Learning Causality in Plant Gene Regulatory Networks from Kinetic Data.

*Synopsis/abstract.*

Due to their sessile mode of life, plants are subject to drastic variations in their environment that leads to rapid adaptation of molecular behavior. A critical question is understanding the relationships among genes that support this adaptation. In the present review we : i) describe experimental approaches to understand such relationships using time series (kinetic) data, ii) review the analytical approaches used to infer causality in the Gene Regulatory Network from this data data, iii) suggest best practices in experimental design and analytical approaches for future efforts of this kind.

Key references of papers that use time series data for this purpose.

**Introduction**

Inferring a causal link is useful in many applications in plant biology, from genomics to ecology. If some A can cause some B to take on a high value (where A could be a gene in our context, a hormone, or a species in ecology), then preventing B from taking such a value can be done by removing some B, by removing some A or by interfering with the link from A to B. Conversely, making B achieve a higher value can be done by adding more B, adding more A, or enhancing the efficiency of the link from A to B. Commonly, causal relationships in biology may involve several elements A1, ..., Ak influencing some B, sometimes positively and sometimes negatively. The influences can be "linear" in which each element has either a positive or negative weight (or coefficient) or "non-linear" in which case the elements work synergistically. An example of synergy would be a dependency of B on the product of the concentrations of A3 and A7. We know that synergy is widespread in biology.

In many cases, however, we simply lack sufficient data to explore all possible synergies. Suppose for example we wanted to explore the effects of all pairs of genes. The most straightforward way to do that would be to over-express or knock out every pair. This would require something like 300 million manipulations. Thus, methods often work in two phases whose first phase consists of finding a good-fitting linear model and whose second phase consists of exploring the synergies among elements that have large positive or negative weights in the linear models. A sort of pre-first phase is to cluster expression patterns in order to create "super-nodes" that can then be analyzed.

Regardless of the analysis that follows, experimental approaches to finding such causal links may entail performing:

A) "Steady state" experiments under multiple different conditions to detect associations between A and other elements. Such associations are bi-directional but may acquire directionality if it is known that, for example, A is an element that can change other elements (in the genomic context A could be a "transcription factor") and B is not.

B) Experiments that increase the quantity of some A to see which other elements are either enhanced (quantity increases) or repressed (quantity decreases).

C) Experiments that decrease the quantity of A or even knock it out (A goes to 0) may also reveal something about the influence of A.

D) Experiments over a closely spaced time course to enable inferences of the form "the state of A at time t may influence B at time t+1."

This article focusses on inference of causality in genomics, but its techniques apply to any setting (whether in plants or other species) in which elements may singly or collectively affect others. The article consists of three parts:

1. A review of efforts to use time series and other data to infer regulatory edges, showing the kinds of results that can be obtained.

2. A description and a categorization of the experimental methods that are used.

3. An in-silico exploration using the DREAM framework to determine questions of practical interest, notably how to plan experiments to gain the maximum insight from each experiment.