An improved method for the rapid isolation of RNA from *Arabidopsis* and seeds of other species high in polyphenols and polysaccharides

Running title: RNA extraction from seeds high in polysaccharides

Steven Footitt, Sajjad Awan, and William E. Finch-Savage*

School of Life Sciences, Wellesbourne Campus, Warwick University, Wellesbourne, Warwick CV35 9EF, United Kingdom

*Corresponding Author: W.E. Finch-Savage

bill.finch-savage@warwick.ac.uk

Tel no. 024 7657 4968

Fax. No. 024 7657 4500

ORCID IDs

S. Footitt: 0000-0001-6644-0621

W.E. Finch-Savage: 0000-0003-2056-3858

Keywords: Arabidopsis thaliana, Brassica, Polysaccharide, RNA extraction, seeds.

Financial support: This work was supported by a grant from the Biotechnology and Biological Sciences Research Council-UK (BB/1022202/1) awarded to WF-S and SF. SA was supported by European Union Framework Programme 7-Knowledge-Based Bio Economy grant 311480 EcoSeed awarded to WF-S.

Abstract

Seeds are notoriously high in polyphenols and polysaccharides, which reduce RNA quality and yield, and interfere with downstream applications. We present simple modifications to a rapid RNA extraction protocol for use with seeds. The inclusion of polyethylene glycol in place of polyvinylpyrrolidone reduced polyphenol and polysaccharide contamination. In addition, replacing NaCl with KCl improved the RNA yield from Arabidopsis seeds still bound by mucilage. On extraction of Arabidopsis seed recovered from field soils clean RNA pellets with no accompanying gelatinous matrix (polysaccharide) were seen, with A260/230 ratios greater than 1.8 confirming the lack of polysaccharide carry over. When *Brassica oleracea* and *Sinapis arvensis* seeds were extracted 260/230 ratios greater than 1.8 were seen. RNA yields in excess of 10 µg /100 mg seed suitable for RT-QPCR were obtained.

Keywords: Arabidopsis thaliana, Brassica, Polysaccharide, RNA extraction, seeds.

Introduction

The isolation of high quality RNA from seeds is notoriously difficult due to the high levels of polyphenols and polysaccharides found in seeds and associated mucilage of many species. The removal of phenolic compounds is important as they bind proteins and may interfere with downstream applications as can polysaccharides and other contaminating compounds that remain. Various published extraction protocols address the issue of high levels of polyphenols, polysaccharides and secondary metabolites present in plant tissues (Asif et al. 2000; Chang et al. 1993; Sharma et al. 2003; Wan and Wilkins 1994; Birtic and Kranner, 2006; Xu et al. 2010). The hot borate method of Wan and Wilkins (1994) has been widely used for the elimination of phenolics and polysaccharides during the extraction of high quality RNA from seeds for microarray studies (Cadman et al. 2006; Carrera et al. 2008). However, despite its widespread use it is time consuming when faced with large numbers of samples. The protocol of Chang et al. (1993) developed for the extraction of RNA from pine trees employed a high salt concentration extraction buffer with cetyltrimethylammonium bromide (CTAB) and polyvinylpyrrolidone (PVP) followed by LiCl precipitation of RNA which combine to reduce polyphenol and polysaccharide carry over. This protocol adapted by Xu et al. (2010) produced a protocol economical on both reagents and time. We report modifications to the protocol of Xu et al., (2010) that retain its speed, economy and single precipitation step while further reducing the presence of polyphenolic and polysaccharide contaminants in the resulting RNA.

Materials and methods

Seed material

Seeds of the *Arabidopsis thaliana* ecotype Cape Verdi Island (Cvi) (glasshouse grown) were treated in two ways. (A) Seeds were hydrated on water for 24 hours at 20°C in boxes (12 X 8 cm) (Stewart Plastics Ltd, UK) containing 2 sheets of 3M chromatography paper, seeds were then removed from germination boxes and surfaced dried with filter paper. (B) Seeds were buried in the field in October at a depth of 5 cm for one month and recovered as described previously (Footitt et al. 2011). In both cases, aliquots of 50-125 mg of seeds were placed in 2 ml Eppendorf tubes, frozen in liquid nitrogen and stored at -80°C until extracted. Seeds of other members of the Brassicaceae: *Brassica olearacea,* the double haploid Chinese kale var. alboglabra, A12DHd, seeds were grown in a glasshouse at 16/22^{o}C, night/day temperature and stored at -20^{o}C soon after harvest (Awan et al. 2018) (80 mg samples); and *Sinapis arvensis* (Herbiseed, UK) (125 mg samples) were hydrated as in (A) above.

Solutions and reagents

All plastic ware and reagents were certified nuclease free. 50% PEG 20,000 solution was obtained from Sigma. Mix CTAB and NaCl or KCl together before adding liquid as this aids solubility. β-Mercaptoethanol is added immediately prior to use and the buffer then heated to 65°C. Perform extractions in 2 ml Eppendorf tubes in a fume cupboard. The protocol is for up to 125 mg hydrated seeds. For larger quantities adjust volumes appropriately.

RNA extraction buffer A: 100 mM TRIS-HCl pH 8.0, 25 mM EDTA, 2 M NaCl, 2% (w/v) CTAB, 2 % (v/v) PEG 20,000, nuclease free water and 2% (v/v) β-Mercaptoethanol.

RNA extraction buffer B: As in (A) but 2 M KCl replaces NaCl.

Other solutions required; Chloroform, 5M NaCl, and 75% ethanol. For RNA precipitation Isopropyl alcohol and a solution of 1.2 M NaCl/ 0.8 M Tri-Na Citrate dihydrate.

RNA extraction protocol

Add 0.45 ml hot extraction buffer to the sample and homogenise immediately using a pellet pestle attached to an electric drill. Immediately add another 0.45 ml of hot extraction buffer, mix and incubate in a water bath or heating block at 65°C for 15 minutes only. Longer incubation times are detrimental. Place samples on ice.

Add 0.5 ml of chloroform, mix well (vortex briefly) and centrifuge at 12,000 rpm/ 4°C/10 min. Transfer upper supernatant to new tube and add 0.133 ml 5M NaCl and 0.4 ml of chloroform mix well and centrifuge at 12,000 rpm/ 4°C/10 min. Transfer upper supernatant to new tube and repeat the previous step.

Collect upper supernatant and add to fresh tube. Add ½ volume of isopropanol and ½ volume of 1.2M NaCl/0.8M tri-Na citrate dihydrate mix gently and store at room temperature for 15 minutes. Recover RNA by centrifugation at 12,000 rpm/ 4°C/10 min. Carefully discard the supernatant and wash the pellet with ice cold 75% ethanol, centrifuge as above and remove ethanol. Air dry pellet for 10 minutes and re-dissolve in 120 µl nuclease free water and store at -80°C.

RNA analysis

The quantity and quality of RNA was determined by measuring the absorbance at 230, 260 and 280 nm using a Nanodrop spectrophotometer (Thermo Scientific, UK). Quality was evaluated by comparing A260/A280 ratio (protein) and A260/A230 ratio (polysaccharide). The quality of RNA extracted with buffer B was also determined using an Agilent 2100

Bioanalyzer (Agilent Technologies, Santa Clara, USA). Arabidopsis data was subjected to analysis of variance to determine significant differences in 260/230 ratios and RNA yields.

Arabidopsis RNA was treated with RNase-free DNase I (Roche Diagnostics) to remove contaminating genomic DNA prior to cDNA synthesis and QPCR which was performed as described elsewhere (Footitt et al 2011 and 2015). Interference by genomic DNA during QPCR was further guarded against by using intron spanning primer pairs designed using Roche Life Science, Assay design centre web site (https://lifescience.roche.com/en_gb/brands/universal-probe-library.html#assay-design-center). Reference genes used for normalisation of the data were At4g34270 (*Tip41-like*) and At4g12590 (see Footitt et al. 2015). Expression of the following genes whose proteins are involved in dormancy regulation was determined in Arabidopsis seeds; *CYP707A2* (At2g29090), *DOG1* (At5g45830), *GA2ox2* (At1g30040), *GA3ox1* (At1g15550), *MFT* (At1g18100), *NCED6* (At3g24220). For primer sequences see Footitt et al. (2011). Data are presented as the ratio of the geometric mean of the gene of interest/ geometric mean of both reference genes**.**

Results and Discussion

Previously, we found that commercially available RNA extraction kits produced low RNA yields and low A260/230 ratios indicating contamination with phenolics and polysaccharides. This is problematic when the quantity and availability of seeds is limited. Traditionally poplyvinylpyrrolidone (PVP) reduces the polyphenol contamination during RNA extraction. Whereas, high molecular weight polyethylene glycol (PEG) reduces both polyphenolic and polysaccharide contamination of RNA (Gehrig et al. 2000). We therefore compared the effect of PVP and PEG20000 on the resulting quality of RNA from hydrated

Arabidopsis seeds extracted using extraction buffer A incorporating 2% PVP-K90 or 2% PEG20000. The resulting RNA yields were similar, but the A260/230 ratio was significantly higher indicated PEG20000 was more efficient in reducing polyphenol and polysaccharide contamination (Table 1).

The RNA pellets using this protocol (Buffer A) were still gelatinous indicating the presence of polysaccharides. Additional post extraction steps to precipitate polysaccharides include precipitation with 3 M Na acetate (pH 5.2) and ethanol (Asif et al. 2000). However, the introduction of additional steps potentially reduces yield and increases the extraction time. To address this we looked at alternative protocols for the precipitation of polysaccharides. High NaCl concentrations have been proposed as a method for polysaccharide removal from DNA (Fang et al 1992). However, KCl is a more efficient precipitant of both neutral and acidic polysaccharides. The efficiency of this increases under alkaline conditions (ionisation of acidic polysaccharides) and in the presence of ethanol as alcohols reduce polysaccharide solubility (Smidsrod and Haug, 1967; Smidsrod et al. 1967). CTAB is an effective precipitant of acidic polysaccharides (Bera et al 1955; Smidsrod and Haug, 1967; Murray and Thompson, 1980). We therefore modified the extraction buffer by replacing NaCl with KCl to increase the efficiency of polysaccharide precipitation to produce RNA extraction buffer B. The presence of β-mercaptoethanol may well further increase the efficiency of KCl.

When buried Arabidodpsis seeds were extracted with Buffer A (NaCl containing) the A260/230 ratio was 1.6, this increased to 1.88 ± 0.04 following precipitation with ½ volume of 5M NH4oAc and 2.5 volumes of ethanol at -20°C for at least 1 hour, but this did not remove the gelatinous matrix. Only extraction with Buffer B (containing KCl) achieved this resulting in a fragile white RNA pellet and eliminating the NH4oAc precipitation step. Using

buffer B significantly increased RNA yields, but not the A260/230 ratio when unburied seeds still bearing mucilage were extracted. In the case of buried Arabidopsis seeds where the mucilage coat is lost RNA yield was not increased, but the A260/230 ratio was greater than 2. Similarly, when *Brassica oleracea* and *Sinapis arvensis* seeds were extracted the A260/230 ratio approached 2.0 resulting in a fragile white RNA pellet (Table 2). The addition of KCl to the extraction buffer effectively removes polysaccharides and other contaminants prior to the final RNA precipitation step. Whereas they would normally co-precipitate in the RNA precipitation buffer.

Further analysis of RNA quality using an Agilent, 2100 Bioanalyzer produced RNA integrity numbers (RIN values; 1= poor quality, 10 = high quality) of 6.4 - 7.0 for unburied Arabidopsis and 7.2 – 7.5 for Arabidopsis recovered from the field soils. *Brassica oleracea* and *Sinapis arvensis* had values ranging from 6.4 – 6.7 and 6.4 – 6.5 respectively. When electropherograms were compared for RNA from unburied and buried Arabidopsis seeds increased background signal is seen in the unburied example reflecting the level of polysaccharide remaining in these samples (Fig. 1). Electropherograms of *B.oleracea* and *S. arvensis* RNA are similar to that of buried Arabidopsis seeds reflecting their low level of polysaccharide. Gel images generated from these electropherograms also reflect this reduced contamination level. (Fig. 2). The yield of RNA was within the range reported previously for Arabidopsis seeds (Vincent and Delseny, 1999; Birtic and Kranner, 2006). Results similar to those of Birtic and Kranner (2006) were obtained for the reduction in polysaccharide contamination for buried Arabidopsis seeds , and for *Brassica oleracea* and *Sinapis arvensis*. However, the advantage of the protocol outlined here is the speed with which large numbers of samples can be processed. The Arabidopsis RNA produced was of

168 suitable quality for downstream application such RT-QPCR (Fig. 3). The pattern of gene
169 expression shown in the example is representative of seeds in low dormancy prior to burial
170 i.e. relatively high *CYP707A2* expression (ABA catabolism) and low *NCED6* expression (ABA
171 synthesis). The buried seeds show the reverse pattern of expression consistent with induced
172 deep dormancy following burial (Footitt et al. 2011).

173 In conclusion, we show that adaptation of the CTAB protocol to minimise polyphenols and
174 polysaccharide contaminants in the early extraction steps retains the speed and high
175 throughput aspects of the protocol, but also has the advantage of highly improved quality of
176 the resultant RNA. Furthermore, we show the successful application of the protocol to seeds
177 of other Brassica species.

178 **Financial support:** This work was supported by a grant from the Biotechnology and
179 Biological Sciences Research Council-UK (BB/1022202/1) awarded to WF-S and SF. SA was
180 supported by European Union Framework Programme 7-Knowledge-Based Bio Economy
181 grant 311480 EcoSeed awarded to WF-S.

182 **Author contributions:** Protocol devised by SF, Experiments performed by SF and SA.
183 Manuscript written by SF, SA, and WEF-S.

184 **Conflicts of interest:** None

185 **References**

186 Asif MH, Dhawan P and Nath P. (2000) A simple procedure for the isolation of high quality
187 RNA from ripening banana fruit. *Plant Molecular Biology Reporter* **18,** 109-115.

188 Bera BC, Foster AB and Stacey M. (1955) Observations on the properties of
189 cetyltrimethylammonium salts of some acidic polysaccharides. *Journal of the Chemical*
190 *Society (Resumed)* 3788-3793.

191 Birtic S and Kranner I. (2006) Isolation of high-quality RNA from polyphenol-,polysaccharide-
192 and lipid-rich seeds. *Phytochemical Analysis* **17,** 144–148.

193 Cadman CS, Toorop PE, Hilhorst HW and Finch-Savage WE. (2006) Gene expression profiles
194 of *Arabidopsis* Cvi seeds during dormancy cycling indicate a common underlying dormancy
195 control mechanism. *Plant Journal* **46,** 805-822.

196 Carrera E, Holman T, Medhurst A, Dietrich D, Footitt S, Theodoulou FL and Holdsworth MJ.
197 (2008) Seed after-ripening is a discrete developmental pathway associated with specific
198 gene networks in *Arabidopsis*. *Plant Journal* **53,** 214-224.

199 Chang S, Puryear J and Cairney J. (1993) A simple and effcient method for isolating RNA from
200 pine trees. *Plant Molecular Biology Reporter* **11,** 113-116.

201 Fang G, Hammar S and Grumet R. (1992) A quick and inexpensive method for removing
202 polysaccharides from plant genomic DNA. *Biotechniques* **13,** 52-4.

203 Footitt S, Douterelo-Soler I, Clay H and Finch-Savage WE. (2011) Dormancy cycling in
204 *Arabidopsis* seeds is controlled by seasonally distinct hormone-signaling pathways.
205 *Proceedings of the National Academy of Sciences of the USA* **108,** 20236-20241.

206 Footitt S, Huang Z, Clay HA, Mead A and Finch-Savage WE. (2013) Temperature, light and
207 nitrate sensing coordinate *Arabidopsis* seed dormancy cycling, resulting in winter and
208 summer annual phenotypes. *Plant Journal* **74,** 1003-1015.

209 Footitt S, Müller K, Kermode AR and Finch-Savage WE. (2015) Seed dormancy cycling in
210 *Arabidopsis*: chromatin remodelling and regulation of DOG1 in response to seasonal
211 environmental signals. *Plant Journal* **81,** 413-425.

212 Gehrig HH, Winter K, Cushman J, Borland A and Taybi T. (2000) An improved RNA isolation
213 method for succulent plant species rich in polyphenols and polysaccharides. *Plant Molecular*
214 *Biology Reporter* **18,** 369-376.

215 Murray MG and Thompson W. (1980) Rapid isolation of high molecular weight plant DNA.
216 *Nucleic Acids Research* **8,** 4321-4326.

217 Sharma AD, Gill PK and Singh P. (2003) RNA isolation from plant tissues rich in
218 polysaccharides. *Analytical Biochemistry* **314,** 319-321.

219 Smidsrød O and Haug A. (1967) Precipitation of acidic polysaccharides by salts in ethanol–
220 water mixtures. In *Journal of Polymer Science, Polymer Symposia* **16,** 1587-1598. Wiley
221 Subscription Services, Inc., A Wiley Company.

222 Smidrod O, Larsen B, Pernas AJ and Haug A. (1967) The Effect offAlkali Treatment on the
223 Chemical Heterogeneity and Physical Properties of Some Carrageenans. *Acta Chemica*
224 *Scandinavica* **21,** 2585-2598.

225 Vicient CM and Delseny M. (1999) Isolation of Total RNA from *Arabidopsis thaliana* Seeds.
226 *Analytical Biochemistry* **268,** 412–413.

227 Xu J, Aileni M, Abbagani S and Zhang P. (2010) A reliable and efficient method for total RNA
228 isolation from various members of spurge family (*Euphorbiaceae*). *Phytochemical Analysis*
229 **21,** 395-398.

230 Wan CY and Wilkins TA. (1994) A modified hot borate method significantly enhances the
231 yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). *Analytical Biochemistry* **223,**
232 7-12.

233

Table 1. Comparing the impact of Polyvinylpyrrolodone (PVP-K90) and Polyethylene glycol (PEG 20000) on polyphenol and polysaccharide contamination of RNA. RNA extracted from unburied Arabidopsis seeds using buffer A. Polyphenol and polysaccharide contamination as reflected by the 260/230 ratio. Mean values for A260/230 and Total RNA with different letters are significantly different ($P < 0.05$).

Test	A260/280	A260/230	Total RNA (µg)
PVP-K90 N = 6	2.105 ± 0.02	1.315 ± 0.11	13.6 ± 1.94
PEG 20000 N = 6	2.130 ± 0.01	1.610 ± 0.03[a]	14.5 ± 2.23

Table 2. Comparison of RNA extraction with NaCl and KCl containing buffers from Arabidopsis and other species. Extration used buffer A (NaCl conatianing) and Buffer B (KCl containing). Data from buried Arabidopsis seeds extracted with buffer A is from RNA prior to precipitation with ½ volume of 5M NH4oAc and 2.5 volumes of ethanol at -20°C for at least 1 hour for RT-QPCR in Footitt et al., 2013. In Arabidopsis, mean values for A260/230 and Total RNA with different letters are significantly different ($P < 0.05$).

Species	Extraction method	A260/280	A260/230	Total RNA (µg)	Sample size (mg)
Arabidopsis thaliana (Cvi) Seeds unburied	Buffer A (NaCl) n = 4	2.14 ± 0.14	1.63 ± 0.04	15.53 ± 3.2	100
	Buffer B (KCl) n = 4	2.09 ± 0.01	1.35 ± 0.04	29.5 ± 2.3[a]	100
Arabidopsis thaliana (Cvi) Seeds buried	Buffer A (NaCl) n = 36	1.99 ± 0.01	1.6 ± 0.04	20.4 ± 1.2	50 - 75
	Buffer B (KCl) n = 6	2.07 ± 0.01	2.19 ± 0.09[a]	16.4 ± 0.7	50 - 75
*Brassica oleracea (*var. alboglabra)	Buffer B (KCl) n = 4	2.11 ± 0.03	1.96 ± 0.08	10.86 ± 2.01	80
Sinapis arvensis	Buffer B (KCl) n = 4	2.13 ± 0.01	1.89 ± 0.03	39.04 ± 6.94	125

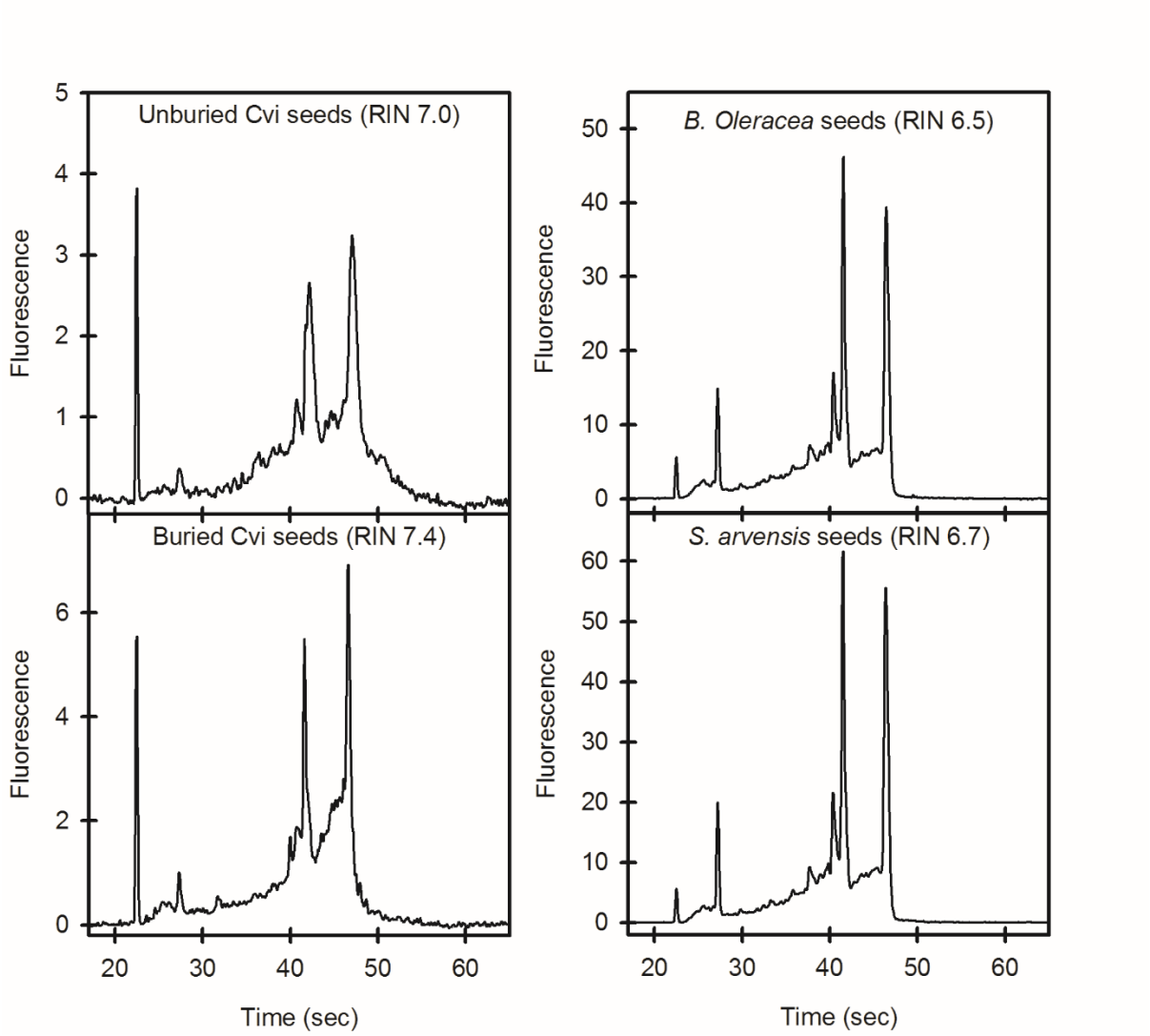
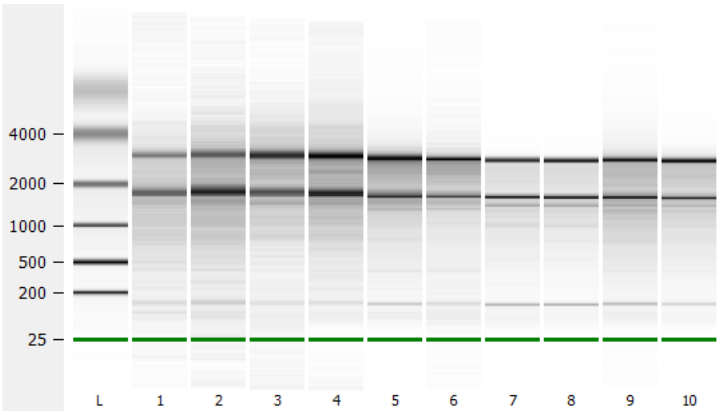


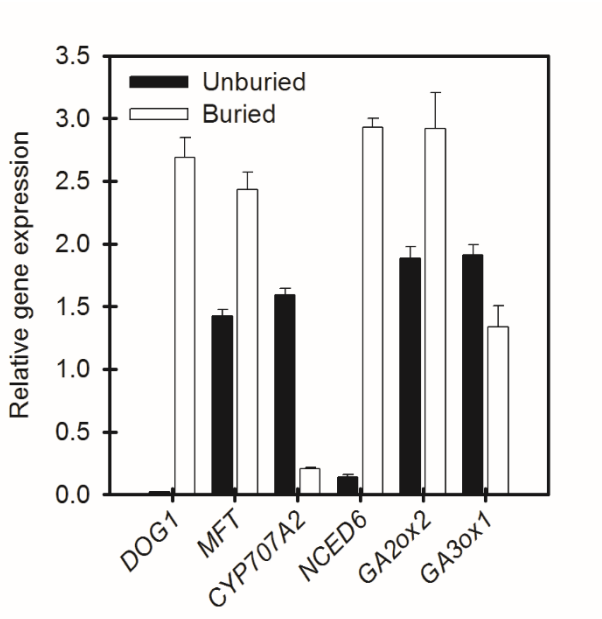
Figure 1. Electropherograms of RNA extracted from seeds with buffer B (KCl containing).

Electropherograms of Arabidopsis Cvi seeds used 40 ng μl^{-1} of RNA, and those of B. Oleracea and S. arvensis used 200 ng μl^{-1} of RNA.



255

256 **Figure 2. Electropherogram generated gel image of RNA from Arabidopsis, *B.oleracea*, and**
257 ***S. arvensis* seeds extracted with buffer B (KCl containing).** Ladder (L), unburied Arabidopsis
258 Cvi seeds (1 -3), buried Arabidopsis Cvi seeds (4 -6), *B. Oleracea* seeds (7 – 8), *S. Arvensis*
259 seeds (9 – 10).



260

261 **Figure 3. Relative gene expression in Arabidopsis Cvi seed RNA extracted using KCl based**
262 **extraction buffer prior to (unburied) and after recovery from one months burial in field**
263 **soils (buried).** Samples are representative of seeds in low dormancy prior to burial and deep
264 dormancy following burial as seen in Footitt et al 2011. The genes *DOG1* and *MFT* are

265 involved in dormancy induction and maintenance. *CYP707A2* and *NCED6* are genes involved
266 abscisic acid catabolism and synthesis respectively. While *GA2ox2* and *GA3ox 1*are genes of
267 gibberellin catabolism and synthesis respectively (See Footitt et al. 2011; 13). Data are
268 means ± standard error of the mean (n = 4).