Dynamics of Genes Regulatory Network Governing de novo Lateral Root Primordium Development in Arabidopsis thaliana

Mikaël Lucas - CR IRD
Kevin Bellande - Post-doctorant IRD
Context: Root System

Development

Root Architecture

Nutrition

Anchorage

Interaction with microorganisms
Context: Secondary organogenesis

From a single embryonic root ... 

... to a mature root system ... 

... with a complex architecture resulting from interactions between genetics and environment.
In Arabidopsis, LR initiation occurs in the pericycle. And LR develop through the tissues to finally emerge.
Context: Lateral root organogenesis

Well described sequence of morphogenesis

Is this development highly regular?
Let’s have a look at lots of LRP

Context: Lateral root organogenesis (Lucas et al. 2013)
Let’s have a look at lots of LRP

Number of cells increases linearly

(Lucas et al. 2013)
Let's have a look at lots of LRP

LRP patterning is not stereotypical

Multiple ways of building a LRP

Sequence of division events actually varies between LRP

(Lucas et al. 2013)
Context: Lateral root organogenesis

Plastic development, but fixed ending

DYNAMIC PATTERNING?
Context: Lateral root organogenesis

Elementary dynamic organogenesis processes

Complex, random (?) root architecture

How to study this process to understand and control RSA?
Modulation of lateral root initiation

LRP initiation correlated with root bending

Gravitropism induces root bending

Can new lateral roots be induced using gravistimulation?

(De Smet et al. 2007)

(Rosen, 1999)
Modulation of lateral root initiation

(Lucas et al. 2008)
Induction of rhizogenesis by gravistimulation

Gravistimulation induces initiation…

… within a tightly controlled spatio-temporal window

DR5:GFP and propidium iodide

Repeated Gravistimulation

Normal growth

(Lucas et al. 2008)
Control of rhizogenesis by gravistimulation

Gravistimulation induces initiation...

... within a tightly controlled spatio-temporal window

Can we use this to access the dynamics of LR morphogenesis regulation?

(Péret et al. 2012)
Transcriptomics of LRP development

3 days after germination

90° rotation

Harvest bends with synchronized LRP

Pool RNA and run transcriptomics analysis

(Voss et al 2015)
Transcriptomics of LRP development

(Voss et al 2015)
Transcriptomics of LRP development

300 to 400 bends per timepoint

18 timepoints: 3 hours apart, from before initiation to after emergence

4 replicates

Database of ~8500 differentially expressed genes (inc. 700 TFs)

(Voss et al 2015)
Transcriptomics of LRP development

Illustration of some transcription factors expression profiles from the database.
Extracting information from the LR dataset?

↓

Use statistical network inference methods

Developed a new algorithm in our lab (J. Lavenus thesis)
**Time Delay Correlation - TDCor**

- Implemented in R (CRAN)
- Runs on expression profiles extracted from the LR dataset (or any other transcriptomic kinetics dataset)
- Looks for non-combinatorial linear interactions
- Uses Pearson’s correlation with time delay computation to produce a preliminary network
- Uses bootstrap and statistical filters to eliminate false positive and refine the network topology
TDCor - data treatment pipeline

Computation of TPI and DPI distributions

Computation of delays and associated correlations

Correlation filter

Delay filter

Index of directness filter

Prior knowledge filter

Overlap filter

Triangle filter (based on TPI)

Diamond filter (based on DPI)

Looking for MRST

MRST filter

Looking for self-regulations

Maximum # of regulators filter

TPI = Triangle Pruning Index

DPI = Diamond Pruning Index

Lavenus et al., Plant Cell 2015
Using TDCor on the LR dataset

Selection of genes involved in

- lateral root formation
- root meristem organization and activity
- hormonal transduction
- cell division
- cell differentiation

Possibility to include any other gene present on the Affymetrix chip (e.g. selected because of interesting features of its expression profile ...)

Perilli et al. 2012
Using TDCor on the LR dataset

- Selected a list of 261 genes
- Not only transcription factors
- A “prior” data is given to each gene, based on the literature, to indicate if transcriptional regulation activity has been reported or not
- This “prior” information helps the inference procedure by authorizing or not the algorithm to draw outward edges from the node. However, indicating a prior is not compulsory (prior = 2)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Name</th>
<th>Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT1G02850</td>
<td>BGLU11</td>
<td>0</td>
</tr>
<tr>
<td>AT1G03430</td>
<td>AHP5</td>
<td>0</td>
</tr>
<tr>
<td>AT1G03840</td>
<td>MGP</td>
<td>2</td>
</tr>
<tr>
<td>AT1G04220</td>
<td>KCS2</td>
<td>0</td>
</tr>
<tr>
<td>AT1G04240</td>
<td>SHY2</td>
<td>-1</td>
</tr>
<tr>
<td>AT1G04550</td>
<td>BDL</td>
<td>-1</td>
</tr>
<tr>
<td>AT1G04610</td>
<td>YUC3</td>
<td>0</td>
</tr>
<tr>
<td>AT1G04880</td>
<td>AT1G04880</td>
<td>2</td>
</tr>
<tr>
<td>AT1G10470</td>
<td>ARR4</td>
<td>-1</td>
</tr>
<tr>
<td>AT1G12820</td>
<td>AFB3</td>
<td>0</td>
</tr>
<tr>
<td>AT1G15580</td>
<td>IAA5</td>
<td>-1</td>
</tr>
<tr>
<td>AT1G15670</td>
<td>KMD2</td>
<td>-1</td>
</tr>
<tr>
<td>AT1G16060</td>
<td>WRI3</td>
<td>2</td>
</tr>
<tr>
<td>AT1G17950</td>
<td>MYB52</td>
<td>2</td>
</tr>
<tr>
<td>AT1G19050</td>
<td>ARR7</td>
<td>-1</td>
</tr>
<tr>
<td>AT1G19180</td>
<td>JAZ1</td>
<td>2</td>
</tr>
<tr>
<td>AT1G19220</td>
<td>ARF19</td>
<td>2</td>
</tr>
<tr>
<td>AT1G19850</td>
<td>MP</td>
<td>1</td>
</tr>
<tr>
<td>AT1G20700</td>
<td>WOX14</td>
<td>2</td>
</tr>
<tr>
<td>AT1G23080</td>
<td>PIN7</td>
<td>0</td>
</tr>
<tr>
<td>AT1G25410</td>
<td>IPT6</td>
<td>0</td>
</tr>
<tr>
<td>AT1G25470</td>
<td>CRF12</td>
<td>1</td>
</tr>
<tr>
<td>AT1G26680</td>
<td>B3</td>
<td>2</td>
</tr>
<tr>
<td>AT1G26870</td>
<td>FEZ</td>
<td>2</td>
</tr>
<tr>
<td>AT1G27320</td>
<td>AHK3</td>
<td>0</td>
</tr>
<tr>
<td>AT1G27450</td>
<td>APT1</td>
<td>0</td>
</tr>
<tr>
<td>AT1G28130</td>
<td>GH3.17</td>
<td>0</td>
</tr>
<tr>
<td>AT1G30330</td>
<td>ARF6</td>
<td>1</td>
</tr>
<tr>
<td>AT1G30490</td>
<td>PHV</td>
<td>2</td>
</tr>
<tr>
<td>AT1G30580</td>
<td>IBB4</td>
<td>2</td>
</tr>
</tbody>
</table>
Using TDCCor on the LR dataset

Generated a full network (~3h computation on standard PC)
With indices of confidence and directness for each interaction.
But are we confident in the predictions?
Validating the network - the ARF7 case

Transcription factor ARF7 experimental profile in the LR data set

Inference by TDcor algorithm

ARF7 is predicted to occupy a upstream position in the network

ARF7 is predicted to positively regulate a LOB/PUCHI genetic module

Lavenus et al., 2015
Validating the network - the ARF7 case

Lavenus et al, Plant Cell 2015

arf7 arf19
pARF7:ARF7::GR

Treated for 4h with:

NAA
DEX
CHX
NAA+DEX
NAA+CHX
DEX+CHX
NAA+DEX+CHX
Validating the network - the ARF7 case

22/31 targets confirmed by transcriptomics
Validating the network - the ARF7 case

4 primary targets confirmed by ChIP PCR
Moving forward with the network

Having validated our inference approach, we went forward with the network exploration.
Expert (i.e. by hand) analysis of the network structure revealed a modular organisation.
Having a look at some of the genes in those two modules...
What are their expression profile like?

Transcript accumulation vs. Time after gravistimulation.

Voß et al., 2015
Topology of the LR GRN - biological meaning?

Module 1
- LBD16
- PUCHI
- PLT7
- PLT5

Module 2
- PLT3
- PLT4
- PLT1
- PI
- PLT2

Where are they expressed?

Voß et al., 2015
Topology of the LR GRN - biological meaning?

Voß et al., 2015
Topology of the LR GRN - biological meaning?

Flanking region

Module 1

Quiescent center establishment

Module 2

Center region

Voß et al., 2015
There appears to be biological meaning behind this modular topology.

Can we investigate the dynamics of this patterning event (establishment of QC / definition of boundary)?
We wanted to investigate the precise dynamics of our GRN.

With several hundred of genes and interactions to consider and no already available solution to simulate such a system easily, we opted to develop our own software.
Modeling GRN dynamics - PANTHEON

PANTHEON

A PYTHON -BASED GENERIC BOOLEAN NETWORK SIMULATOR

Based on Boolean formalism

Automatically model large-scale genes network

Designed to work from simple network description (list of genes and interactions)
GUI: no need to code to simulate your gene network behavior

Import your network or generate a random one / Export simulation results as csv files

Tools included: in silico mutants study with a click among other things

Modular structure: base library of regulation models can be extended at will with your own
Working on a subset of 134 genes / 495 interactions, full simulation of the network behavior using pure logical or algebraic model

Prediction of majority stable state corresponding to meristematic state (genes from module 2 active, genes from module 1 inactive)
ARGOS Module - Mass *in-silico* mutagenesis and computation of a score of impact on network behavior for each gene (mean hammond distance between wild-type and mutants stable states for all model and mutation combinations)

Highlight the most important genes for the network behavior with no a priori
**ARGOS Module - Mass *in-silico* mutagenesis and computation of a score of impact on network behavior for each gene**

<table>
<thead>
<tr>
<th>Genes</th>
<th>mean distance for KO</th>
<th>Genes</th>
<th>mean distance for OA</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLT1</td>
<td>0.168600746</td>
<td>PLT7</td>
<td>0.346343284</td>
</tr>
<tr>
<td>ARF6</td>
<td>0.136735075</td>
<td>PUCHI</td>
<td>0.326567164</td>
</tr>
<tr>
<td>LRP1</td>
<td>0.13113806</td>
<td>CRF1</td>
<td>0.272742537</td>
</tr>
<tr>
<td>PHB</td>
<td>0.124869403</td>
<td>ARF2</td>
<td>0.248246269</td>
</tr>
<tr>
<td>TMO5</td>
<td>0.103973881</td>
<td>PLT5</td>
<td>0.245466418</td>
</tr>
<tr>
<td>SHR</td>
<td>0.098656716</td>
<td>ARF9</td>
<td>0.226529851</td>
</tr>
<tr>
<td>SCR</td>
<td>0.087817164</td>
<td>ARF17</td>
<td>0.179328358</td>
</tr>
<tr>
<td>SHP1</td>
<td>0.081100746</td>
<td>U.box</td>
<td>0.153955224</td>
</tr>
<tr>
<td>ATML1</td>
<td>0.063302239</td>
<td>ARF19</td>
<td>0.148022388</td>
</tr>
<tr>
<td>PID2</td>
<td>0.060970149</td>
<td>WRKY43</td>
<td>0.13369403</td>
</tr>
</tbody>
</table>

TOP10 predicted as most significant genes when KO or OA
Most impactful genes when OA are in module 1
(early genes which we need to be repressed later on)
Most impactful genes when KO are in module 2 (late genes which we need to be expressed for LRP development)
LRP morphogenesis - In summary

Arabidopsis LR as an excellent model system of organogenesis: simple, controllable, accessible

Creation of the LRP database covering the full development of the organ

Creation of the TDCor algorithm and inference of the LRP development GRN
LRP morphogenesis - In summary

Topological analysis revealed a modular structure tied to biological function and a possible bifurcation switch between flank/organizing center identities.

Creation of an automated Boolean modeling software which predicted that the topology of the GRN was enough to generate a meristematic identity and was able to retrieve modular organization with no *a priori*

Once now, back to biology to confirm the prediction of the model (i.e. working on generating and characterizing mutants...)
Acknowledgements

Vincent Vadez (DR IRD)
Soazig Guyomarc’h (MC UM)
Mikaël Lucas (CR IRD)
Antony Champion (CR IRD)
Alexandre Grondin (CR IRD)
Daniel Moukouanga (TR IRD)
isabelle Bourrié (TR UM)
Julien Lavenus (former PhD)
Trinh Duy Chi (PhD USTH)
Jérémy Lavarenne (PhD CIFRE)
Mathieu Gonin (PhD UM)
Le Khanh Nguyen (PhD GRiSP)
Carla de la Fuente (Postdoc ANR)
Awa Faye (PhD USAID)
Marie-Thérèse Mofini (PhD DAAD)
Donald Tchouomo (PhD DAAD)
Trang Hieu Nguyen (PhD USTH)
Kevin Bellande (PostDoc ANR)
Maguette Mbaye (Master UCAD)
Fadel Ndiaye (Master UCAD)

M. Bennett
K. Hill
U. Voss
K. Kenobi
M. Wilson

H. Fukaki

Y. Boutté
J.-D. Faure
F. Tellier
C. Godin
E. Farcot

Supporting agricultural research for sustainable development