[Letter to the Editor]
Comparison of small RNA next-generation sequencing with and without isolation of small RNA fraction

The role of small RNAs (sRNAs) such as miRNAs and siRNAs (1) in the post-transcriptional regulation of developmental processes and stress responses in plants is becoming well-established (2,3) through studies in model species, such as *Arabidopsis thaliana* and *Nicotiana benthamiana* as well as in several crop species (4,5), including wheat (6,7).

Next-generation sequencing (NGS) techniques are increasingly used to identify and discover sRNAs. For these sequencing methods, preparation of libraries from isolated nucleic acids is required (8). For sRNA sequencing, the library preparation can be performed using either total RNA or the isolated sRNA fraction, as, for example, the Illumina library preparation protocol indicates (http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqsmallrna/truseq-small-rna-library-prep-kit-reference-guide-15004197-01.pdf).

Here we compare data obtained by sequencing small RNA libraries prepared from both the isolated total RNA and the sRNA fraction. For the experiments, developing grain samples were collected from three wheat lines (*Triticum aestivum* "Bánkuti" B35, B52, and *Triticum spelta*) at 10, 20, and 30 days after pollination in 2 biological replicates.

To isolate total RNA, grains were ground to fine powder in liquid nitrogen. The powder (100 mg) was suspended by vortexing in 600 µL extraction buffer (0.1 M glycine-NaOH, pH 9.0, 100 mM NaCl, 10 mM EDTA, 2% SDS, and 1% sodium lauryl-sarcosine) and then mixed with an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1). After centrifugation at 18,000 × *g* for 10 min at 4°C, total RNA was purified from the upper aqueous phase using the Direct-Zol RNA Miniprep Kit (Zymo Research, Freiburg, Germany) according to the manufacturer's instructions. The integrity and purity of the total RNA were assessed using a Bioanalyzer 2100 instrument and the RNA 6000 Nano Assay Kit (both from Agilent, Santa Clara, CA).

To purify small RNAs, 30 µg total RNA from each sample was run on an 8% denaturing polyacrylamide gel (19:1 acrylamide/bis-acrylamide, 8 M urea, 1× TBE buffer), and the 15–30 nucleotide fraction was isolated from the gel as described in the Illumina TruSeq Small RNA Library Prep Kit Reference Guide.

Libraries for small RNA sequencing were prepared from either 1 µg of total RNA or 75 ng of the small RNA fraction using the Illumina TruSeq Small RNA Library Preparation Kit (Illumina, San Diego, CA) and strictly following the manufacturer's instructions. Briefly, after ligation of the 3′ and 5′ RNA adapters, reverse transcription was performed to synthesize cDNA, which was then amplified by PCR. Since the total length of the adapter sequences was 126 nucleotides, the obtained products centered between 140 bp and 160 bp, corresponding to 14–34 nucleotide–long sRNAs. The products were purified from 6% Novex TBE-PAGE gels (Thermo Fisher Scientific, Waltham, MA) according to the protocol described in the Illumina TruSeq Small RNA Library Prep Kit Reference Guide, quantified using a Qubit Fluorometer (Thermo Fisher Scientific), and verified on a Bioanalyzer 2100 instrument using a DNA1000 chip. Nine libraries were pooled to create two total RNA and two isolated sRNA pools (hereafter referred to as the T and P libraries, respectively), which were denatured at a concentration of 10 pM. Clusters were generated using a cBot instrument, and the TruSeq SR Cluster Kit v3 (Illumina) according to the manufacturer's protocol. Fifty cycles of 50-bp single-read sequencing were run on an Illumina HiScan SQ instrument using the TruSeq SBS Kit v3. Image processing and base calling were performed using the off-line base-caller and Illumina's CASAVA software on a Linux platform.

Data analysis was performed according to the steps described below. Adaptor sequences were trimmed from the reads using cutadapt version 1.2.1 (9). Reads that were 16–28 nucleotides long were selected, and low abundance sequences (<3 reads) were removed using custom Python scripts. tRNA and rRNA sequences were also removed using the Rfam database (10). To predict miRNAs, the reads were aligned to the wheat genome using Bowtie (11), and they and the 250-bp flanking sequences on both sides were folded into a secondary structure using the ViennaRNA Package 2.0 (12). Folded structures were analyzed to determine whether they met the criteria for plant miRNAs (13) using the software miRDeep-P (14). The reads that passed the test were aligned to sequences in the miRBase Release 21 database (15). Unique sequences among the aligned reads with and without full homology (length and sequence) to the miRBase entries were considered to be conserved or potentially new miRNAs, respectively. The number of reads was normalized for each library using the following formula: (number of reads/number of raw reads) × 10^6.

In addition to the number of raw reads, the average value and variation for four measures from successive steps during the data analysis [the number of reads after adapter removal/selection of 16–28 nucleotide–long reads; the number of reads after removing low abundance sequences (<3 reads); the number of reads after removing tRNA and rRNA sequences; and the number of sequences matching to the wheat genome] were compared between the T and P libraries. The potential of the two library preparation methods to detect known miRNAs and to discover new miRNAs [the number of reads predicted as potential miRNAs; the number of reads matching miRNAs of any species in miRBase; and the number of reads matching *T. aestivum* miRNAs in miRBase] was also compared.

Sequencing of the T and P libraries resulted in very similar average numbers of raw reads (9,192,098 and 9,244,612, respectively), although the variation was much larger for the T than the P libraries (Supplementary Figure S1A). In the successive

Table 1. The number and proportion of reads in small RNA sequencing of wheat developing grain samples.

Library name*	Raw**	Adaptor removal/ size selection***	%	Low abundance removal (<3)***	%	t/rRNA removal***	%	Match to genome**	%	Predicted miRNA**	%	Match to miRBase with zero mismatches**	%	Match to miRBase tae with zero mismatches**	%
T_B35_10D_I	15156563	5985082	39.49	4198781	27.70	2922908	19.28	1943177	12.82	150830	1.00	14371	0.095	6108	0.040
T_B35_10D_II	10560703	4138679	39.19	2843829	26.93	2012860	19.06	1544520	14.63	114439	1.08	12705	0.120	6176	0.058
T_B35_20D_I	4978036	1802685	36.21	1234103	24.79	631452	12.68	466224	9.37	27902	0.56	2379	0.048	1075	0.022
T_B35_20D_II	12665049	4972945	39.27	3647128	28.80	2040604	16.11	1716032	13.55	126954	1.00	10762	0.085	4493	0.035
T_B35_30D_I	3153188	967205	30.67	681665	21.62	308882	9.80	239019	7.58	11585	0.37	1217	0.039	549	0.017
T_B35_30D_II	10282980	3320674	32.29	2622299	25.50	1331992	12.95	1152858	11.21	72929	0.71	7823	0.076	3727	0.036
T_B52_10D_I	8066911	3167427	39.26	2129093	26.39	1494550	18.53	1014480	12.58	70347	0.87	7422	0.092	3365	0.042
T_B52_10D_II	10643935	3625301	34.06	2718402	25.54	1780370	16.73	1456167	13.68	77877	0.73	9521	0.089	3958	0.037
T_B52_20D_I	2658471	1012539	38.09	700319	26.34	308417	11.60	235783	8.87	11220	0.42	1409	0.053	565	0.021
T_B52_20D_II	9527805	3309063	34.73	2495972	26.20	1316431	13.82	1149283	12.06	55273	0.58	6572	0.069	2543	0.027
T_B52_30D_I	3224967	990137	30.70	705035	21.86	331417	10.28	256921	7.97	11875	0.37	1174	0.036	508	0.016
T_B52_30D_II	13318595	4034106	30.29	3337243	25.06	1520049	11.41	1362650	10.23	62131	0.47	5533	0.042	1976	0.015
T_Spelta_10D_I	6922111	2724663	39.36	1870765	27.03	1214871	17.55	850036	12.28	54924	0.79	5801	0.084	2928	0.042
T_Spelta_10D_II	10244620	3847448	37.56	2790599	27.24	1799502	17.57	1450158	14.16	79568	0.78	10330	0.101	5529	0.054
T_Spelta_20D_I	18755713	4546016	24.24	3434077	18.31	1699611	9.06	1362869	7.27	41243	0.22	4274	0.023	1865	0.010
T_Spelta_20D_II	10435412	3664643	35.12	2651772	25.41	1403163	13.45	1223047	11.72	51720	0.50	6134	0.059	2547	0.024
T_Spelta_30D_I	3544375	1142490	32.23	832051	23.48	385715	10.88	305947	8.63	8036	0.23	971	0.027	386	0.011
T_Spelta_30D_II	11318342	3556470	31.42	2857925	25.25	1347476	11.91	1208358	10.68	47446	0.42	5059	0.045	2164	0.019
P_B35_10D_I	10680324	9951178	93.17	8091518	75.76	5759787	53.93	4646015	43.50	447959	4.19	44029	0.412	18038	0.169
P_B35_10D_II	10414069	9682160	92.97	7400984	71.07	5630382	54.07	4152810	39.88	681013	6.54	50909	0.489	29588	0.284
P_B35_20D_I	9874139	7932669	80.34	6516468	66.00	3254787	32.96	4347295	44.03	242183	2.45	26290	0.266	6550	0.066
P_B35_20D_II	13091631	12573317	96.04	8024391	61.29	5287615	40.39	2864145	21.88	467109	3.57	25019	0.191	11003	0.084
P_B35_30D_I	9483375	8782266	92.61	7447694	78.53	3740447	39.44	3463523	36.52	189385	2.00	11228	0.118	6717	0.071
P_B35_30D_II	9484795	8133769	85.76	6948940	73.26	3380046	35.64	2985161	31.47	249120	2.63	18942	0.200	9920	0.105
P_B52_10D_I	9555964	9139351	95.64	6760205	70.74	5528573	57.85	4126986	43.19	305772	3.20	19632	0.205	9376	0.098
P_B52_10D_II	9948877	9517941	95.67	7165287	72.02	5249757	52.77	4159426	41.81	358728	3.61	32257	0.324	14553	0.146
P_B52_20D_I	6828887	6410809	93.88	5224370	76.50	2565220	37.56	2264419	33.16	161640	2.37	11060	0.162	4371	0.064
P_B52_20D_II	8826441	8480154	96.08	6177125	69.98	3955177	44.81	3278862	37.15	264828	3.00	15359	0.174	7289	0.083
P_B52_30D_I	8723205	7518644	86.19	6574122	75.36	3139374	35.99	2934558	33.64	151602	1.74	7661	0.088	4559	0.052
P_B52_30D_II	7157139	6586173	92.02	5433831	75.92	2541607	35.51	2343319	32.74	127136	1.78	10256	0.143	4504	0.063
P_Spelta_10D_I	11401908	10376732	91.01	9064974	79.50	5936665	52.07	4829776	42.36	523112	4.59	40767	0.358	23875	0.209
P_Spelta_10D_II	7362199	7034430	95.55	5218290	70.88	3917853	53.22	3085648	41.91	294228	4.00	18981	0.258	12756	0.173
P_Spelta_20D_I	8600522	8009596	93.13	6041576	70.25	3192767	37.12	2814586	32.73	135092	1.57	12892	0.150	5446	0.063
P_Spelta_20D_II	9370866	8918178	95.17	5033947	53.72	3373970	36.00	2831803	30.22	201720	2.15	16829	0.180	8967	0.096
P_Spelta_30D_I	8408441	7553635	89.83	6758574	80.38	3188603	37.92	2956476	35.16	115875	1.38	7852	0.093	4059	0.048
P_Spelta_30D_II	7190250	6336904	88.13	5273237	73.34	2483147	34.53	2270714	31.58	98280	1.37	7861	0.109	3707	0.052

*T: sequencing library prepared from total RNA; P: sequencing library prepared from purified small RNA fraction; B35 and B52: *Triticum aestivum* (tae) cv. Bánkuti lines; Spelta: *Triticum spelta*; 10D, 20D, 30D: developing grain samples collected 10, 20, and 30 days after pollination, respectively; I, II: biological replicates; **number of reads; ***number of reads after the selection step; %: percentage of reads relative to raw reads.

steps of the analysis, the proportion of the above-described four measures, relative to the number of the raw reads, was higher in the P than in the T libraries (Table 1). Their mean value and variability, in terms of both absolute and normalized values, was also higher in the P than in the T libraries (Supplementary Figure S1A).

Several measures were included in the above-described library preparation and data analysis in order to identify only potential miRNAs. First, the 15–30 nucleotide sRNA fraction was isolated to prepare the P libraries; thus, the purified sRNAs would contain only minimal contamination from snoRNAs and snRNAs, which are >60 nucleotides (1). Second, the amplified libraries with the appropriate size corresponding to 14–34 nucleotide sRNAs were also purified. Third, after removing the adaptor sequences, only 16–28 nucleotide–long raw reads were processed further. Finally, reads that matched to the wheat genome were subjected to a rigorous in silico test to eliminate those that did not meet the criteria for plant miRNAs (14). The portion of the reads that were identified as potential miRNAs after that selection varied from 2.63% to 7.76% and 3.92% to 16.4% of the genome-mapped reads in the T and P libraries, respectively, which is quite similar to the miRNA discovery rate described in other wheat studies (16,17).

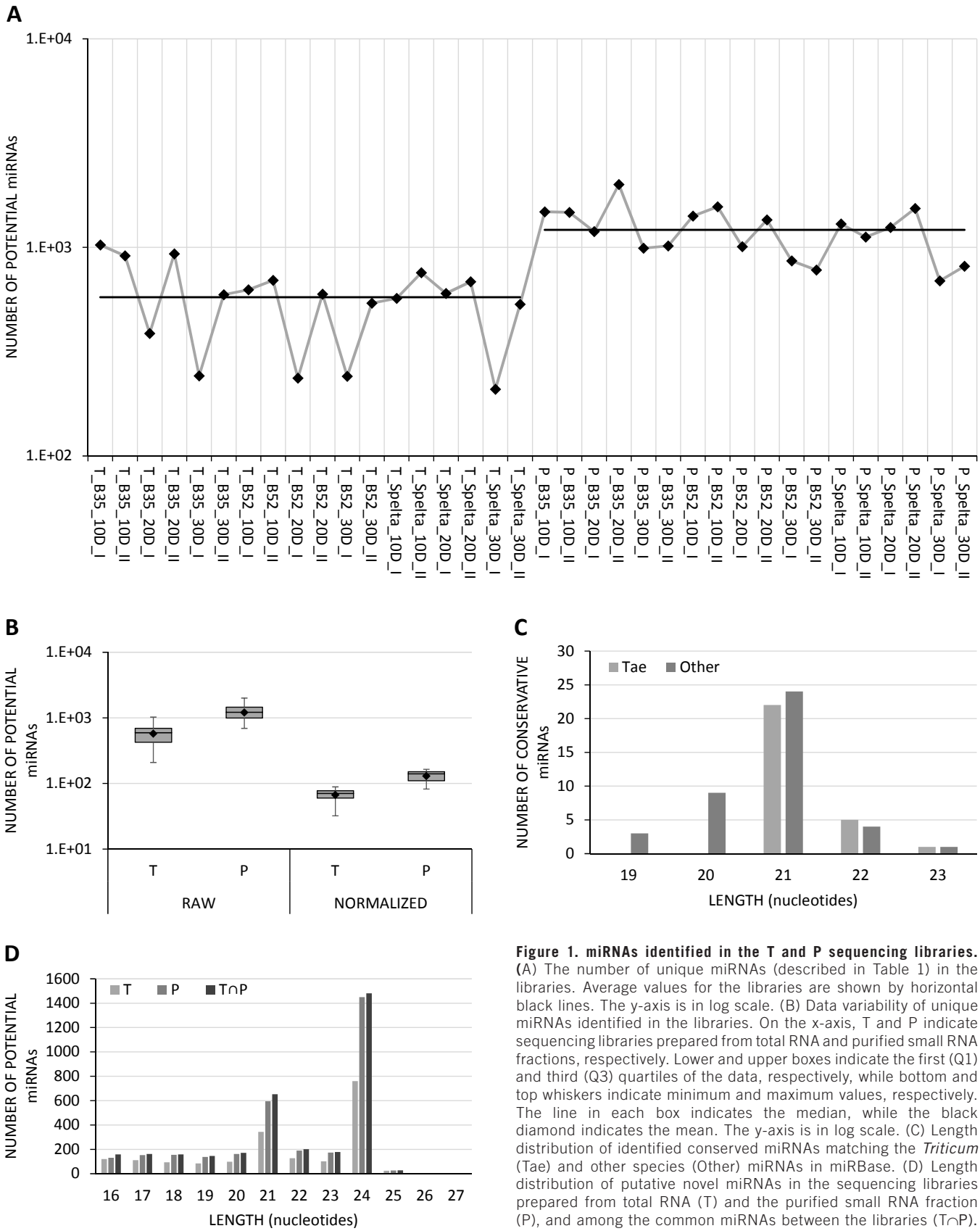

Figure 1. miRNAs identified in the T and P sequencing libraries. (A) The number of unique miRNAs (described in Table 1) in the libraries. Average values for the libraries are shown by horizontal black lines. The y-axis is in log scale. (B) Data variability of unique miRNAs identified in the libraries. On the x-axis, T and P indicate sequencing libraries prepared from total RNA and purified small RNA fractions, respectively. Lower and upper boxes indicate the first (Q1) and third (Q3) quartiles of the data, respectively, while bottom and top whiskers indicate minimum and maximum values, respectively. The line in each box indicates the median, while the black diamond indicates the mean. The y-axis is in log scale. (C) Length distribution of identified conserved miRNAs matching the *Triticum* (Tae) and other species (Other) miRNAs in miRBase. (D) Length distribution of putative novel miRNAs in the sequencing libraries prepared from total RNA (T) and the purified small RNA fraction (P), and among the common miRNAs between the libraries (T∩P).

Selection of unique sequences among those reads, which met with the miRNA selection criteria (Table 1), resulted in 1934 and 3247 16–28 nucleotide–long potential miRNAs in the T and P libraries, respectively. The average number of the identified potential miRNAs was about 2-fold higher in the P than in the T libraries (Figure 1A), while data variability was about the same for both the raw and normalized data (Figure 1B). Due to the overlap between the T and P library sets, 3338 unique potential miRNAs in total were identified. Among these, 69 had perfect matches (length and sequence) to entries in miRBase Release 21; therefore, we presumed them be conserved miRNAs. The number of such miRNAs was very similar in the T and P libraries: 67 and 68, respectively. Of these, 26 and 28 were *T. aestivum* miRNAs, and 41 and 40 were miRNAs of other species in the T and P libraries, respectively. Between the T and P libraries, 26 and 40 miRNAs of *T. aestivum* and of other species, respectively, were in common. In contrast, among the 3269 potential miRNAs that did not have a 100% match to miRBase, 1775 were common between the libraries, while 91 and 1403 were specific for the T and P libraries, respectively. Twenty-one nucleotide–long miRNAs were the most numerous among the detected 69 miRNAs, while among the 3269 potential miRNAs, 24-nucleotide–long miRNAS were present in the largest numbers (Figure 1, C and D).

Here we demonstrated that sequencing the libraries prepared from isolated small RNA fractions and from total RNA resulted in very similar numbers of raw reads, but the absolute and normalized values of both the raw and processed data had higher variability among the libraries made from isolated small RNA than from total RNA. Predictably, libraries constructed from isolated low molecular weight RNAs resulted in more sRNA reads than libraries constructed from total RNA. Known miRNAs were detected in both types of libraries in almost equal number, but libraries prepared from isolated small RNA outperformed total RNA libraries with respect to the discovery of novel miRNAS. Because of this latter observation, we recommend isolating the small RNA fraction prior to the preparation of sequencing libraries since this step does not add much extra cost and labor to a project.

Project data can be found in the NCBI BioProject database under accession number PRJNA309061.

Author contributions

T.N. and F.M. analyzed the data. A.K. grew plants, isolated RNA, and prepared libraries. S.P. prepared libraries and performed sequencing. E.B. supervised data analysis. Z.H. supervised plant and RNA work. F.M. conceived the study and wrote the manuscript.

Acknowledgments

This work was supported by Hungarian Research, Development and Innovation Fund grant AGR_PIAC13-2013-0074.

Competing interests

The authors declare no competing interests.

References

1. **Röther, S. and G. Meister.** 2011. Small RNAs derived from longer non-coding RNAs. Biochimie *93*:1905-1915.
2. **Hofmann, N.R.** 2014. A world beyond *Arabidopsis* : updates on small RNAs in plant development. Plant Cell *26*:4564.
3. **Guleria, P., M. Mahajan, J. Bhardwaj, and S.K. Yadav.** 2011. Plant small RNAs: biogenesis, mode of action and their roles in abiotic stresses. Genomics Proteomics Bioinformatics *9*:183-199.
4. **Arikit, S., R. Xia, A. Kakrana, K. Kun Huang, J. Zhai, Z. Zhe, O. Valdés-López, S. Prince, et al.** 2014. An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. Plant Cell *26*:4584-4601.
5. **Thompson, B.E., C. Basham, R. Hammond, Q. Ding, A. Kakrana, T.-F. Lee, S.A. Simon, R. Meeley, et al.** 2014. The dicer-like1 homolog fuzzy tassel is required for the regulation of meristem determinacy in the inflorescence and vegetative growth in maize. Plant Cell *26*:4702-4717.
6. **Han, R., C. Jian, J. Lv, Y. Yan, Q. Chi, Z. Li, Q. Wang, J. Zhang, et al.** 2014. Identification and characterization of microRNAs in the flag leaf and developing seed of wheat (*Triticum aestivum* L.). BMC Genomics *15*:289.
7. **Meng, F., H. Liu, K. Wang, L. Liu, S. Wang, Y. Zhao, J. Yin, and Y. Li.** 2013. Development-associated microRNAs in grains of wheat (*Triticum aestivum* L.). BMC Plant Biol. *13*:140.
8. **Head, S.R., H. Kiyomi Komori, S.A. LaMere, T. Whisenant, F. Van Nieuwerburgh, D.R. Salomon, and P. Ordoukhanian.** 2014. Library construction for next-generation sequencing: Overviews and challenges. Biotechniques *56*:61-68.
9. **Martin, M.** 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*:10-12.
10. **Nawrocki, E.P., S.W. Burge, A. Bateman, J. Daub, R.Y. Eberhardt, S.R. Eddy, E.W. Floden, P.P. Gardner, et al.** 2015. Rfam 12.0: updates to the RNA families database. Nucleic Acids Res. *43*:D130-D137.
11. **Langmead, B., C. Trapnell, M. Pop, and L.S. Salzberg.** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*:R25.
12. **Lorenz, R., S.H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P.F. Stadler, and I.L. Hofacker.** 2011. ViennaRNA Package 2.0. Algorithms Mol. Biol. *6*:26.
13. **Meyers, B.C., M.J. Axtell, B. Bartel, D.P. Bartel, D. Baulcombe, J.L. Bowman, X. Cao, J.C. Carrington, et al.** 2008. Criteria for Annotation of Plant MicroRNAs. Plant Cell *20*:3186-3190.
14. **Yang, X. and L. Li.** 2011. miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plant. Bioinformatics *27*:2614-2615.
15. **Kozomara, A. and S. Griffiths-Jones.** 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. *42*:D68-D73.
16. **Sun, F., G. Guo, J. Du, W. Guo, H. Peng, Z. Ni, Q. Sun, and Y. Yao.** 2014. Whole-genome discovery of miRNAs and their targets in wheat (Triticum aestivum L.). BMC Plant Biol. *14*:142.
17. **Zhou, J., Y. Cheng, M. Yin, E. Yang, W. Gong, C. Liu, X. Zheng, K. Deng, et al.** 2015. Identification of novel miRNAs and miRNA expression profiling in wheat hybrid necrosis. PLoS ONE *10*:e0117507.

Tibor Nagy[1,†,*], András Kis[1,*], Szilárd Poliska[2], Endre Barta[1], Zoltán Havelda[1], and Ferenc Marincs[1]

[1]Agricultural Biotechnology Institute, National Agricultural Research and Innovation Centre, Gödöllő, Hungary and [2]Genomic Medicine and Bioinformatic Core Facility, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary

* T.N. and A.K. contributed equally to this work.
† T.N. present affiliation: Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom.

BioTechniques 60:273-278 (June 2016)
doi 10.2144/000114423

Keywords: wheat; miRNA; Illumina sequencing; library preparation

Supplementary material for this article is available at www.BioTechniques.com/article/114423.

Received 03 November 2015; accepted 25 February 2016.

Address correspondence to Ferenc Marincs, Agricultural Biotechnology Institute, NARIC, Szent-Györgyi Albert u. 4., Gödöllő, Hungary. E-mail: marincs.ferenc@abc.naik.hu