

A Rapid and Efficient Method To Prepare Di-tagged cDNA Libraries from Small RNA for Discovery and Profiling by Deep Sequencing

Ramesh Vaidyanathan, Joanne Decker, Jerry Jendrisak, Nick Caruccio, and Jim Pease

EPICENTRE Biotechnologies, 726 Post Road, Madison, WI 53713, USA

Introduction

Identification and analysis of small RNA by deep sequencing requires preparation of a di-tagged cDNA library. Most library preparation methods capture small-RNA molecules that have a 5'-monophosphate and a 3'-hydroxyl group and exclude RNA molecules, such as 5'-triphosphorylated secondary siRNA and small capped-RNA molecules, which have other 5'-terminal structures.

We present a method to generate libraries of di-tagged cDNA from small RNA for discovery and profiling. The methods utilize enzymes to select the desired RNA molecules of interest that are used to generate the di-tagged cDNA library based on the structure of their 5' ends. The streamlined protocols do not require gel purification to remove excess sequencing adaptors, and generate libraries of amplified di-tagged cDNA templates in less than 1 day. Multiple samples can be handled without danger of cross-contamination. The di-tagged cDNA libraries can be used for next-generation sequencing, construction of cloned libraries in *E. coli*, quantitative RT-PCR, microarray analysis, and other applications.

Methods

Enzymatic manipulation to create 5'-monophosphorylated RNA for tagging

A widely used method to prepare small-RNA libraries relies on sequential ligation of adaptors (tags) to the ends of small RNAs. This method selects only 5' monophosphorylated RNA (p-RNA), while RNA having other 5'-end groups — such as a cap (Gppp-RNA) or triphosphate (ppp-RNA)—are excluded. We present methods to tag RNA molecules depending on the nature of the 5'-end group (p-RNA, ppp-RNA, Gppp-RNA, HO-RNA). The methods rely on selective treatment with RNA modifying enzymes (Table 1).

Table 1. Activities of various enzymes applicable to the selection process.

RNA Modifying Enzyme	Substrate	End Product
Tobacco Acid Pyrophosphatase (TAP)	Gppp-RNA and ppp-RNA	p-RNA
RNA 5' Polyphosphatase	ppp-RNA	p-RNA
T4 Polynucleotide Kinase (PNK)	HO-RNA	p-RNA
Terminator™ 5'-Phosphate-Dependent Exonuclease	p-RNA	NMPs

Procedure to enrich for small RNAs that contain a 2'-OMe group at the 3'-terminus

The 3'-terminal nucleotide of plant miRNAs are methylated at the 2'-hydroxyl group. These end-methylated RNAs are not polyadenylated by *E. coli* poly(A) polymerase (Ebhardt HA *et al.*, [2005] *Proc. Natl. Acad. Sci. USA* 102 13398-403). We have used the absence of polyadenylation as a method to enrich for end-methylated small RNAs as shown in Fig. 1.

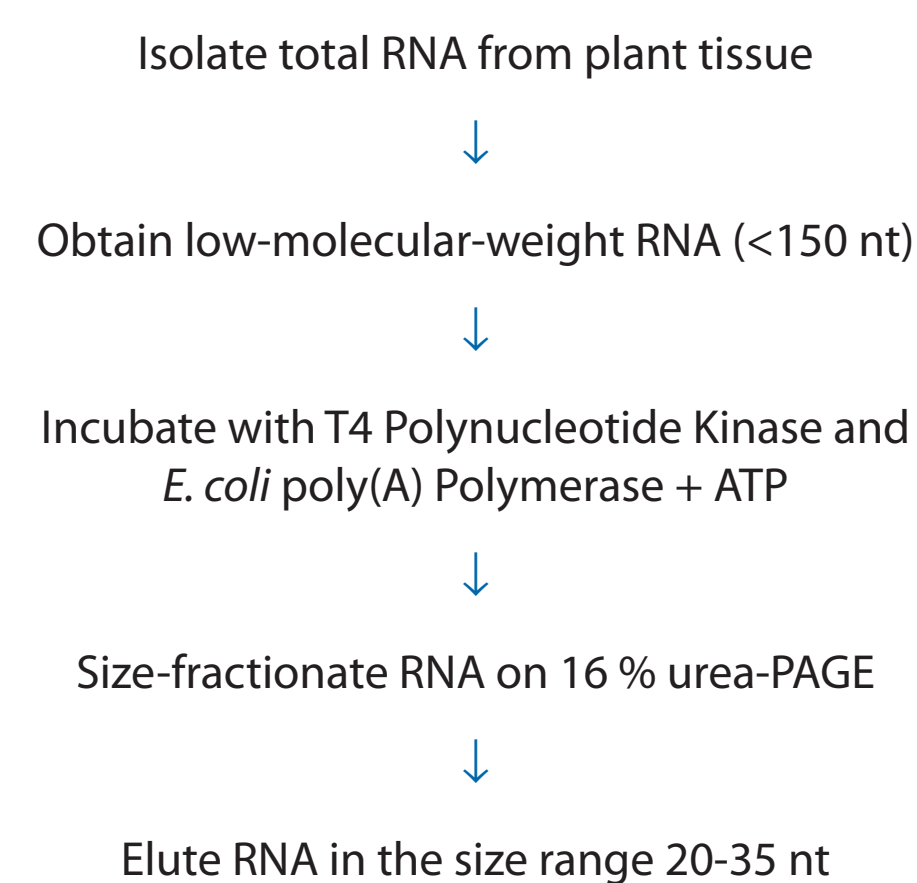


Figure 1. Enrichment for 3'-end-methylated small RNAs.

Preparation of di-tagged RNA and double stranded cDNA library

Adaptors are ligated to the 5' and 3' ends of RNA using the scheme shown in Fig. 2 to prepare a di-tagged RNA library. To tag the 3' end of the RNA, a 5'-activated, 3'-blocked DNA oligonucleotide is used in a reaction mediated by a deletion mutant of T4 RNA Ligase 2. Excess activated oligo is removed by a simple enzymatic reaction*, which does not require gel purification and can be easily automated. The 5' end of the RNA is then tagged by standard methods. First-strand cDNA is prepared from the RNA and amplified to create a dual end-tagged, double-stranded (ds) cDNA library.

*Covered by existing and/or pending patents.

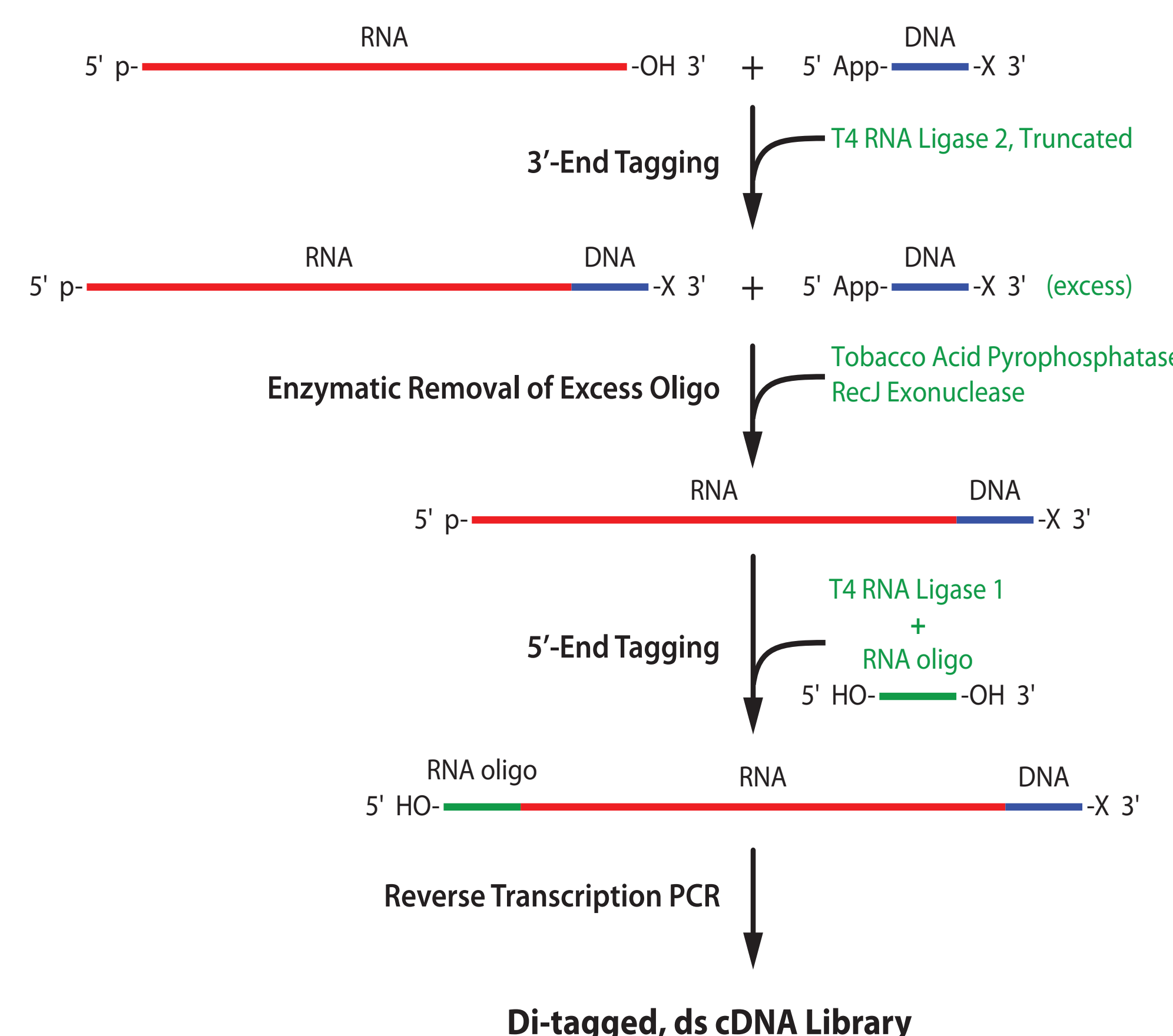


Figure 2. Overview of the method to prepare a di-tagged ds cDNA library.

Preparation of a wheat-germ small-RNA library

Total RNA was isolated from wheat germ and enriched for end-methylated small RNAs using the procedure described in Fig. 1. A ds cDNA library was prepared using the scheme in Fig. 2. The cDNA was cloned into the cloning-ready pCDC1-K™ vector as shown in Fig. 3 and the plasmids from the transformants were sequenced.

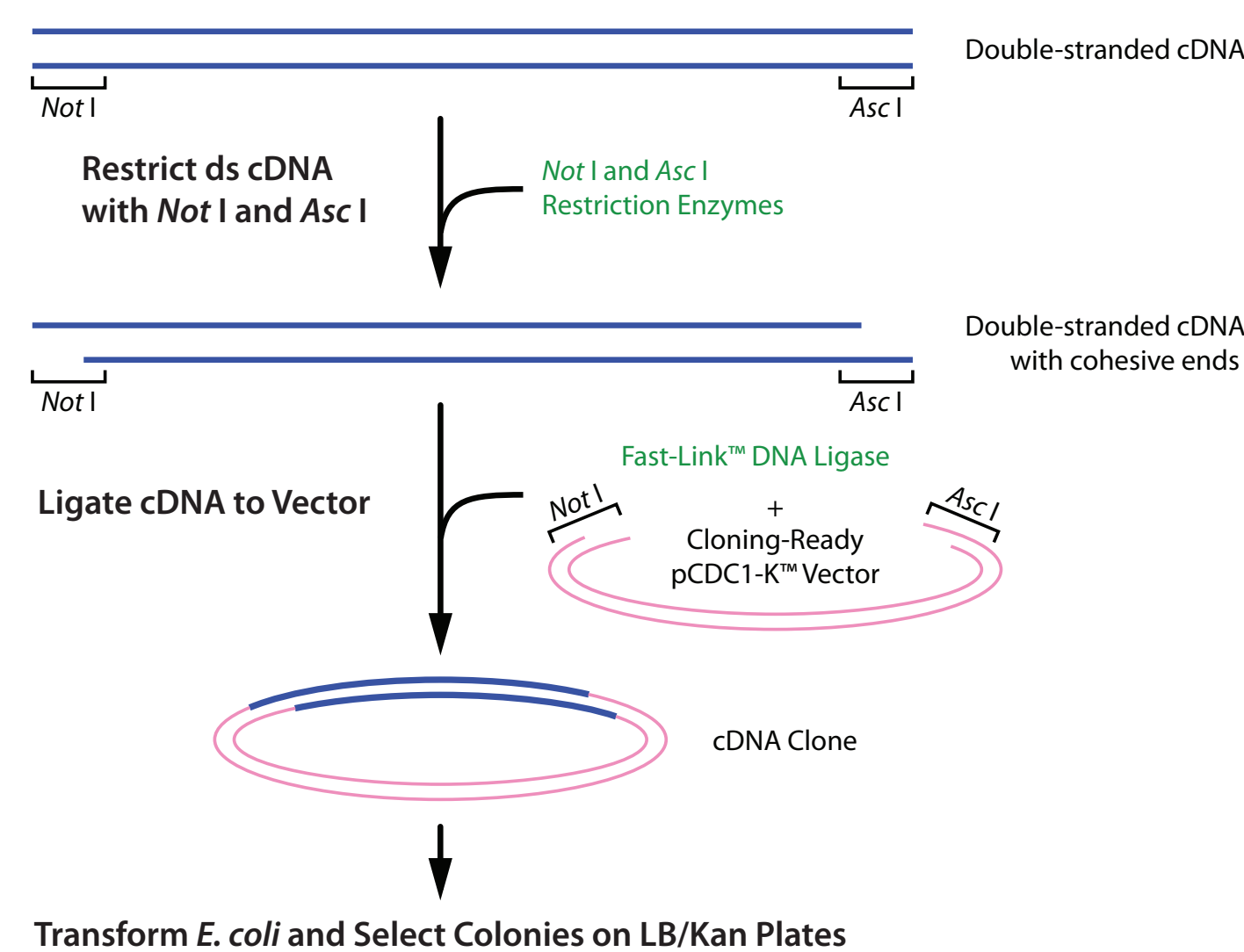


Figure 3. Construction of cDNA library derived from wheat-germ small RNA.

Results

- The majority of the clones had inserts in the 20- to 25-bp size range, with 20 bp and 22 bp being the most abundant (~33%).
- A wheat expressed sequence tag (EST) database [*Triticum aestivum* L. (taxid:4565)] was searched to find the matches to the insert sequence using BLASTn algorithm.
- 77% of the clones mapped to a known EST in the database (Fig. 4).
- 57% of these matches were complementary to the EST (shown in red, Fig. 4).

Clone ID	Sequence of insert	Length	EST No	Match in EST
A1	GGGTGTTGTGTAC	14	CK216011.1	181-165
A2	CTCCCTATATCTAAATCACAAGC	23	GH732540	238-225
A3	CGTTTGCACGCCCGGTTCCGT	20	DR734032.1	373-390
A4	CGTCTCACACCTGGTTCTTAC	22	BM135428.1	170-184
A5	GTCCGATACACTGCTCATC	20	CD891341.1	7-23
A7	CGCTCCCGACGTTCTCA	19	CJ639314.1	208-223
A8	GTTCAGGATAGCTTCACTCCG	22	GR304342.1	570-555
A9	TCGACGAGGGCACACCAAGGT	22	CJ54980.1	466-448
A10	GTTCATATTAGCAGGAGCCGG	23	CK170331.1	519-533
A11	CGTCTGCTCATCTTCTAGCAGGACGGCG	30	CJ64568.1	97-83
A12	GTTCAGATCTTGAGTCTCCTC	22	CJ926800.1	464-450
B1	CACITTTCTCGCATTTGACGTCTTT	26	BJ269146.1	36-50
B2	CGGGCCCGAACCCGTTCCGC	21	DR738394.1	938-955
B3	CTATCGATCCTTTAGAAAA	19	GR884580.1	376-391
B4	TCGACGAGGGCACACCAAGGT	22	CJ54980.1	466-448
B5	GCCTGAGCGGGCGAAGCCAGAG	24	GH729189.1	562-539
B6	CACGGCCCTATCAGCTCAGC	20	CA642933.1	141-155
B7	TACTGTTGATCCTGCGCAGA	21	GH728288.1	194-213
B8	GTCCGACGGATCTGTAATAAA	20	GH732397.1	368-387
B9	CGCCCGGTCAGTGA	14	CJ862113.1	183-195
B10	CCTGCAACATATTTTC	16	CJ551335.1	111-96
B12	GCCTGGGCCTCGGATC	17	CJ718825.1	543-527
C1	CGCTTTCGGGGGCTTCGGCCGGT	25	CD988643.1	269-284
C3	AATACATTTCTCGTACATTTCC	23	CJ959595.1	70-87
C4	TAAACCTTCACAATTCCTT	20	CJ629474.1	251-270
C5	TACCGGTTTCGATCCCGCCAGT	22	BQ38157.1	606-623
C6	GAGACTCGAAAGCGGTTGAAAGC	24	AL29563.1	94-109
C7	GGACGCTTAATCAGCAGGTC	21	BQ171149.1	240-252
C10	TACCGGTTTCGATCCCGCCAGT	22	BQ38157.1	606-623
C11	GTCCGACGGATCTGTAATAAA	22	GH732397.1	368-389
D1	CGAGCTCGGGCCAGCCAGTTCGGA	27	CV522333.1	598-585
D5	CGCTGAAATCTACTCAAG	20	CJ948536.1	286-272
D6	AGCTCGGCCGACAGCCCGGGTTA	25	GH729233.1	365-343
D7	CAAGAGGATCAATTAAGGA	20	CJ848949.1	308-289
D9	TCACCCAGATAGCTCAGTTGGTAGAGCA	29	CK163011.1	772-752
D11	GCCTGGGTTTCGGCCCGCAACCCGAAA	35	GR884582.1	166-144
D12	TCAGCTAATTAGG	14	BJ263746.1	572-584
E5	GGGGCCGACGCTCAATACGG	20	CJ967040.1	229-2316
E6	CGCCGCTGTTCCAAAAA	21	CK211522.1	24-8
E7	CGCCACGATCCACTGAAA	19	GH732397.1	340-354
E8	CGGATACCCCGGAGGTTCCGGCT	24	CV761706.1	442-456
E9	CTGACATCGGTTCCCGGAGTCCGAGGGTT	31	CJ895222.1	253-235
E11	CGTTCCGCTATCCGG	15	CK153851.1	176-188
E12	TCGCGGCGACGGGGCGGTT	20	GH722114.1	273-254
F3	TCACCAAAAGTTTCAAGCCGA	22	BJ298696.1	516-502
F4	CACGACCAACATCAGCGCTCATG	23	CJ832935.1	365-350
F5	GAACTCGGGAACCGCTTCAATTAATCAGTT	32	CJ680174.1	1-27
F8	GGCCGCCGAATTCGAAATCGAGCAGC	25	CJ835070.1	268-254
F9	CGTTCTCGCCACGATTAG	19	DY761224.1	531-518
F11	TTTTCTTCGGGCATTTTCAGACAA	24	BQ579118.1	583-567
F12	TACTGGTCTCGATCCTCCAGAA	24	CO348033.1	382-368
G1	CGTAGACCCGATCAACATGAGAAAAA	31	DR731615.1	722-706
G2	GGAGCATCGGTCGGGATACGG	23	CA871759.1	405-392
G3	GCTCGGGTTCGGCCCGCCACCCCGTCGGC	31	GH729102.1	417-387
G4	TCAGAGCTCATCGA>CAAC	25	CV762554.1	514-494
G5	TCOTAGCTCCGCTTCAGAAAGC	25	CK214723.1	616-601
G6	GTTFAGAGCCGCAAGGCACTTCTCGCA	28	CV768973.1	542-529
G11	TCGATGCCCCGATCGCTCAAGGA	26	EB515442.1	135-122
G12	CGTTCTGTTCCCGCCAGCAC	20	CJ958808.1	655-642
H7	AACAAACCCGACTTCCGGAGGGG	25	CJ965545.1	401-377
H8	CGTCCGACCGGGGCGCTTCC	22	GH723364.1	273-253
H11	GCCTCGCCGACAGCCCGGTTA	23	GH729233.1	364-342
H12	TCCCGGCCCGAACCCTCGGC	22	GH729102.1	408-387

Figure 4. Sequence analysis of di-tagged cDNA clones that mapped to EST database. Clones whose sequences were complementary to those in the EST database are shown in red.

Conclusions

- Using a simple and rapid procedure, di-tagged cDNA libraries can be prepared from small RNA for identification and analysis.
- The small-RNA library preparation method described in this work includes RNA containing capped and triphosphorylated 5' ends in addition to those containing 5'-monophosphorylated ends.
- A simple procedure is described to enrich for 3'-end-methylated small RNA.
- Using these methods, a wheat-germ small-RNA library was constructed. Analysis of sequences revealed that a significant number of small RNAs were complementary to ESTs in a *Triticum aestivum* L. (taxid:4565) database.