Egyptian J. of Sheep and Goat Sciences (Special Issue, 2nd Inter. Sci. Conf. on SR Production, 2008) Vol. 3(1): 188-220

Utilization of Functional Genomics Techniques in Sheep Breeding, A Recent Approach For Genetic Improvement

Ahmed R. Elbeltagy, PhD
Dept. Animal Biotechnology
Animal Production Research Institute, Cairo, Egypt

What is the Genome?

Genome of an organism is the whole hereditary information that is encoded in the DNA (or, for some viruses, RNA). This includes both the genes and the non-coding sequences.

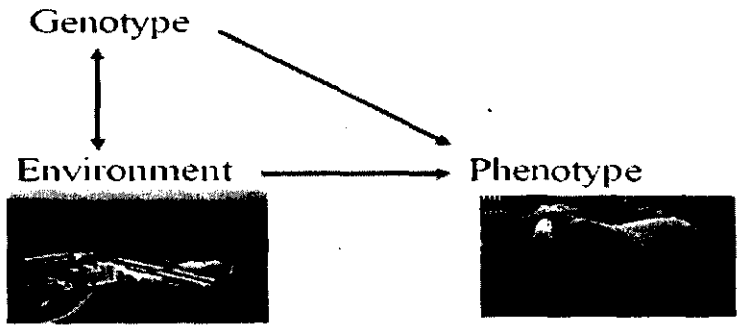
What is structural Genomic?

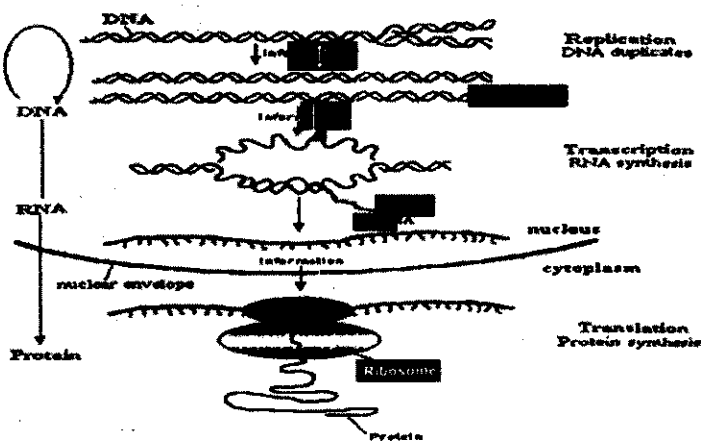
- Structural genomics or structural bioinformatics refers to the analysis of macromolecular structure particularly DNA and proteins, using computational tools and theoretical frameworks.
- The main goal of studying structural genomics is the extension of idea of genomics, to obtain accurate structural models for DNA and related proteins.
- Structural alignment is a tool of structural genomics.

What is functional Genomic?

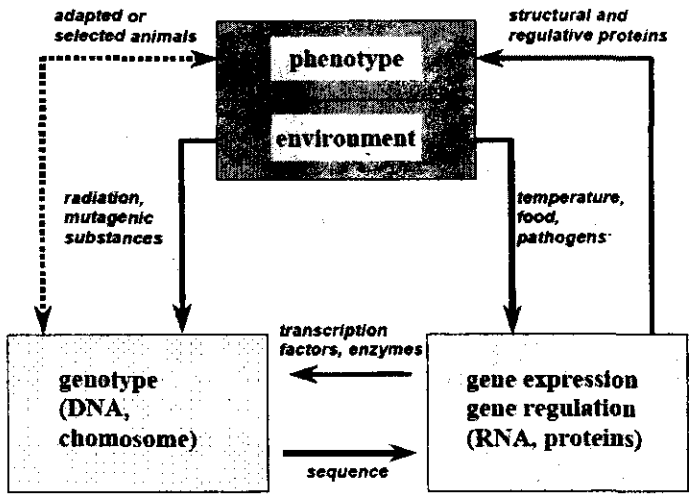
- Functional genomics is concerned with understanding the function of genes and other parts of the genom.
- Functional genomics is a field of molecular biology that is attempting to make use of the vast wealth of data produced by genome mapping and sequencing to describe genome function.
- Functional genomics uses high-throuput techniques like DNA microarrays, proteomics, metabolomics and mutation analysis to describe the function and interactions of genes.
- The fundamental strategy behind functional genomics is to expand the scope of biological investigation from studying single genes, transcripts or proteins to studying all genes, transcripts or proteins simultaneously, in a systematic fashion.
- Functional genomics thus promises to narrow the gap between sequence and function to yield new insights into the behavior of biological systems.

- The billions of bases of DNA sequence do not tell us what all the genes do, how cells work, how organisms form, what is going wrong in disease states, production performance, or how an organism ages. This is where functional genomics comes into play.
- The overall goal is not simply to provide a catalog of all the genes and the information about their functions (Structural Genomics), but to understand how the components work together to comprise functioning cells and organisms.
- Gene expression can be analyzed at the RNA or Protein level or, in cases in which expression is affected by alterations in the copy number of genes, also at the level of DNA.
- Although in most cases protein production is the ultimate output of the gene, protein analysis techniques are presently more difficult to perform, less sensitive, and have a lower throughput than RNA-based ones.

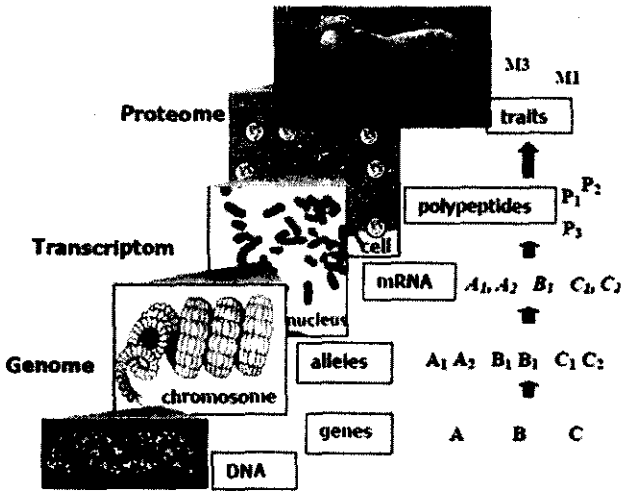




The Central Dogma of Molecular Biology

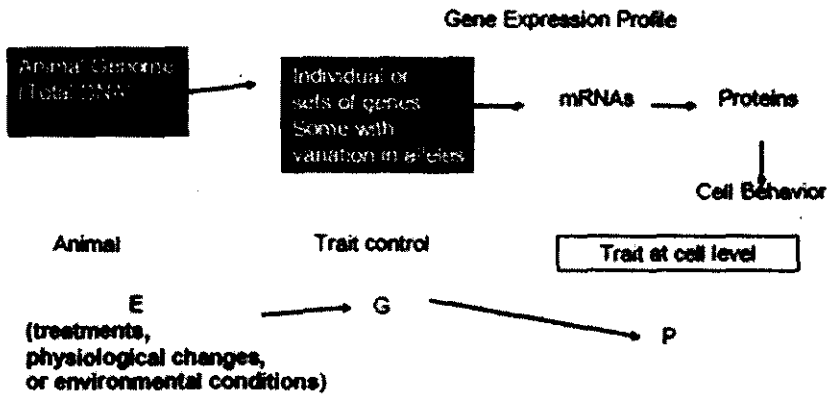


Genotype-Phenotype Interaction



Molecular Basis of Phenotype

Molecular Basis of Phenotype



Central concept of functional Genomics

To identifying key genes in physiological pathways.

To develop methods to up/down regulate the expression of the appropriate major genes.

Functional Genomics Association to Animal Breeding

- For years genetic ability (G) of an animal has been estimated for the collective effects of genes yielding values called Predicted Breeding Value for traits such as milk production and growth performance. Predicted Breeding Value used to be computed without knowing what genes are involved, their function, or their DNA sequence.

- **This approach has resulted in significant improvement for many economic traits.**
- On the other hand, much animal research has focused on understanding the effects of management and environmental factors (E) on physiological responses of animals (P), ignoring (G).
- Functional genomics allows us to put it all together by looking at phenotypes at the cellular and molecular levels and linking genes with their function. The results will further help scientists understand animal biology and lead to new ways to improve health, well-being, and performance.
- Functional genomics provides a peak into the "black box" by linking a trait or characteristic (P), such as disease resistance or milk/meat production, with specific genes, yielding new knowledge about physiological functions such as milk secretion, immune response, digestion, and metabolism at the cell level.
- This is done by determining which genes are "turned on" or "off" in specific types of cells as a result of treatments, physiological changes, or environmental conditions.
- By studying gene expression profiles and known pathways, we can learn more about economic traits (performance); growth, milk secretion, immune response, digestion, and numerous other physiological functions in the sheep.

Functional Genomics Tools (applicable in FAnGR)

High-throughput marker genotyping

DNA hybridization to an oligonucleotide chip
Mass spectrometry
Pyrosequencing

Transcriptome analysis

Serial analysis of gene expression (SAGE)
Massively parallel signature sequencing (MPSS)
Microarray-based transcript analyses

Bioinformatics and statistical modeling in transcriptome analysis

Data normalization and calibration
Detecting differential expression
Multivariate statistical analysis

Proteome analyses

Two-dimensional polyacrylamide gel electrophoresis
Mass spectrometry
Chip-based methods
Bioinformatics in proteome analyses

Strategies for defining gene functions *in vivo*

Transgenesis
Ribonucleic acid interference
Short interfering RNAs (siRNAs)
MicroRNAs (miRNAs)

Tools for studying gene expression at the Transcriptome level

Hybridization-based techniques

- **Northern blotting**
- **Subtraction cloning**
- **DNA arrays**
- **RNase protection assays**
- **RNA invader assay**

PCR-based techniques

- **qRT-PCR**
- **differential display (DD, DDRT-PCR)**
- **Representational Difference Analysis (RDA)**

Sequence based techniques

- **Expressed sequence tags (ESTs)**
- **Serial analysis of gene expression (SAGE)**
- **DNA sequencing chip**
- **Massive parallel signature sequencing (MPSS)**
- **Mass Spectroscopy sequencing**

Why Transcriptome level?

- Cell behavior is influenced by the proteins generated in the cell. These proteins are created by digital instructions called mRNA (messenger Ribonucleic Acid) which are produced by copying the DNA sequences of genes when genes are "turned on".

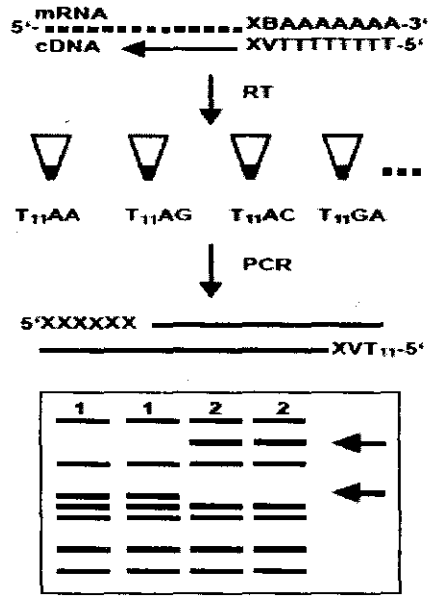
- It is the collection of synthesized proteins which ultimately define the behavior and function within a cell and are responsible for the biological responses that underlie phenotypic traits of an animal.
- Each gene may come in several configurations or allele(s) with one allele received from each parent. Different alleles can change the mRNA message or the amount of message and thereby change the amount or make up of the proteins generated, causing variation in cell behavior.
- Conversely, when genes are "turned off" by specific stimuli, the DNA is not copied into RNA instructions and their corresponding proteins are not synthesized. This can modify the behavior and function of the cell as well.
- So genes or gene profiles are like a set of hundreds of switches with some "turned on" and others "turned off" at various times to regulate development or physiological functions. And these switches are influenced by various stimuli resulting from treatments, physiological status, the genetic make up of an animal, and environment.

The most common recent tools for studying gene expression at the transcriptome level applicable in FAnGR

- **Differential Display**
- **Microarray**
- **qRT-PCR**

Differential display (DD)

- Differential Display (DD RT-PCR differential display reverse transcription PCR) is a multiplex RT-PCR technique.
- The process begins with reverse transcription using one of three primers that anchor to the poly(A) tail at the 3'-end of mRNA (e.g., TnG, TnC, Tn A, which define the three possible junctions of the poly(A) tail).
- Multiplex PCR is then carried out using a defined arbitrary oligomer as the up-stream primer. Each RT-PCR amplifies a set of fragments corresponding to the 3'-ends of various mRNA species.
- Resolving the fragments on a high resolution polyacrylamide gel gives a "fingerprint" of gene expression/ expression profile and allows a side-by-side comparison of different mRNA populations.
- The procedure requires only small amounts of RNA and is therefore a perfect method for studying changes in gene expression from scarce biological materials.

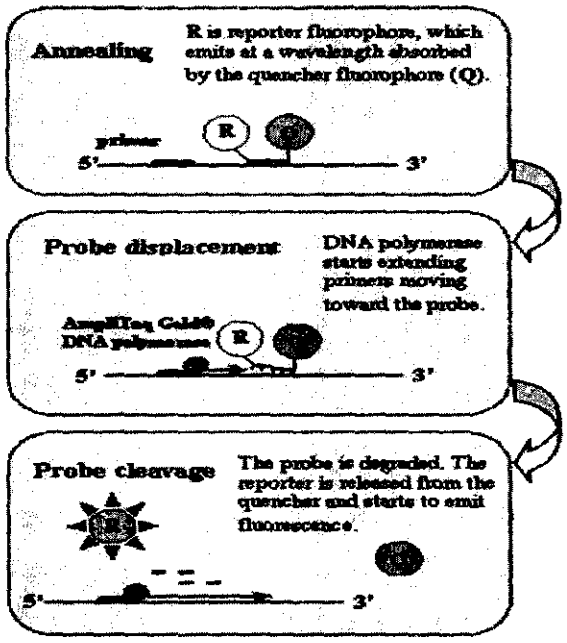


Differential Display: Reverse transcription, PCR amplification, fragment separation. Critical steps and potential modifications.

Quantitative RT-PCR

- For quantification of mRNA, the RT-PCR signal from a target mRNA is compared to that of a housekeeping mRNA, (TaqMan assay).
- The quantification of mRNA transcripts is estimated indirectly, by determining the accumulation of PCR products in real time.

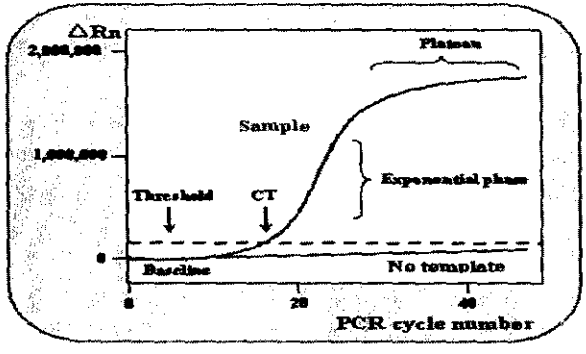
TaqMan® Applied Biosystems



TaqMan The method exploits the 5' endonuclease activity of Taq DNA polymerase to cleave an oligonucleotide probe during PCR, thereby generating a detectable signal. The probes are fluorescently labeled at their 5' end and are non-extendable at their 3' end by chemical modification.

Specificity is conferred at three levels: via two PCR primers and the probe. Probes also include a minor groove binder for added specificity.

Model of real time quantitative PCR plot

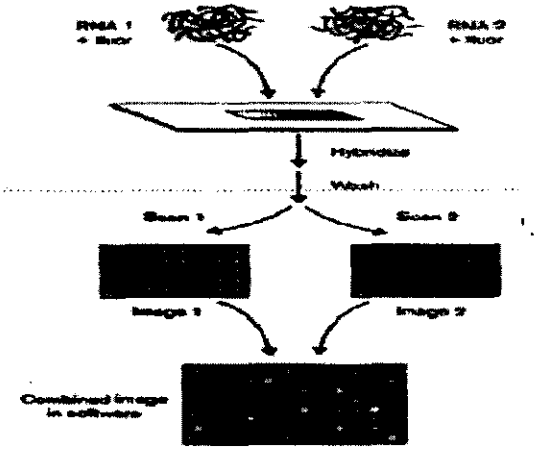


Baseline **is defined as PCR cycles in which a reporter fluorescent signal is accumulating but is beneath the limits of detection of the instrument.**

ΔRn **is an increment of fluorescent signal at each time point. The ΔRn values are plotted versus the cycle number.**

Threshold **is an arbitrary level of fluorescence chosen on the basis of the baseline variability. A signal that is detected above the threshold is considered a real signal that can be used to define the threshold cycle (Ct) for a sample. Threshold can be adjusted for each experiment so that it is in the region of exponential amplification across all plots.**

Ct **is defined as the fractional PCR cycle number at which the reporter fluorescence is greater than the threshold. The Ct is a basic principle of real time PCR and is an essential component in producing accurate and reproducible data.**



Microarray experiments

are basically genome-scale surveys of the collection of genes expressed in a given cell type during a defined physiological change, environmental insult, or treatment scenario.

The combined pattern of

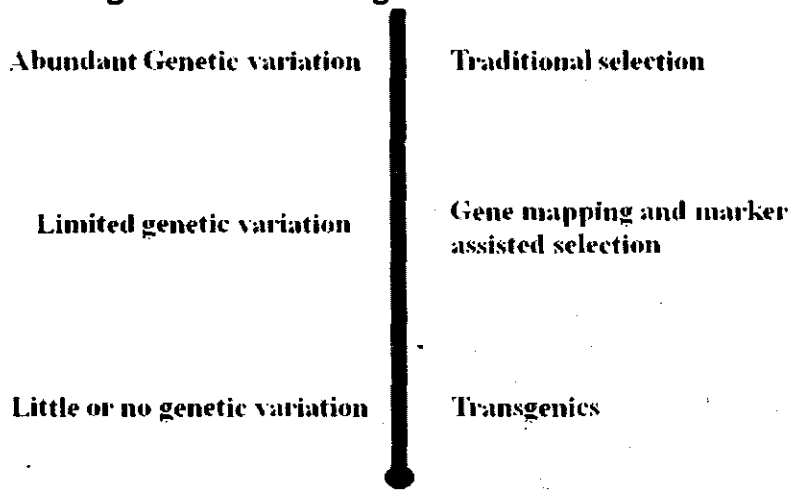
red, green and yellow spots on the microarray reveal which genes are turned "on" (green), "off" (red) or unchanged (yellow) by a "treatment".

Outlook

- **Understand quantitative genetics, and selection theory.**
- **Understand the biology and physiology of economically/biomedically important traits.**
- **Develop new methods/tools for diagnostics, selection, pharmaceutics, and transgenics in the future.**

Genetic Improvement:

Challenges and Technologies



Opening the Genetics Black Box

Old: Infinitesimal model

New: Major gene or Candidate gene models

How many genes?

Sizes of gene effects?

Major Gene or Candidate Gene Model

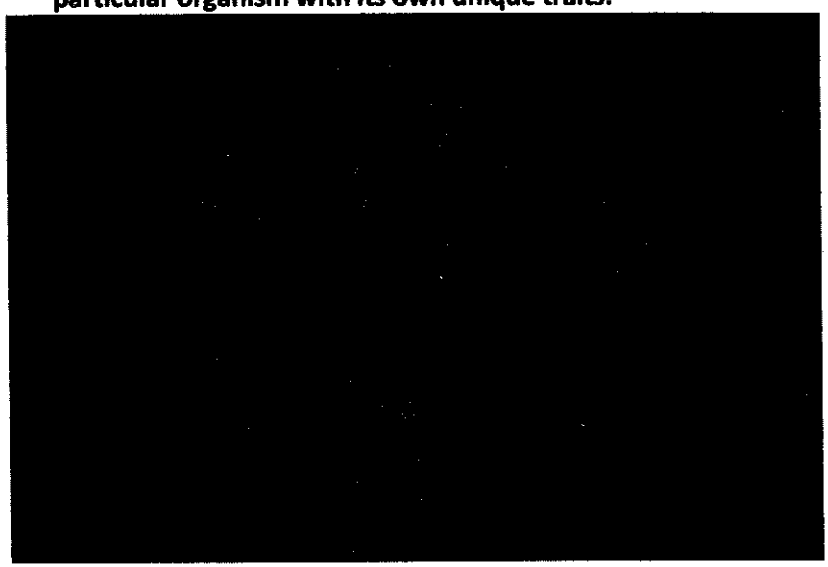
- **There are individual genes that contribute a significant proportion of the variation in economically important traits.**
- **Molecular biology and genomics are used to find and use such genes.**

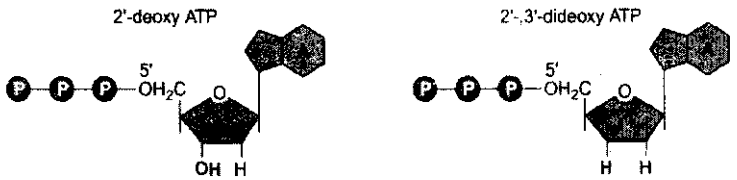
Molecular biology *review*

- Molecular genetics review
  - DNA/RNA
  - Genes/alleles
  - Chromosomes
- Molecular biology review
  - Restriction enzymes
  - PCR
  - Genotyping methods

DNA

- a DNA molecule consists of two strands that wrap around each other to resemble a twisted ladder whose sides, made of sugar and phosphate molecules, are connected by rungs of nitrogen-containing chemicals called bases.
- Each strand is a linear arrangement of repeating similar units called nucleotides, which are each composed of one sugar, one phosphate, and a nitrogenous base
- Four different bases are present in DNA adenine (A), thymine (T), cytosine (C), and guanine (G).
- The particular order of the bases arranged along the sugar- phosphate backbone is called the DNA sequence;
- the sequence specifies the exact genetic instructions required to create a particular organism with its own unique traits.





Genes/alleles

- A gene is a specific coding sequence, or unit of transmission, recombination, and function.
- Each specific sequence of three DNA bases (codons) directs the cells protein-synthesizing machinery to add specific amino acids. For example, the base sequence ATG codes for the amino acid methionine.
- The gene is made up of coding regions - called exons and non coding regions called introns.
- Gene structure runs as below

(5' -> 3') ATGGAATTCTCGGTC (sense strand)

(3' <- 5') TACCTTAAGAGCCAG (antisense strand)

(5' -> 3') AUGGAAUUCUCGGUC (mRNA made from Template strand) - use for cDNA

Alleles/Haplotypes

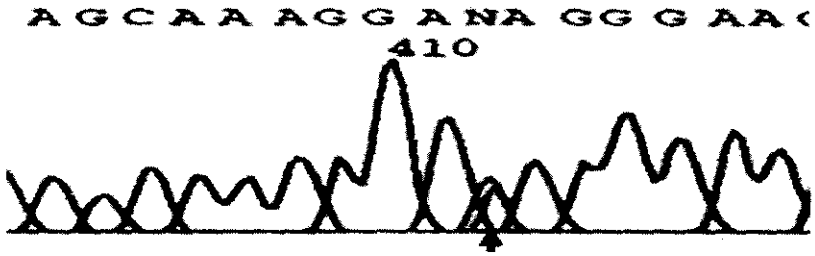
An allele is an alternate form of a gene or a polymorphism within a gene.
Genes or genetic markers may be monomorphic - no alleles or polymorphic - many alleles.
A haplotype is a set of linked alleles on one chromosome.
A b or a B are 2 haplotypes

Genotype

Genotypes are combinations of alleles at a given locus.
AA or Aa or aa
11 or 12 or 22

Locus/Loci

A specific mutation or polymorphism within a given gene (locus) or genes (loci).



Locus/Loci

- While in conventional animal breeding each gene has one locus, in molecular genetics there can be many loci within the gene.
- So care needs to be taken when we describe a locus as to whether we are discussing a gene or a particular site within the gene.

Finding a Gene - Analogy

- The genome
- The chromosome
- The general location
- The gene
- The mutation

Molecular methods

Some methods:

Restriction enzymes and PCR

Types of genetic markers

Restriction Enzymes

- Restriction enzymes recognize short DNA sequences and cut the DNA molecules at those specific sites.
- Since hundreds of different restriction enzymes have been characterized, DNA can be cut into many different small fragments.

Schematic of a double stranded DNA fragment with an *Eco*RI restriction site

5′ AGCTTGACCGTACTA*GAATTC*TCGTACGTAGGCAATTTGCTA 3′

3′ TCGAACTGGCATGAT*CTTAAG*AGCATGCATCCGTTAAACGAT 5′

Restriction enzyme cut site: A specific nucleotide sequence of DNA at which a particular restriction enzyme cuts the DNA

Schematic of a double stranded DNA fragment with an *Eco*RI restriction site

5′ AGCTTGACCGTACTA*GAATTC*TCGTACGTAAATTTGCTA 3′

3′ TCGAACTGGCATGAT*CTTAAG*AGCATGCATTTAAACGAT 5′

+

***Eco*RI restriction endonuclease**

↓

AGCTTGACCGTACTA*G* *AATTC*TCGTACGTAAATTTGCTA

TCGAACTGGCATGAT*CTTAA* *G*AGCATGCATTTAAACGAT

Restriction Enzymes

On average, restriction enzymes with

4-base recognition sites will yield pieces 256 bases long

6-base recognition sites will yield pieces 4,000 bases long

8-base recognition sites will yield pieces 64,000 bases long

The Polymerase Chain Reaction (PCR)

A technique for creating millions of copies of a particular segment of DNA. PCR can be used to amplify the amount of a specific DNA sequence from a pool of billions of base pairs until there are enough copies to permit analysis and manipulation.

The Polymerase Chain Reaction (PCR)

PCR amplification generally consists of three steps:

1. **Heat denaturation of the complementary DNA strands**
2. **Annealing of single stranded DNA primers to the denatured target sequence**
3. **Extension of the primers with *Taq* DNA polymerase to make a "copy" of the target sequences.**

PCR

The DNA target or template

GTCATAGCATTATTATTATTATTCAGGACTA

CAGTATCGTAATAATAATAATAAGTCCTGAT

A double stranded DNA molecule

The Polymerase Chain Reaction

The DNA target or template is heat denatured

GTCATAGCATTATTATTATTATTCAGGACTA

CAGTATCGTAATAATAATAATAAGTCCTGAT

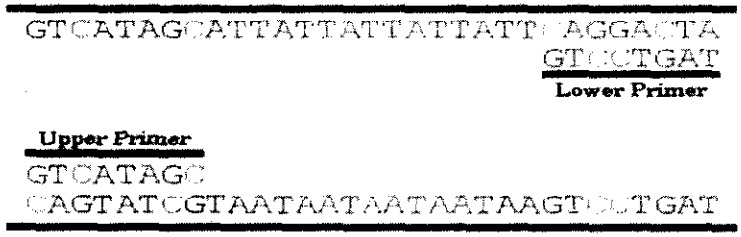
Heat denatures the original template strand

The Polymerase Chain Reaction

The DNA target or template is heat denatured

Denatured DNA molecule

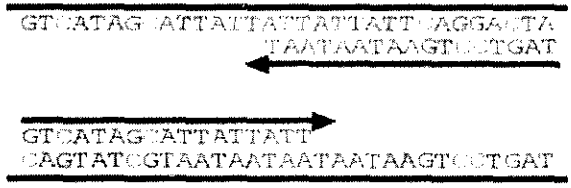
PCR primers anneal to the DNA target



The primers can then anneal to their complementary sequences

The Polymerase Chain Reaction

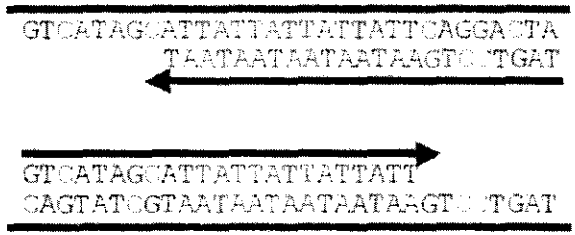
"Extension" of PCR primers in the presence *Taq* DNA polymerase and nucleotides



Extension of the primers using *Taq* polymerase amplifies the target sequence

The Polymerase Chain Reaction

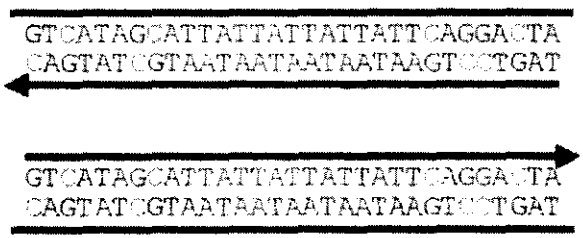
"Extension" of PCR primers in the presence *Taq* DNA polymerase and nucleotides continues



Extension of the primers using *Taq* polymerase amplifies the target sequence

The Polymerase Chain Reaction

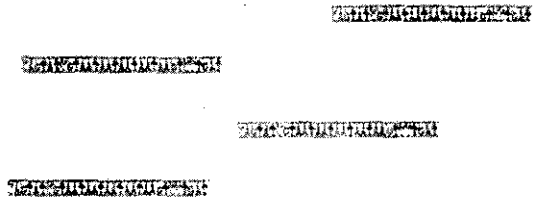
"Extension" of PCR primers in the presence *Taq* DNA polymerase and nucleotides is completed



Extension of the primers using *Taq* polymerase amplifies the target sequence

The Polymerase Chain Reaction

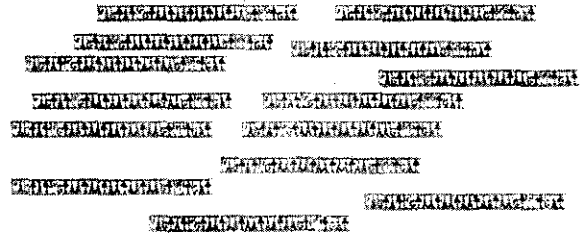
Geometric amplification of the target sequence



Several million copies of DNA can be made in just a few hours

The Polymerase Chain Reaction

Geometric amplification of the target sequence



Several million copies of DNA can be made in just a few hours

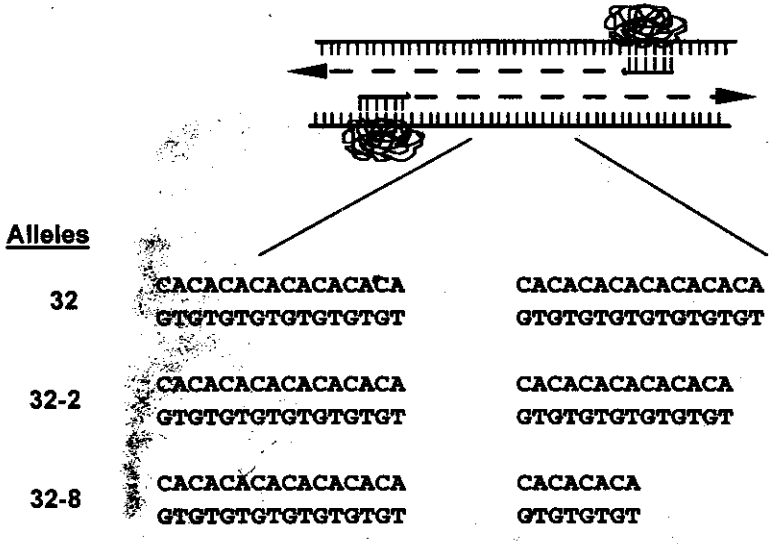
DNA markers and testing

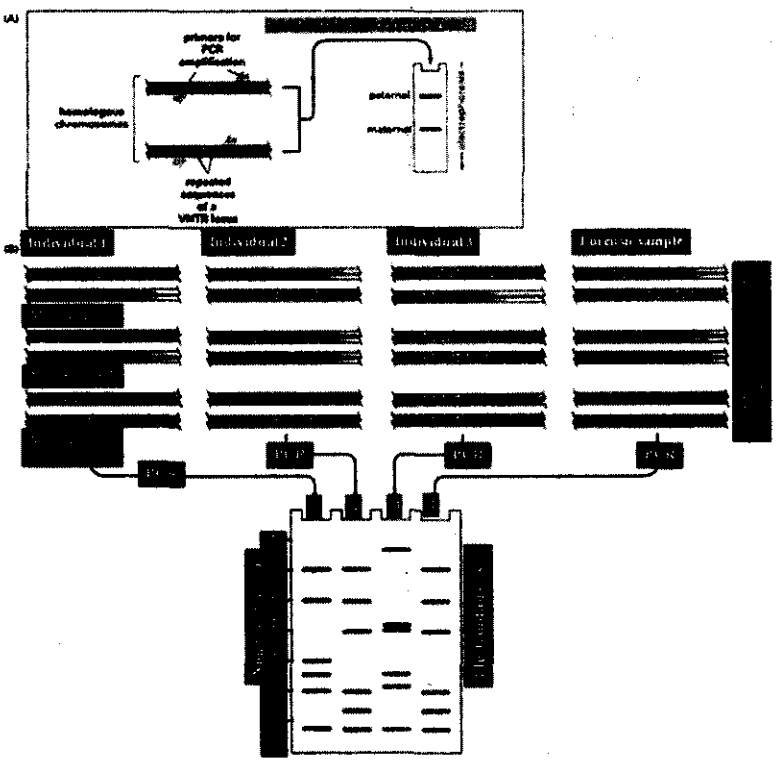
- A DNA marker is a sequence of bases at a unique physical location in the genome, which varies sufficiently between individuals that its pattern of inheritance can be tracked through families and/or it can be used to distinguish among cell types. A genetic marker may or may not be part of a gene.
- Markers are used for testing and mapping.

Types of Markers

- PCR - RFLP (Restriction Fragment Length Polymorphism) based upon the absence or presence of one or more restriction enzyme cut sites in a specific point in the genome - codominant.
- RAPD (Random Amplified Polymorphic DNA) Based upon the presence or absence of PCR amplification using a single arbitrary PCR primer - dominant.
- SSCP (Single Stranded Conformation Polymorphism) - codominant.
- SSLP (Simple Sequence Length Polymorphism) based upon differences in the number of short DNA repeated sequences at a specific point in the genome detected via PCR - codominant.
- Microsatellite - simple sequence repeats (SSR) - often random di or tri nucleotide repeats - anonymous marker - codominant.

PCR Amplification of Polymorphic Microsatellite Region





Example of microsatellites

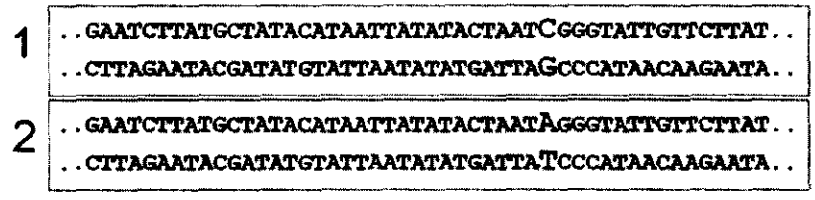
Types of Markers

- AFLP (Amplified Fragment Length Polymorphism) - based upon the absence or presence of a restriction enzyme cut site at a specific point in the genome detected via PCR - dominant.
- SNPs (Single Nucleotide Polymorphisms) Based upon single changes of the DNA bases at a specific point in the genome - codominant.

SNP example

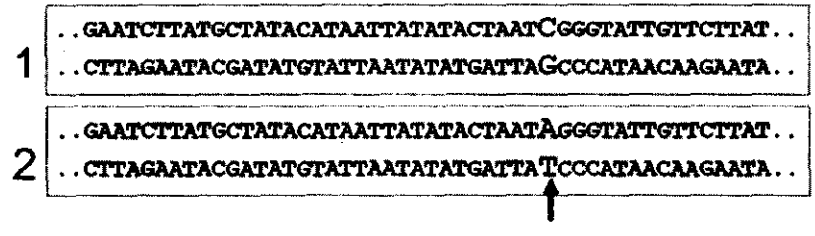
A single base difference in the DNA of two individuals

Often referred to as point mutations



SNP

- These may be detected by sequencing or by RFLPs, SSCP or AFLP.



SNP

Gene Mapping

Linkage mapping
Physical mapping
Comparative mapping

Genome Project - Phases

Construct genome maps and develop mapping tools
QTL mapping / genome scan
Isolate trait genes from within QTL

Gene Mapping Uses

- Find genes (quantitative trait loci) that are associated with traits of economic importance.
- Discover genes causing major physiological defects.
- Use genetic markers for marker assisted selection and introgression programs.
- Develop comparative maps.

Types of Genetic Maps

- Genetic linkage map
  linear description of markers/genes on a given chromosome with markers closer being inherited together more often.
- Physical Map
  The physical location on the chromosome - two types : Cytogenetic and physical
- Sequence Map – entire set of genome sequence

Linkage and Physical Mapping

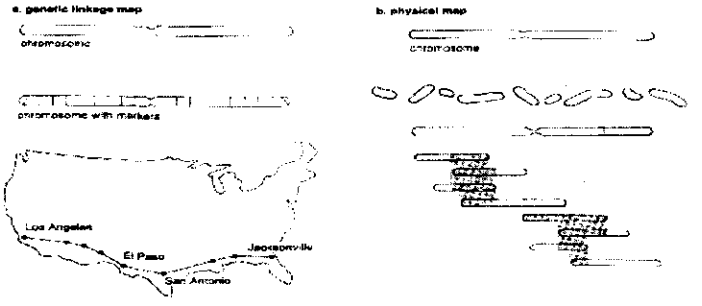


Figure 1. Genetic linkage maps of chromosomes (a) Are similar to establishing the positions of towns on a road map. Interstate 10 runs from Florida to California. If there were no towns mapped along that highway, it would be difficult to locate a landmark between, for example, Jacksonville and Los Angeles. It is easier to locate a landmark if it is located between markers that are closer together, such as El Paso and San Antonio. Physical maps (b) are created by chemically cutting the chromosomes into pieces and reordering them the way they were on the chromosome. Because researchers will know where on the chromosome each piece came from, they can tell where any gene in that piece is located on the chromosome.

Gene Mapping Tools/Resources

- **Linkage maps**
  - anonymous genetic markers
  - genes
- **Physical maps**
  - Radiation/fluorescence hybrid panels
  - somatic cell hybrid panels
- **Libraries**
  - cDNAs, BACs, YACs
- **Gene sequences and DNA chips**
- **Defined populations**
- **Databases**

Linkage mapping

Reference

http://www.genome.iastate.edu/edu/doe/

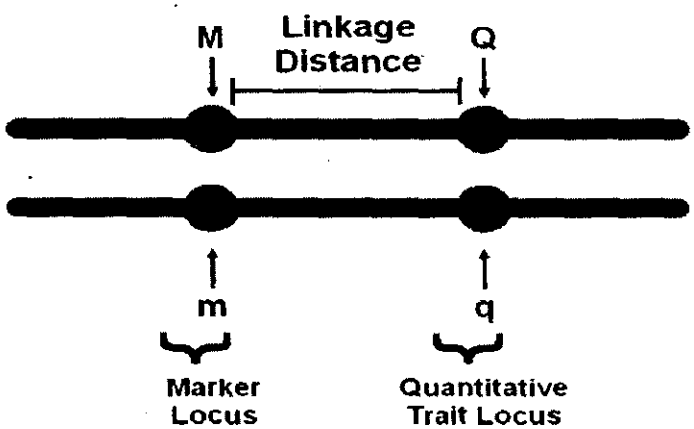
- **A genetic linkage map shows the relative locations of specific genes and DNA markers along the chromosome.**
- **Any inherited physical or molecular characteristic that differs among individuals and is easily detectable in the laboratory is a potential genetic marker.**

Genetic (linkage) Mapping Requirements

- Pedigrees
  - large families - usually three generations
- Polymorphic genetic markers
  - many markers and genes

Constructing a Linkage Map

- Two markers located near each other on the same chromosome will tend to be passed together.
- Frequency of how often this occurs allows the researcher the ability to estimate distance between the markers.



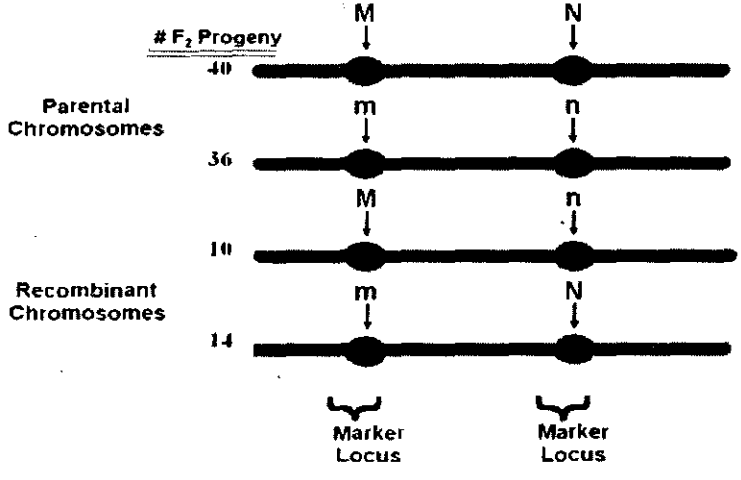
Linkage Mapping

- Recombination Frequency

$$\frac{}{\text{Total progeny}} = \frac{}{p + r}$$

- 0% < recombination frequency < 50%
- centimorgans (cM)

$$= \frac{}{p + r} \times 100$$



Linkage Mapping

- Recombination Frequency

$$\frac{\text{Total progeny}}{} = \frac{p+}{} = \frac{(40+36)+}{}$$

$$= 0.24$$

- centimorgans (cM)

$$= \frac{p+r}{} \times 100 = \frac{(40+36)+}{} \times 100 = 24\ cM$$

Distances: Linkage mapping

Building a map we could use all pair wise analyses for linkage and then try to build a map. However, with more than 2 loci distances measured in recombination rates are not additive because of the occurrence of double recombinants and interference.

Types of Linkage Maps

- **Framework map**
  - **markers are ordered in a high degree of certainty**
  - **usually use LOD score >3.0 (1000 times more likely.**
  - **choice of markers includes those physically mapped or those evenly spaced and very polymorphic.**
- **Comprehensive map**
  - **includes all markers**

Variation in Recombination Frequency

- **True order of the markers will be the same for the sexes**
- **Distances may differ between males and females (Haldane, 1922).**
- **Distances may vary between families but order should not.**
- **Most maps average the distances so sex of the animal is important.**

Programs

CRIMAP

Documentation:
http://www.genome.iastate.edu/~max/lab/ CRIMAPwkshp/crimap-doc.html

Tutorials:

http://www.biobase.dk/Embnetut/Crimap/
http://www.genome.iastate.edu/~max/lab/CRIMAPwkshp/index.html

Genetic (linkage) Mapping Situation in Cattle

Pedigrees

International reference mapping pedigrees - Bovmap, Nordic and USDA

Polymorphic genetic markers

>5,000 mapped polymorphic markers
Gene sequences

Physical Mapping

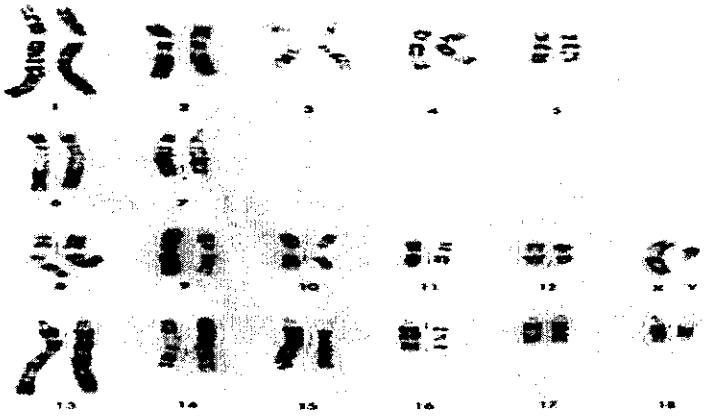
Several levels of physical mapping

Chromosome, chromosome arm, band
specific region within an arm
a certain stretch of bases

Cytogenetic map or a physical map

genes and markers - several techniques

Karyotyping - method of staining and separating chromosomes



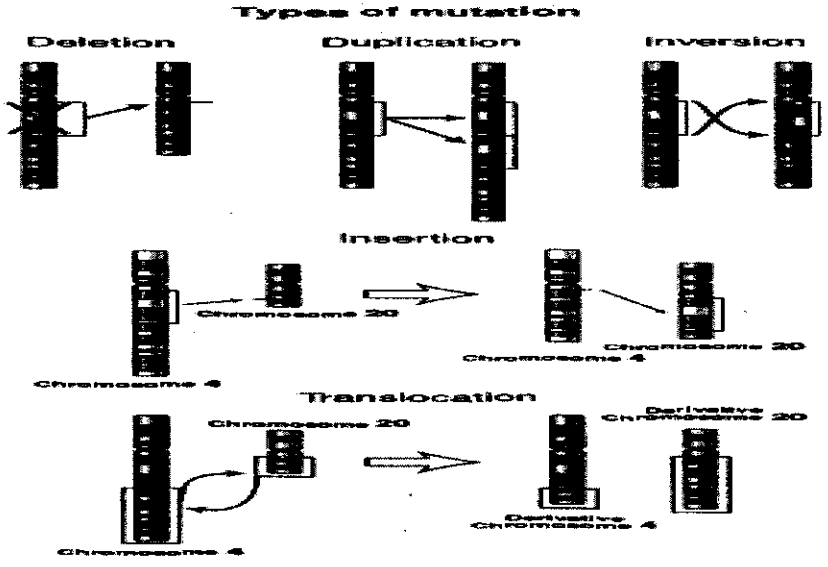
Physical Mapping

- In most species there are standard karyotypes based on the banding pattern.
- This is useful for determining chromosome abnormalities.

  Numerical aberrations

  XO, XXY or even XXXY

  Structural aberrations - translocations or duplications or deletions.



Physical Mapping

Over 50 structural abnormalities have been reported in animals.

Cause sterility

Cause reduced fertility (25-50% in litter size)

Other abnormalities

Methods:

- **Somatic Cell Hybrid Mapping**
- **Radiation Hybrid Mapping**
- ***In Situ* Hybridization**
- **Radioactive**
- **Fluorescent**
- **Primed In Situ PCR,**
- **Chromosome painting**
- **BAC contigs and sequencing**

In Situ Hybridization

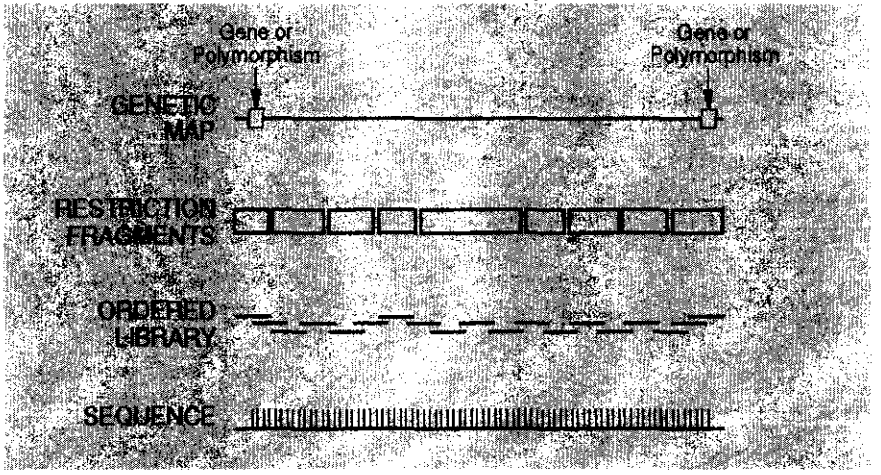
Somatic Cell Hybrid Mapping

- Method involves fusion of two cell lines from different species.
- The hybrids cells unilaterally lose chromosomes of the species under investigation.
- The cells stabilize with only some chromosomes or parts of chromosomes.
- Cell lines can be examined and parts of chromosomes determined.
- Using PCR and gene primers DNA from the clones can be used to see if certain genes are in in certain clones.
- Using this information and information as to which cell lines have which parts of chromosomes a gene can be cytogenetically mapped.
- DNA from the cell lines are used to form a panel.
- No polymorphisms needed.

Radiation Hybrid Mapping

- Basically the same approach as SCH mapping.
- Panel is created by breaking the chromosome of interest or the whole genome up into several fragments by a high dose of X-ray irradiation.
- The broken fragments are recovered in recipient cells (eg. rodent) and the clones are analyzed for presence or absence of markers.
- Usually 100 - 120 clones.
- Use PCR to see if genes are in certain clones.
- Assumption is that the closer the genes the more likely they are in the same fragment.
- Distances are presented in Ray (R) units. These distances are converted to cR units on the RH map. Linkage groups, or single unlinked markers, are ordered according to marker position on the genetic map.

Comparison of Genetic Maps



Comparative Mapping

Reasons for comparative mapping

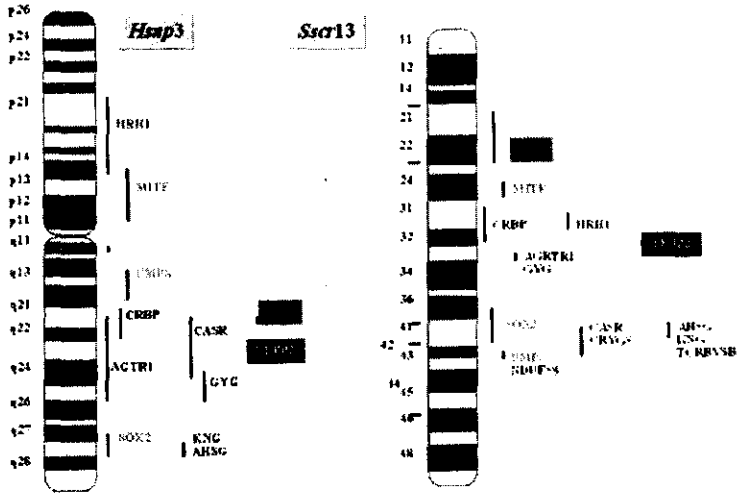
- **Study of evolution**
- **Prediction of genes from "map rich" species to "map poor" species**
  - **mouse, human ⇨ livestock**
- **Real mapping and virtual mapping**

Comparative Genome Mapping - Definitions

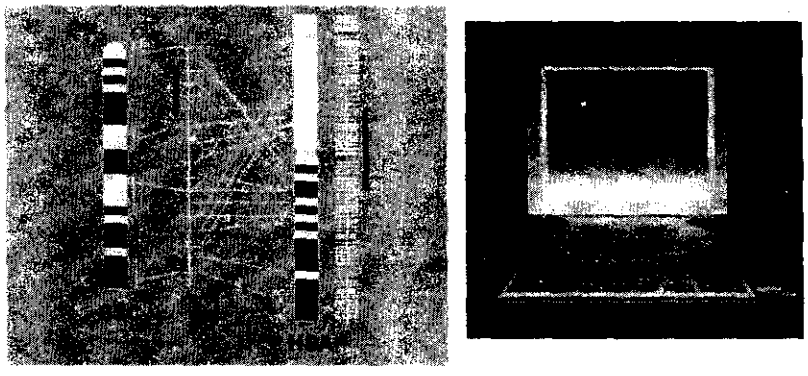
'synteny' = on the same chromosome.

'conservation of synteny' = genes found together on one chromosome in species A are also found together in species B.

Chromosome 13 Comparative Map



Virtual Fine Mapping



Bovine and Ovine Genome Mapping Databases Sources

- **(Roslin, NAL, Iowa)**
  http://www.genome.iastate.edu
- **USDA-MARC data base**
  http://www.marc.usda.gov