LEXOGEN
The RNA Experts

237P00017V0100

Don’t go wasting all your reads – Sequence what really matters with RiboCop rRNA Depletion Kit for Plants

Pamela Moll, Lukas Faltner, Kathrina Radl, Cynthia Wollner, Musashi Tsujita, Mira Schaupper, Michael Moldaschl, Eva-Maria Rösmann and Torsten Reda

Lexogen GmbH, Campus Vienna Biocenter 5, 1030 Wien, Austria

ABSTRACT

Current Next Generation Sequencing (NGS) technologies are routinely applied in the context of RNA-Seq. RNA extracted from plant species comprises **up to 98 % of the highly abundant ribosomal RNAs (rRNA)**. These vast amounts of rRNA represent a unique challenge especially when analyzing the transcriptome capacity from plant communities. **Efficient removal of rRNA substantially reduces sequencing costs and enables comprehensive analyses of plant transcriptomes**. Lexogen’s RiboCop rRNA Depletion Kit for Plants was designed using an optimized algorithm to effectively remove cytoplasmic (5.8S, 18S and 25S), mitochondrial (5S, 18S and 26S) and plastid (4.5S, 5S, 16S and 23S) rRNA from complex plant samples. RiboCop works without RNA fragmentation leaving intact full-length mRNA and long ncRNA to be recovered. The probe mix can be used on **plant leaf, seedling, or root RNA** for efficient rRNA removal, while maintaining a consistent transcript expression across a wide range of input amounts and plant species. Other high-abundant, undesired sequences in plants, like transcripts of the **photosynthesis system consuming up to 25 % of the read depth in plant leaf RNA**, can be depleted upon request as well. Although NGS sequencing costs have been decreasing in recent years, removal of rRNAs before sequencing leads to significant cost savings on sequencing and data storage and empowers efficient data analysis.

Workflows

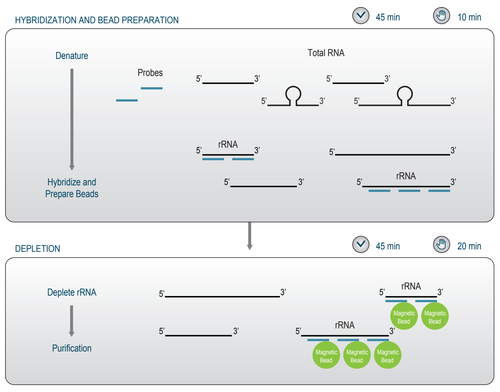

Figure 1 | Schematic overview of the RiboCop rRNA workflow.

RiboCop Plant rRNA Depletion

RiboCop rRNA depletion uses a set of affinity probes for specific depletion of rRNA or other unwanted sequences. The probes are designed for compatibility with **intact and fragmented input RNA.**

RiboCop probes **efficiently remove rRNA** and afford a comprehensive view of transcriptome composition without off-target effects. Samples void of **cytoplasmic** (5.8S, 18S and 25S), **mitochondrial** (5S, 18S and 26S) and **chloroplast** (4.5S, 5S, 16S and 23S) ribosomal sequences are obtained within 1.5 hours of total processing time. **No enzymatic reactions** or mechanical shearing steps are involved, **leaving full-length transcripts intact** for downstream processing.

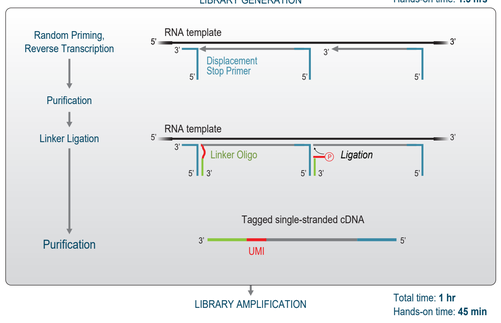

Figure 2 | Schematic overview of the CORALL RNA-Seq V2 library preparation workflow.

CORALL Whole Transcriptome RNA-Seq

CORALL RNA-Seq covers the whole length of transcripts. CORALL uses Lexogen’s proprietary displacement-stop technology to generate NGS library inserts **omitting any RNA fragmentation steps** (Fig. 2). Paired with **RiboCop for enzyme-free rRNA depletion**, CORALL is **ideal for plant transcriptomics** and any application requiring coverage uniformity, including coverage analysis, alternative splicing, and analysis of non-coding RNA (e.g., lncRNA).

Experiment and Results

rRNA Depletion Redirects Reads to RNAs of Interest, Increasing Gene Detection

Plant total RNA is comprised of large amounts of undesired ribosomal RNA (rRNA) which accounts for up to ~98 % of all transcripts. **By depleting rRNA before Next Generation Sequencing (NGS) gene detection can be significantly increased** (Fig. 3). The number of reads required for sequencing can be reduced enabling a higher degree of multiplexing on an NGS sequencing lane, therefore saving sequencing cost.

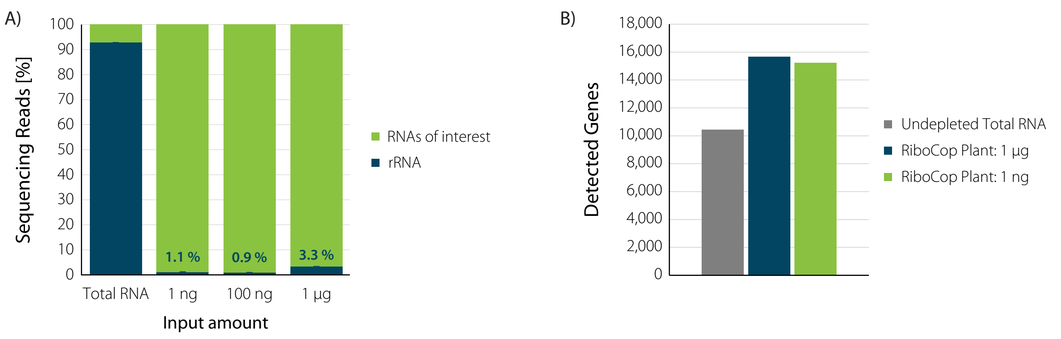

Figure 3 | A) RiboCop rRNA depletion for plants efficiently removes rRNA across a wide range of input amounts. NGS libraries were prepared using Lexogen’s CORALL RNA-Seq V2 Library Prep. Successful depletion was monitored by NGS (NextSeq2000, 1×90 bp) and analysis of remaining rRNA reads from 10 ng untreated (Total RNA) and depleted *A. thaliana* (*Ath*) RNA (1 ng, 100 ng, and 1 µg). The percentage of reads mapping to rRNA is plotted in blue. **B) Increased gene detection upon rRNA depletion.** The number of genes detected at 3M reads per sample can be significantly increased from 10.4K for the undepleted sample to 15.6K and 15.2K genes rRNA depleted samples (1 µg and 1 ng *Ath* total RNA inserted into RiboCop rRNA depletion prior to NGS library prep).

RiboCop for Plants is Suitable for Depletion of a Wide Range of Plant Species

RicoCop for Plants was successfully used on a wide range of plants, such as *A. thaliana*, *Zea mays*, and *Gylcine max* (Fig. 4), *Lactuca sativa* (**reduction from 97.7 % rRNA to 0.9 % rRNA** post depletion), *Humulus lupulus* (leaf, seed, and flower, **reduction from 98.6 % rRNA to 1.1 % rRNA** post depletion).

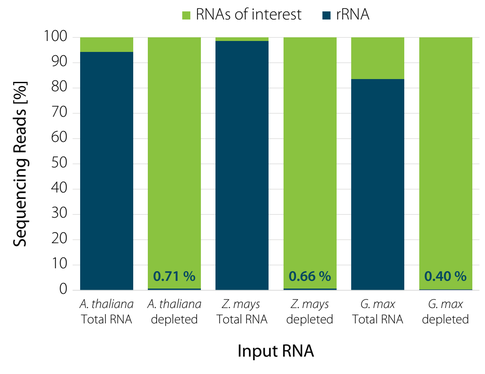

Figure 4 | RiboCop rRNA Depletion for Plants efficiently removes rRNA from *A.thaliana*, *Z. mays*, and *G. max*. NGS libraries were prepared using Lexogen’s CORALL RNA-Seq V2 Library Prep Kit. Successful depletion was monitored by NGS sequencing (NextSeq2000, 1×90 bp) and subsequent analysis of remaining rRNA reads from 5 ng untreated (Total RNA). 50 ng RNA from *A. thaliana*, *Z. mays,* and *G. max* were used for rRNA depletion. CORALL libraries were generated from 5 ng total RNA equivalents for all conditions. Reads were mapped to the respective reference genomes and the percentage of reads mapping to rRNA is plotted in blue.

Leaf RNA Contains a Substantial Fraction of Photosynthesis-related Transcripts

Plant RNA traces differ depending on which tissue the RNA was extracted from. Compared to RNA extracted from seeds, **RNA from leaves contains many additional peaks** corresponding to **chloroplast-encoded rRNA** and **highly abundant transcripts, e.g., photosynthesis-related transcripts** which appear as a distinct fraction after depletion (Fig. 5).

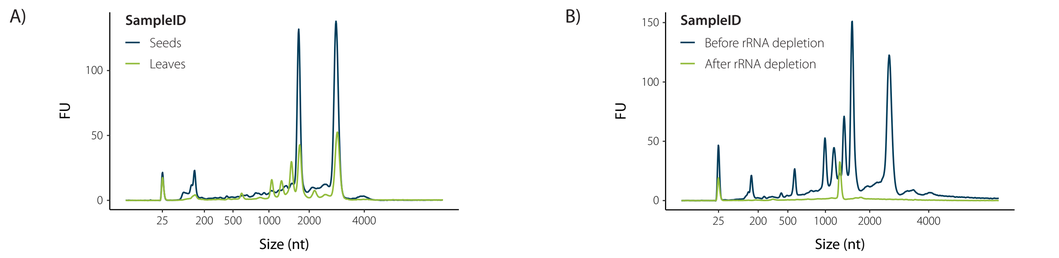

Figure 5 | Bioanalyzer traces. A) Hop (*Humulus lupulus*) RNA was extracted either from seeds (blue trace) or leaves (green trace) **B)** Leaf RNA before (blue trace) and after (green trace) RiboCop for Plants rRNA depletion. While rRNA peaks are efficiently removed, a distinct peak corresponding to *psbA* becomes clearly visible following rRNA deletion from leaf RNA.

Following rRNA depletion from leaf, a significant fraction of **up to 30 % of sequencing reads** map to photosynthesis-associated transcripts such as *psbA*. The *psbA* transcript alone can take up ~25 % of the sequencing space.

Additional Depletion of Abundant Sequences Increases Gene Detection by ~12 %

Removal of rRNA, which consumes ~94 % of reads in undepleted samples, frees up sequencing space for other high abundant transcripts such as *psbA*. **Hence, depletion of these highly abundant transcripts allows a more comprehensive view into the transcripts of interest and increases gene detection by an additional ~12 %** (Fig. 6).

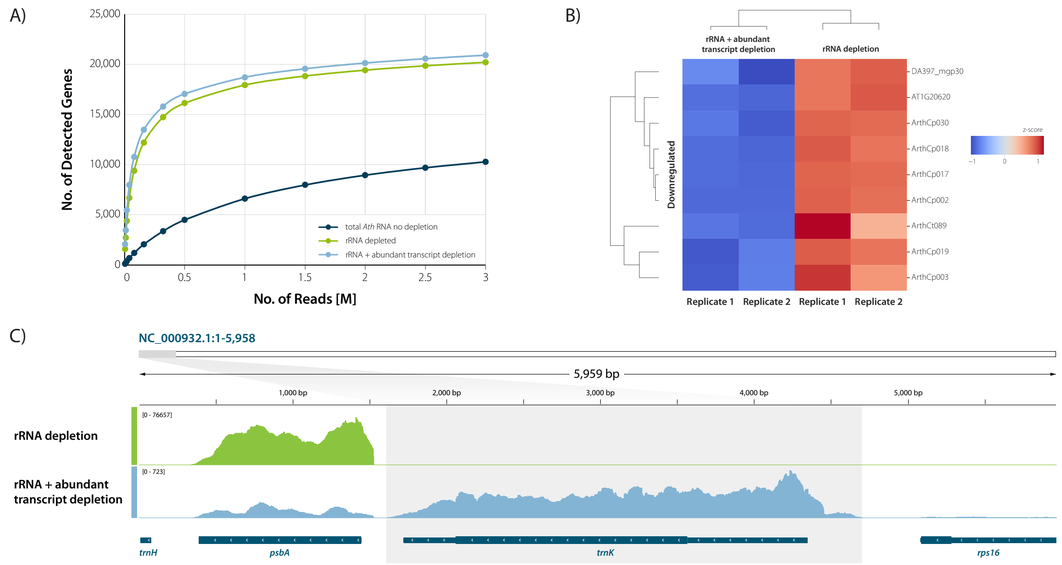

Figure 6 | Removal of highly abundant transcripts increases gene detection. A) **Gene discovery.** Without rRNA depletion gene detection is significantly reduced. Removal of rRNA significantly increases gene detection (green). Simultaneous depletion of rRNA and abundant transcripts increases gene detection additionally by 12 %. **B) Heatmap.** Comparison of rRNA depletion vs simultaneous depletion of rRNA and abundant transcripts shows a reduction in highly abundant transcripts, e.g., ArthCp002 = *psbA*. **C) Coverage of *trnK* increases upon *psbA* depletion.**

Annotations of Plant Genomes can be Significantly Improved with CORALL

Even model organisms such as *Arabidopsis* still show significant gaps in the annotation. Although rRNA is highly abundant the latest annotation **arabidopsis_thaliana_ath_TAIR10.49 still misclassified various rRNA transcripts as intergenic regions leading to only 54 % rRNA in total RNA**. However, following manual curation of known rRNA sequences, the **~43 % intergenic reads are reduced to 8 %.** Concomitantly, the **rRNA content observed with the curated annotation accurately reflects the 94 % rRNA expected** in total RNA samples. Also for other highly abundant transcripts such as *psbA,* **the annotation misses genuine 3’ and 5’ ends**. With **CORALL’s improved 5’ coverage**, alternative splice variants as well as **alternative transcription start** (TSS) and **transcription end sites** (TES) can be properly annotated generating invaluable additional information for plant scientists (Figs. 7 - 8).

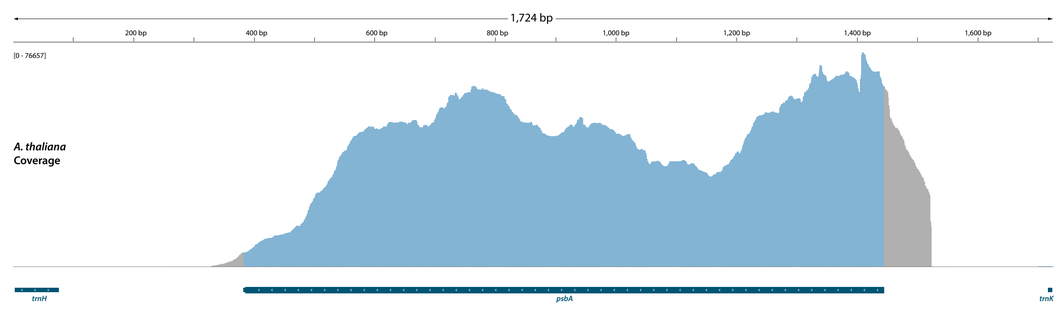

Figure 7 | *PsbA* shows incomplete annotation. Even highly abundant transcripts such as *A. thaliana psbA* show incomplete annotation at 3’ and 5’ ends. CORALL significantly improves 5’ coverage of transcripts and is ideal to determine genuine TSS, improving annotations.

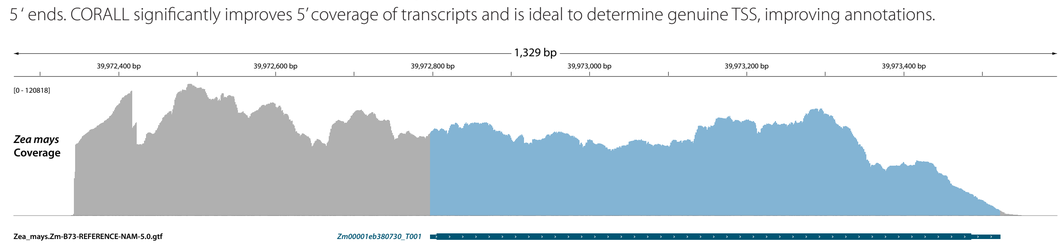

Figure 8 | *Zea mays* annotation lacks almost 40 % of the transcript length/gene (Zm00001eb380730 P48183 Photosystem II protein D1). CORALL library tracks show continuous coverage of the transcript extending up to 40 % into the non-annotated upstream region. Hence, the annotation lacks the genuine 5’ end of the transcript.

Conclusion

RiboCop efficiently removes rRNA from a wide range of input amounts and plant species. The **enzyme-free depletion** allows the recovery of full length transcripts. Intact and degraded input material can be used. **Depletion of rRNA significantly improves gene detection** (8x more genes detected within the first 1M reads), therefore saving costs on sequencing and data analysis. Depending on the tissue source (e.g., leaves) **additional depletion of abundant transcripts can further increase gene detection** by 12 %. Plant genome annotations still show surprising gaps and could profit from additional RNA-Seq data. RiboCop for Plants and CORALL whole transcriptome library preps can be used to significantly improve gene and transcript annotations for all non-rRNA (including lncRNAs), and enable splice variant analysis.

For more information please visit our website: www.lexogen.com/ribocop-rrna-depletion-kit/plants