, where few species have been investigated at genomic level so far: *Arabidopsis thaliana*⁷, ⁸, *Oryza sativa*⁹, *Zea mays*¹⁰, ¹¹, *Gossypium* ssp¹², *Populus trichocarpa*¹³ and *Solanum lycopersicum*¹⁴.

LncRNAs are generally defined as a heterogeneous family of long transcripts greater than 200 nucleotides likely transcribed from RNA polymerase (Pol) II and subjected to 5′ capping, 3′ polyadenylation and splicing events, although plant specific Pol IV and Pol V have been also reported to be associated with lncRNAs¹⁵. In addition, Wu *et al.*¹⁶ described Pol III transcribed lncRNAs in *Arabidopsis thaliana*. Moreover, even if the main feature of lncRNAs is the lack of evident open reading frames (ORFs), a large fraction of IncRNAs has been shown to be associated with ribosomes in ribosome profiling experiments, showing the potentiality to code for small functional peptides¹⁷. The features to code for regulatory peptides were also found in the primary transcripts of plant microRNAs, revolutionizing the common idea of coding transcripts¹⁸. Recently a newly discovered class of endogenous noncoding RNAs, termed circular RNAs (circRNAs), has been described in animal and plant species, adding to the transcriptome complexity¹⁹.

In general, depending on their genomic origin, IncRNAs are classified in long intergenic ncRNAs (lincRNAs) or longgenic ncRNAs, which are further subgrouped in exonic ncRNAs, intronic ncRNAs (incRNAs) and natural antisense transcripts (NATs), transcribed from the complementary DNA strand of the associated coding genes²⁰, ²¹. Because of their mRNA-like features, lncRNAs are the most abundant class of eukaryotic IncRNAs found in poly(A)+ RNA-seq data²².

¹Institute of Life Sciences, Scuola Superiore Sant’Anna, Piazza Martiri della Libertà 33, 56127, Pisa, Italy. ²Present address: Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, MO, 63132, USA. Correspondence and requests for materials should be addressed to E.B. (email: ebertolini@danforthcenter.org)
The mechanisms of action of long noncoding RNAs are not fully understood yet, but several studies in animals have shown a broad range of action23. The spatio-temporal expression patterns linked to the high tissue/organ specificity of these regulatory molecules, together with the ability to regulate gene expression both through cis and trans-acting mechanisms, make IncRNAs more plastic and capable to act within a broad range of biological pathways14,29.

A large number of IncRNAs were found to be transcribed within the plant genomes26 but few clear functional examples of gene regulation mediated by IncRNAs have been described. In Arabidopsis thaliana, COOLAIR and COLDAIR were shown to regulate the expression of the FLOWERING LOCUS C (FLC) during vernalization27,28. In particular, COOLAIR, a natural antisense transcript, causes a transient transcriptional silencing of FLC in early cold exposure, whereas COLDAIR, an intronic IncRNA, is transcribed from the intronic region of the FLC locus and acts repressing FLC expression. This determines a stable change at FLC chromatin level through the association with Polycomb Repressive Complex 2 (PRC2)28. Also in Arabidopsis the first example of target mimicry, INDUCED BY PHOSPHATE STARVATION 1 (IPS1), was discovered and shown to act as a decoy of miR399. This leads to the attenuation of the post-transcriptional repression of the target gene PHOSPHATE 2 (PHO2)29. In cereals, a clear example of IncRNAs implicated in the male fertility is the rice LONG DAY SPECIFIC MALE FERTILITY ASSOCIATED RNA (LDMAR) that is required for pollen development under long day conditions30.

Here we describe for the first time a large catalogue of IncRNAs in Brachypodium distachyon (Bd), produced by analyzing public and proprietary transcriptome data sets that include RNA-seq from 26 experiments carried out in the reference inbred lines Bd21, Bd21-3 and in the divergent line Bd1-1. The reliability of our IncRNA discovery pipeline was also independently confirmed by public full-length ESTs. Our catalogue includes 25,338 Bd bona fide IncRNAs expressed in various organs and tissues, at different developmental stages and in response to biotic stresses. We investigated IncRNAs expression patterns, highlighting organ, tissue and stress-specific expression profiles. We also discuss the regulatory roles of IncRNA-mediated gene regulation, their potential targets and target mimicry activity.

Results

Identification of long noncoding RNAs. Data from twenty-six public and proprietary poly(A)+ RNA-seq libraries deposited and available at the National Centre for Biotechnology Information (NCBI) Sequence Read Archive (SRA)31 were collected (Supplementary Table 1). The quality of the raw reads was assessed and subjected to adaptors removal and quality filter (see Methods). We retained about 1.5 billion high quality reads (Supplementary Table 2) that were used to re-assemble the transcriptome using a genome guided approach. The various steps of the bioinformatics pipeline used to identify bona fide IncRNAs are summarized in the Supplementary Fig. 1. We applied stringent filters based on the main features of IncRNAs currently recognized25,26. By using the program TopHat233 and allowing 2 mismatches we mapped onto the Bd21 reference genome version 2.1.10 654 million and 435 million trimmed reads from Bd21 and Bd21-3 respectively. Similarly, 456 million trimmed reads from the Bd1-1 libraries were aligned to the equivalent re-sequenced genome35. In total we aligned 1.39 billion reads using two iteration mapping steps. Transcripts were re-assembled using Cufflinks and an unique transcriptome for each inbred line was generated using Cuffmerge29 (Supplementary Fig. 1). Overall 77,016; 59,083; 15,764 million reads were trimmed from the Bd21, Bd21-3 and Bd1-1 libraries respectively (Supplementary Files 1, 2 and 3). The complete set of 16,679 full length ESTs produced from several Bd21 tissues37 was also included in our analysis. These four transcripts data sets were subjected to the pipeline for the IncRNAs identification.

After transcripts reconstruction, the full set of sequences of each inbred line was subject to six consecutive filters (Supplementary Fig. 1). Concisely, only transcripts longer than 200 bp and encoding ORF smaller than 100 amino acids were kept; transcripts with protein domain annotated in Pfam database and transcripts with low ability to encode proteins were discarded and structural RNAs transcripts were discarded based on sequences similarity with the Brachypodium housekeeping RNAs deposited in the Rfam database. The selection steps resulted in a final set of putative IncRNAs: 7,252 in Bd21; 3,715 in Bd21-3; 17,179 in Bd1-1 and 858 in Bd21-EST.

To identify small RNAs associated with IncRNAs we mapped the small RNA sequences available at the plant MPSS database38 (http://mpss.udel.edu) and from Bertolini et al.39 to the four sets of putative IncRNAs (Supplementary Fig. 1). We found on average 5% of IncRNAs associated to small RNAs. Thereby, we classified IncRNAs as bona fide IncRNAs that do not share sequence similarity with small noncoding RNAs and IncRNAs associated with small RNAs. Finally, IncRNAs with no counts in all samples were removed from subsequent analyses, resulting in a subset composed by 5,851 in Bd21, 2,681 in Bd21-3 and 15,948 in Bd1-1 (Supplementary Files 4, 5 and 6). We could not apply the selection based on counts level to the 858 IncRNAs found in Bd21-EST (Supplementary File 7), because they were sequenced using the traditional Sanger approach.

To validate the in silico IncRNA sequence reconstruction from short reads RNA-seq data we performed a cluster analysis between the IncRNAs found in Bd21 and the complete set of full length Bd21-EST IncRNAs37. We found 277 EST sequences (32% of the IncRNAs found in Bd21-EST) clustered with the Bd21 EST IncRNAs with a 95% sequence identity (Supplementary Fig. 2 and Supplementary File 8). Moreover, we performed RT-PCR amplifications on nine IncRNAs expressed at different level in six Bd21 tissues/organs, such as: third leaf, leaves 20 days, early inflorescence, emerging inflorescence, seed 5 DAP and seed 10 DAP. For all the 9 randomly selected IncRNAs the desired PCR product was obtained (Supplementary Methods).

Structural and genomic features of long noncoding RNAs. The complete set of long noncoding RNAs identified was examined at sequence and genomic level. We took into account transcripts length, GC content and number of exons in order to highlight possible difference among Bd inbred lines, protein coding genes and plant species.

The majority of Brachypodium IncRNAs clustered in the range between 200 nt and 700 nt in length (Fig. 1a). The IncRNAs median transcript length in Bd21, Bd21-3, Bd1-1 was respectively 458, 621 and 481 nucleotides, with the first and third quartile ranging between 317 and 868 nt and maximum transcript length of 8,581 nt (Fig. 1a). The GC mean content was between 46% and 50% (Fig. 1b). These features were also confirmed in the
pool of lncRNAs identified in Bd21-EST (data not shown). Considering the exons number, a large set of lncRNAs was composed by a single exon with Bd1-1 showing the richest number of monoexonic lncRNAs (97%) (Fig. 1c).

Differently protein coding mRNAs annotated in Bd21 were characterized by longer sequences with a median of 1062 nt, a higher CG bases composition (mean: 53.63%) and a higher number of exons (Fig. 1). We found lncRNAs equally distributed along the five chromosomes, suggesting a pervasive transcription of the genome and a few lncRNAs clustered in centromeric regions (Fig. 2).

Figure 1. Genomic features of Bd lncRNAs and Bd21 reference protein coding transcripts. The plots show transcript length (a), percentage of GC content (b) and exons number (c) of the lncRNAs annotated in Bd21, Bd21-3 and Bd1-1 and of Bd21 reference protein coding transcripts.
Focusing on the reference inbred line Bd21, for which a high quality genome assembly and annotation are available, Bd21 lncRNAs were classified according to their genomic location as long intergenic noncoding RNAs (LncRNAs) (4,487), exonic lncRNAs (319), intronic lncRNAs (67) and putative antisense LncRNAs (611) (Fig. 3).

Transposon elements, tandem repeats and microRNA associated with LncRNAs. Recent studies have shown that transposable elements (TEs) permeate long intergenic noncoding RNAs (LncRNAs). To explore the involvement of TEs in the origin of LncRNAs, we intersected the genomic coordinates of both LncRNAs and annotated TE loci to identify genomic relationship. Based on current TEs classification, we found 466 LncRNAs (7.96%) associated...
with TE sequences, of which the vast majority (80.47%) belonging to class I retrotransposon. LTRs Gypsy (RLG) and LTR Copia (RLC) were the most abundant families for class I and CACTA (DTC) for class II (Fig. 4a).

We also investigated the presence of simple tandem repeats (TRs) within lncRNA sequences, by overlapping the current Brachypodium TRs annotation with lncRNAs coordinates. We found a positive overlap between 331 minisatellite and lncRNAs (Fig. 4b). To further investigate, at the genomic level the relationship between MIR genes and lncRNAs, we interpolated the coordinates of pre-miRNAs with those of lncRNAs and obtained 8 hits corresponding to the hairpins of miR167c, miR172d, miR395c, miR395j, miR399, miR509, miR7737, and miR7744. Interestingly, the conserved miR167d and miR399a were found within the intronic regions of the lncRNAs TCONS_00007390 and TCONS_00016338 respectively, suggesting a spliced intron origin from the long noncoding transcripts (Supplementary Fig. 3). Differently, miR395c and miR395j were found within the monoexonic lncRNA TCONS_00072591 (Supplementary File 18), suggesting a simultaneous transcription of a polycistronic transcript. These findings are in agreement with miRNA biogenesis in animals, where miRNAs were discovered in introns and exons of coding or noncoding genes43.

**LncRNAs conservation among Brachypodium inbred lines.** In this study, we explored the evolutionary conservation of long noncoding RNAs at intraspecific level through sequence similarity searches. Internal lncRNAs

---

**Figure 4.** Bd21 lncRNAs associated with transposable elements (TE) and tandem repeats (TR). (a) TE and lncRNAs: Y axis is lncRNA length in nucleotides, X axis is TE families. DNA Class II: Helitron (DHH) [7]; Hat (DTA) [5]; CACTA (DTC) [60]; PIF/Harbinger (DTH) [11]; Mutator (DTM) [5]; Tc1/Mariner (DTT) [3]. TE Class I: LINE (RIX) [36], Ty1/copia (RLC) [105], Ty3/gypsy (RLG) [198], Unclassified LTR (RLX) [36]. Round brackets indicate TE code based on the classification system proposed by Wicker et al.42. Square brackets show number of lncRNA loci associated with TEs. (b) TR and lncRNAs: Y axis shows number of lncRNAs, X axis shows types of tandem repeats according to the classification by monomer length into: micro (microsatellites: monomer length 2-9); mini (minisatellites: monomer length 10–99), sat (satellites: monomer length ≥100); vntr (variable number tandem repeats: homopolymeric stretches and mixed types).
redundancy in each inbred line was estimated using the program CD-HIT with a sequence similarity cutoff of 95%. This resulted in 5,698, 2,610, 15,639 uniquely expressed lncRNA in Bd21, Bd21-3 and Bd1-1 respectively (Supplementary Files 9, 10 and 11). A similar cutoff was applied to the lncRNAs BD21-EST data set and we found a high level of sequence redundancy (~50% of the transcript sequences), reflecting cloned mRNA abundance.

LncRNAs conservation at intraspecific level was investigated by searching for sequence homology with a 90% identity among the three Bd inbred lines. We found 135 lncRNAs highly conserved among the three inbreds, showing a sequence conservation that spans almost the entire sequence length. Notably, the majority of these transcripts was found in syntenic and collinear genomic locations (Supplementary Fig. 4).

Expression profiles of lncRNAs. We found that Bd lncRNAs were expressed at a low level in all three inbred lines (Fig. 5a). Fifty percent of the lncRNA transcripts showed an expression level greater than 1 read per kilobase per million reads (RPKM) and a maximum expression level at 17,664.4 RPKM (TCONS_00027261) in Bd21 (Fig. 5a). TCONS_00027261 was the most abundant in all Bd21 samples with a minimum TPM of 141.71 and a peak in developing kernels (6957.88 TPM in embryos 25 days after pollination and 3514.14 TPM in endosperm 25 days after pollination) and Pistil (5420.72 TPM) (Supplementary File 19).

To better explore the complexity of the lncRNA transcripts found we did not apply any further filter based on RPKM values since lncRNAs have been shown to be low expressed, extremely heterogeneous in their expression patterns and with short half-life7, 9, 11. In addition, due to the intrinsic nature of lncRNAs, it was shown that large data sets need to be investigated in order to capture the complete constitutive and tissue specific expression of the noncoding transcriptome45. Moreover, some works used a low expression threshold, e.g. 0.3 RPKM to define genes constitutively expressed45. In our data set lncRNAs with RPKM < 0.3 were 301, 394 and 3,218 in Bd21, Bd21-3 and Bd1-1 respectively. In addition some lncRNAs with expression level < 1 RPKM were also retrieved in the Sanger EST collection (see Methods).

Tissue/organ specific expression was assessed in Bd21 and Bd21-3 by computing the Shannon Entropy46, 47 (see Methods). LncRNAs were generally found uniformly expressed in all samples (entropy = 0) but, on average, 10% were specifically detected in only one tissue/organ (entropy > 3 in Bd21 and entropy = 2 in Bd21-3) (Fig. 5b).
Overall, the expression levels of lncRNAs vary significantly in the different samples, from ubiquitous (837 in Bd21) to tissue/organ specific (590 in Bd21), showing divergent and specific degrees of expression (Fig. 5c,d).

In Bd21 the highest number of specificity was observed in anther (44%), in the third growing leaf (32%) and in the endosperm 25 days after pollination (9.76%) (Fig. 5c and Supplementary Fig. 5). Interestingly, we found that lncRNAs specifically expressed in anthers (254) had the highest expression (Fig. 5e).

Concerning Bd21-3 line, shoot samples infected with *Panicum mosaic virus* (PMV) and its satellite virus (SPMV) showed the strong modulation of lncRNA expression during the plant-virus interactions (Fig. 5d). This result highlighted a core set of 1,743 lncRNAs constitutively expressed and a treatment specific set expressed in response to the pathogen infection (78 Mock specific, 73 PMV specific and 121 PMV + SPMV specific) (Fig. 5f and Supplementary Fig. 5).

**Interactions between microRNAs and long noncoding RNAs.** To understand the link between microRNAs and lncRNAs in *Brachypodium distachyon* we investigated the crosstalk between lncRNAs and microRNAs by identifying lncRNAs targeted by miRNAs and lncRNAs acting as miRNA decoy.

We identified miRNAs able to target lncRNAs using the program TargetFinder. A total of 3,753 targets were retrieved. Among these, 25% of the predictions had a target score ≤2 with an almost perfect sequence complementarity, resembling *bona fide* targets (Fig. 6a). Notably, within this subgroup of targets the miRNAs mostly belong to lineage-specific miRNA families (miR1122, miR1135, miR5171, miR5174, miR5175, miR5180, miR5181, miR5183, miR5185, miR7758, miR9493) with the three families miR1122, miR5174 and miR5181 that account for the vast majority of putative lncRNA targets that are characterized by a high degree of conservation in the first 10 nucleotides at the 5′ end of the miRNA target site (Supplementary Figs 6–8). MiR5174 and miR5181 were previously found to originate from repetitive regions rich in heterochromatic 24 nt small RNAs whereas miR1122 was found associated with biotic and abiotic stresses in wheat and barley.

Target mimicry (TM) lncRNAs were predicted using the algorithm psMimic, that led to the identification of 228 TM hits with complementary score ranging between 0 and 4. The majority of miRNA sequences retrieved in this screening belonged to lineage-specific miRNAs and was characterized by the typical central-nucleotide bulge/mismatch surrounding the 10th–11th nucleotides of the miRNA (Fig. 6b), which is the canonical cleavage site catalyzed by ARGONAUTE. We also found TMs related to large miRNA families such as miR156, miR395, miR399 and miR5174. LncRNAs TCONS_00021517 and TCONS_00016924 have sequence complementarity in the TM site with the mature sequence of miR399a,b,c,d (Fig. 7a). miR399 is involved in the phosphate homeostasis and takes part in the systemic signaling pathways to communicate phosphorus availability and demand between shoot and root. In *A. thaliana* the lncRNA IPS1 (*INDUCED BY PHOSPHATE STARVATION 1*) was
found to actively attenuate the expression of miR399, mediating a TM activity and increasing the expression of the phosphate homeostasis target gene PHOSPHATE 2 (PHO2)49. Although we did not find any sequence homology between the *B. distachyon* IncRNAs (TCONS_00021517 and TCONS_00016924) and *A. thaliana* IPS1 along the entire transcript we observed strong conservation in the TM site sequence. Moreover, miR395, which is involved in the regulatory network of sulfate assimilation48, was found in our Bd1-1 data set associated through a TM mechanism with the two IncRNAs: TCONS_54981.1 and TCONS_11459.1 (Fig. 7b).

MiR156, which plays a role in controlling flowering and leaf development59 appeared targeted by several IncRNAs. In particular, the 3’ of miR156g,i was found to be potentially targeted by the TM IncRNA TCONS_00042468 that is highly expressed in Bd21 early inflorescence (44.6579 TPM) and pistil (32.062 TPM). A similar pattern of expression has been highlighted for TCONS_00067507 which shows sequence complementarity with bdi-miR156b-3p (Fig. 7c). This computational approach allowed us to identify a similar TM Bd21-3 and Bd1-1 likely to bind both bdi-miR156g,i-3p and bdi-miR156a,j-5p, and bdi-miR156d,h-3p, respectively (Supplementary Files 16, 17). Similarly, the lineage-specific miR5174f was found associated with a perfect complementarity in the TM region to several IncRNA transcripts that might act to sequestrate the mature miRNA (Fig. 7d).

**Discussion**

In recent years an increasing number of studies has revealed the complexity of the eukaryotic transcriptome and the important role of noncoding RNAs in regulating gene expression60. This complexity has been deeply investigated in animal models and it is now emerging also in plant species due to the advent of high-throughput technology and the extensive application of RNA-seq approaches61–63.

In this work we conducted an extensive annotation of IncRNAs in *Brachypodium distachyon*, using the data from public and proprietary poly(A)+ RNA-seq experiments produced from several tissues/organs in different Bd lines. Overall our study led to the identification of 25,338 bona fide IncRNAs in the genotypes 21, 21-3 and 1-1. IncRNAs were mainly associated to the intrinsic limit of the poly(A)+ RNA-seq libraries, in our study we could not identify IncRNAs transcribed by Pol III, IV, V since they do not contain a poly(A) tail and are mostly enriched in ribo-minus RNA-seq experiments64. Bd IncRNAs were characterized by a median length ranging between 458–621 bp and were mostly monocexonic transcripts. Among the IncRNAs here identified, the large majority was classified as long intergenic RNAs, confirming lincRNA as the most abundant class of long noncoding genes present in poly(A)+ RNA-seq experiments21,22.

Sequencing depth of RNA-seq and fragmented genome assembly lead to the identification of a large number of IncRNA44,65. Thereby, mRNA-seq depth in Bd1-1 and the current draft version of its reference genome35 could in part explain the higher number of IncRNAs found in this divergent inbred line.

According to their functional importance, IncRNAs are expected to be conserved at different levels: sequence, structure, function and expression profiles66. Although at primary sequence level IncRNAs showed a low conservation67, other IncRNAs features such as syntenic relationships, microhomology and secondary structure resulted in higher degree of conservation68,69.

Investigation at primary sequence level among inbred lines suggests a low intraspecific IncRNAs conservation. In fact, although Bd21 and Bd21-3 lines were expected to be phylogenetically closed, only 47% of IncRNA sequences showed high degree of sequence conservation. Natural diversity study based on SSR markers showed Bd1-1 as the most divergent line within the Bd genotypes, whereas Bd21-3 was considered genetically distinct from Bd21 even if both originated in the same location site69. Moreover, current knowledge suggests that IncRNAs diverge rapidly, being subjected to different selective pressure as compared to protein coding genes, accumulating base substitutions and indels. A possible explanation for the lack of extensive conservation is that most IncRNAs could be subjected to exaptive instead of adaptive evolution70.

In vertebrates, transposable elements and tandem repeats elements were shown to be strongly enriched in IncRNAs. This suggests their role in the origin of lineage-specific IncRNAs71. In addition, human TEs, mostly LTR retrotransposons class, were found to pervade lincRNAs, suggesting a role in controlling IncRNAs transcription and specific expression41,72. *Brachypodium* genome is characterized by a relatively low number of retrotransposons44, and class I TEs is principally composed by *Copia* and *Gypsy* elements. Indeed, in our data set we retrieved Bd21 IncRNAs that were mainly associated to *Copia/Gypsy* LTR and to *CACTA* DNA transposons, which have been also found abundant in many *Triticeae* genomes where they are involved in gene regulation73. In this frame, our results could support the novel hypothesis named Repeat Insertion Domains of IncRNAs (RIDL)74, according to which TE-derived fragments of IncRNA act as structural and regulatory domains. Tandem repeats (TRs) occur frequently in many eukaryotic genomes where they are interspersed in different genomic locations, including coding and noncoding genes75. In humans, TRs have been found enriched in genes and regulatory regions participating in transcriptional regulation and development processes48. Although a few studies in plants have shown the highest density of TRs in 5′-UTRs, promoters and intergenic regions76,77, the link between TRs and IncRNAs has not been clearly examined so far. Our work shows the strong association of IncRNAs with minisatellite TR.

Higher eukaryote transcriptomes are diverse and dynamic, with IncRNA loci exceeding those of miRNAs and exhibiting specificity, context and time dependent expression, a shorter half-life and a low expression78,80. We found all those features in the Bd IncRNAs we identified. Bd IncRNA transcripts showed lower abundance than protein coding genes, even though some IncRNAs exhibited extremely high, dynamic and specific expression profiles. Bd male gametophyte was characterized by the greatest number of specific IncRNAs, similarly to rice and maize8.11. Phased small-interfering RNAs (phasiRNAs) originated from noncoding RNA precursors were found abundant in panicle of several grass species, including *Brachypodium*48,81,82. Accordingly, in Bd the expression pattern of anther specific IncRNAs might correlate with the presence of phased loci, being part of the phasiRNA biogenesis machinery in which the synthesis requires both miRNAs and phasiRNA precursors. In particular, Zhai et al.83
highlighted the synthesis of miR2118 dependent 21 nt and miR2275 dependent 24 nt phasiRNAs, putatively implicated as mobile signals in anther development, coordinating cell-type specific expression.

Figure 7. In silico endogenous target mimics and their corresponding microRNAs. (a) Bd IncRNAs target mimic of miR399. The figure shows the conservation of the target mimic binding site between Bd and Arabidopsis IPS1 IncRNA. (b) Bd IncRNAs target mimic of miR395. (c) Bd IncRNAs target mimic of miR156. (d) Bd IncRNAs target mimic of miR5174. Grey regions show sequence complementarity between microRNAs and IncRNAs sequences. The dash line indicates the bulge between IncRNA and microRNA molecules. The genomic coordinates are provided next to each IncRNA.
systems, including mouse, where regulatory RNAs such as microRNAs, siRNAs, Piwi-interacting RNAs and long noncoding RNAs participate in the strict developmental process giving rise to mature spermatozoa5, 85.

MicroRNAs (miRNAs) have been clearly shown to act as post-transcriptional regulators of gene expression, whereas long noncoding RNAs only recently emerged as new regulatory molecules involved in several biological pathways using a plethora of mechanisms46, 47. In plant and animal systems the influence of lncRNAs upon microRNAs, termed target mimicry (TM) or competing endogenous RNA (ceRNA), has been reported in several works8, 29, 54, 88, 89 that showed the ability of noncoding transcripts to serve as endogenous sponge able to sequeste miRNAs. Differently, the influence of microRNAs on gene expression through the targeting of long noncoding RNAs via complementary sequence site is now only emerging90.

The first evidence of the TM was observed in Arabidopsis thaliana, where miR399 is efficiently modulated by IPS1 harboring a complementary sequence site to miR39999. This characteristic, highlighted in other TM lncRNAs, led us to assume that the regulatory mechanism controlling miR399 through TM mechanism could be conserved between dicot and monocot species54, 91. Moreover, the polycistronic miR395 family, involved in the sulfate assimilation64, was found in our Bd1-1 data set associated through a TM mechanism with two lncRNAs. Clustered miRNAs have been shown to be transcribed into polycistronic transcripts encoding homologous miRNAs. In this context, the presence of lncRNAs with miRNA target mimic pairing could potentially compete with the RISC-complex, sequestering the entire miR395 cluster.

Here for the first time we generated a collection of in silico predicted target mimic lncRNAs that are a valuable starting point to investigate miRNA repression based on miRNA target mimicry. In general, we observed that the majority of TM interactions is between lineage-specific miRNAs and lncRNAs, suggesting a convergent evolution of the entire regulatory RNA noncoding component of Brachypodium for the control of specific biological processes. Accordingly, the interaction of lncRNAs, miRNAs and their direct mRNA targets might be linked in regulatory nodes (lncRNA-miRNA-targets) that coordinate gene expression programs at specific developmental stages. With this genome-wide characterization of the Brachypodium lncRNAs component we provide novel clues and tools to speed up the identification and validation of RNA based regulatory nodes in grasses that could be useful targets for biotechnology applications.

Methods

Data sets used in this study. This work is based on the data produced in several RNA-seq experiments carried out in the three Brachypodium distachyon inbred lines Bd21, Bd21-3 and Bd1-1, whose list is reported in the Supplementary Table 1. All data were downloaded from the National Centre for Biotechnology Information (NCBI) Sequence Read Archive (SRA)82. In addition, we included the data from a proprietary RNA-seq library produced from a pool of third leaves from the reference inbred Bd21. As reported in Supplementary Table 1, the public Bd21, Bd21-3 and Bd1-1 poly(A) RNA-seq data include several tissues and organs. Specifically, the Bd21 data set from Davidson et al.92 comprises leaf collected at 20 days after sowing, early and emerging inflorescence, pistil, anther, seed at 5 and 10 days after pollination (DAP), embryo at 25 DAP, endosperm at 25 DAP. The Bd21-3 data set includes two experiments consisting in: pooled libraries of parenchymatic tissue and shoots plants collected 7 days post infection with Panicum mosaic virus78. The Bd1-1 data set comprises a pooled RNA-seq experiment produced from third leaves harvested from 3-week-old plants78. Within the three data set, the samples of leaf collected 20 days after sowing and embryo at 25 DAP included in Bd21 data set have two biological replicates. While each library produced from Bd21-3 infected plants included three mixed biological replicates. We included also full length ESTs dataset derived from several Bd21 tissues/organisms and treatments (seed at germination, leaf at vegetative stage and after flowering, shoot, crown, spikes at flowering and at different stages after pollination, callus and leaf at 2 weeks after germination treated with different stresses and compounds)88 to provide an independent validation to our computational pipelines for lncRNA discovery and to increase our catalogue of Brachypodium lncRNAs. We used the Bd21 version 2.1 genome and Bd1-1 re-sequenced genome, downloaded from Phytozome version 10 (https://phytozome.jgi.doe.gov) as reference genome sequences.

Plant material. Bd21 third leaves were grown as previously described by Verelst et al.93. cDNA libraries were prepared from three independent experiments, and each experiment consisted of a collection of 400 plants. Third leaves were collected after 24 hours from their emergence, stored in RNA later solution (Thermo Fisher Scientific) and successively frozen in liquid nitrogen. Total RNA samples were extracted using the Plant/Fungi total RNA purification kit (Cat. 25800) from Norgen Biotek Corp. Poly(A) RNA-seq libraries were produced according to the Illumina TruSeq RNA library preparation kit and sequenced (50 bp single-read) using Illumina HiSeq 2000. Data are available on SRA BioProject PRJNA386608.

Transcriptome reconstruction. Raw reads were processed by removing the sequencing adapters using the program Cutadapt version 1.2.194 and by filtering low quality reads (Phred score ≥30) with ERNE-FILTER version 1.395. Trimmed reads from the Bd21 and Bd21-3 were mapped against the reference Bd21 genome (version 2.1), whereas reads from the inbred line Bd1-1 were mapped against the re-sequenced genome (version 1)35, using the spliced aligner TopHat31 version 2.0.9. To exploit the ability of using splice site information derived from the first alignments, a second round of mapping was carried as suggested by Cabili et al.96. For each library, the transcriptome was independently re-assembled using Cufflinks 36 version 2.0.9. Subsequently we used Cuffmerge to obtain a unique non-redundant transcriptome for each inbred line. Finally, the reconstructed transcripts sequences were retrieved using the gffread tool.

LncRNAs identification pipeline. LncRNAs were filtered out from the entire collection of assembled transcripts by applying a stringent stepwise filtering pipeline based on the currently established lncRNAs features22, 32. Our pipeline described in De Quattro et al.97 is composed by six consecutive filters: (i) a size cutoff based on the
assumption that lncRNAs are longer than 200 nucleotides; (ii) an open reading frame (ORF) putatively coding for a peptide sequence shorter than 100 amino acids. Since lncRNAs can code for small peptides10, this selection allows to retain a significant level of stringency without losing many potential lncRNAs; (iii) known protein domain identification using protein sequences downloaded from the Pfam database99 version 27 (BlastX with E-value ≤ 0.001) to eliminate transcripts encoding protein; (iv) a coding potential calculator (CPC) to test the protein coding potential of the remaining transcripts100; (v) housekeeping noncoding RNAs to exclude all transcripts homologous to Brachypodium genomic and plastidial tRNA, rRNA, snRNA and snoRNA retrieved from the Rfam database101 version 12 (http://rfam.xfam.org/) and (vi) the precursor of small noncoding RNAs to filter out transcripts with noncoding features associated to small RNAs. For this final step of the pipeline, nineteen Bd21 small RNA-seq data from several tissues, organs and treatments were downloaded from the Plant MPSS database98, plus eight Bd21 small RNA-Seq libraries produced in our laboratory from young developing leaves39. Bona fide lncRNA identified in Bd21, Bd21-3 and Bd1-1 were named using the prefix TCONS.

Validation of Bd21 IncRNA transcripts. The quality of our in silico lncRNA transcripts was assessed using as benchmark ~16,000 full length ESTs generated by Sanger technology37. The CD-HIT108 program version 4.6 was applied to perform a cluster sequence analysis to check sequence identity between lncRNAs and ESTs with a cutoff equal to 95%. Results were further investigated using Dotter109 and mVISTA105.

Classification of IncRNAs. LncRNAs were classified into four main categories based on their location relative to protein coding genes: (i) intergenic; (ii) genic-intronic; (iii) genic-exonic and (iv) putative antisense lncRNAs, on the condition that the sequence was entirely contained within the above classes. LncRNAs classification was conducted using the Bioconductor package GenomicFeatures100. Genic, exonic, intronic and intergenic regions were selected from the Bd21 annotation version 2.1. To identify antisense lncRNA we used the approach described in Li et al.11.

LncRNAs associated with transposable elements, tandem repeats and microRNAs. Transposable elements (TEs) associated with the lncRNA loci were determined based on the current Bd21 TEs annotations, retrieved from MIPS (ftp://ftpmpis.helmholtz-muenchen.de/plants/brachypodium/repeats). LncRNAs coordinates were intersected with TEs coordinates using the Bioconductor package GenomicFeatures84. We retained only lncRNAs contained within the TEs genomic coordinates. In addition, we also investigated the presence of tandem repeats (TRs) and precursors of microRNAs within the putative IncRNA sequences.

Sequence similarity analysis. We performed a clustering analysis to assess lncRNA sequence similarity among the three inbred lines using CD-HIT version 4.6. LncRNA sequences of each inbred line were first clustered separately, considering a cutoff of 95% to reduce sequence redundancy. Unique lncRNA sequences in each inbred line were then compared pairwise, with the threshold of sequences nucleotides similarity set at 90%.

Expression level and tissue specificity of lncRNAs. Two normalization methods were used to investigate the lncRNAs expression level: (i) RPKM (Reads Per KiloBase per Million mapped reads) for the intra-sample quantification, by counting reads mapped on the reconstructed transcript models with the script htsseq-count109 version 0.5.4p5 (with option intersection-noempty to exclude multi-mapping reads). The resulting matrix of counts was used to discard transcripts not expressed (those with zero counts in all samples), and the expression level was quantified in RPKM using the Bioconductor package edgeR105. (ii) TPM (Transcripts Per Million) for inter-sample quantification, by using the program Salmon106 in the quasi-mapping based mode. Specific expression of lncRNAs was determined using the Shannon Entropy method with the BioQC package from Bioconductor.

Identification of lncRNA targets of microRNAs and competing endogenous target mimics. We searched for the lncRNAs as potential target of microRNAs and lncRNAs target mimic activity. We collected B. distachyon miRNA mature sequences from the miRBase version 21. TargetFinder (https://github.com/carrington-lab/TargetFinder) with default options was used to identify lncRNAs potentially targeted by miRNAs, applying default parameters. According to the TargetFinder score based criterion49 we considered potential target of microRNAs for in silico assessment, by counting reads mapped on the reconstructed transcript models with the script htseq-count109 version 0.5.4p5 (with option intersection-noempty to exclude multi-mapping reads). The resulting matrix of counts was used to discard transcripts not expressed (those with zero counts in all samples), and the expression level was quantified in RPKM using the Bioconductor package edgeR105. (ii) TPM (Transcripts Per Million) for inter-sample quantification, by using the program Salmon106 in the quasi-mapping based mode. Specific expression of lncRNAs was determined using the Shannon Entropy method with the BioQC package from Bioconductor.

Competing endogenous target mimics (eTMs) activity was determined by locally running the software psMimic54 version 1.1 with default parameters (http://omicslab.genetics.ac.cn/psMimic).

Data availability. All supplementary files are available at Figshare https://doi.org/10.6084/m9.figshare.3423635.v1 through the link https://figshare.com/s/35435039f2ea3d7bb4ef.

References
1. Mattick, J. S. & Rinn, J. L. Discovery and annotation of long noncoding RNAs. Nat. Struct. Mol. Biol. 22, 5–7 (2015).
2. Qian, X., Ba, Y., Zhuang, Q. & Zhong, G. RNA-Seq technology and its application in fish transcriptomics. OMICS 18, 98–110 (2014).
3. Carninci, P. et al. The transcriptional landscape of the mammalian genome. Science 309, 1559–1563 (2005).
4. Cheng, J. et al. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science 308, 1149–1154 (2005).
5. Kim, E.-D. & Sung, S. Long noncoding RNA: unveiling hidden layer of gene regulatory networks. Trends Plant Sci. 17, 16–21 (2012).
6. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 42, D68–73 (2014).
7. Wang, H. et al. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis. Genome Res. 24, 444–453 (2014).
57. Liu, T.-Y., Chang, C.-Y. & Chiou, T.-J. The long-distance signaling of mineral macronutrients. *Curr. Opin. Plant Biol.* 12, 312–319 (2009).

58. Kawashima, C. G. *et al.* Interplay of SL1M1 and miR395 in the regulation of sulfate assimilation in Arabidopsis. *Plant J.* 66, 863–876 (2011).

59. Chen, X. *et al.* SQUAMOSA promoter–binding protein–like transcription factors: star players for plant growth and development. *J. Integr. Plant Biol.* 52, 946–951 (2010).

60. Morris, K. V. & Mattick, J. S. The rise of regulatory RNA. *Nat. Rev. Genet.* 15, 423–437 (2014).

61. McGettigan, P. A. Transcriptionics in the RNA-seq era. *Curr. Opin. Chem. Biol.* 17, 4–11 (2013).

62. Abdel-Ghany, S. E. *et al.* A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* 7, 11706 (2016).

63. Wang, B. *et al.* Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* 7, 11708 (2016).

64. Di, C. *et al.* Characterization of stress-responsive IncRNAs in Arabidopsis thaliana by integrating expression, epigenetic and structural features. *Plant J.* 80, 848–861 (2014).

65. Hezroni, H. *et al.* Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* 11, 1110–1122 (2015).

66. Diederichs, S. The four dimensions of noncoding RNA conservation. *Trends Genet.* 30, 121–123 (2014).

67. Ulltitsky, L., Shukumata, A., Jan, C. H., Sive, H. & Bartel, D. P. Conserved function of IncRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147, 1537–1550 (2011).

68. Quinn, J. J. *et al.* Rapid evolutionary turnover underlies conserved IncRNA–genome interactions. *Genes Dev.* 30, 191–207 (2016).

69. Vogel, J. P. *et al.* Development of SSR markers and analysis of diversity in Turkish populations of Brachypodium distachyon. *BMC Plant Biol.* 9, 88 (2009).

70. Milligan, M. J. & Lipovich, L. Pseudogene-derived IncRNAs: emerging regulators of gene expression. *Front. Genet.* 5 (2014).

71. Kapusta, A. *et al.* Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* 9, e1003470 (2013).

72. Kelley, D. & Rinn, J. Transposable elements reveal a stem cell–specific class of long noncoding RNAs. *Genome Biol.* 13, R107 (2012).

73. Wicker, T., Guyot, R., Yahiaoui, N. & Keller, B. CACTA transposons in Triticeae. A diverse family of high–copy repetitive elements. *Plant Physiol.* 132, 52–63 (2003).

74. Johnson, R. & Guigó, R. The RfID hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA* 20, 959–976 (2014).

75. Usdin, K. The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. *Genome Res.* 18, 1011–1019 (2008).

76. Legendre, M., Pochet, N., Pak, T. & Verstrepen, K. J. Sequence-based estimation of minisatellite and microsatellite repeat variability. *Genome Res.* 17, 1787–1796 (2007).

77. Morgante, M., Hanafey, M. & Powell, W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat. Genet.* 30, 194–200 (2002).

78. Zhao, Z. *et al.* Genome-wide analysis of tandem repeats in plants and green algae. *G3* 4, 67–78 (2014).

79. Gloss, B. S. & Dinger, M. E. The specificity of long noncoding [RNA] expression. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1859, 16–22 (2016).

80. Dinger, M. E., Amaral, P. P., Mercer, T. R. & Mattick, J. S. Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications. *Brief. Funct. Genomic. Proteomic.* 8, 407–423 (2009).

81. Zheng, Y., Wang, Y., Wu, J., Ding, B. & Fei, Z. A dynamic evolutionary and functional landscape of plant phased small interfering RNAs. *BMC Biol.* 13, 32 (2015).

82. Johnson, C. *et al.* Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. *Genome Res.* 19, 1429–1440 (2009).

83. Zhai, J. *et al.* Spatiotemporally dynamic, cell-type–dependent premeiotic and meiotic phased small RNAs in maize anthers. *Proceedings of the National Academy of Sciences* 112, 3146–3151 (2015).

84. Axtell, M. J. Non-coding RNAs: The small mysteries of males. *Nature Publishing Group* 1–2, doi:10.1038/nplants.2015.55 (2015).

85. Lu, A.-C.-S., Chan, W.-Y., Rennert, O. M. & Lee, T.-L. Long noncoding RNAs in spermatogenesis: insights from recent high-throughput transcriptomics studies. *Reproduction* 147, R131–R141 (2014).

86. Axtell, M. J. Classification and Comparison of Small RNAs from Plants. *Annu. Rev. Plant Biol.* 64, 137–159 (2013).

87. Gardini, A. & Shiekhattar, R. The many faces of long noncoding RNAs. *FEBS J.* 282, 1647–1657 (2015).

88. Fan, C., Hao, Z., Yan, J. & Li, G. Genome-wide identification and functional analysis of lncRNAs acting as miRNA targets or decoys in maize. *BMC Genomics* 16, 793 (2015).

89. Song, C. *et al.* Construction and analysis of cardiac hypertrophy-associated IncRNA–mRNA network based on competitive endogenous RNA reveal functional IncRNAs in cardiac hypertrophy. *Oncotarget* 7, 10827–10840 (2016).

90. Yoon, J.-H., Abdelmoisen, K. & Gorospe, M. Functional interactions among microRNAs and long noncoding RNAs. *Semin. Cell Dev. Biol.* 34, 9–14 (2014).

91. Meng, Y., Shao, C., Wang, H. & Jin, Y. Target mimics: an embedded layer of microRNA-involved gene regulatory networks in plants. *BMC Genomics* 13, 197 (2012).

92. Davidson, R. M. *et al.* Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. *Plant J.* 71, 492–502 (2012).

93. Veerels, W. *et al.* Molecular and physiological analysis of growth-limiting drought stress in Brachypodium distachyon leaves. *Mol. Plant* 6, 311–322 (2013).

94. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EM Bnet. Journal* 17 (2011).

95. Del Fabbro, C., Scalabrin, S., Morgante, M. & Giorgi, F. M. An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis. *PLoS One* 8, e58024 (2013).

96. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927 (2011).

97. De Quattro, C., Mica, E., Pè, M. E. & Bertolini, E. In *Brachypodium Genomics* (eds Gaurav, S., Hikmet, B. & Peter, R.) 1667, Springer (2017).

98. Juntawong, P., Girke, T., Bazin, J. & Bailey-Serres, J. Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis. *Proc. Natl. Acad. Sci. USA* 111, E203–E212 (2014).

99. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D30 (2014).

100. Kong, L. *et al.* CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35, W345–W349 (2007).

101. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. & Eddy, S. R. Rfam: an RNA family database. *Nucleic Acids Res.* 31, 439–441 (2003).

102. Sonnhammer, E. L. & Durbin, R. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* 167, GCI–10 (1995).

103. Frazer, K. A., Pachter, L., Poliaak, A., Rubin, E. M. & Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32, W273–W279 (2004).
104. Lawrence, M. et al. Software for computing and annotating genomic ranges. PLoS Comput. Biol. 9, e1003118 (2013).
105. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2010).
106. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417–419 (2017).
107. Qi, L. et al. The compact Brachypodium genome conserves centromeric regions of a common ancestor with wheat and rice. Funct. Integr. Genomics 10, 477–492 (2010).
108. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28(23), 3150–3152 (2012).
109. Anders, S., Pyl, P. T. & Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31(2), 166–169 (2015).

Acknowledgements
This work was supported by the International Doctoral Programme in Agrobiodiversity - Scuola Superiore Sant'Anna (www.santannapisa.it).

Author Contributions
E.B. conceptualized, designed and coordinated the research project work. E.B. and C.D.Q. carried out all data collection and computational data analysis. E.B. and C.D.Q. wrote the manuscript. MEP provided the computational infrastructure for data analysis. All authors contributed to the discussion of the data. The final version of the manuscript was edited and approved by all authors.

Additional Information
Supplementary information accompanies this paper at doi:10.1038/s41598-017-11206-z

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017