Methods 67 (2014) 74–83

Contents lists available at SciVerse ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

High throughput characterizations of poly(A) site choice in plants

Liuyin Ma [a,1], Pratap Kumar Pati [b,c,1], Man Liu [d], Qingshun Q. Li [d,e,*], Arthur G. Hunt [a,*]

[a] Department of Plant and Soil Sciences, University of Kentucky, Lexington, KY, USA
[b] Department of Horticulture, University of Kentucky, Lexington, KY, USA
[c] Department of Biotechnology, Guru Nanak Dev University, Amritsar, India
[d] Department of Botany, Miami University, Oxford, OH, USA
[e] Key Laboratory of Ministry of Education for Coastal Wetland Ecosystem, College of the Environment and Ecology, Xiamen University, Fujian, China

ARTICLE INFO

Article history:
Available online 11 July 2013

Keywords:
Alternative polyadenylation
High throughput sequencing
Gene expression

ABSTRACT

The polyadenylation of mRNA in eukaryotes is an important biological process. In recent years, significant progress has been made in the field of mRNA polyadenylation owing to the advent of the next generation DNA sequencing technologies. The high-throughput sequencing capabilities have resulted in the direct experimental determinations of large numbers of polyadenylation sites, analysis of which has revealed a vast potential for the regulation of gene expression in eukaryotes. These collections have been generated using specialized sequencing methods that are targeted to the junction of 3′-UTR and the poly(A) tail. Here we present three variations of such a protocol that has been used for the analysis of alternative polyadenylation in plants. While all these methods use oligo-dT as an anchor to the 3′-end, they differ in the means of generating an anchor for the 5′-end in order to produce PCR products suitable for effective Illumina sequencing; the use of different methods to append 5′ adapters expands the possible utility of these approaches. These methods are versatile, reproducible, and may be used for gene expression analysis as well as global determinations of poly(A) site choice.

1. Introduction

The introduction of the second generation sequencing technology has greatly facilitated the study of mRNA polyadenylation. This is indicated by numerous studies of high through-put studies of poly(A) sites [1–3]. The results provide detailed cellular transcript poly(A) profiles in different developmental stages, cell types, or disease states. From this growing body of literature, the extent of alternative polyadenylation and its relevance to gene expression has been revealed.

Among next generation sequencing platforms, RNA-seq using Illumina sequencing-by-synthesis technology was designed to efficiently characterize transcriptomes. However, regular RNA-seq protocols can only collect a small proportion of poly(A) sites due to the randomness of RNA fragmentation. To enhance the recovery of poly(A) sites, a number of poly(A) tag sequencing (PAT-seq) protocols have been developed that target only the junctions of the 3′-UTR and poly(A) tails. Systematic characterization of almost all transcripts in a sample is made possible by such technologies, the impact of which can be easily seen by comparing how many poly(A) sites were available before [4–6] and after the advent of PAT-seq methodologies (e.g., [3,7,8]). High-throughput determinations of poly(A) site usage has in turn led to a growing appreciation for the role of alternative polyadenylation (APA) in numerous processes including oncogene regulation [9,10], development [11–14], and cellular differentiation [15,16].

In plants, APA has been shown to contribute to the regulation of flowering time [17–21], oxidative stress responses [22,23], and the expression of genes involved in RNA processing [24]. However, additional biological processes that involve APA are awaiting discovery. With the sequences of many plant genomes now available, the application of PAT-seq technologies to global transcriptome analysis of numerous crop plants and other species is now possible. We have developed a set of PAT-seq protocols that takes the advantage of several properties of Illumina sequencing, namely short (75–100 nts) and massive amounts of reads (10–100 million/lane). In this report the performance and reproducibility of these methods are discussed.

2. Methods

2.1. Biological sample preparation and RNA isolation

Arabidopsis leaf and seed RNA was isolated from young plants as described previously [8]. For RNA from imbibed seeds,

* Corresponding authors. Address: Department of Botany, Miami University, Oxford, OH 45056, USA (Q.Q. Li). Address: Department of Plant and Soil Sciences, University of Kentucky, Lexington, KY 40546-0312, USA (A.G. Hunt).
E-mail addresses: liq@miamioh.edu (Q.Q. Li), aghunt00@uky.edu, aghunt@aol.com (A.G. Hunt).
[1] The first two authors made equal contributions.

1046-2023/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymeth.2013.06.037

approximately 0.10 g of wild type *Arabidopsis thaliana* (Col-0) seeds were placed on top of two layers of Whatman no.1 filter paper and wetted with distilled water in a BD Falcon Bacteriological Petri dish (standard style dish, 100 × 15 mm). The seeds were incubated at 4 °C for 3 days to alleviate dormancy and then transferred to 25 °C under a light intensity of ~130 μmole m^{-2} sec^{-1}. After 48 h, RNA was isolated for further manipulation.

Three related strategies for PAT preparation are described in the following. These strategies share a common first step, the isolation of RNA from an appropriate source. For the studies described in this report, RNA was isolated from *A. thaliana* seed or from young *A. thaliana* leaf tissue. These isolation procedures follow those described previously [25–28]. Briefly, leaf RNA was isolated using the Trizol reagent (Life Technologies, Carlsbad, CA) for extraction and subsequent precipitation using ethanol [25–27]. Seed RNA was isolated using a hot borate extraction [28]. Subsequently, total RNA (between 2 and 10 μg, in a total volume of 50 μl) was treated with 2 U of RNase-free DNase I (Thermo Scientific, #EN0521) following the manufacture's protocol. DNAse I-treated RNA was purified using the RNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA; #74904). RNA samples were quantified using a Nanodrop instrument (Biotek, Winooski, VT, USA; Model: Synergy™ HT), and were occasionally assessed using a Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA; Model 2100).

2.2. Preparation of poly(A) tags (PATs)

Three related procedures are described in the following. These procedures are termed as Method A, Method B1, and Method B2, respectively (Fig. 1). Methods B1 and B2 share a common set of steps up to a point, after which they diverge. The rationales for the different variations are discussed in the Results and Discussion.

2.2.1. Method A – preparation of PATs using restriction enzymes

The overall goal for PAT-Seq is to generate short cDNA tags that query the mRNA-poly(A) junction. One approach to this end is to anchor a sequencing tag to the 5′ end of a cDNA (corresponding to the polyadenylated 3′ end of the mRNA), then digest full-length cDNAs with restriction enzymes that recognize four base pair sequences, and subsequently append linkers to the ends left by the respective restriction enzymes. The truncated and adapted cDNA tags are then amplified submitted for sequencing. This approach is illustrated in Fig. 1; the details have been published elsewhere [8] and are only summarized here.

Briefly, the DNase I-treated RNA (2–10 μg) is poly(A)-enriched using oligo-dT magnetic beads. The poly(A) RNA is used as a template to produce cDNA. For this, one of the RT-PE1 or RT-PE2 series of primers (Table 1) is used, along with the SWITCH1.1 primer (Table 1). The cDNA is purified using a QIAquick PCR Purification Kit (Qiagen, #28106). This purified cDNA is converted to double-stranded form by PCR or the Klenow fragment of DNA polymerase I. For PCR the first strand cDNA is used as a template for PCR using the PE-RTbio and SWITCH1.2 primers (Table 1). Two 25 μl reactions using 1 μl (each) of the first strand cDNA are set up. Second strands are synthesized using 12–18 cycles of PCR (each cycle consisting of 95 °C melting for 15 s, 60 °C annealing for 30 s; 2 min extension at 72 °C, with the cycling preceded by a dissociation step of 95 °C for 30 s, and followed by an extension step of 72 °C for 10 min). For Klenow reactions, all of the first strand cDNA is added to a 50 μl reaction containing 100 pmol of the second strand primer (SWITCH1.2), 1X NEB restriction enzyme buffer 2, 1 mM dNTPs, and five units of Klenow DNA polymerase (NEB). These reactions are incubated for 2 h at 37 °C. In both cases, double-stranded cDNA is purified using a QIAquick kit and eluted into 50 μL.

The purified cDNA is then digested with one of two restriction enzymes – Nla III or Tai I (Fermentas). The digested cDNA is ligated to one of two Y-adapters that consisted of primers with partial

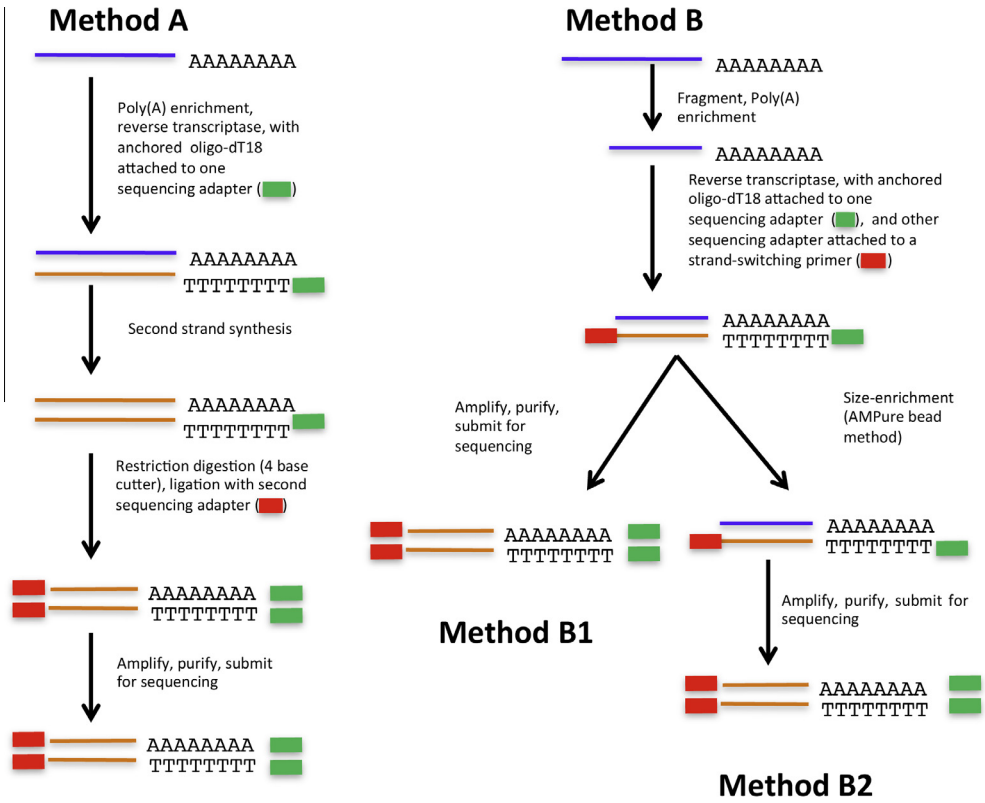

Fig. 1. Illustration of the three methods of PAT-Seq that are described in this report.