RNA-seq module: Functional analysis of Phytophtora effectors in yeast

W. Morgan and M. Reeder, The College of Wooster

Student Learning Goals

After completing the module, students should be able to…

- Explain how changes in an organism's physiology can produce changes in gene expression.
- Explain the challenges of working with a natural system to determine effector function and the advantages of using a model system.
- Diagram the series of steps (workflow), indicating inputs and outputs, required to perform transcriptome analysis using high-throughput (next generation) DNA sequencing technology, *i.e.* RNA-seq.
- Diagram the series of steps (workflow), indicating inputs and outputs, required to identify differentially expressed genes, starting with RNA-seq raw sequence files.
- Describe the effects of altering key parameters of relevant computation tools.
- Interpret tabular data presenting gene expression ratios (both raw and log2-transformed) and p-values.
- Explain the need to adjust p-values with large data sets.
- (Describe the advantage of using advanced gene set-based bioinformatics tools over the analysis of single genes.)

Pre-requisite knowledge

Before undertaking this module, students should understand the following concepts:

- The process of gene expression (Central dogma) and the mechanism of gene regulation
- Fundamental cellular processes: metabolic pathways, signaling pathways, cell structure, etc.
- Basics of DNA technology:
    - reverse transcription
    - DNA cloning
    - expression vectors
    - inducible promoters

Accompanying Module/Lecture

While undertaking this module, students should complete an accompanying module (not included) on RNA-seq technology that describes (1) methods for preparing cDNA for high throughput sequencing and (2) next generation sequencing technology.

Introduction

Host-pathogen interactions

Organisms are constantly interacting with each other and the physical environment. Interactions that are beneficial to one organism, the pathogen, while harming the other, the host, are defined as pathogenic. Pathogenic organisms include bacteria, fungi, and protists.

Medical and agricultural significance of pathogens

Many pathogens are of great significance to human well-being. In addition to pathogens that attack humans, others attack animals and plants that are of agricultural and ecological importance. Pathogens of plants (or *phytopathogens*) affect a wide array of economically important crop species. The risk posed by these pathogens has become exacerbated due to the extensive use of monoculture (the agricultural practice of growing a single plant species), in which the whole crop may be susceptible to a single pathogen.

The phytopathogens *Phytophthora sojae* and *P. infestans* are of particular importance as they cause approximately $1-2 billion and $5 billion in annual worldwide crop losses in soybean and potato respectively (Kamoun, 2003 Molecular Genetics of Pathogenic Oomycetes). As a result it has become increasingly necessary to elucidate the molecular and biochemical mechanisms that govern *Phytophthora spp.* pathogenicity.

Literature Exploration

To learn more about these agriculturally important pathogens, complete the following questions, referring to the indicated scientific articles.

Tyler (2007) Molecular Plant Pathology **8** : 1–8. DOI : 10.1111/ J .1364-3703.2006.00373.Xkam

1. *What is the host range of* P. sojae*? Why is this organism being studied?*
2. *Why were* Phytophthora *species originally classified as fungi? Looking at Figure 2 of the paper, which "kingdoms" are more closely related to fungi than Stramenopiles (which include* Phytophthora *and other oomycetes)?*
3. *Why would this misclassification lead to ineffective treatments against* Phytophthora *species?*

Kamoun and Smart (2005) Plant Disease **89**: 692-9. DOI: 10.1094/PD-89-0692

1. *What is the host range of* Phytopthora infestans*? Why is this organism being studied?*
2. *What are pathogen effectors?*

Host-Pathogen Interactions

During the infection process, the pathogen must gain access to the host's nutrients and other resources. Not surprisingly, hosts have evolved multi-layered defense systems to combat invading pathogens. To counteract this, many pathogens produce dozens to hundreds of effector proteins to compromise the host immune system. Some of these pathogen-encoded effector proteins are secreted into the external environment of the host; others are delivered, or translocated, either directly or indirectly into the host cell.

Literature Exploration

To better understand how the pathogen compromises the host's defense responses, researchers are particularly interested in proteins secreted by the pathogen that are then transported into the host cell. To better understand the relevance of such host-cell targeted proteins, complete the following questions, referring to the indicated scientific articles.

Birch *et al.* (2006) Trends Microbiology **14**: 8-11. DOI:10.1016/j.tim.2005.11.007

1. *What is the proposed role of the RXLR motif?*
2. *Why are proteins with this motif of interest for understanding how the pathogen attacks the host?*

Kramer *et al.* (2007) PLoS Pathogens **3**: 0179-190**.** DOI:10.1371/journal.ppat.0030021

Referring to the Author Summary on the second page, answer the following questions:

1. *Why do bacterial pathogens deliver effector proteins into host cells?*
2. *In general, why are these effector proteins poorly characterized?*

One way to predict the function of a protein is based on its homology to previously studied proteins. Unfortunately, most of the effector proteins with the RXLR motif share no additional amino acid homology with previously characterized proteins. Consequently their functions during the infection process remain largely unknown.

Use of Yeast as a Model System for Functional Genomics

Since these RXLR effector proteins are generally not homologous to enzymes, they are thought to function as molecular mimics that bind to host proteins and thereby block the target's activity within a biological pathway. But what are the targets of individual effector proteins and what biological pathways in the host are blocked?

Several approaches have been developed for exploring the function of individual effectors. However, each approach has its limitations. One approach has been to genetically engineer the pathogen genome to mutate individual genes that encode effector proteins and then to examine the effects on infection. Another is to express individual effector proteins ectopically in the plant host and then examine the effects on the plant's phenotype, such as its ability to withstand subsequent infection. However, these approaches are costly, take much time, and can be technically difficult.

A cheaper, faster alternative is to use yeast as a model eukaryotic system to study the action of individual effector proteins, although this approach also has its limitations.

Literature Exploration

To learn more about how the yeast *Saccharomyces cerevisiae*, a single celled eukaryote, can be used as a model for a plant cell, complete the following questions, referring to the indicated scientific article.

Munkvold *et al.* (2008) MPMI **21**: 490-502. DOI:10.1094/MPMI-21-4-0490

Referring to the Introduction, answer the following questions:

1. What are the three premises for using yeast to study the function of pathogen effectors?

2. The authors note that less than 1% of GAL1-inducible expression of yeast genes severely inhibit yeast growth. Why is this noteworthy? (Hint: Would this system be a good model if most (>50%) genes inhibited growth?)
3. What examples are presented that what we learn in yeast are relevant to pathology in the natural host (that is, that the processes targeted in a yeast cell are actually perturbed in the natural host during infection)?

Kramer *et al.* (2007) PLoS Pathogens **3**: 0179-190. DOI:10.1371/journal.ppat.0030021

Previously the authors discovered a pathogen effector protein, OspF, that inhibited yeast cell growth, and so they assumed that this effector is blocking a cellular process in the natural host that is conserved among eukaryotes, including yeast. To determine exactly which pathway was being perturbed, the authors used two yeast functional genomic approaches. The first screened for yeast mutants that were hypersensitive to the effector protein; the second approach used transcriptome analysis to examine how yeast cells changed gene expression to compensate for the inhibitory effects of the effector protein.

1. What criterion was used to identify yeast genes that were differentially regulated?
2. How many genes met this criterion after induction of OspF expression?
3. What pathway was identified as important in this analysis?

Module research goals (Specific Aims)

Transcriptome analysis will be conducted in yeast expressing a pathogen effector protein to identify the biological pathway blocked. The specific aims of this project are to…

- First identify yeast genes that are differentially regulated upon expression of a specific effector protein in yeast, and
- Subsequently determine the common biological pathway of these differentially expressed genes.

Protocol (Materials & Methods)

- Accessing data set in Galaxy (loading RNA-seq raw files, etc.)
- Introduction to quality scores & RNA-seq raw files
- Quality control analysis of RNA-seq files
    - Conversion of quality scores (FASTQ Groomer)
    - Examining overall sequence quality (FASTQC)
    - Trimming and filtering of sequences (FASTQ Trimmer, ??)
    - Output: trimmed, filtered reads
- Mapping of RNA sequence reads to reference genome or transcriptome
    - ~~Tophat (for identifying splicing variants)~~
    - Bowtie (mapping to reference)
    - Cufflinks (differential gene expression)
    - Output: frequency of gene hits
- ~~Statistical analysis of differential gene expression~~
- Identify targeted biological pathway
    - ~~Overrepresentation of GO terms, Gene sets, Pathway analysis~~

Which V&C core competencies are being addressed and how

- Ability to apply the process of science:
  - Must understand experimental design
  - Evaluate experimental evidence by analyzing RNAseq data set
- Ability to use quantitative reasoning
  - Interpreting graphs (examine FASTQC output)
  - Applying statistical methods to diverse data (quality scores, differential gene expression, pathway analysis)
  - Managing and analyzing large data sets (all steps)
- Ability to use modeling and simulation
  - Apply informatics tools (throughout)
  - Managing and analyzing large data sets (all steps)
- Ability to communicate…
  - Scientific writing
  - Team participation
- Ability to understand the relationship between science and society
  - Evaluating the relevance of social contexts to biological problems

GCAT-SEEK sequencing requirements

None unless wet lab module is incorporated

Computer/program requirements for data analysis

Access to Galaxy (with programs listed above)

Assessments (content and/or student attitude)

Survey of student attitudes pre- & post activity:

- Student has greater confidence…
  - in pursuing bioinformatics projects.
  - selecting appropriate parameters for individual computational tools.
  - processing large data sets in sequential steps.
  - working in teams or with peers.
  - presenting complex information in tables and figures, and highlighting relevant trends in accompanying text.

Course-embedded assessment either as written report, homework activity, or exam questions:

- Students will be able to…
  - Diagram the series of steps (workflow), indicating inputs and outputs, required to perform transcriptome analysis using high-throughput (next generation) DNA sequencing technology, *i.e*. RNA-seq.
  - Diagram the series of steps (workflow), indicating inputs and outputs, required to identify differentially expressed genes, starting with RNA-seq raw sequence files.

Time line of module

- Lab Period 1 (approx. 3 hours)
    - Overview of RNA-seq analysis to find differentially expressed genes
    - Introduction to Galaxy suite
    - Introduction to quality scores and RNA-seq raw (FASTQ) format
    - Quality control analysis of RNA-seq data
    - Mapping of RNA sequence reads to reference genome; begin DGE analysis
- Lab Period 2 (approx. 3 hours)
    - Differential gene expression analysis
    - ~~Overview of Pathway analysis~~
    - ~~Pathway analysis~~

Discussion Topics for lab

- Overview of RNA-seq analysis to find differentially expressed genes
- Introduction to Galaxy suite
- Introduction to quality scores and RNA-seq raw (FASTQ) format

Lecture Topics for class

- RNA-seq technology
    - preparing cDNA for high throughput sequencing
    - next generation sequencing technology

References

To be added later