

Introduction to Plant Genomics and Online Resources

Manish Raizada
University of Guelph

Genomics Glossary

http://www.genomenetwork.org/articles/06_00/sequence_primer.shtml

Annotation

Adding pertinent information such as gene coded for, amino acid sequence, or other commentary to the database entry of raw sequence of DNA bases.

Arrayed library

Individual primary recombinant clones (hosted in phage, cosmid, YAC, or other vector) that are placed in two-dimensional arrays in microtiter dishes. Each primary clone can be identified by the identity of the plate and the clone location (row and column) on that plate. Arrayed libraries of clones can be used for many applications, including screening for a specific gene or genomic region of interest.

Assembly

Putting sequenced fragments of DNA into their correct chromosomal positions

Bacterial artificial chromosome (BAC)

A vector used to clone DNA fragments (100- to 300-kb insert size; average, 150 kb) in *Escherichia coli* cells. Based on naturally occurring F-factor plasmid found in the bacterium *E. coli*.

BLAST

A computer program that identifies homologous (similar) genes in different organisms, such as human, fruit fly, or nematode.

cDNA library

A collection of DNA sequences that code for genes. The sequences are generated in the laboratory from mRNA sequences.

Clone

An exact copy made of biological material such as a DNA segment (e.g., a gene or other region), a whole cell, or a complete organism.

Cloning

Using specialized DNA technology to produce multiple, exact copies of a single gene or other segment of DNA to obtain enough material for further study. This process, used by researchers in the Human Genome Project, is referred to as cloning DNA.

The resulting cloned (copied) collections of DNA molecules are called clone libraries.

Cloning vector

DNA molecule originating from a virus, a plasmid, or the cell of a higher organism into which another DNA fragment of appropriate size can be integrated without loss of the vector's capacity for self-replication; Examples are plasmids, cosmids, and yeast artificial chromosomes;

Conserved sequence

A base sequence in a DNA molecule (or an amino acid sequence in a protein) that has remained essentially unchanged throughout evolution.

Contig

Group of cloned (copied) pieces of DNA representing overlapping regions of a particular chromosome.

Contig map

A map depicting the relative order of a linked library of overlapping clones representing a complete chromosomal segment.

Cosmid

Artificially constructed cloning vector containing the cos gene of phage lambda. Cosmids can be packaged in lambda phage particles for infection into E. coli; this permits cloning of larger DNA fragments (up to 45kb) than can be introduced into bacterial hosts in plasmid vectors.

Expressed sequence tag (EST)

A short strand of DNA that is a part of a cDNA molecule and can act as identifier of a gene. Used in locating and mapping genes.

GC-rich area

Many DNA sequences carry long stretches of repeated G and C which often indicate a gene-rich region.

Gene mapping

Determination of the relative positions of genes on a DNA molecule (chromosome or plasmid) and of the distance, in linkage units or physical units, between them.

Gene prediction

Predictions of possible genes made by a computer program based on how well a stretch of DNA sequence matches known gene sequences

Genetic marker

A gene or other identifiable portion of DNA whose inheritance can be followed.

Genome

All the genetic material in the chromosomes of a particular organism; its size is generally given as its total number of base pairs.

Genomic library

A collection of clones made from a set of randomly generated overlapping DNA fragments that represent the entire genome of an organism.

Linkage map

A map of the relative positions of genetic loci on a chromosome, determined on the basis of how often the loci are inherited together. Distance is measured in centimorgans (cM).

Long-Range Restriction Mapping

Restriction enzymes are proteins that cut DNA at precise locations. Restriction maps depict the chromosomal positions of restriction-enzyme cutting sites. These are used as biochemical "signposts," or markers of specific areas along the chromosomes. The map will detail the positions where the DNA molecule is cut by particular restriction enzymes.

Macrorestriction map

Map depicting the order of and distance between sites at which restriction enzymes cleave chromosomes.

Megabase (Mb)

Unit of length for DNA fragments equal to 1 million nucleotides .

Oligonucleotide (Oligo)

A molecule usually composed of 25 or fewer nucleotides; used as a DNA synthesis primer for PCR.

P1-derived artificial chromosome (PAC)

One type of vector used to clone DNA fragments (100- to 300-kb insert size; average, 150 kb) in *Escherichia coli* cells. Based on bacteriophage (a virus) P1 genome.

Physical map

A map of the locations of identifiable landmarks on DNA (e.g., restriction-enzyme cutting sites, genes), regardless of inheritance. Distance is measured in base pairs. For the human genome, the lowest-resolution physical map is the banding patterns on the 24 different chromosomes; the highest-resolution map is the complete nucleotide sequence of the chromosomes.

Positional cloning

A technique used to identify genes, usually those that are associated with diseases, based on their location on a chromosome.

Proteome

Proteins expressed by a cell or organ at a particular time and under specific conditions.

Proteomics

The study of the full set of proteins encoded by a genome.

Pseudogene

A sequence of DNA similar to a gene but nonfunctional; probably the remnant of a once-functional gene that accumulated mutations.

Reciprocal translocation

When a pair of chromosomes exchange exactly the same length and area of DNA. Results in a shuffling of genes.

Restriction enzyme, endonuclease

A protein that recognizes specific, short nucleotide sequences and cuts DNA at those sites. Bacteria contain over 400 such enzymes that recognize and cut more than 100 different DNA sequences.

Restriction fragment length polymorphism (RFLP)

Variation between individuals in DNA fragment sizes cut by specific restriction enzymes; polymorphic sequences that result in RFLPs are used as markers on both physical maps and genetic maps.

Scaffold

In genomic mapping, a series of contigs that are in the right order but not necessarily connected in one continuous stretch of sequence.

Sequence assembly

A process whereby the order of multiple sequenced DNA fragments is determined.

Sequence tagged site (STS)

Short (200 to 500 base pairs) DNA sequence that has a single occurrence in the human genome and whose location and base sequence are known. Detectable by polymerase chain reaction, STSs are useful for localizing and orienting the mapping and sequence data reported from many different laboratories and serve as landmarks on the developing physical map of the human genome. Expressed sequence tags (ESTs) are STSs derived from cDNAs.

Sequencing

Determination of the order of nucleotides (base sequences) in a DNA or RNA molecule or the order of amino acids in a protein.

Shotgun method

Sequencing method that involves randomly sequenced cloned pieces of the genome, with no foreknowledge of where the piece originally came from. This can be contrasted with "directed" strategies, in which pieces of DNA from known chromosomal locations are sequenced. Because there are advantages to both strategies, researchers use both random (or shotgun) and directed strategies in combination to sequence the human genome.

Single nucleotide polymorphism (SNP)

DNA sequence variations that occur when a single nucleotide (A, T, C, or G) in the genome sequence is altered.

Syntenly

Genes occurring in the same order on chromosomes of different species.

Tandem repeat sequences

Multiple copies of the same base sequence on a chromosome; used as markers in physical mapping.

Transformation

A process by which the genetic material carried by an individual cell is altered by incorporation of exogenous DNA into its genome.

Yeast artificial chromosome (YAC)

Constructed from yeast DNA, it is a vector used to clone large DNA fragments.

Homolog

* A gene related to a second gene by descent from a common ancestral DNA sequence. The term, homolog, may apply to the relationship between genes separated by the event of speciation (see ortholog) or to the relationship between genes separated by the event of genetic duplication (see paralog).

Ortholog

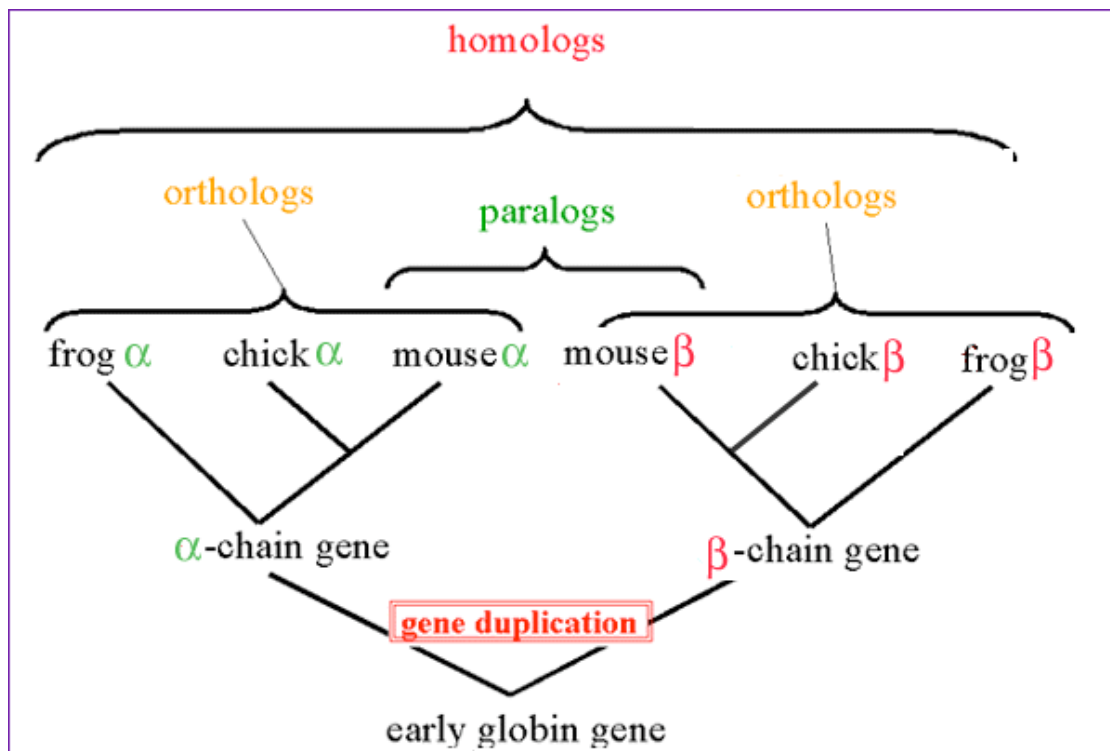
* Orthologs are genes in different species that evolved from a common ancestral gene by speciation. Normally, orthologs retain the same function in the course of evolution. Identification of orthologs is critical for reliable prediction of gene function in newly sequenced genomes. (See also Paralogs.).

Paralog

* Paralogs are genes related by duplication within a genome. Orthologs retain the same function in the course of evolution, whereas paralogs evolve new functions, even if these are related to the original one.

See the following NCBI diagram Orthology

<http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/Orthology.html>



Genome Tutorials - Basics

1. What does the cystic fibrosis gene look like if I want to PCR/sequence myself?

<http://www.ensembl.org/index.html>

Click Homo sapiens, Search Ensembl, then type Cystic Fibrosis

2. Check out human disease homologs in Arabidopsis

http://mips.gsf.de/proj/thal/db/tables/tables_comp_frame.html

3. By user the browser function for Chromo1, find out the number of amino acids in PhytochromeA photoreceptor protein

http://mips.gsf.de/proj/thal/db/gv/gv_frame.html

4. In TAIR, find the genetic map position of PhyA using the search function:

<http://www.arabidopsis.org/servlets/sv>

5. Find PhyA in rice Oryzabase (click genes, type phyA in Keyword)

<http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp>

6. On which chromosome is PhyA in rice? (TIGR database)

(Go to Search Functions, Putative Function Search, type Phytochrome)

<http://www.tigr.org/tdb/e2k1/osa1/>

7. Find chromosome location of phytochrome in maize (corn) (MaizeGDB)

(Go to gene products, Type in Phytochrome in Gene Products with name)

<http://www.maizegdb.org/>

8. Find out microarray expression of Phytochrome:

<http://www.arabidopsis.org>

(go to Microarray Expts, type Phytochrome under description, then shade avoidance...)

9. Examine Portions of the Arab genome which are redundant

http://mips.gsf.de/proj/thal/db/gv/rv/rv_frame.html

10. Examine all the genes in the chloroplast genome – what do you notice?

http://mips.gsf.de/proj/thal/db/gv/gv_frame.html

11. Find the syntenic regions between rice and maize in Gramene.

<http://www.gramene.org/>

12. Find all the genes on rice BAC clone P0436E04 on chromo1 at 0.3cM (TIGR)

http://www.tigr.org/tigr-scripts/e2k1/irgsp_orderedBAC.spl?db=osa1&chr=1

13. Find the number of protein-coding genes in humans

(click on Homo sapiens and check out cover page on right side)

<http://www.ensembl.org/index.html>

How does the human gene number compare to Drosophila? What the...!?!?!?

Links to Plant Genome Sequencing Project Databases

1. General

Genome News Network (see Quick Guide)

<http://www.genomenewsnetwork.org/>

TIGR Plant Genomics Projects

<http://www.tigr.org/plantProjects.shtml>

MIPS (Arabidopsis, Maize, Medicago, Lotus, Rice, tomato)

<http://mips.gsf.de/projects/plants>

*****ENSEMBL (non-plant eukaryotic genomes)

<http://www.ensembl.org/index.html>

NCBI Plant Genome

<http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html>

2. Arabidopsis

****MIPS MATDB Genome Viewer

http://mips.gsf.de/proj/thal/db/gv/gv_frame.html

TAIR Sequence Viewer

<http://www.arabidopsis.org/servlets/sv>

TIGR Arabidopsis Genome Database

<http://www.tigr.org/tdb/e2k1/ath1/ath1.shtml>

MPSS - Arabidopsis small RNA Database

<http://mpss.udel.edu/at/>

Salk Institute Mapped TDNA Insertion Database

<http://signal.salk.edu/about.html>

Arabidopsis GDB (in development)

<http://www.plantgdb.org/AtGDB/index.php>

3. Rice

International Rice Genome Sequencing Project

<http://rgp.dna.affrc.go.jp/IRGSP/index.html>

TIGR Rice Genome Annotation

<http://www.tigr.org/tdb/e2k1/osa1/index.shtml>

Oryzabase

<http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp>

Links to Plant Genome Sequencing Project Databases (continued)

4. Maize (corn)

<http://www.maizegdb.org/>

5. Comparative Grass Genomics

***Gramene

<http://www.gramene.org/>

GrainGenes - wheat and oat

<http://wheat.pw.usda.gov/GG2/index.shtml>

6. Poplar (tree)

<http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>

7. Potato (and Solanaceae)

<http://www.tigr.org/tdb/potato/>

8. Tomato

http://www.sgn.cornell.edu/help/about/tomato_sequencing.pl

9. Lotus japonicus (legume)

<http://www.kazusa.or.jp/lotus/>

10. Medicago (alfalfa, legume)

<http://mips.gsf.de/proj/plant/jsf/medi/index.jsp>