Lecture 15:

Introduction to Plant Genomics and Microarrays

BIOL 754 Fall 09

1

Genomics?

- **Genomics** is the study of all of the genes in an organism…
- **Proteomics** is the study of all proteins….
- **Metabolomics** is the study of all metabolic pathways…

All of these areas of study try to unravel the bigger picture of what is going on in an organism, beyond the individual genes.

BIOL 754 Fall 09

2

Lecture Outline

- Model species in plant biology
- Research in the field of plant science
- Microarray technology and microarray experiment animation

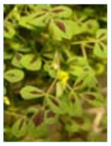
BIOL 754 Fall 09

3

Plant Model Organisms

- ***Arabidopsis thaliana*** (thale cress)
  --model flowering plant and dicot

- ***Oryza sativa*** (rice)
  --model monocot



- ***Medicago truncatula*** (barrel medic)
  --model legume

- ***Lycopersicon esculentum*** (tomato)
  --model fruit-bearing plant

Also maize, tobacco, *Chlamydomonas*, wheat, etc....

Plant Genome Research

- Plant-pathogen interactions and plant-insect interactions
- Determining the evolutionary history of plants using sequence data from conserved genes
- Light perception to set circadian rhythms and determine the developmental pattern of plants
- Increasing the nutrient value of crop plants
- Determining the genetics behind fruit ripening and nutrient accumulation

Tools of Genomics

- Advanced molecular biology techniques
- Quantitative Trait Loci analysis; Linkage and association mapping
- Large-scale sequencing
- Microarrays
- Protein arrays

Genome sequencing

- The first draft of the sequence of the human genome was finished in 2000
- Arabidopsis genome was finished in 2000, representing the first flowering plant (**www.arabidopsis.org** ; **www.tigr.org**)
- Rice is complete (**www.gramene.org**) and initiatives are underway for sequencing Medicago, tomato, soybean

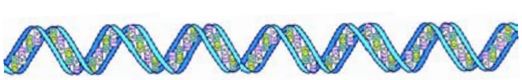
As of October 17, 2009, 1117 genomes (plant, animal, bacterial, and viral) had been sequenced! (was 298 genomes exactly four years ago)
(http://www.genomesonline.org)

Why sequence genomes?

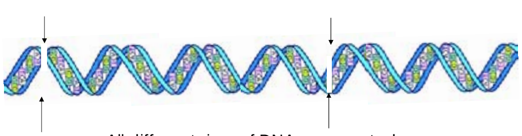
- **Provides information about how genes work**
  Example: Understanding how proteins fold may help us "see" where the catalytic site of an enzyme is. Genes responsible for causing disease.
- **To understand the structure of the genome**
  Example: Are all genes related to photosynthesis grouped together?
- **Makes it much easier to identify the gene of a phenotypic mutation**
  Example: I have a plant with a flower mutation. Using map-based cloning, I can narrow down the options of what it could be.
- **To compare similar genes between different species**
  Example: Flowers in maize and tomato look very different. Are the genes for flower architecture similar in sequence? What does this mean evolutionarily?
- **Discover the locations of genes on chromosomes for plant breeding purposes**
  Example: With a known location of a gene, marker-assisted breeding for drought tolerance is a lot quicker and easier.

How is sequencing done?

First, the genome needs to be broken into smaller pieces

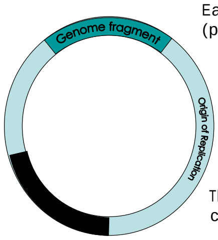


This can be done by sonicating the sample to randomly sheer the DNA



All different sizes of DNA are created

Creating the library



Each fragment is ligated into a vector (plasmid)

Transform each vector into bacteria and select for transformants

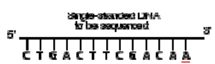
The collection of these vector-containing colonies is called a library

Colonies are grown, DNA is extracted from the bacteria, and sequencing reactions are performed....

Sequencing reaction

- All of the same components as a PCR (Polymerase Chain Reaction) reaction: buffer, enzyme, DNA template, primers, dNTPs (A, T, C, G)
- Two major differences between PCR and a sequencing reaction: use only one primer and in addition to normal dNTPs, there are terminating bases (ddNTPs, dideoxynucleotides)
- Terminating bases have a large fluorescent dye molecule (a different color for each base), which stops the addition of more nucleotides and provides an identifier for the nucleotide

Sequencing Reactions



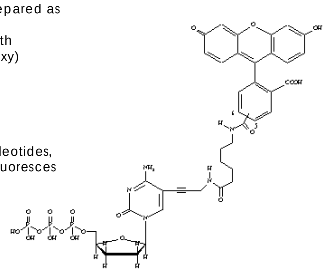
The DNA to be sequenced is prepared as a single strand.
This template DNA is supplied with a mixture of all four **normal** (deoxy) nucleotides in ample quantities

- dATP
- dGTP
- dCTP
- dTTP

a mixture of all four **dideoxy**nucleotides, each labeled with a "tag" that fluoresces a different color:

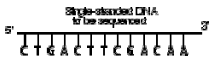
- **ddATP**
- **ddGTP**
- **ddCTP**
- **ddTTP**

DNA polymerase I
Buffer
$MgCl_2$
primers



Fluorescein-12-ddCTP

Sequencing Reactions



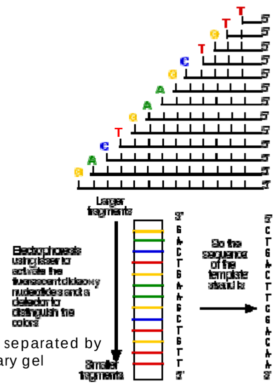
The DNA to be sequenced is prepared as a single strand.
This template DNA is supplied with
a mixture of all four **normal** (deoxy) nucleotides in ample quantities

- dATP
- dGTP
- dCTP
- dTTP

a mixture of all four **dideoxy**nucleotides, each labeled with a "tag" that fluoresces a different color:

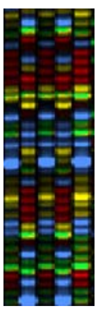
- **ddATP**
- **ddGTP**
- **ddCTP**
- **ddTTP**

DNA polymerase I
Buffer
$MgCl_2$
primers

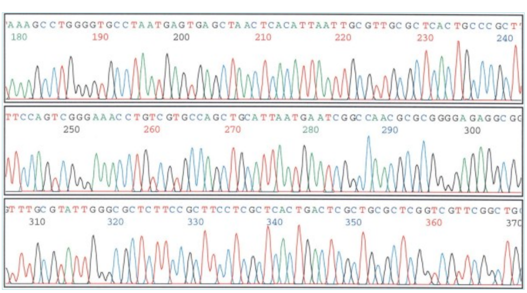


Fragments are separated by size in a capillary gel

Sequencing gel



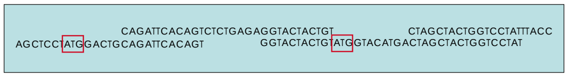
Electropherogram



Animated cycle sequencing reaction
(http://www.dnalc.org/ddnalc/resources/cycseq.html)

Assembly and annotation

- Once all of the DNA has been sequenced and contiged (contig - the DNA sequence reconstructed from a set of overlapping DNA segments), computer software searches for Open Reading Frames (ORFs)



- ORFs are defined by an ATG start codon followed by enough bases before a stop codon to indicate that there is a potential gene (called “putative gene”)

- Can use other software to identify motifs that provide clues to the function and localization of the gene in the cell

- Information is deposited in a database for other researchers to use

Arabidopsis sequencing facts:

- **Arabidopsis** has a 125 Mb sized-genome on 5 chromosomes

  -**human** has 3,000 Mb on 23 chromosomes
  -**maize** has 2,500 MB on 10 chromosomes
  -**Medicago** has 520 Mb on 8 chromosomes
  -**rice** has 430 Mb on 12 chromosomes
  -**lily** has 50,000 Mb on 12 chromosomes

- Arabidopsis has about 25,500 genes
  -humans have slightly fewer, about 24,000

For more information...

Go to the National Center for Biotechnology Information (NCBI) website: **http://www.ncbi.nlm.nih.gov/**. At that site you can:

- Search for literature

- Look for genes and protein sequences (they are deposited in the database)

- Find updates on genome sequencing projects

- lots more!

Microarrays: large-scale observation of gene expression

- Gene expression indicates what is going on in a cell or structure at a given time
- Microarrays allow scientists to look at the gene expression of literally thousands of genes all at once
- Comparing two different conditions on a microarray

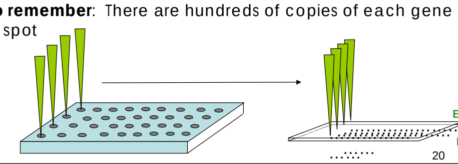
Examples:
1. Leaf in the dark vs. a leaf in the light
2. Diseased plant vs. a normal plant
3. Ripe vs. unripe tomato

Printing the microarray slides

Printed on the microarray slide is a collection of thousands of genes, with a known location. To make the slides:

- First, must do large-scale PCR reactions in multi-well plates
- An automated machine dips into the wells and spots on a glass slide in a specified pattern
- DNA is single stranded on the slide
- Each spot can be DNA, cDNA, or oligonucleotides (short fragment of a single-stranded DNA that is typically 30 to 70 nucleotides long)

Important to remember: There are hundreds of copies of each gene within each spot

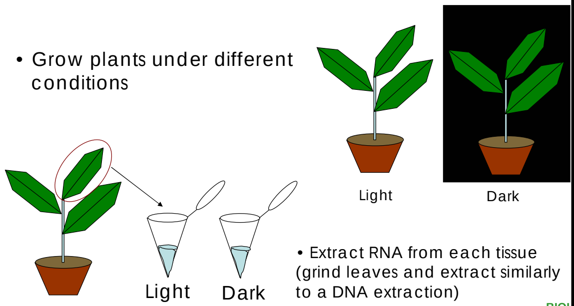


Arrayer: spots DNA on the glass slides

Paul Debbie from the Center for Gene Expression Profiling (CGEP) at The Boyce Thompson Institute for Plant Research, Ithaca, NY (2004)

Steps for Doing a Microarray Experiment

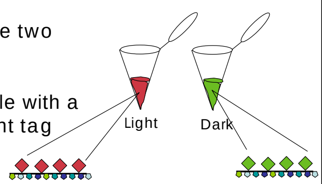
- Grow plants under different conditions

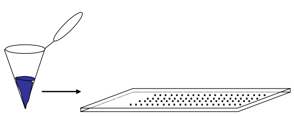


- Extract RNA from each tissue (grind leaves and extract similarly to a DNA extraction)

Microarray experiment, cont.

- Generate cDNA from the two samples
- Label each cDNA sample with a different color fluorescent tag (red and green)



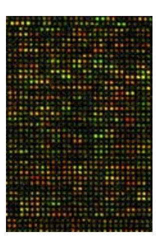


Both red and green tagged cDNA together

- Mix the solutions together and put the mixture on the slide

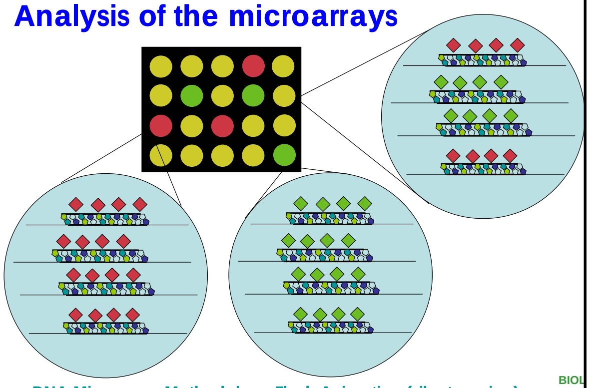
Analysis of the microarrays

- The slide is put in a machine that scans the slide to individually detect the fluorescent dyes
- The computer superimposes the two images
- Statistical software identifies patterns in expression



Superimposed scans

Analysis of the microarrays



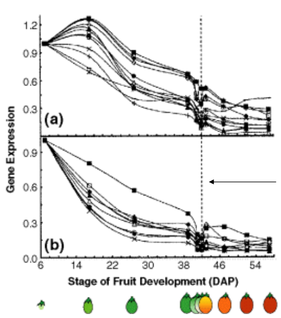
[DNA Microarray Methodology Flash Animation (silent version)](http://www.bio.davidson.edu/courses/genomics/chip/chipQ.html)

http://www.bio.davidson.edu/courses/genomics/chip/chipQ.html

Usefulness of Microarrays

- Previously, gene expression studies had to be done with **blots**
- Blots are time consuming if you are looking at more than a few genes
- The use of microarrays allow scientists to observe gene expression for thousands of genes at once

Microarrays to Dissect plant development and physiology



Expression profile of 21 genes encoding components of the photosynthetic apparatus.

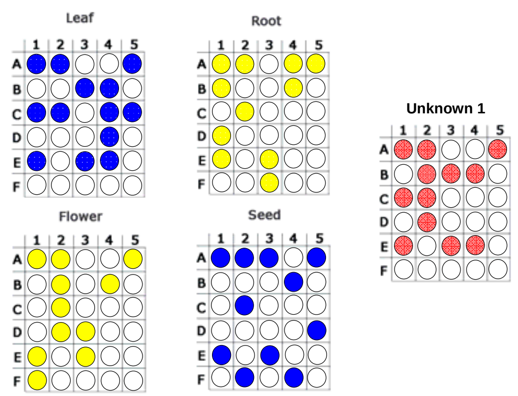
Decrease in the accumulation of transcripts during fruit ripening

Breaker stage of fruit development (i.e. the point at which the fruit begins to turn red)

Alba et al. (2004) Plant J. 39: 697

Microarray Activity

- Each student will have 2 slides representing microarray data sets from 2 different plant organs (a blue and a yellow slide) and one red slide with microarray data from an unknown structure (organ)
- First, lay over the blue and the yellow slide and identify the genes they have in common (the green spots; these are the housekeeping genes)
- Can you make a guess as to what the unknown structure is?



Some important databases

http://www.ncbi.nlm.nih.gov/
-> scroll down to gene -> type in accession number or Atg# -> GO -> links on the right

TAIR = The Arabidopsis Information Resource
www.arabidopsis.org

TIGR = The Institute for Genome Research
www.tigr.org -> Databases -> Plant Genomics

MIPS = Munich Information Center for Protein Sequences
http://mips.gsf.de/