Plant Non-coding RNAs

- **General background and many details of the topic from:**

The Plant Transcriptome Landscape Puzzle

- Plant genomes very greatly in size
  - 0.1 Gb - 25.0 Gb
- Most species rich in TE, repeats, and satellite DNA
- Early on this DNA was considered Junk DNA
  - Junk DNA = non-functional (protein coding) DNA
- But 90% of the entire genome is transcribed
  - But only 2% of RNA used to encode protein

- **What are the other RNAs?**
- **What is the function/relevance of the other RNA products?**

Transcription Products Known for A Long Time

- **mRNA for Protein Coding Genes**
  - Functions are to encode
  - Enzymatic, structural, regulatory functions
  - Responsible for nearly all phenotypes

- **Traditional non-coding RNA**
  - tRNAs = Transfer RNAs
  - rRNAs = Ribosomal RNAs
  - snRNAs = Small nuclear RNA
    - Processing of hnRNA
  - snoRNAs = Small nucleolar
    - Involved in chemical modification of other RNAs
      - rRNAs, tRNAs, snRNAs
        - Methylation
        - Pseudouridinylation
Non-coding RNAs: Genomic Era Discovery

Major features of plant ncRNAs

• Provide no (or very little) protein coding function
• Act as regulatory elements
• Primarily located in the intergenic (between genes) regions of the genome

General Classes of Plant Non-coding (nc) RNAs

• **Small RNAs**
  o 18-30 nucleotides (nt) in length
• **Medium-sized ncRNAs**
  o 31-200 nt in length
• **Long ncRNA (lncRNA)**
  o >200 nt in length
Small, Non-coding RNAs of Plants

- Regulatory RNAs that act through gene silencing
- Function via
  - Transcriptional silencing
  - Posttranscriptional gene silencing
- Three classes of small RNAs (sRNAs)
  - microRNA (miRNA)
    - Encoded by miRNA genes in the genome
    - Typically, 21nt long
    - Regulate expression by
      - Cutting mRNA transcript or repressing translation
  - small interfering RNA (siRNA)
    - Derived from double stranded RNA produced by transposable elements and repeat elements
    - Typically, 24nt long
    - Regulate expression by
      - DNA methylation OR histone methylation
  - Phased small interfering RNAs (phasiRNA)
    - Derived from
      - Cleavage products of mRNA
        - Dicots
          - Nucleotide-binding leucine-rice repeat (NLR) encoding genes
          - Pentatricopeptide repeat (PRR) encoding genes
        - Monocots
          - Long-non coding RNAs (lncRNA)
      - Non-coding trans-acting siRNAs TAS transcripts (tasiRNA)
        - Typically 21-24nt long

- Common miRNA, siRNA, and phasiRNA functional biochemical steps
  1. Double-stranded RNA (dsRNA) molecules are induced
  2. dsRNAs are processed into 18-25 nt sRNA
  3. sRNA is incorporated into functional effector AGO (Argonaute) complexes that affect gene expression
Synthesis and Evolution of Plant Small RNAs

Figure 1  Biogenesis and modes of action of plant small RNAs. (a) A MIR gene is transcribed into a pri-miRNA, which is sequentially processed first into a pre-miRNA and then into a miRNA/miRNA* duplex. The duplex is methylated by HEN1, and the miRNA strand is loaded into AGO1 in the nucleus. The miRNA-AGO1 complex is transported to the cytoplasm and regulates target gene expression through transcript cleavage and/or translation repression. (b) Pol IV generates single-stranded siRNA precursors, which are converted into dsRNAs and processed into 24-nt siRNA duplexes. Methylated siRNAs are loaded into AGO4 in the cytoplasm and are transported to the nucleus, followed by the recruitment of these siRNA-AGO4 complexes to Pol V transcripts. The subsequent recruitment of DRM2 catalyzes DNA methylation at RdDM target loci. (c) TAS or PHAS loci are transcribed into single-stranded RNAs that are targeted by an miRNA-AGO1/7 complex. The 5’ or 3’ cleavage fragment is protected by SGS3 and converted into dsRNA by RDR6. DCL proteins process these dsRNAs into 21- or 24-nt phasiRNAs. The 21-nt tasiRNAs, which are phasiRNAs from TAS loci, are primarily loaded into AGO1 and guide transcript cleavage of their targets.

Abbreviations: AGO, ARGONAUTE; AMP1, ALTERED MERISTEM PROGRAM 1; CBC, CAP-BINDING COMPLEX; DCL, DICER-LIKE; DDL, DAWDLE; DRM2, DOMAINS REARRANGED METHYLASE 2; dsRNA, double-stranded RNA; HEN1, HUA ENHANCER 1; HST, HASTY; HYL1, HY Ponastic LEAVES 1; Me, methylated; phasiRNA, phased siRNA; Pol, RNA polymerase; pre-miRNA, precursor miRNA; pri-miRNA, primary miRNA; RdDM, RNA-directed DNA methylation; RDR2/6, RNA-DEPENDENT RNA POLYMERASE 2/6; SE, SERRATE; SGS3, SUPPRESSOR OF GENE SILENCING 3; siRNA, small interfering RNA; tasiRNA, trans-acting siRNA; TRN1, TRANSPORTIN 1.
Plant miRNAs

- Encoded by ~100s of *MIR* genes
  - Low abundance relative to siRNAs
- 20-22 nt in size
  - Contain
    - 2-nt overhangs
    - 5’phosphate and 3’hydroxyl groups
  - On BOTH ends
- Act as regulatory RNAs

Defining Plant MicroRNAs

- Requirements for acceptance by the Plant MicroRNA Database (miRbase)
  - [http://www.mirbase.org/](http://www.mirbase.org/)

**Table 1. Updated Criteria for Plant miRNA Annotations**

<table>
<thead>
<tr>
<th>2008 Criteria</th>
<th>2018 Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 One or more miRNA/miRNA* duplexes with two-nucleotide 3’ overhangs</td>
<td>Add requirements that exclude secondary stems or large loops in the miRNA/miRNA* duplex and limit precursor length to 300 nucleotides</td>
</tr>
<tr>
<td>2 Confirmation of both the mature miRNA and its miRNA*</td>
<td>Disallow confirmation by blot; sRNA-seq only</td>
</tr>
<tr>
<td>3 miRNA/miRNA* duplex contains ≤4 mismatched bases</td>
<td>Up to five mismatched positions, only three of which are nucleotides in asymmetric bulges</td>
</tr>
<tr>
<td>4 The duplex has at most one asymmetric bulge containing at most two bulged nucleotides</td>
<td>Up to five mismatched positions, only three of which are nucleotides in asymmetric bulges</td>
</tr>
<tr>
<td>5 ≥75% of reads from exact miRNA or miRNA*</td>
<td>Include one-nucleotide positional variants of miRNA and miRNA* when calculating precision</td>
</tr>
<tr>
<td>6 Replication suggested but not required</td>
<td>Required; novel annotations should meet all criteria in at least two sRNA-seq libraries (biological replicates)</td>
</tr>
<tr>
<td>7 Homologs, orthologs, and paralogs can be annotated without expression data, provided all criteria met for at least one locus in at least one species</td>
<td>Homology-based annotations should be noted as provisional, pending actual fulfillment of all criteria by sRNA-seq</td>
</tr>
<tr>
<td>8 miRNA length not an explicit consideration</td>
<td>No RNAs &lt;20 nucleotide or &gt;24 nucleotides should be annotated as miRNAs. Annotations of 23- or 24-nucleotide miRNAs require extremely strong evidence.</td>
</tr>
</tbody>
</table>
Examples of 2018 Valid/Invalid Plant miRNAs

Figure 1. Examples of Valid and Invalid miRNA Loci. Small RNA-seq data were from GSE105262 (Polydore and Axtell, 2017). Left: Arabidopsis (ath) MIR399b, a locus judged valid under the new criteria. Right: ath-MIR405a, a locus judged invalid under the new criteria.
Number of miRNAs in Plants and Animals

- Data from miRbase database ([http://www.mirbase.org/](http://www.mirbase.org/))
- April 20, 2020

<table>
<thead>
<tr>
<th>Common name</th>
<th>Genus/species</th>
<th># miRNA primary transcript</th>
<th># miRNAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grape</td>
<td><em>Vitis vinifera</em></td>
<td>159</td>
<td>178</td>
</tr>
<tr>
<td>Arabidopsis</td>
<td><em>Arabidopsis thaliana</em></td>
<td>326</td>
<td>428</td>
</tr>
<tr>
<td>Corn</td>
<td><em>Zea mays</em></td>
<td>172</td>
<td>325</td>
</tr>
<tr>
<td>Rice</td>
<td><em>Oryza sativa</em></td>
<td>588</td>
<td>654</td>
</tr>
<tr>
<td>Mosquito</td>
<td><em>Aedes aegypti</em></td>
<td>155</td>
<td>192</td>
</tr>
<tr>
<td>Nematode</td>
<td><em>Caenorhabditis elegans</em></td>
<td>253</td>
<td>450</td>
</tr>
<tr>
<td>Drosophila</td>
<td><em>Drosophila melanogaster</em></td>
<td>267</td>
<td>495</td>
</tr>
<tr>
<td>Chimpanzee</td>
<td><em>Pan troglodytes</em></td>
<td>697</td>
<td>707</td>
</tr>
<tr>
<td>Human</td>
<td><em>Homo sapiens</em></td>
<td>918</td>
<td>2883</td>
</tr>
</tbody>
</table>

Functional Components of Plant miRNAs


Details of complexes involved in miRNA synthesis function.
Synthesis of miRNAs

1. miRNAs transcribed from nuclear MIR genes by RNA polymerase II
   - Primary transcript forms imperfect fold back structure
     - Called pri-miRNAs
       - pri-mRNAs are capped and methylated

2. pri-miRNAs converted to
   - pre-miRNA
     - folds into stem-loop structure
       - then
     - miRNA/miRNA* dsRNA molecule formed
   - Both steps performed in D-bodies
     - D-bodies
       - Nuclear processing center
         - DCL1/HYL1/SE complex
           - DCL1 = dicer
             - RNA folding protein
               - Null mutants are lethal
           - HYL1 = HYPONASTIC LEAVES 1
             - RNA binding function
           - SE = SERRATE
             - C2H2 zinc finger protein
               - Null mutant is lethal
       - pre-miRNA formation aided by
         - Cap-binding proteins
           - CAP80
           - CAP20

3. miRNA/miRNA* complex methylated at each end
   - Methylation by HEN1
     - Methyltransferase protein

4. miRNA/miRNA* complex moves to cytosol
   - Movement through the HST1 protein
     - HASTY1 protein

See: Yu et al. page 419 for details
Functional regulation of gene expression by miRNAs

- Process called **Post Transcriptional Gene Silencing**
  - PTGS
- Functional unit
  - Called **RNA-Induced Silencing Complex**
    - RISC
- miRNA incorporated into ARGONAUTE 1 (AGO1) protein
  - Arabidopsis AGO family
    - 10 family members
- miRNA* stand degraded
- **AGO1/miRNA complex action**
  - Complex binds to exact or nearly-exact target mRNA
    - Binding by base pair complementarity
- **AGO1 functions by**
  - Slice at the phosphodiester bond between nucleotides 10 and 11 of the miRNA
    - OR
      - Binding to mRNA prevents translation
- AGO1/miRNA functions aided by
  - SQUINT
  - HSP90
Plant Biological Processes Controlled by miRNAs


**Figure 2** Functions of miRNAs in plant development and stress responses and an overview of the current understanding of miRNA-mediated regulation during development (inner circle) and responses to biotic and abiotic stresses (outer circle) in Arabidopsis and rice. Red font indicates miRNA-target modules that act in rice, but not in Arabidopsis.

**miRNAs involved in ALL stages of plant growth, development, and abiotic/biotic stress response!!!**
Stress-Regulated miRNAs in Plants

- Khraiwesh et al. 2012 Biochimica et Biophysica Acta 1819:137

Fig. 2. Summary of stress-regulated small RNAs and their target families. Small RNAs are categorized based on the stress that they respond to. Grey box: abiotic stress, dark grey: biotic stress, green boxes: upregulated small RNA, light green boxes: slightly upregulated small RNA, and red boxes: downregulated small RNA.


**Many species utilize miRNAs during the life cycle!!!**
Plant miRNA Stress Regulatory Pathway

- Khraiwesh et al. 2012 Biochimica et Biophysica Acta 1819:137

Fig. 3. Regulatory network of stress-responsive miRNAs in *Arabidopsis*. A network is proposed that describes the molecular mechanisms underlying the response of *Arabidopsis* plants to different biotic and abiotic stresses. The network is based on the changes in expression profiles of miRNA and subsequent target transcripts in plants under stress. Green boxes: upregulated RNAs; red boxes: downregulated RNAs.

Many signals activate miRNAs to regulate plant growth and development!!!
Plant siRNAs Background

- Derived (mostly) from
  - Heterochromatin repeats
  - Transposable elements
- Modify gene expression by
  - DNA methylation
  - Histone methylation
- Process called
  - RNA-directed DNA methylation (RdDM)
Maintenance of Heterochromatin Methylation by siRNA

- **Source sequences**
  - TEs and repeats
    - **Goat:** Add methyl groups to source loci
- **Steps of siRNA production**
  1. **Creation of siRNA precursors**
     - **Pol IV** recruited by CLASSY remodeling factors
     - Move to RdDM target loci
     - Create single-stranded siRNA precursor molecules
  2. **Synthesis of functionally mature 24-nt siRNAs complex**
     - Double stranded RNA (dsRNA) produced by RNA-DEPENDENT RNA POLYMERASE 2 (**RDR2**)
     - dsRNA processed by DCL3 into 24-nt siRNA
     - 24-nt siRNA loaded into AGO4 (Argonaut 4)
  3. **Meanwhile,**
     - Noncoding transcripts at the same RdDM loci produced
       - Product of RNA polymerase V
         - A plant-specific RNA polymerase
  4. **Formation of the methylation complex**
     - Non-coding transcript complexes with 24nt siRNA/AGO4 complex
     - DRM2 (DNA REARRANGED METHYLASE 2) recruited to the complex
     - Methylation state established at RdDM loci by action of DRM2
- **Additional Features**
  - A. Histone H3 lysine 9 methylation and DNA methylation
    - A repressive state
      - Assists in recruiting Poly IV (for siRNA production)
  - B. Maintains methylation at CG, CHG, and CHH sequences in
    - **Heterochromatin**
    - TEs
Phasi-RNAs

- Unique feature of Phasi-RNA
  - Source nucleic acid
    - mRNA fragments from miRNA slicing
    - Cleavage products of miRNA RISC complex action
    - TAS (Trans-acting siRNA) loci

Steps of Phasi-RNA production

1. **PHAS or TAS loci transcribed**
   - Loci cleaved by AGO1/7 RISC complex
   - Product bound on 5’ end by SGS3 protein
   - SGS3-bound nucleotide converted into double-stranded RNA
     - Conversion to dsRNA mediated by RDR6

2. **dsRNAs cleaved**
   - By DCL proteins

3. **Cleavage product is methylated**
   - By HEN1
   - Methylated product loaded into AGO1 complex

4. **AGO1 loaded complex cleaves target loci**

Phasi-RNA:
A method to control a genes expression by using slicing products of its mRNA pool to slice other mRNA in its pool.
Unique Functions of Phasi-RNAs in Regulating Disease Defense in Dicots

- **Plant Resistance Associated Genes**
  - **Nucleotide-binding Site, Leucine Rice Repeat Sequences**
    - NB-LRR genes
      - Exists as gene families
        - NB-LRR largest class of genes that PhasiRNAs
    - NB-LRR transcripts
      - Cleaved by 22-nt miRNAs
        - Generate
          - 21-nt DCL4-dependent phasiRNAs
    - NB-LRR derived phasiRNAs
      - Post-transcriptionally regulate NB-LRR expression
  - **Pentatricopeptide repeat (PPR) gene family**
    - Some members regulated by tasiRNAs
      - Transcripts cleaved by miRNAs
    - Derived phasiRNAs of PPR
      - Silence pathogen transcripts
Plant Long Non-coding RNA (IncRNA)

Types of Plant IncRNAs

- **Linear**
  - Derived from:
    - Intergenic regions
      - Long intergenic ncRNAs = lincRNAs
    - Introns
      - Intronic ncRNAs = incRNAs
    - Coding regions
      - Natural antisense transcripts = IncNATs

- **Circular**
  - Derived from:
    - Coding or intronic regions
      - Circular RNAs = circRNAs

Linear IncRNAs

- Synthesized by
  - RNA Polymerase II
  - Show typical mRNA features
    - 5’ methylated CAP
    - PolyA+ tail
  - Biochemical roles
    - As RNA molecule with other RNAs or proteins
    - Not Translated

IncNATs

- Expression CORRELATED with level of sense strand mRNA expression
  - 70% of Arabidopsis mRNAs have associated IncNATS
- IncNATs associated with histone methylation
  - Components of scaffolds that recruit
    - DNA-modifying and histone-modifying enzymes
- Induce epigenetic changes that
  - Regulate (up or down) gene expression
lincRNAs

• Expression NOT CORRELATED with neighboring gene expression
• Function in trans
  o Target gene elsewhere in the genome
• Example:
  o lincRNA ELENA1
    ▪ Interacts with transcription module
      • Activates pathogen resistance protein PR1

incRNA Example

• COLDAIR incRNA
  o Transcribed from first intron of flowering gene FLC
  o Interacts with lincRNA COLDWARP a FLC promoter IncRNA
    ▪ FLC gene expression is silenced

Circular IncRNAs

• Non-polyadenylated
• Synthesized by
  o Back-splicing of internal exons
    ▪ Exported to cytoplasm
  o Intron lariats
    ▪ Nuclear localized
• Few functions discovered
Figure 1. A schematic drawing to illustrate the diversity of lncRNAs in mammalian genomes. cirRNA, circular intronic RNA derived from intron lariats; circRNA, circular RNA produced from back-splicing of exons; eRNA, enhancer RNA; lincRNA, large intervening/intergenic noncoding RNA; NAT, natural antisense transcript; PROMPT, promoter upstream transcript; sno-lncRNA, small nucleolar RNA (snoRNA)-ended lncRNA; SPA, 50 snoRNA-ended and 30-polyadenylated lncRNA; metastasis-associated lung adenocarcinoma transcript 1 (MALAT1)/Nuclear enriched abundant transcript 1, the long isoform (NEAT1_2), lncRNA with the 30 end alternatively processed by ribonuclease P (RNase P). The number of lncRNAs identified in each category in mammals is shown beneath.
Functions of Plant IncRNAs


Table 1. List of plant long noncoding RNAs identified on the basis of their biological function and mode of action.

<table>
<thead>
<tr>
<th>SPECIES</th>
<th>Name</th>
<th>Mechanism of action</th>
<th>Biological function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis thaliana</td>
<td>IPS1</td>
<td>Target mimicry: interacts with an miRNA, ath-mir399</td>
<td>Regulates phosphate homeostasis</td>
<td>[22]</td>
</tr>
<tr>
<td></td>
<td>COLDAIR</td>
<td>Epigenetic silencing: modification of histones</td>
<td>Regulates flowering</td>
<td>[27]</td>
</tr>
<tr>
<td></td>
<td>COOLAIR</td>
<td>Epigenetic silencing: Promoter interference</td>
<td>Regulates flowering</td>
<td>[88]</td>
</tr>
<tr>
<td></td>
<td>HID1</td>
<td>Stage specific: Associates with chromatin and represses transcription of phytohormone interaction factor 3 (PIF3)</td>
<td>Regulates photomorphogenesis</td>
<td>[103]</td>
</tr>
<tr>
<td></td>
<td>asHSFB2a</td>
<td>Hijacks nuclear Antisense transcription regulators</td>
<td>Regulates the gametophytic and vegetative development</td>
<td>[111]</td>
</tr>
<tr>
<td></td>
<td>ASCO-IncRNA</td>
<td>Regulators of alternate splicing</td>
<td>Regulates the development of lateral roots</td>
<td>[5]</td>
</tr>
<tr>
<td></td>
<td>APOLO</td>
<td>Causes the formation of Chromatin loop and regulates the expression of its neighboring genes</td>
<td>Regulated plant development through auxin signaling</td>
<td>[2]</td>
</tr>
<tr>
<td></td>
<td>ELENA1</td>
<td>Increases the expression of genes involved in plant immunity, such as PR1 and PR2, B 1,3glucanase and salicylic acid genes</td>
<td>Upregulated in disease caused by Pseudomonas syringae</td>
<td>[61]</td>
</tr>
<tr>
<td></td>
<td>Drought induced RNA (DRIR)</td>
<td>Regulates stress responses such as stomata closure</td>
<td>Upregulated in drought and salt stress</td>
<td>[72]</td>
</tr>
<tr>
<td></td>
<td>CDF5 LONG NONCODING RNA (FLORE)</td>
<td>Natural anti-sense transcript of CDF5, upregulates flowering</td>
<td>Photomorphogenesis</td>
<td>[26]</td>
</tr>
<tr>
<td>Glycine max (soybean)</td>
<td>GmENOD40</td>
<td>Involved in the development of nodules</td>
<td>Regulates the re-localization of proteins from nucleus to cytoplasm</td>
<td>[117]</td>
</tr>
<tr>
<td>Hordeum vulgare (Barley)</td>
<td>HvCesA6 Inc-NAT</td>
<td>Acts as a precursor for siRNA</td>
<td>Involved in the synthesis of cell wall</td>
<td>[25]</td>
</tr>
<tr>
<td>Medicago truncatula</td>
<td>MeENOD40</td>
<td>Involved in the development of nodules</td>
<td>Regulates the re-localization of proteins</td>
<td>[83]</td>
</tr>
<tr>
<td>Oryza sativa (Rice)</td>
<td>OsPi1</td>
<td>Unknown</td>
<td>Regulates phosphate homeostasis</td>
<td>[106]</td>
</tr>
<tr>
<td></td>
<td>OsENOD40</td>
<td>Involved in the development of nodules</td>
<td>Unknown</td>
<td>[42]</td>
</tr>
<tr>
<td></td>
<td>LDMAR (P/TMS12-1)</td>
<td>Acts by causing promoter methylation</td>
<td>Male fertility regulation</td>
<td>[17]</td>
</tr>
<tr>
<td></td>
<td>Cis-NATPHO1;2</td>
<td>Acts by enhancing translation of PHOSPHATE1;2</td>
<td>Regulates phosphate homeostasis</td>
<td>[30]</td>
</tr>
<tr>
<td></td>
<td>IncRNAs: 2224 lincRNAs: 1624 IncNATS: 600</td>
<td>Tissue and stage specific: Anthers, pistils, seeds five days after pollination, and shoots 14 days after germination</td>
<td>Regulates sexual reproduction</td>
<td>[128]</td>
</tr>
<tr>
<td>Petunia hybrid</td>
<td>SHO Inc-NAT</td>
<td>Degrades dsRNA</td>
<td>Regulates the synthesis of local cytokines</td>
<td>[138]</td>
</tr>
<tr>
<td>Populus trichocarpa</td>
<td>incRNA20 lincRNA2752</td>
<td>target mimic of ptc-miR476 target mimic of ptc-miR169</td>
<td>Upregulated in abiotic stress (drought)</td>
<td>[80]</td>
</tr>
<tr>
<td>Solanum lycopersicum</td>
<td>TPS11</td>
<td>Unknown</td>
<td>Regulates phosphate homeostasis</td>
<td>[53]</td>
</tr>
<tr>
<td></td>
<td>SlyInc0195 slyInc1077</td>
<td>target mimic of miR166 target mimic of miR399</td>
<td>Upregulated in infections caused by tomato yellow leaf curl virus</td>
<td>[97,101]</td>
</tr>
</tbody>
</table>
Functional Roles of lncRNAs

Figure 3 Representative models for the roles of plant long noncoding RNAs (lncRNAs). lncRNAs can serve as scaffolds, molecular mimics and sponges, and small interfering RNA precursors. They can also be translated into small peptides to regulate target genes in cis or in trans during plant development. The expression of most reported plant lncRNAs is induced by diverse environmental conditions. COLDAIR, COOLAIR, MAS, and LAIR are lncRNAs regulating mRNA transcription in cis. PMS1T is a lncRNA acting as a phasiRNA precursor. IPS1 regulates PHO2 by acting as the endogenous target mimic of miR399 and affects phosphate homeostasis together with another lncRNA, NATpho1;2. Pri-miR171b is a peptide-encoding lncRNA. PROMPTs and ASCO-lncRNA are protein-binding lncRNAs that suppress the function of target proteins.