Molecular characterization of the acquisition of longevity during seed maturation in soybean

Juliana Joice Pereira Lima¹,²*, Julia Buitink²*, David Lalanne², Rubiana Falopa Rossi¹,², Sandra Pelletier², Edvaldo Aparecido Amaral da Silva¹*; Olivier Leprince²*

¹ Faculdade de Ciências Agronômicas, Universidade Estadual Paulista Júlio de Mesquita Filho, Botucatu, São Paulo State, Brazil, ² Institut de Recherche en Horticulture et Semences, INRA, Agrocampus Ouest, Université d’Angers, SFR QUASAV, Beaucouzé, France

These authors contributed equally to this work.
* olivier.leprince@agrocampus-ouest.fr (OL); amaraldasilva@fca.unesp.br (EAADS)

Abstract

Seed longevity, defined as the ability to remain alive during storage, is an important agro-economic factor. Poor longevity negatively impacts seedling establishment and consequently crop yield. This is particularly problematic for soybean as seeds have a short lifespan. While the economic importance of soybean has fueled a large number of transcriptome studies during embryogenesis and seed filling, the mechanisms regulating seed longevity during late maturation remain poorly understood. Here, a detailed physiological and molecular characterization of late seed maturation was performed in soybean to obtain a comprehensive overview of the regulatory genes that are potentially involved in longevity. Longevity appeared at physiological maturity at the end of seed filling before maturation drying and progressively doubled until the seeds reached the dry state. The increase in longevity was associated with the expression of genes encoding protective chaperones such as heat shock proteins and the repression of nuclear and chloroplast genes involved in a range of chloroplast activities, including photosynthesis. An increase in the raffinose family oligosaccharides (RFO)/sucrose ratio together with changes in RFO metabolism genes was also associated with longevity. A gene co-expression network analysis revealed 27 transcription factors whose expression profiles were highly correlated with longevity. Eight of them were previously identified in the longevity network of Medicago truncatula, including homologues of ERF110, HSF6AB, NFXL1 and members of the DREB2 family. The network also contained several transcription factors associated with auxin and developmental cell fate during flowering, organ growth and differentiation. A transcriptional transition occurred concomitant with seed chlorophyll loss and detachment from the mother plant, suggesting the activation of a post-abscission program. This transition was enriched with AP2/EREBP and WRKY transcription factors and genes associated with growth, germination and post-transcriptional processes, suggesting that this program prepares the seed for the dry quiescent state and germination.
Introduction
Soybean is one of the most important oil crop species for food, feed and a range of industrial applications. Producing highly vigorous seeds is a key lever to increase crop production. Seed longevity, defined as the ability to remain alive during storage under dry conditions, is an important agronomic factor in the preservation of seed fitness after harvest [1]. Poor longevity leads to unexpected losses in seed viability during storage and negatively impacts seedling establishment and crop yield [1, 2]. This is particularly problematic for soybean as seeds have a short lifespan during storage, especially in humid and tropical environment [2–4]. In addition, longevity is pivotal to ensure the preservation of our genetic resources through dry seeds of crops and wild species [5, 6].

Longevity is conferred by the ability to stabilize the biological entity for long periods of time by the formation of an amorphous highly viscous, solid-like matrix (i.e. a glassy state) in the cells that suspends integrated metabolic activities and severely slows down deteriorative reactions [2, 7, 8]. Seed longevity is also attributed to a range of protective compounds [9, 10], including non-reducing soluble sugars (sucrose (Suc) and raffinose (Raf) family oligosaccharides, RFO [11, 12]) and a set of late embryogenesis abundant (LEA) proteins and heat shock proteins (HSP) [13–15]. Together with sugars, both types of proteins act as chaperones and molecular shields to prevent protein denaturation and membrane destabilization during drying and in the dry state. Longevity is also conferred by antioxidant mechanisms that limit oxidation of lipids, proteins and nucleic acids during storage such as glutathione [16 and references therein], tocopherols [17], flavonoids that are present in the seed coat [18] and lipocalins [19]. Several repair mechanisms also contribute to longevity when they are activated during seed imbibition to fix damage that occurred to proteins and DNA during storage [20, 21]. Next to protection and repair, an impaired degradation of chlorophyll appears to negatively affect longevity [12, 22]. The presence of chlorophyll is considered as an indicator of immaturity but how it affects longevity remains unsolved.

To be commercially successful, crop seeds should be harvested when longevity reaches its maximum [1, 10]. In legumes, longevity is progressively acquired during seed maturation from seed filling onwards [10, 23–25]. In soybean, there exist conflicting data as to whether seed longevity reaches a maximum at seed filling [4, 26] or later during maturation [25, 27]. Delaying harvest to obtain maximum longevity increases the risk of exposing mature seeds to rapid deterioration in the field due to high humidity and temperature [25–27]. The mechanisms regulating the acquisition of seed longevity and vigor during late maturation remain poorly understood. Hence, the dearth of knowledge of late seed maturation programs remains an obstacle to commercial production of high quality seeds.

In soybean, transcriptome studies generated a wealth of data describing seed development during embryogenesis and filling [28–33]. However, there is little information on transcriptome changes occurring in soybean during late seed maturation, when longevity is acquired. The purpose of this study was to provide a physiological and molecular characterization of the soybean seed maturation using RNAseq to obtain a comprehensive overview of the regulatory genes that are potentially involved in seed longevity. Our data show that a developmental program is activated during late maturation that is more complex than a simple arrest of seed filling and seed drying.

Material and methods
Plant material and seed physiology
Soybean plants (Glycine max L. cv. BRS284) were grown using standard planting and cultural techniques in the experimental farm of the São Paulo State University in Botucatu (Brazil)
during two consecutive years (2012/2013; 2013/2014). Seed development was monitored using phenological stages [34] and flower tagging (for 2014 only). Intermediate phenological stages between at R7 and R8 (for both years) were incorporated in order to obtain a higher precision of the time course of acquisition of physiological quality attributes after seed filling (S1 Table). The relationship between seed age and phenological stages is shown in Fig 1. Pods were manually removed and seeds were used immediately. For germination assays, 4 replicates of 25 seeds were imbibed in moistened paper rolls at 25°C. To test desiccation tolerance, seeds harvested at different stages were subjected to fast drying by incubation at 40% RH at 30˚C until they reached a moisture content of 10% (dry weight basis, DW) (i.e. after 2 days). Thereafter, seed germination was assayed as described above. To assess longevity, immature artificially dried and mature seeds were stored in the dark at 35˚C at 75% RH using hermetically closed containers containing a saturated NaCl solution. The water content of the seeds at these conditions was 0.13 g H₂O/g DW⁻¹. At different time intervals, 4 replicates of 25 seeds were retrieved and imbibed as described above, and final germination percentage was counted.

### Soluble sugar assay

Soluble sugar contents were assessed separately in cotyledons and embryonic axes from the phenological stage R6 onwards using DIONEX-HPLC according to Rosnoblet et al. [35]. Analysis was performed on triplicates of 6 axis and cotyledons.

### RNA sequencing, quality control and reads alignment

Developing and mature seeds harvested from up to 200 plants of the 2014 crop at each stage were frozen in liquid nitrogen. Total RNA was extracted using the NucleoSpin® RNA Plant kit (Macherey-Nagel) according to the manufacturer instructions. Total RNA from high-quality samples (RIN values > 8.9 evaluated by a 2100 Bioanalyzer, Agilent Technologies, Santa Clara, CA, USA) were used for library preparation sequencing at the Laboratório Central de Tecnologias de Alto Desempenho em Ciências da Vida (LaCTAD) from the University of Campinas, Brazil. cDNA libraries were generated using the TruSeq RNA sample preparation kit (Illumina, San Diego, CA, USA). After estimation of the insert size of the libraries and quantification using quantitative PCR, samples were diluted and pooled. Three lanes were sequenced using a HiSeq2500 (Illumina) with the TruSeq SBS Kit v3-HS, according to the manufacturer instructions. Sequencing adaptors and low complexity reads were removed in an initial filtering step. After quality control, reads were mapped to the ‘Williams 82’ soybean reference genome (assembly Glyma.Wm82.a1.0, annotation v2.0) using Bowtie2 [36]. RNAseq data were deposited in the NCBI Gene Expression Omnibus database (accession no. GSE98199).

### Transcriptome and gene network analysis

Estimation of differential gene expression and statistical analyses were performed using DESeq2, v1.11.21 [37] available as a R Bioconductor package. Genes were retained as differentially expressed when the ratio was at least two-fold and the P-value adjusted for multiple testing using the Benjamini-Hochberg (BH) method <0.05. Relative expression data were normalized by dividing the mean normalized gene expression value obtained of the stage n+1 by the value obtained of stage n. Functional enrichment of gene ontology (GO) was performed using the GO enrichment tool of Soybase (https://www.soybase.org) with the Glyma 2.0 gene model.

Data used for the generation of the transcription factor network corresponded to genes encoding transcription factors (TF) that varied in their expression profiles during development. A total of 754 TF were retained with transcripts levels showing a variance >1
throughout seed development. The gene co-expression network was constructed using the Expression Correlation Plugin for Cytoscape, with a Pearson correlation coefficient (PCC) cutoff of 0.97, only including positive correlations. Gene interactions were visualized using the open source software Cytoscape (version 2.8.1) using an organic layout. Identification of genes correlated with longevity was obtained using the trait based gene significance measure [24, 38] where the gene significance of a gene equals the absolute correlation between the gene expression profile and longevity expressed as P50 (days to obtain 50% germination during storage).

Fig 1. Seed and pod development of soybean. (A) Seed phenological stages during the acquisition of seed longevity (2013 crop). (B) The relationship between seed age and phenological stages (2014 crop). Stage 9 corresponds to dry mature seeds.

https://doi.org/10.1371/journal.pone.0180282.g001
**Quantitative PCR**

RNA extraction was performed as previously described on three biological replicates of 30 seeds. First strand cDNA was synthesized from 2 μg total RNA using High Capacity RNA-to-cDNA kit (Applied Biosystems, place) following manufacturer’s instructions. Quantitative real time PCR was performed on a thermocycler Eco Real-Time (Illumina) with SYBR Green qPCR ReadyMix (Sigma Aldrich, place) using the manufacturer’s instructions. Data were analyzed with EcoStudy program version 5.0 (Illumina). Primer efficiency was calculated as described in Ruiter et al. [39]. Relative expression levels were calculated using the comparative 2^ΔΔCt method [40] using two reference genes, 20S proteasome subunit beta (Glyma.06G078500) and 60S ribosomal protein L6 (Glyma.15G271300) [41]. In our RNAseq dataset, transcript levels of both genes showed little variation (variance of 0.035 and 0.08 respectively). Forward and reverse primers used for these genes are listed in S2 Table.

**Results**

**Seed longevity is acquired at the end of maturation**

Various events related to seed maturation were characterized in two consecutive years between reproductive stages R5 and R9. Seeds were collected at similar phenological stages between the 2013 and 2014 cultures to allow for comparison. Since no major developmental differences were found between both crops (S1 Fig), only data from the 2014 crop are presented here. Fig 2A shows that the end of the seed filling phase occurred around 63 d after flowering (DAF, corresponding to stage 7.2), whereas the onset of maturation drying started at 71 DAF (stage 8.1). The end of the seed filling phase also coincided with the abscission of seeds from the mother plant. At 57 DAF (stage 7.1) and 63 DAF (stage 7.2), 88% and 6% of the seeds were still attached to the fruit, respectively. At stage 7.3, all seeds were detached from the mother plant. In agreement with earlier works [4, 25, 27], germination capacity was acquired early during seed filling, between 25 and 57 DAF (stage 7.1). Desiccation tolerance, i.e. the ability to germinate after fast drying to 10% moisture, was acquired between 57 and 63 DAF (Fig 2B). There was no significant difference in the acquisition of these traits between the 2013 and 2014 crops (S1 Fig). From the loss of viability curves during storage at 75% RH and 35˚C (Fig 2C), the longevity, expressed as the time to obtain 50% germination after storage (P50), was calculated. Seed longevity was progressively acquired shortly after the seed filling phase. During maturation, P50 values increased sharply between 57 and 63 DAF (stage 7.2) from 0 to 28 d (Fig 2B). Thereafter, P50 increased almost two-fold during further maturation, reaching 48 d in mature dry seeds. Longevity data obtained for the 2013 crop followed a similar trend as those of 2014 (S1 Fig), with the P50 being on average 18% lower than the 2014 crop. Monitoring field humidity and temperature revealed that the 2014 crop grew under significant higher average temperatures compared to 2013 (26.5˚C and 30.2˚C for 2013 and 2014, respectively).

**Transcriptome profiling during late maturation highlights dynamic changes during the acquisition of longevity**

To characterize transcriptome changes during the end of seed maturation, total RNA was extracted from six stages from whole seeds during development from stage 7.1 until stage 9. A replication of stage of 7.2 and 9 was nested in the experimental design to assess reproducibility of the data. After sequencing and assembly, RNAseq produced between 14 and 38 million reads per library, with all libraries having >90% reads mapping (Table 1).

From this dataset, 16,248 genes were retained as differentially expressed throughout the samples (i.e. with a variance > 1), S3 Table. PPC calculations and principal component
Fig 2. Physiological characterization of soybean seed maturation. (A) Evolution of seed dry weight (back circle) and water content (blue square). Data are the means (± SE) of 3 to 5 replicates of 20 seeds. (B)
Acquisition of germinability (●, black circle) and desiccation tolerance (blue circle), evaluated after fast drying to 10% moisture and longevity (red triangle) as assessed by P50 (time necessary to obtain a loss of viability of 50% during storage 35˚C and 75% RH). Data are the means (± SE) of 4 replicates of 25 seeds. (C) Loss of seed germination during storage at 75% HR, 35˚C. Data are the mean of 4 replicates of 25 seeds harvested at indicated phenological stages.

https://doi.org/10.1371/journal.pone.0180282.g002

analysis were performed to assess the dynamic response of the transcriptome throughout the maturation stages (Fig 3). Two transcriptional switches occurred during seed maturation, the first one between stage 7.1 and 7.2, visualized by the correlation matrix, and the second one between stage 7.3 and 8.1, apparent from the correlation matrix and the principal component analysis (Fig 3). Data from the two biological replicates were highly similar, with between 98% and 97% similarity between the two biological replicates of stage 7.2 and stage 9.

Using Pageman (MAPMAN, GABI Germany), a gene ontology (GO) enrichment was performed to obtain an overview of the different biological processes that are overrepresented at the different stages of seed maturation (Fig 4). Stages 7.1 (and to a lesser extend stage 7.2) were characterized by an over-representation of functional classes related to growth and metabolic activities, reflecting the active seed filling that still went on in these green seeds. These classes included photosystem light reaction, starch synthesis, lipid metabolism and amino acid degradation, as well as storage proteins, protein targeting secretory pathway and cell vesicle transport. At stage 7.2–7.3, these classes were no longer over-represented, indicating the end of seed filling at the molecular level. From this developmental stage onwards, the class “abiotic stress” became over-represented, as well as functional classes related to raffinose family sugars and protein degradation (ubiquitin), whereas classes corresponding to “signaling” became underrepresented. At stage 8.1, a transition occurred with a temporary overrepresentation of protein synthesis (elongation and ribosomal protein synthesis) and an underrepresentation of cell wall degradation and biotic stress. From stage 8.2 onwards, functional classes that became overrepresented were related to mitochondrial electron transport/ATP synthesis and abiotic stress (wounding/touch) (Fig 4).

**Accumulation of RFO in relation to the acquisition of longevity**

One of the major changes occurring during seed maturation is the accumulation of RFOs. Numerous studies exist on the accumulation of these sugars during soybean seed development

<table>
<thead>
<tr>
<th>DAF</th>
<th>Stage</th>
<th>F/D</th>
<th># reads</th>
<th># Mapped reads</th>
<th>% Mapped reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>7.1</td>
<td>F</td>
<td>25 673,031</td>
<td>24 112,015</td>
<td>93.9</td>
</tr>
<tr>
<td>63</td>
<td>7.2a</td>
<td>F</td>
<td>38 088,839</td>
<td>34 971,702</td>
<td>91.8</td>
</tr>
<tr>
<td></td>
<td>rep 1</td>
<td></td>
<td>34 381,008</td>
<td>31 799,644</td>
<td>92.5</td>
</tr>
<tr>
<td>7.2b</td>
<td>rep 2</td>
<td>F</td>
<td>14 259,082</td>
<td>12 881,662</td>
<td>90.3</td>
</tr>
<tr>
<td></td>
<td>rep1</td>
<td>D</td>
<td>66 764,127</td>
<td>60 626,937</td>
<td>90.8</td>
</tr>
<tr>
<td>69</td>
<td>7.3</td>
<td>F</td>
<td>22 267,371</td>
<td>20 475,504</td>
<td>92.0</td>
</tr>
<tr>
<td>71</td>
<td>8.1</td>
<td>F</td>
<td>14 533,337</td>
<td>13 392,050</td>
<td>92.2</td>
</tr>
<tr>
<td>73</td>
<td>8.2</td>
<td>F</td>
<td>22 057,408</td>
<td>19 621,133</td>
<td>89.0</td>
</tr>
<tr>
<td>77</td>
<td>9a rep1</td>
<td>F</td>
<td>28 824,583</td>
<td>26 457,357</td>
<td>91.8</td>
</tr>
<tr>
<td></td>
<td>rep 2</td>
<td>F</td>
<td>23 064,604</td>
<td>21 097,336</td>
<td>91.5</td>
</tr>
</tbody>
</table>

# reads, number of reads following trimming of the libraries for quality
# mapped reads, number of reads that unambiguously mapped to the soybean genome. Percentages of mapped sequences are also indicated. DAF, days after flowering; Rep, replicates. F: Freshly harvested seeds; D: rapidly dried seeds

https://doi.org/10.1371/journal.pone.0180282.t001

Table 1. Mapping of single-end reads to the soybean genome.
but how this increase relates to the acquisition of seed longevity is unknown. The transcriptome analysis identified an overrepresentation of genes involved in RFO metabolism during maturation (Fig 4). Considering their controversial role in the survival in the dry state [10], we investigated the accumulation of the different soluble sugars in the axes and cotyledons during seed maturation and assessed whether these metabolic changes occurred in relation to the acquisition of longevity (Fig 5). Both in axes and cotyledons, glucose (Glc) and fructose (Fru) contents were highest during the seed filling phase and decreased throughout further maturation to almost undetectable levels. For Glu, this decrease occurred earlier in cotyledons compared to axes (Fig 5A and 5D). In axes, Suc contents remained high during the seed filling phase at around 80 mg/g DW (Fig 5B). At 57 DAF, Suc started to decrease until 73 DAF, when the seed moisture was 28% (DW basis). In cotyledons, Suc levels exhibited a sharp decrease at the end of the seed filling phase from 33 mg/DW at 57 DAF to 13 mg/g DW at 63 DAF (stage 7.2), concomitant with the decrease in Glu content and increase in longevity. Thereafter, Suc content remained constant until the dry state, representing 55% of the total amount of soluble sugars in the cotyledons. Stachyose (Sta) was the preponderant RFO in dry seeds, representing 90% of the total amount of RFO. In axis, its pattern of accumulation coincided with the acquisition of desiccation tolerance rather than longevity (Fig 5C). Sta accumulated during seed filling between 46 and 63 DAF then remained at steady level. In cotyledons, Sta contents increased later during maturation than in axis, along with the increase in longevity until 73 DAF (Fig 5F). Thereafter it decreased by ca. 30% during further maturation drying. The RFO/Suc ratio in the axis increased concomitantly with the increase in longevity (Fig 5B). By comparing Fig 5B and 5C, two successive factors contributed to this increase; first the synthesis of Sta (up to stage 7.3) then a decrease in Suc contents. In the cotyledons, the increase in RFO/Suc was mainly driven by the synthesis of RFO. Thus, the RFO/Suc ratio, especially in the embryonic axis, seems to be a good indicator of the seed longevity acquisition during maturation in this indeterminate cultivar.

Identification of the transcription factor network involved in late seed maturation

To obtain information on the regulatory factors underlying the transcriptional and biochemical changes during maturation, changes in transcript profiles of genes encoding TF, representing 1086 genes belonging to 32 families, were further analyzed. To capture the temporal regulation of the TF transcripts during maturation, a gene co-expression network was constructed. The resulting scale-free network contained 499 nodes and 12,183 edges and was visualized using the cytoscape organic layout algorithm (Fig 6A). To understand the topology of the network, nodes were colored based on their expression profile, with the stage at which transcript levels were maximum. The network consisted of a central cluster of interacting nodes corresponding to genes with highest transcript level at the end of seed filling (Fig 6A, stage 7.1, dark green), followed by a tail composed of probes corresponding to TF with highest transcript level during the transition between seed filling and maturation drying phase (stage 7.2, light green). A second cluster that was disconnected from the main cluster represented TF with transcripts being maximum at stage 8.2 (yellow), 8.3 (orange) and 9 (mature seeds, grey).

Twenty four families of TF were present in the network (S4 Table, Fig 6A). Using a χ² testing the null hypothesis that these families are randomly enriched in the different stages, AP2/EREBP and WRKY were the only families that were significantly enriched in the tail with genes exhibiting a transient expression profile with a maximum level around stage 7.2 and 7.3. At these stages, AP2/EREBP and WRKY represented 20% (p = 0.005) and 11% (p = 0.048) of all TF found differentially expressed, respectively. A closer inspection of these TF revealed
Fig 3. Correlation and principal component analyses of soybean transcriptomes during seed maturation. (A) Pair-wise Pearson correlation coefficients were used to generate the heat map. The color
scale indicates the degree of correlation (blue, low; yellow, high). (B) and (C) Principal component analysis performed using median centering of the transcriptomes of seed phenological stages. 7.2D, transcriptomes of rapidly dried seeds at stage 7.2. The letters a,b, correspond to biological replicates.

https://doi.org/10.1371/journal.pone.0180282.g003

genes involved in germination and organ growth, flowering, chloroplast dismantling, defense and ABA signaling.

Next, we used the trait-based gene significance measure [38] to integrate the acquisition of longevity in the TF gene co-expression network to identify those TF that exhibit transcription profiles highly correlating with P50. For this purpose, PCC were calculated between the

---

Fig 4. Over-representation analysis of functional classes during seed maturation. Functional classes and subclasses statistically affected are indicated according to Mapman ontology. Data were subjected to a Bin-wise Wilcoxon test and resulting p-values were adjusted according to Bonferroni. The scale bar indicates the z-score calculated from p-values (i.e. p-value of 0.05 represents a z-score of 1.96 after adjustment). Over-represented and under-represented classes are indicated respectively in blue and yellow.

https://doi.org/10.1371/journal.pone.0180282.g004
transcript levels of all TF genes and the P50 value during maturation (S4 Table). When genes with a PCC > 0.9 were projected on the TF network, 27 TFs were identified that formed a highly co-expressed module (blue nodes Fig 6B, S4 Table). A total of 8 of the 27 genes belonged to the AP2/EREBP family, including four homologues of ETHYLENE RESPONSE FACTOR (ERF10, a member of the ERF subfamily B4, PCC > 0.98) and homologues of the Arabidopsis DREB2C and DREB2F. Other TF belonged to families such as WRKY, auxin response factor, homeodomain-like proteins (S4 Table) and X-box binding transcriptional repressor family with a homologue of NUCLEAR TRANSCRIPTION FACTOR X-BOX BINDING LIKE 1 (NFXL1), a gene previously found in the longevity modules of *M. truncatula* and Arabidopsis co-expression network during seed maturation [24]. In addition, two heat shock factors (HSF, Glyma.03g191100 and Glyma.03g157300) were present in the list of the 27 TFs. RT-qPCR on additional seed samples from different developmental stages validated their increased transcript level during maturation (Fig 7A and 7B), with expression profiles being comparable to those obtained by RNAseq (r² > 0.8). Interestingly, inspection of the Soybase gene expression profile showed that the paralogs of these two genes are not expressed in seeds. These HSF are
transcriptional regulators known to activate small heat shock proteins (HSP). Consistent with this, analysis of the most differentially expressed genes between stage 7.2 and stage 9 (Int (log2) > 4) that correlated with longevity (PPC > 0.85) shows that many of these genes code for small HSP (Fig 8). RT-qPCR data confirmed the increase of transcript levels during final maturation for three sHSPs (Fig 7C–7E).

Comparative analysis of dried, immature short-lived seeds with mature long-lived seeds identifies developmental indicators linked to longevity

Premature drying of immature seeds (stage 7.2) leads to reduced acquisition of longevity compared to fully mature seeds (Fig 2). A comparison of the transcriptome of these immature seeds with fully mature seeds can be used for further investigation of markers for longevity. First, the impact of enforced drying on the transcriptome of seeds harvested at stage 7.2 was visualized by PCA analysis (Fig 3). In comparison with the freshly harvested immature seeds,
Fig 7. qPCR analysis of selected genes during seed maturation validate RNA-Seq data. (A) Heat shock transcription factor A3 (HSFA3: Glyma.03g191100); (B) Heat shock transcription factor A6B (HSFA6B: Glyma.03g157300); (C) sHSP17a (Glyma.17g224900); (D) sHPS17b (Glyma.14g099900), (E) sHSP21 (Glyma.08g318900). Data (±SE) are the average of three biological replicates of 30 seeds.

https://doi.org/10.1371/journal.pone.0180282.g007

Fig 8. Heat map of most differentially expressed genes correlating with longevity. Genes were retained when the log2 intensity > 4 between stage 7.2 and 9 and that correlate with longevity (PPC P50 > 0.85). Genes were log2 mean centered and colored from the lowest (green) to highest values (red).

https://doi.org/10.1371/journal.pone.0180282.g008
the transcriptome of the dried immature seeds was positioned between those of seeds harvested at stage 7.3 and 8.1 in the PCA plot (Fig 3). Apparently, premature drying accelerated the maturation of the seeds: in planta, the time lapse between stage 7.2 and 7.3 was 6–7 days (Fig 1), whereas it took only two days to dry the seeds harvested at stage 7.2.

Transcriptome comparison between immature, rapidly dried seeds and mature seeds identified 742 genes with transcripts that were significantly higher in seeds with high longevity compared to those with low longevity, whereas 1525 genes had transcripts that were lower in the seeds with high longevity (S5 Table). Next, we investigated how many of these transcripts also changed during enforced drying. For this purpose, a statistical analysis was performed to identify the differentially expressed genes in seeds of stage 7.2 before and after enforced drying (S5 Table). A total of 5423 transcripts were up-regulated and 4931 down-regulated upon enforced drying of the seeds. The Venn diagram (Fig 9) shows that out of the 742 genes with higher transcript level at stage 9 compared to stage 7.2D, 139 (19%) also increased significantly during enforced drying. For the down-regulated transcripts, 207 out of the 1525 (14%) also responded to the enforced drying of the immature seeds (Fig 9, S5 Table).

A Gene Ontology (GO) enrichment analysis of the 139 differentially expressed genes with higher transcript level in both dried immature seeds and dry mature seeds revealed an overrepresentation of biological functions related to response to heat, hydrogen peroxide and high light intensity (Table 2). A closer look at the expression profiles of these genes showed that they increased sharply between stage 7.2 and stage 8.1, with a further gradual increase until stage 9 (S3 Table). Considering that enforced drying resulted in a transcriptome that was comparable to stages 7.3–8.1, the increased transcript levels between 7.2F and 7.2D could be the result of an accelerated maturation.

An analysis of the enrichment of biological functions of the 603 genes that were up-regulated during maturation drying in planta but not upon enforced drying of freshly harvested seeds of stage 7.2 revealed functions related to pyrimidine ribonucleotide biosynthetic process, protein import into nucleus and protein targeting to mitochondrion, and nucleosome assembly (Table 3). Another biological function that was overrepresented is RNA methylation.

Regarding the 207 genes that were down-regulated between stage 7.2D and stage 9 as well as stage 7.2F and stage 7.2D, no GO enrichment was found. In contrast, the enrichment analysis of differentially expressed genes with lower transcript levels in seeds with high longevity (1318) revealed an overrepresentation of biological functions related to response to generation of precursor metabolites, transcription, as well as cellular respiration (Table 4). In addition, biological processes to photosynthesis were down-regulated. A closer inspection of the list of genes belonging to these GO categories revealed that all of them were associated with various chloroplast functions such as photosynthesis and starch and lipid synthesis. Interestingly, the list also contained 69% of the soybean chloroplast genes (77 out 111, S5 Table), including those involved in the photosystem II reaction, ATP synthase sub-units, Calvin cycle (Rubisco large subunit), chlorophyll binding and translation (RNA polymerase, ribosomal proteins, maturase K). In addition, we found two homologues of STAY-GREEN1; an important gene involved in chlorophyll catabolism and photosystem degradation [45]. A large number of these genes were strongly upregulated in seeds of stage 7.2 that were submitted to enforced drying compared to freshly harvested seeds (S5 Table). Considering the acceleration of maturation in the dried seeds, this increase can be explained in part by their transitory expression profile, with transcript levels being higher in seeds at stage 8.1 compared to stage 7.2 (S3 Table).

Furthermore, an overrepresentation was found of processes related to water deprivation and salt stress (Table 4), which included genes involved in RFO metabolism such as galactosyl synthase (Glyma.03g229800, Glyma.20g094500), raffinose synthase (Glyma.06g179200).
GO term “response to water deprivation”; we noticed two Arabidopsis homologues of a RING domain ligase, namely RGLG1 and RGLG2, that have a ubiquitin E3 ligase activity and mediate the transcription of AtERF53 in response to drought [46].

**Discussion**

Poor longevity results in economic losses due to the impossibility of carry-over of seed lots, having lost their vigor and viability so that they are no longer marketable. Identification of the underlying regulatory factors should provide information to design marker for
prebreeding aiming to improve soybean seed quality. In developing soybean seeds, physiological maturity corresponds to the stage when final seed weight is reached, germination/desiccation tolerance and seed vigor are acquired [4, 25, 27]. In this study using an indeterminate cultivar, physiological maturity corresponded to stage 7.2, in agreement with previous works [4]. At this stage, most seeds were detached from the mother plant and had lost most of their chlorophyll. However, our physiological, sugar and transcriptome data show that the seed maturation program has not yet ended at physiological maturity. An additional period of 14 days after physiological maturity is necessary to acquire maximum longevity (Fig 2), in agreement with previous data on other genotypes [25, 26]. During this period, we detected 16,248 transcripts being differentially expressed until the developing seeds reached the dry state. Our RNAseq study complements and extends previous transcriptome characterization of soybean seed development [28–33]. These studies focused mostly from fertilization to end of seed filling whereas here we characterized the phase from end of seed filling to final maturation drying. Our RNAseq data and co-expression network analysis suggest that complex transcriptome changes occur after the so-called physiological maturity until the dry state, identifying several TFs associated with seed longevity. Several of these TFs were previously identified in a gene co-expression network associated with longevity in M. truncatula and Arabidopsis [24] and thus provide new resources for marker of seed development.

Table 2. GO enrichment analysis of the 139 differentially expressed transcripts that are significantly higher at stage 9 compared to 7.2 after drying and common between 7.2D/7.2F and 9/7.2D.

<table>
<thead>
<tr>
<th>GO term</th>
<th>GO description</th>
<th>Genome GO count</th>
<th>Expressed GO</th>
<th>Corrected P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overrepresented</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0009408</td>
<td>response to heat</td>
<td>659</td>
<td>36</td>
<td>2.43E-22</td>
</tr>
<tr>
<td>GO:0042542</td>
<td>response to hydrogen peroxide</td>
<td>511</td>
<td>29</td>
<td>3.13E-18</td>
</tr>
<tr>
<td>GO:0009644</td>
<td>response to high light intensity</td>
<td>582</td>
<td>29</td>
<td>1.06E-16</td>
</tr>
<tr>
<td>GO:0006457</td>
<td>protein folding</td>
<td>779</td>
<td>29</td>
<td>2.24E-13</td>
</tr>
<tr>
<td>GO:0034976</td>
<td>response to endoplasmic reticulum stress</td>
<td>483</td>
<td>18</td>
<td>6.19E-08</td>
</tr>
</tbody>
</table>

Analysis was performed using the GO enrichment tool of Soybase using the Glyma 2.0 gene models.

https://doi.org/10.1371/journal.pone.0180282.t002

Table 3. GO enrichment analysis of the differentially expressed transcripts that are significantly higher at stage 9 compared to 7.2 after drying and not induced upon drying between 7.2F and 7.2D.

<table>
<thead>
<tr>
<th>GO term</th>
<th>GO description</th>
<th>Genome GO count</th>
<th>Expressed GO</th>
<th>Corrected P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overrepresented</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0009220</td>
<td>pyrimidine ribonucleotide biosynthetic process</td>
<td>315</td>
<td>28</td>
<td>4.63E-21</td>
</tr>
<tr>
<td>GO:0006606</td>
<td>protein import into nucleus</td>
<td>243</td>
<td>23</td>
<td>2.69E-16</td>
</tr>
<tr>
<td>GO:0006626</td>
<td>protein targeting to mitochondrion</td>
<td>235</td>
<td>19</td>
<td>9.13E-15</td>
</tr>
<tr>
<td>GO:0006334</td>
<td>nucleosome assembly</td>
<td>128</td>
<td>12</td>
<td>2.62E-13</td>
</tr>
<tr>
<td>GO:0006164</td>
<td>purine nucleotide biosynthetic process</td>
<td>75</td>
<td>9</td>
<td>3.09E-08</td>
</tr>
<tr>
<td>GO:0001510</td>
<td>RNA methylation</td>
<td>418</td>
<td>22</td>
<td>9.26E-07</td>
</tr>
<tr>
<td>Underrepresented</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0006952</td>
<td>defense response</td>
<td>1116</td>
<td>1</td>
<td>3.84E-05</td>
</tr>
</tbody>
</table>

Analysis was performed using the GO enrichment tool of Soybase using the Glyma 2.0 gene models.

https://doi.org/10.1371/journal.pone.0180282.t003
Important transcriptional changes occur after mass maturity before the onset of desiccation

A salient feature of the genetic program that occurs during late seed maturation is a transitory phase that is delimited by two transcriptional switches, one from stage 7.1 to stage 7.2, marked by the end of seed filling, and the second between stages 7.3 and 8.1 (Fig 3). Phenotypically, this phase is accompanied by the loss of chlorophyll, acquisition of desiccation tolerance and the biggest increase in RFO sugars (Figs 2 and 6). A transcriptome shift was also observed in Arabidopsis during maturation and attributed to the desiccation of the seed [47]. This is not the case in soybean, since seed moisture content at stage 7.3 was around 1.2 g water/g DW, thus before the desiccation phase per se. However, during this transitory phase, seeds loose their connection to the mother plant. Concomitantly, changes in the transcriptome suggest the activation of a post-abscission program to prepare for the dry state and germination by synthesizing mRNA that will be stored until seed imbibition [48]. Several observations support this hypothesis. Genes with expression profiles that show maximum transcript levels at stage 7.2 or 7.3, i.e. represented by the tail of the co-expression network (Fig 6), are significantly enriched in members of the AP2/EREBP and WRKY families. AP2/EREBP transcription factors play an important role in controlling developmental processes and in hormone, sugar and redox signaling in relation with abiotic stresses [49]. Their over-representation in developing soybean seeds was also reported by Jones & Vodkin [32]. Several homologues were found to be co-regulated with the induction of longevity in the legume M. truncatula [23]. Most of the TFs were related to germination and growth, such as a homolog of WRKY6 (Glyma.13g310100) that in Arabidopsis acts as a positive regulator of ABA signaling during seed germination and early seedling development [50], the homolog of SOMNUS (Glyma.12g205700), the homolog of HOMEOBOX 1 (HB-1), involved in hypocotyl growth under short days [51] and a homolog of INDETERMINATE DOMAIN1/ENHYDROUS (Glyma.02g058500), that in Arabidopsis promotes the transition to germination by regulating light GA effects and ABA signaling during seed maturation [52]. We also found a B3 transcriptional repressor (Glyma.02g36090) whose Arabidopsis homologue is a negative regulator of seed size in developing seeds [53]. Transcripts associated with protein degradation via the SCF family of modular E3 ubiquitin pathway increased during maturation drying at stage 7.3. This observation is in agreement with previous studies [28, 32]. These proteins are known to filter the proteome by degrading key
regulatory proteins as main targets [54]. This suggests that the drying seeds are setting up a machinery for post-translational regulation before entering in the dry state that will presumably serve upon imbibition.

An emerging picture of longevity-related genes identifies confirmed regulators and newcomers in legume seeds

The network analysis revealed 27 TF whose transcript profiles are correlated with P50 (Fig 6, S4 Table), thereby forming a longevity module similar to that found in developing Arabidopsis and M. truncatula [23, 24]. Indeed, the homologues of 7 soybean genes were also present in the list of the 9 TF belonging to Medicago longevity module [24], namely three homologues of an ERF110 of Arabidopsis (Glyma.06g148400, Glyma.04g217400, Glyma.08g145300), a homologue of the ETHYLENE INSENSITIVE PROTEIN 3 family (Glyma.05g180300), two homologues of the DREB2 family (Glyma.07g156200, Glyma.14g056200), and a homologue of NF-X-like1 gene (Glyma.09g173000). The implication of two of these genes in longevity has been demonstrated [55, 24]. A DREB2 from sunflower enhanced seed longevity of tobacco when ectopically over-expressed with a heat shock factor, HaHSFA9 [55]. In Arabidopsis, seeds of nfxl1 exhibited impaired acquisition of longevity during maturation [24]. When overexpressed in vegetative tissues, NFXL1 induces a higher survival upon salt stress, drought and high light intensity [56].

Several TF that highly correlated with P50 are associated with auxins and gibberellins (S3 and S4 Tables), such a homologue of HECATE2 (Glyma.11g055300), that affects auxin responses in Arabidopsis during flower development [57], a homologue of AUXIN RESPONSE FACTOR19 (ARF19, Glyma.13g112600, Glyma.09g072200), ARF4 (Glyma. 12G171000), ARF8 (Glyma.10G 210600) ARF9 (Glyma.03g36710), ARF10 (Glyma.13g325200), ARF16 (Glyma.10g210600). This reinforces the link between auxin and longevity previously found in the Medicago and Arabidopsis maturation network, where 60% of genes belonging to the longevity were significantly enriched in binding sites for auxin response factor [24]. The putative role of auxins in inducing longevity remains to be investigated. These auxin-related genes exhibit functions that are associated with embryogenesis, meristem maintenance, stem cell specification, positioning of lateral organs and organ growth meristem in connection with GA signaling [57–59]. A further inspection of the TF present in the network tail (Fig 6) confirmed the enrichment in genes with the above mentioned functions such as the homologue of JACKDAW (Glyma.10g051500), AINTEGUMENTA-LIKE 6 (Glyma.01g022500), SEPALA3 (Glyma.20g153700) and GATA (Glyma.17g228700). We speculate that the presence of these transcripts in dry seeds may be necessary to anticipate the restoration of the developmental fate of specific cells during germination of seeds that were damaged by ageing during storage. Consistent with this, stem cell niches are hypersensitive to DNA damage [60], which is a known cause for decrease in seed viability after storage [21].

The transcriptome comparison between immature short-lived and mature seeds exhibiting maximal life span further highlights key mechanisms that could be involved in longevity. Seed longevity correlated with the synthesis of HSPs and several chaperones implicated in protein-protein interactions and protein folding (Table 2, Fig 8, S5 Table). They are known to assist in creating proper-folding conditions during abiotic stress [61] and protect against oxidative stress during storage [62] that could be conducive to seed longevity. Our transcriptome also revealed the presence of HSFA6B (Glyma.03G157300), whose transcript profile was correlated with P50 (S4 Table). This is consistent with the observation that the over-expression of sunflower HaHSFA9 in tobacco led to an increased stability against accelerated ageing of the seeds [13]. This TF also interacts with the drought-responsive factor HaDREB2 in a seed-specific manner to enhance stability against accelerated aging [55].
Concomitant with the acquisition of longevity, there was an overrepresentation of genes involved in RFO synthesis (Table 3). This was confirmed by the observation that RFO content and the ratio Suc/RFO increased during late seed maturation (Fig 5), reinforcing the idea that the regulation of RFO metabolism occurs at the transcriptional level [10, 12]. The role of RFO in seeds is unclear, particularly in soybean. No defects in desiccation tolerance, seed germination or seedling emergence were reported in soybean lines with low RFO content [44, 63]. However, recent literature suggests that RFO metabolism plays a role in the acquisition of longevity. Galactinol, the precursor of RFO was found to be a marker for seed longevity in Arabidopsis, cabbage and tomato [64]. In Arabidopsis, seeds of galactinol synthase mutants (gols2 and gols1gols2) were more sensitive to accelerated aging (85% RH, 40°C) while seeds of raffinose synthase (rs) and stachyose synthase (sts) mutants did not differ significantly from those of wild type [64]. Further genetic proof for a role for RFO metabolism in longevity came from the impaired shelf-life of Arabidopsis α-galactosidase2 (agal2) seeds stored at 75% RH and 35°C [24]. Suc levels were significantly higher in these mutants but no change was detected in RFO levels. Seed-specific overexpression of CaGolS1 and CaGolS2 of chickpea in Arabidopsis resulted in an improved resistance against controlled deterioration [11]. However, the moisture content of these seeds during ageing was 24%, a value high enough to lead to increased metabolic activity and accumulation of RFO, while this would not occur during dry storage at RH below 75%. Additional correlative evidence of a link of RFO metabolism and longevity comes from the analysis of mutants of two regulatory genes of seed maturation in *M. truncatula*: Mt-abi5 and Mt-snf4b [12, 35]. Seeds of these mutants show decreased RFO accumulation and increased Suc accumulation during maturation together with a decreased in longevity. The underlying mechanisms that explain the role of galactinol or RFO in conferring longevity remain elusive. A specific role of RFO in the protection of membranes or other macromolecules seems unlikely. Suc molecules are known to protect membranes just as efficiently as RFO and furthermore make denser glasses (reviewed in [10]). Some studies suggest a direct or indirect role of RFOs in the protection against oxidative damage during storage [11, 65]. Overall, our data show that the RFO/Suc ratio in the embryonic axis is a good indicator of the progress in the acquisition of longevity during maturation. However, one should keep in mind that although this relationship is valid during seed development, it might be blurred when comparing mature seeds from different genotypes, environments and date of planting.

Typically, late seed maturation is characterized by a degreening process resulting from the degradation of chlorophyll via a multi-step catabolic pathway that is characterized mostly during leaf senescence [41, 45]. Here, chlorophyll was lost during the transition phase, between stage 7.1 and 7.3. Genes encoding enzymes that are required for the initiation of the degradation of chlorophyll and light harvesting complexes such as chlorophyll b reductase (NYC1, Glyma.07g085700, Glyma.09g191200) and STAYGREEN1 (Glyma01g241600, Glyma011g027480) were already highly expressed at stage 7.1 and increased slightly at stage 7.2. This suggests that on a transcriptional level, the onset of degreening is activated prior to the induction of longevity. Furthermore, a high number of transcripts associated with photosynthesis and chloroplast activity were down-regulated when artificially dried immature seeds at stage 7.2 were compared with mature dry seeds at stage 9 (Table 4). This decrease involved both nuclear and chloroplastic genes among which 69% of the genome was concerned. It might reflect a major shutdown of chloroplast metabolism and dismantlement, which requires a close coordination between the nucleus and chloroplastic genomes. These events might be conducive to seed longevity. The presence of green seeds in mature seed lots has been associated with a decreased shelf-life during storage in various species, including soybean [12, 22, 66]. In Arabidopsis, seeds of mutants affected in chlorophyll degradation contained 10-fold more chlorophyll than the wild type and had a strongly reduced longevity [22]. Chlorophyll degradation
and longevity were affected during maturation of pea and *M. truncatula* seeds defective in ABC-SISIC ACID INSENSITIVE5 [12]. The more chlorophyll was retained in mature in *abi5* seeds, the more longevity was affected. In addition, the repression in photosynthesis-associated genes during maturation was also impaired in these *abi5* mutants [12]. Thus, these data reinforce the idea that degreening and chloroplast dismantling appears a pre-requisite for longevity.

In this study, the storage conditions were set at a RH of 75% and a temperature of 35˚C that bring the seed tissues to a water content of 0.13 g H$_2$O/g DW, equivalent to 11.5% on a fresh weight basis. In commercial practice, soybean seeds are harvested at 12–14% moisture to avoid mechanical damage and dried to 10–12% for short-term storage [4]. Drying to lower values is known to make the seeds more susceptible to cracking. The combination 75%RH/35˚C also represents a good compromise between conditions that are deleterious enough to induce a loss of viability to allow measurements within a reasonable experimental time but low enough to be near the glassy state where metabolism no longer occurs. Whereas a storage environment of 75% RH/35˚C brings the cytoplasm out of the glassy state, it is still in an amorphous rubber with solid-like properties [7, 8, 67, 68]. It has been demonstrated that measurements of longevity in more humid conditions (i.e. ≥ 85% RH) are unreliable to predict life span in storage conditions corresponding to the dry state [68–70]. In soybean, this RH will bring the water content of the tissues around 0.24 g H$_2$O/g DW (19% FW basis). Under these conditions, the cytoplasm will no longer be in a rubbery state, where molecular movement will still be restricted, but rather in a liquid state, allowing for metabolism to occur [7, 8, 67]. Quantitative trait loci analysis of seed aging in *M. truncatula* at 60%/35˚C and 75% RH/35˚C revealed similar loci and data between both aging conditions were very well correlated (r = 0.71), suggesting similar mechanisms of deterioration between these two RH [12]. However, comparison of aging at 30% RH/9˚C and 75% RH/50˚C in different genotypes of lettuce showed poor correlation [70]. Thus, one should remain careful with extrapolating our soybean data to dry storage conditions (<50% RH). Indeed, in soybean water properties change cotyledons below 8% (DW basis), most likely due to the formation of a glassy state in the cells whereas the respiration rate becomes detectable at 24% [71].

**Supporting information**

S1 Fig. Comparison of soybean seed development for the 2013 and 2014 harvest at indicated phenological stages. (A) Evolution of seed dry weight (black circles) and water content (blue squares). Data are the means (± SE) of 3 to 5 replicates of 20 seeds. (B) Acquisition of germination (black circles) and desiccation tolerance (blue squares), evaluated after fast drying to 10% moisture. (C) Acquisition of longevity as assessed by P50 (time necessary to obtain a loss of viability of 50% during storage 35˚C and 75% RH). Data are the means (± SE) of 4 replicates of 25 seeds. Data are presented for 2013 (open symbols) and 2014 (closed symbols).

(TIF)

S1 Table. Description of the reproductive phenological stages of soybean seeds.

(PDF)

S2 Table. List of primers combination used for gene expression validation of target genes by qRT-PCR.

(DoCx)

S3 Table. List of differentially expressed genes during seed maturation.

(XLSX)
S4 Table. Transcription factors present in the co-expression network of soybean seed maturation. Genes with expression profiles that correlate with longevity (P50, PCC>0.9) are indicated in bold with a blue background.

S5 Table. List of differentially expressed genes associated with longevity. List of differentially expressed genes that are higher/lower in mature seeds exhibiting maximal longevity (stage 9) compared to dried immature seeds (stage 7.2D) with low longevity.

Acknowledgments
The authors thank the staff of the Genomics section of the Life Sciences Core Facility (LaCTAD), part of the University of Campinas (UNICAMP), for their contributions to RNA sequencing.

Author Contributions
Conceptualization: Julia Buitink, Edvaldo Aparecido Amaral da Silva, Olivier Leprince.

Formal analysis: Juliana Joice Pereira Lima, Julia Buitink, David Lalanne, Sandra Pelletier.

Funding acquisition: Edvaldo Aparecido Amaral da Silva, Olivier Leprince.

Investigation: Rubiana Falopa Rossi.

Project administration: Julia Buitink, Edvaldo Aparecido Amaral da Silva, Olivier Leprince.

Resources: David Lalanne, Sandra Pelletier, Edvaldo Aparecido Amaral da Silva.

Supervision: Julia Buitink, Edvaldo Aparecido Amaral da Silva, Olivier Leprince.

Validation: Juliana Joice Pereira Lima.

Visualization: Juliana Joice Pereira Lima, Olivier Leprince.

Writing – original draft: Julia Buitink, Olivier Leprince.

Writing – review & editing: Juliana Joice Pereira Lima, Sandra Pelletier, Edvaldo Aparecido Amaral da Silva, Olivier Leprince.

References


