Conifer Translational Genomics Network
Coordinated Agricultural Project

CTGN CAP

Genomics in Tree Breeding and Forest Ecosystem Management

Module 2 – Genes, Genomes, and Mendel

Nicholas Wheeler & David Harry – Oregon State University

Quick review: Genes and genomes

- In eukaryotes, DNA is found in the...
  - *Nucleus*
  - *Mitochondria*
  - *Chloroplasts (plants)*
- Organelle inheritance is often uniparental, making it powerful for certain types of applications

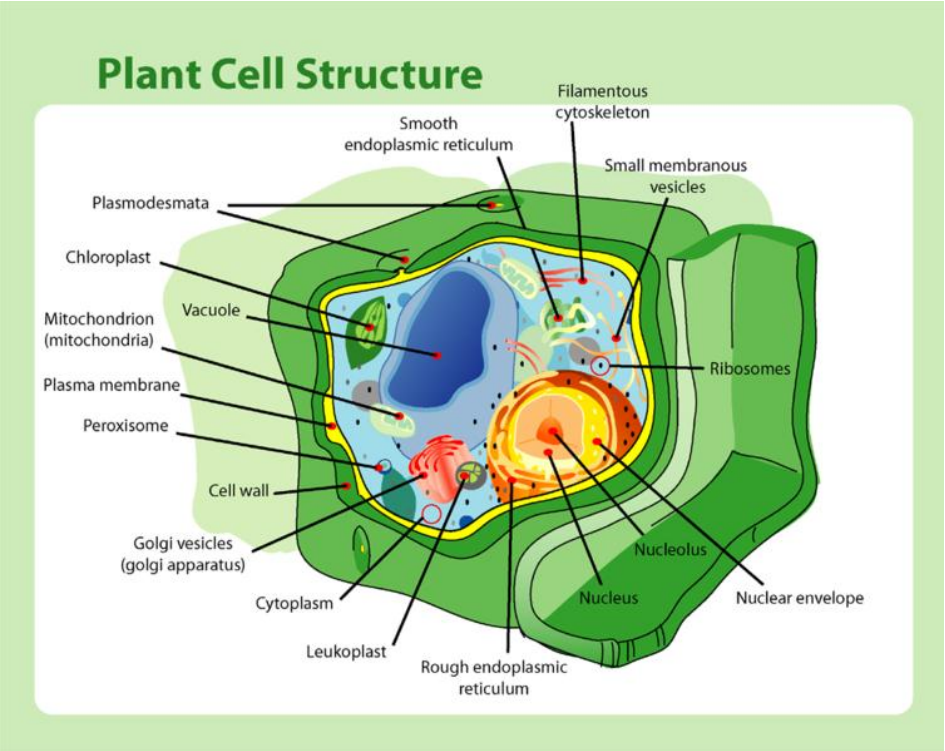


Figure Credit: Wikipedia Contributors, http://commons.wikimedia.org/w/index.php?title=File:Plant_cell_structure.png&oldid=45093602

Chromosomes

Linear strands of DNA and associated proteins in the nucleus of eukaryotic cells

- Chromosomes carry the genes and function in the transmission of hereditary information
- Diploid cells have two copies of each chromosome
- One copy comes from each parent
- Paternal and maternal chromosomes may have different alleles

Image Credit: Jane Ades, National Human Genome Research Initiative (NHGRI)

Genes

Units of information on heritable traits

- In eukaryotes, genes are distributed along chromosomes
- Each gene has a particular physical location: **a locus**
- Genes encompass regulatory switches and include both coding and non-coding regions
- Genes are separated by intergenic regions whose function is not understood

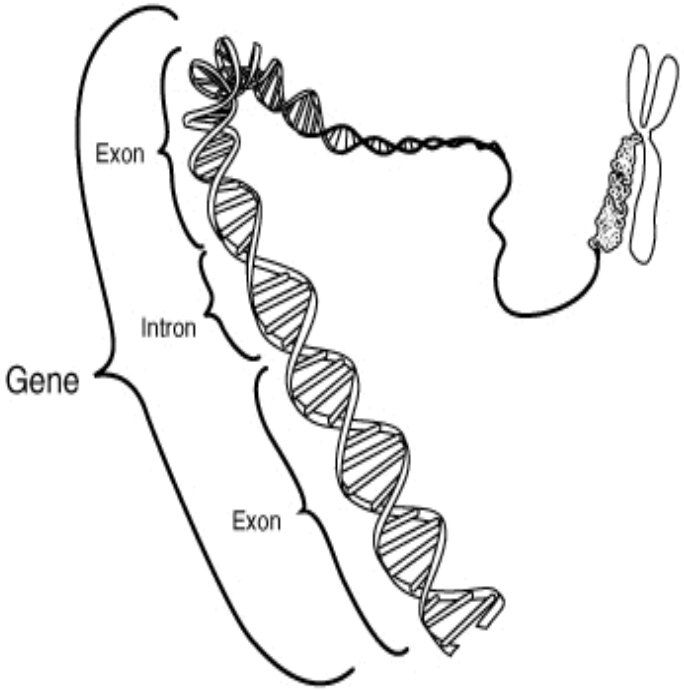


Figure Credit: Darryl Leja, National Human Genome Research Initiative

The central dogma of molecular biology

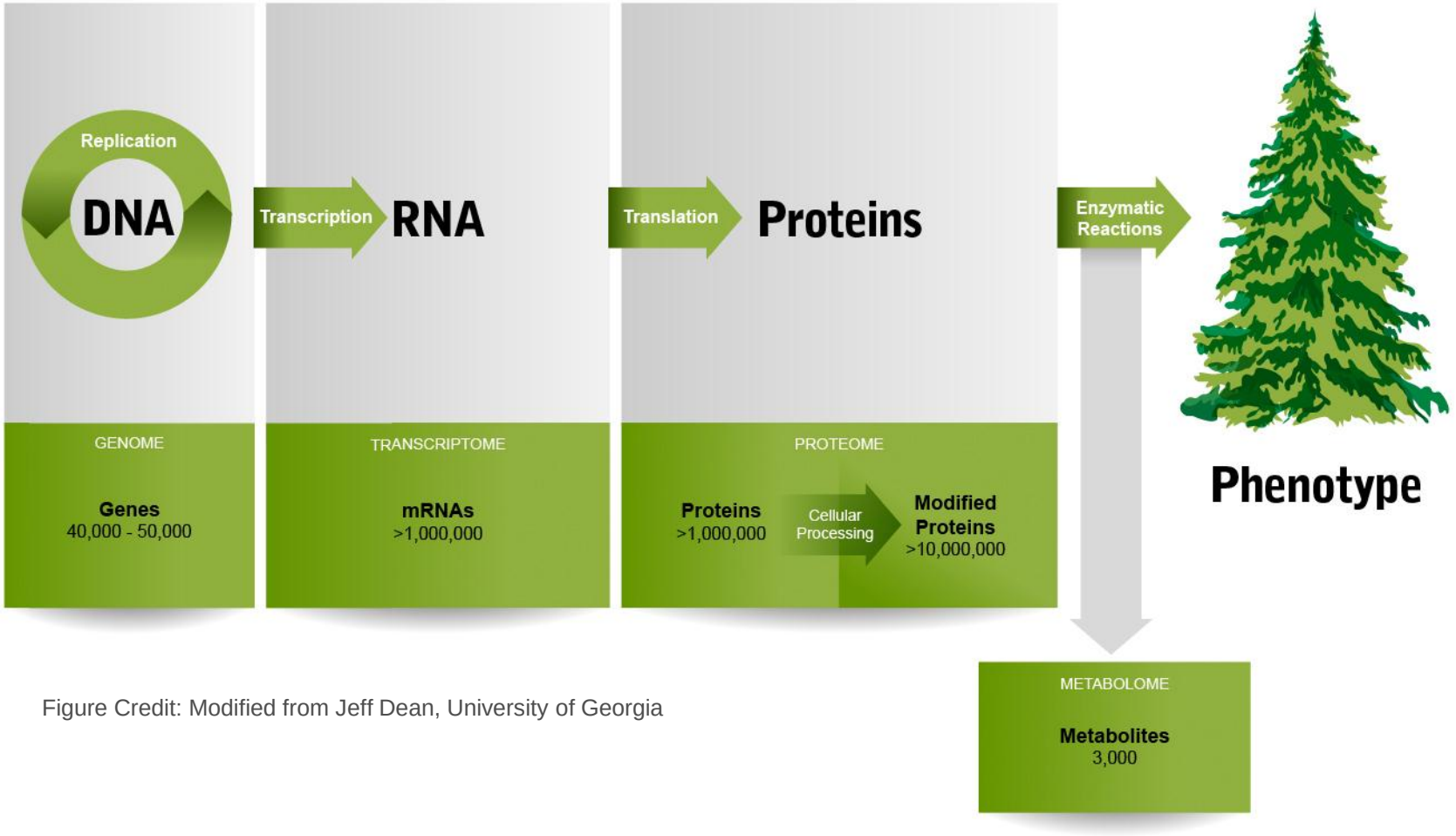


Figure Credit: Modified from Jeff Dean, University of Georgia

Alleles

Alternative forms of a gene

- A diploid cell has two copies of each gene (i.e. two alleles) at each locus
- New alleles arise through mutation
- Alleles on homologous chromosomes may be the same or different (homozygous vs. heterozygous)

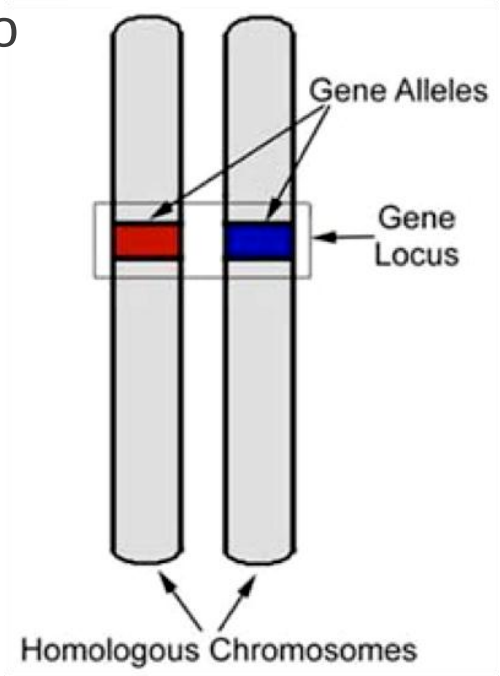


Figure Credit: Megan McKenzie-Conca, Oregon State University

Markers reflect genetic polymorphisms that are inherited in a Mendelian fashion

- DNA markers 'mark' locations where DNA sequence varies (2 or more alleles)
  - *Such polymorphisms can vary within and among individuals (e.g. heterozygotes vs. homozygotes) and populations*
- Markers may be located in genes or elsewhere in the genome
  - *Historically, we've had too few markers to inform breeding*
- Genomics tools provide an almost unlimited supply of markers

Mutations may take many forms

- Some simple, single nucleotide mutations can totally alter a protein product by producing a frameshift. This results in a new amino acid being produced
- Insertions and deletions: The addition or loss of one or more nucleotide(s) in coding sequence

Frameshift mutations show that the genetic code is read in triplets from a fixed starting point.

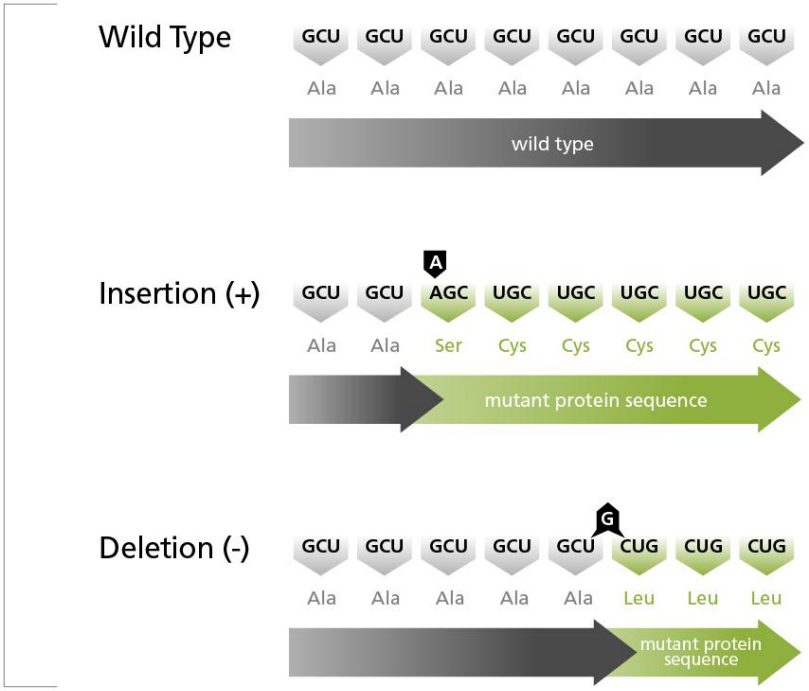


Figure Credit: Modified from Lewin, 2000.

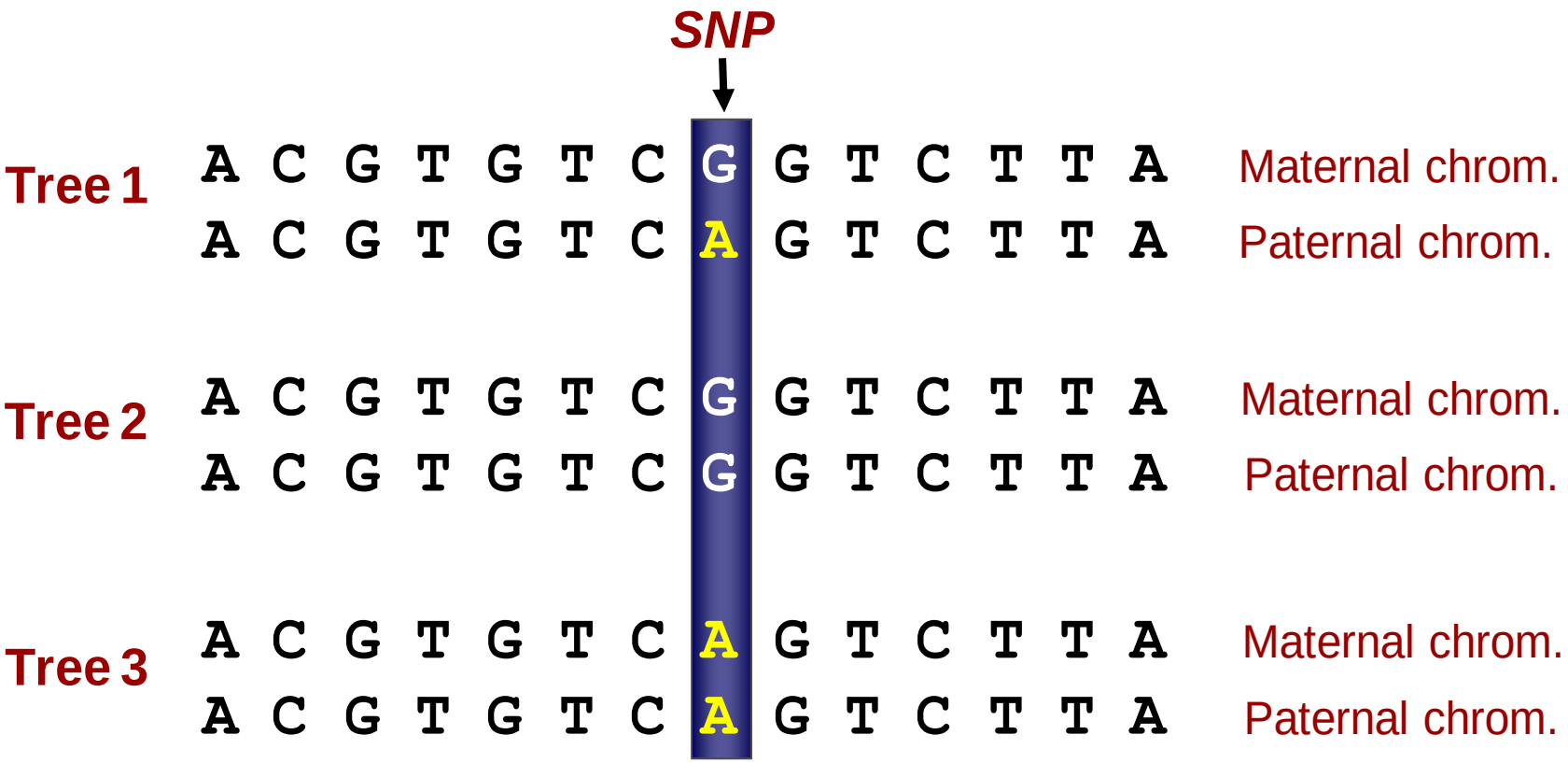
*S*ingle *N*ucleotide *P*olymorphisms (SNPs) embedded within a DNA sequence

- DNA sequences are aligned
- Polymorphic sites are identified
- Haplotypes (closely linked markers of a specific configuration) are deduced by direct observation or statistical inference

```
   *                    *     *
atggctacctgaactggtcaactcatcgaaagctaa  1
atggctacctgaactggtcaactcatcgaaagctaa  1
atgcctacctgaactggtcaactcatcgaaagctaa  2
atgcctacctgaactggtcaactcatcgaaggctaa  3
atgcctacctgaactggtcaacacatcgaaggctaa  4
```

Figure Credit: David Harry, Oregon State University

Single Nucleotide Polymorphism (SNP)



Tree 1 is ***heterozygous*** Trees 2 and 3 are ***homozygous***

Figure Credit: Glenn Howe, Oregon State University

The genome

An individual's complete genetic complement

- For eukaryotes, a haploid set of chromosomes
- For bacteria, often a single chromosome
- For viruses, one or a few DNA or RNA molecules
- Genome size is typically reported as the number of base pairs (nucleotide pairs) in one genome complement (i.e. haploid for eukaryotes)
- Until recently, we studied genes and alleles one, or a few at a time (genetics)
- Aided by high throughput technologies we can now study virtually all genes simultaneously (genomics)

Genome size

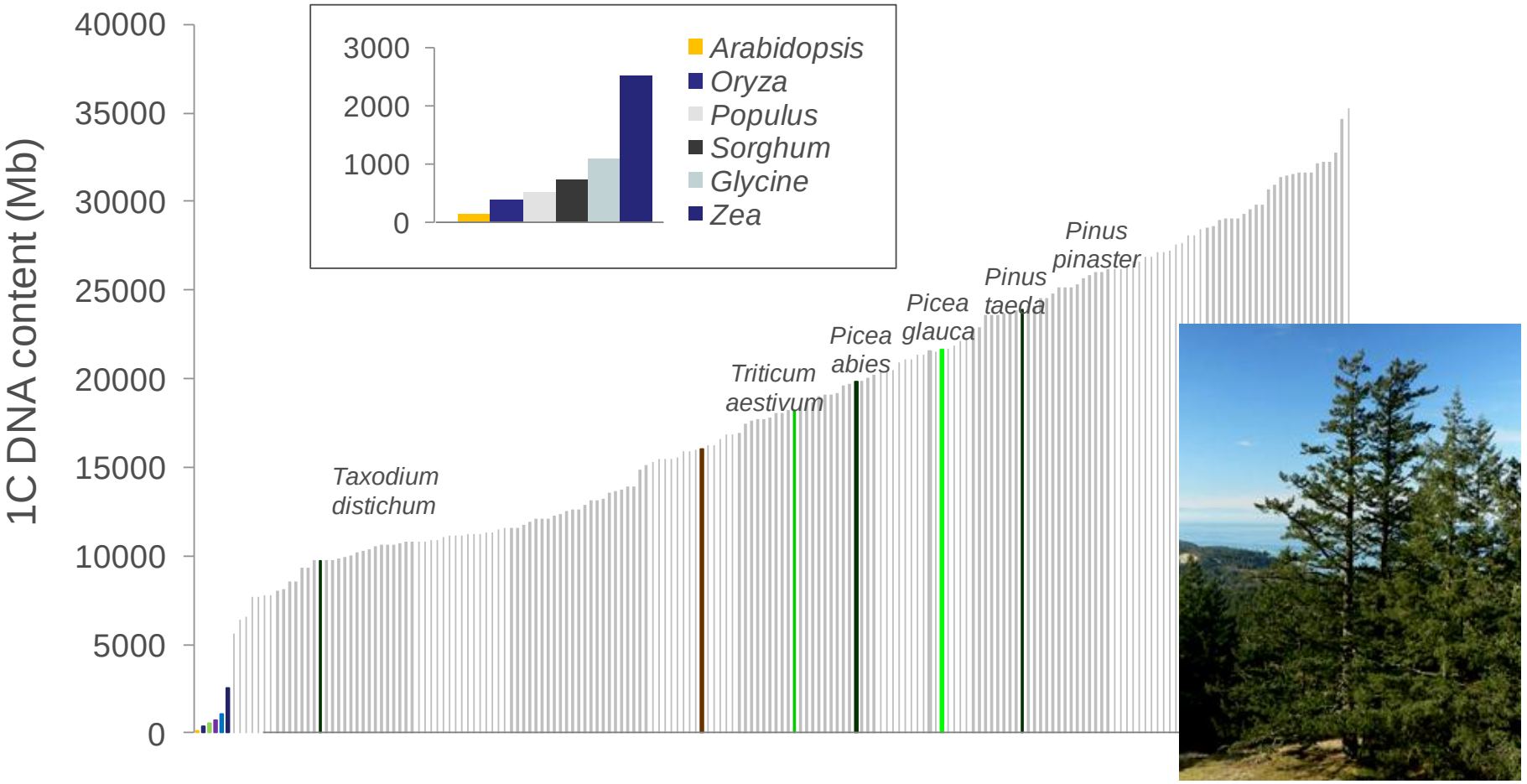


Image Credit: Modified from Daniel Peterson, Mississippi State University

How large are genes?

- The **longest** human gene is **2,220,223** nucleotides long. It has 79 exons, with a total of only 11,058 nucleotides, which specify the sequence of the **3,685** amino acids and codes for the protein dystrophin. It is part of a protein complex located in the cell membrane, which transfers the force generated by the actin-myosin structure inside the muscle fiber to the entire fiber
- The **smallest** human gene is **252** nucleotides long. It specifies a polypeptide of **67** amino acids and codes for an insulin-like growth factor II

How many genes are there in a genome?

Organism	Number of genes (predicted)
C. Elegans (Roundworm)	20,000
Ancestral flowering plants	12,000 to 14,000
Arabidopsis	26,500
Medicago	40,000
Poplar	45,000
Humans	25,000

Why so much DNA?

Genome size / Gene number do not add up

- Duplicated genes
- Repetitive elements
- Regulatory elements
- (Other?)

(Gibson and Muse, 2004)

The central dogma of molecular biology

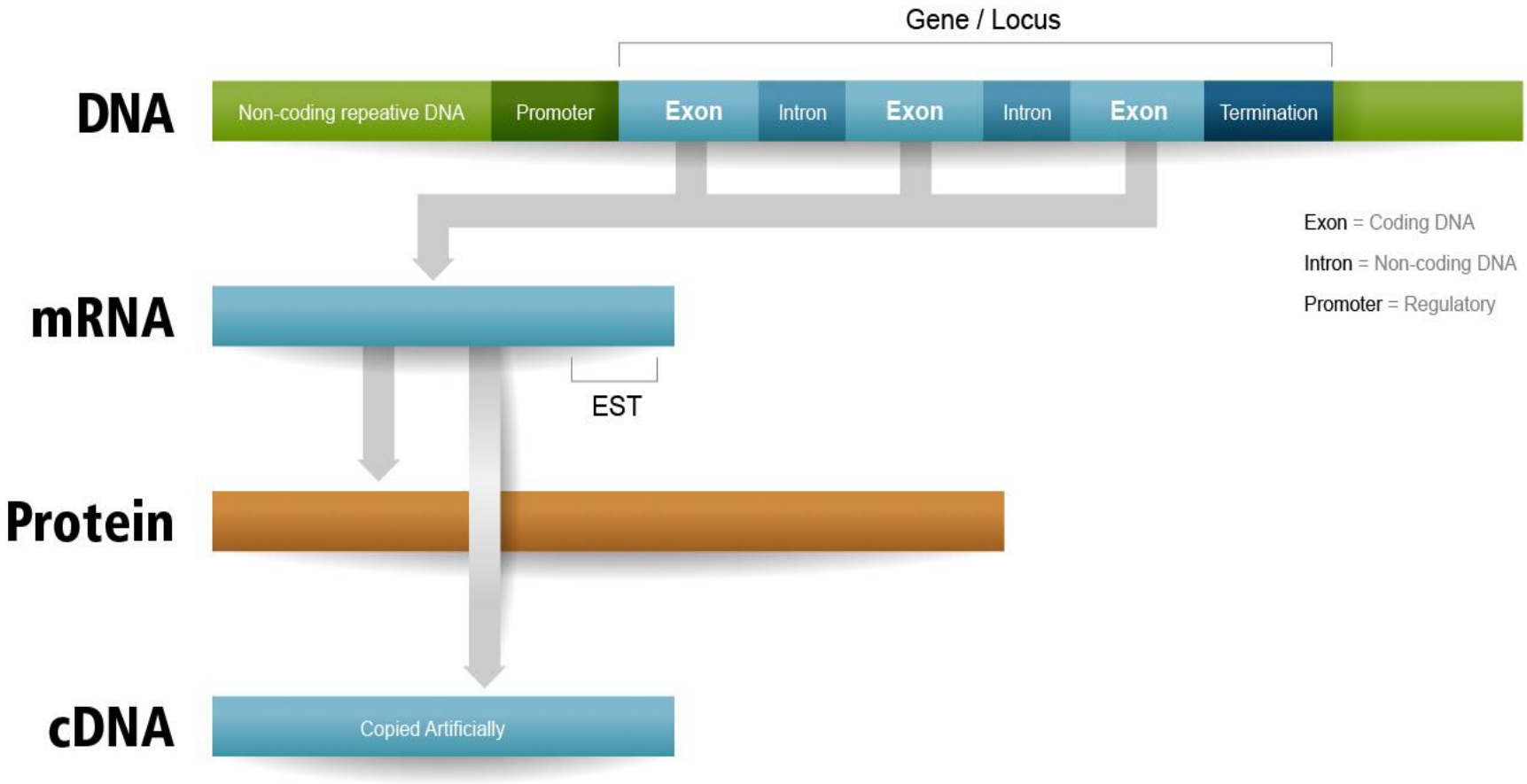


Figure Credit: Modified from Randy Johnson, U.S. Forest Service

Applied genetics

Applying genomics to breeding and resource management requires an introduction to sub-disciplines of genetics…

- Mendelian genetics describes inheritance from parents to offspring
  - *Discrete qualitative traits (including genetic markers)*
  - *Predicts frequencies of offspring given specific matings*
- Population genetics describes allele and genotype frequencies over space and time, including
  - *Changes in allele frequencies between generations*
  - *Environmental factors contributing to fitness*
  - *Models are limited to a small number of genes*
  - *Analyzes variation within and among populations*
- Quantitative genetics describes variation in traits influenced by multiple genes (continuous rather than discrete attributes)
  - *Relies on statistical tools describing correlations among relatives*
  - *Many genes, each with a small effect, influence a specific trait*

Mendelian genetics
(and the basis of genetic markers)

Mendelian inheritance

How do we explain family resemblance?

Mendel's experiments

- Selected 14 garden pea varieties to conduct breeding trials
- These represented alternative forms of seven distinct traits
- Cross bred alternative forms, followed by inbreeding the progeny of those crosses (F_2)

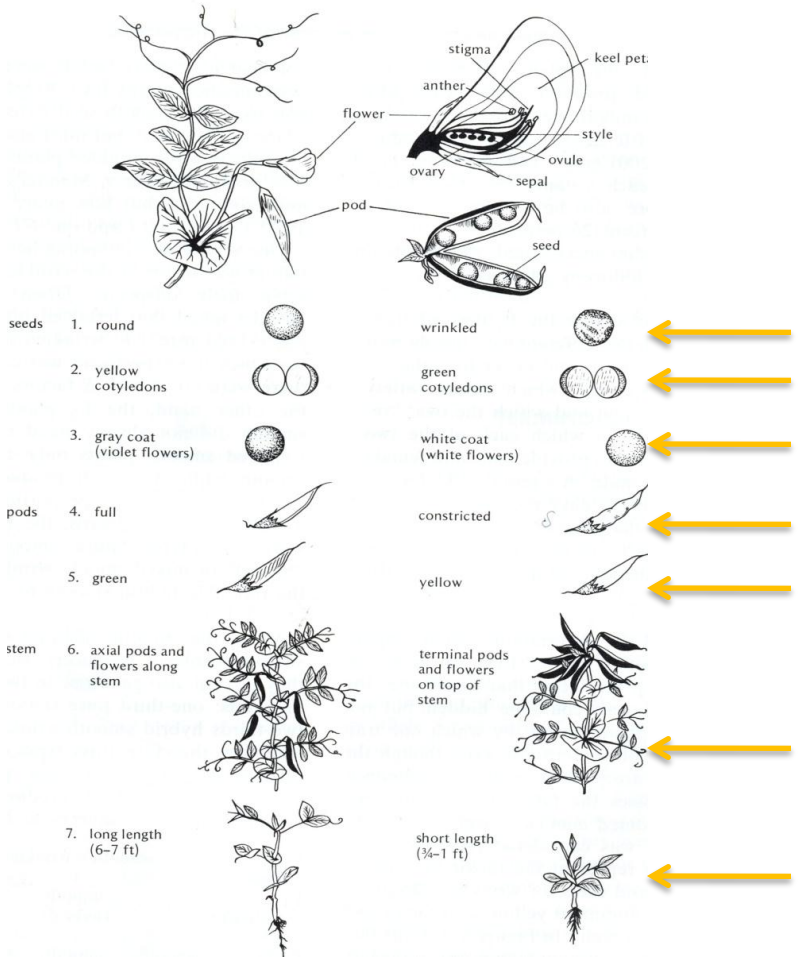
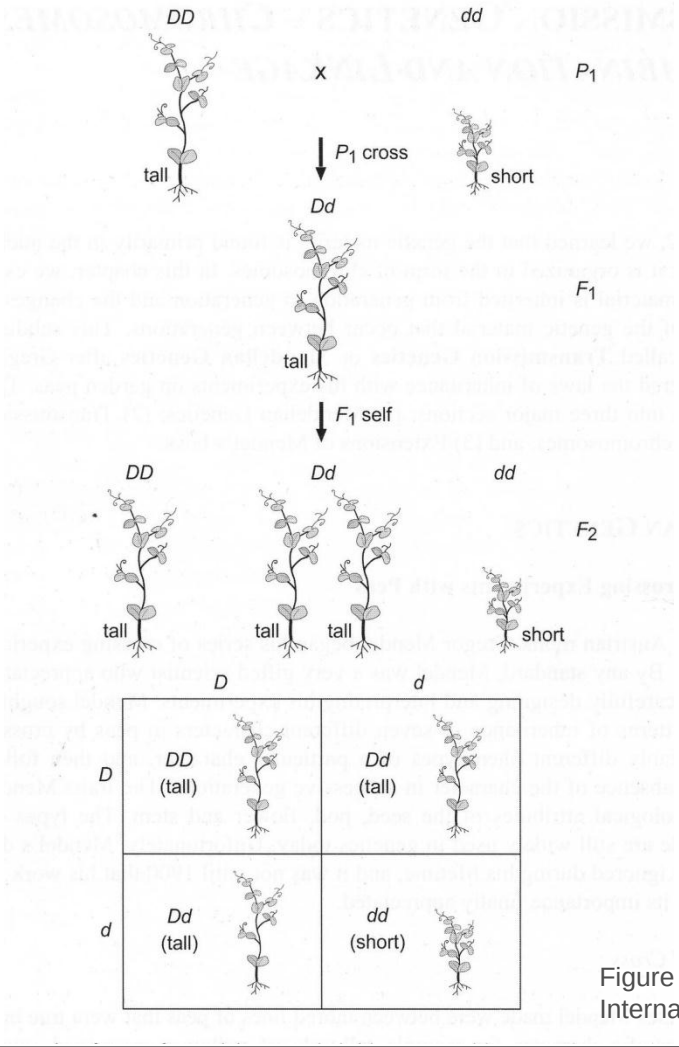


Figure Credit: Modified from Strickberger, Genetics. 1976.

Segregation



- Segregation – Members of an allelic pair separate cleanly from each other in germ cells

Figure Credit: White, T. L., W. T. Adams, and D. B. Neale. 2007. Forest Genetics. CAB International, Wallingford, United Kingdom. Used with permission.

Independent assortment

- Independent assortment – Members of different pairs of alleles assort independently of each other
- Exceptions:
  - *When loci are linked*
  - *When loci are in linkage disequilibrium*

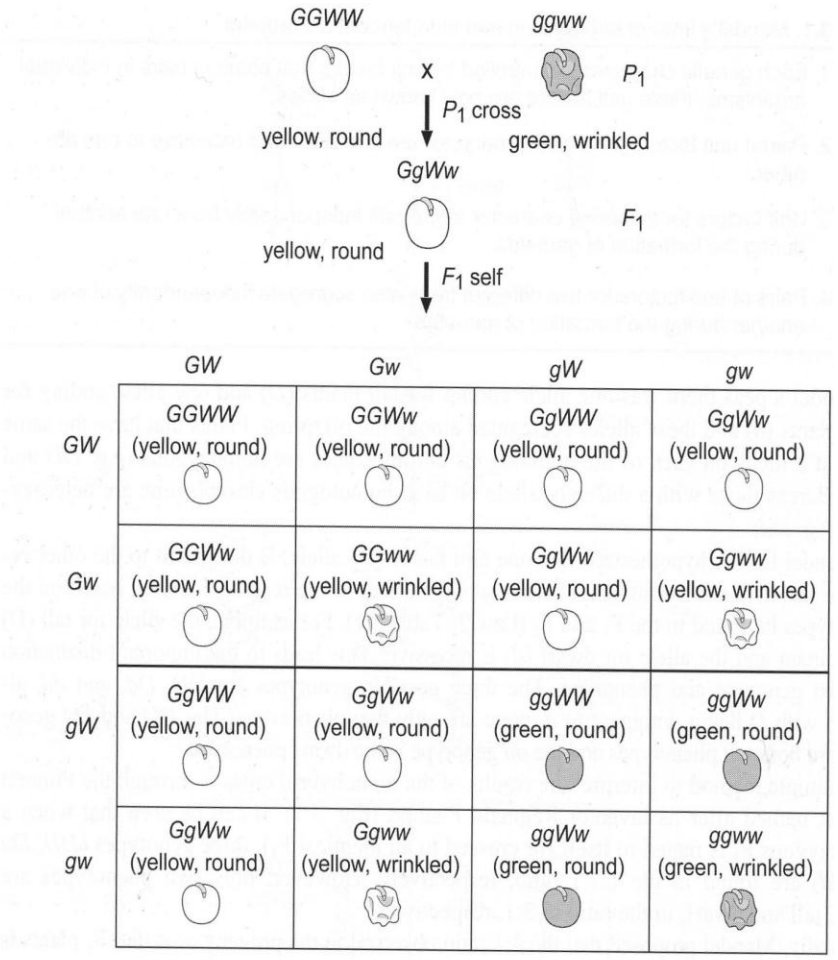
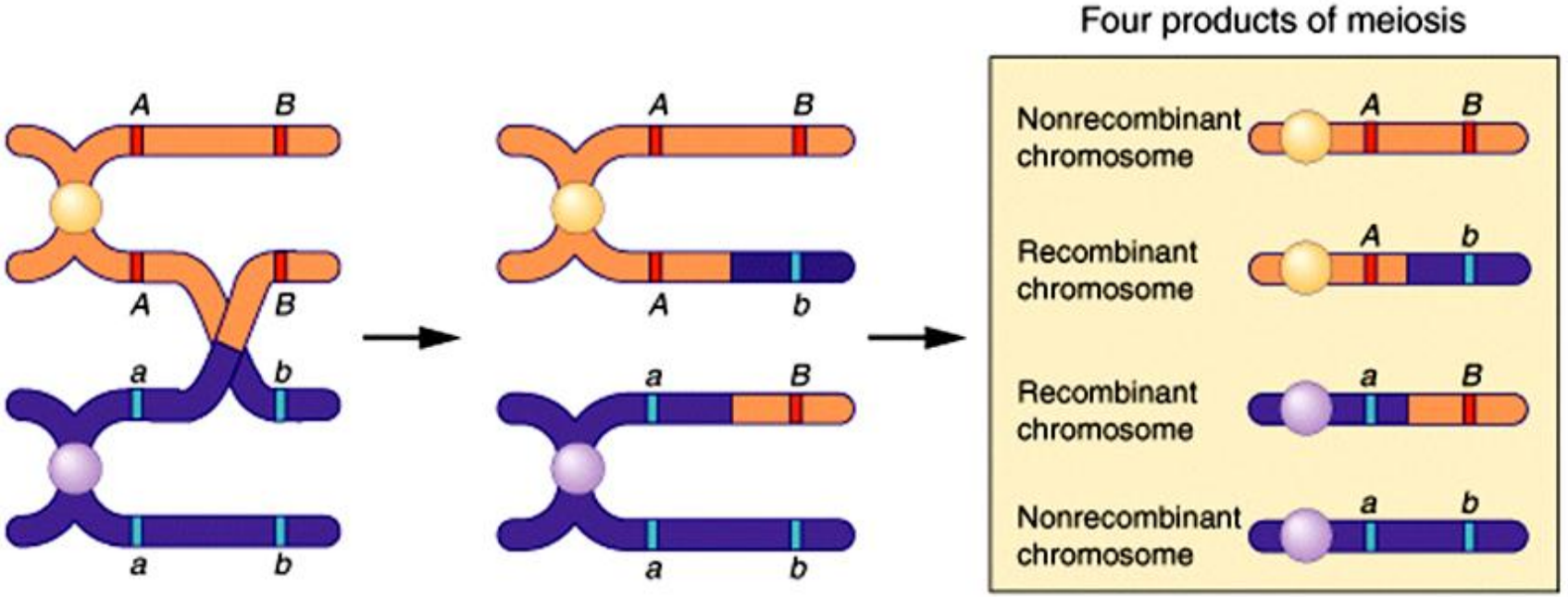


Figure Credit: White, T. L., W. T. Adams, and D. B. Neale. 2007. Forest Genetics. CAB International, Wallingford, United Kingdom. Used with permission.

Genetic linkage and recombination

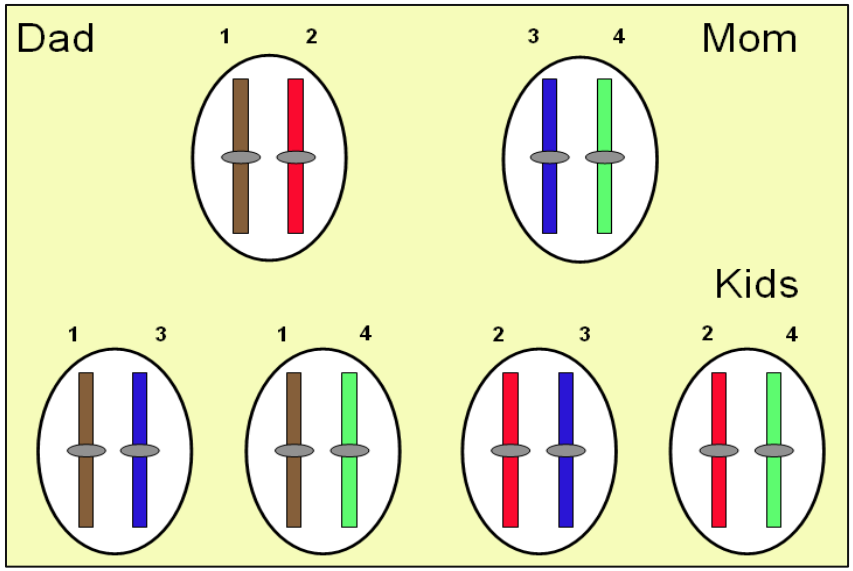
- Genes on different chromosomes are inherited independently
- Genes located on the same chromosome tend to be inherited together because they are physically linked – except that widely separated genes behave as if they are unlinked
- Recombination during gametogenesis breaks up parental configurations into new (recombinant) classes
- The relative frequency of parental and recombinant gametes reflects the degree of genetic linkage
- Genetic mapping is the process of determining the order and relative distance between genes or markers

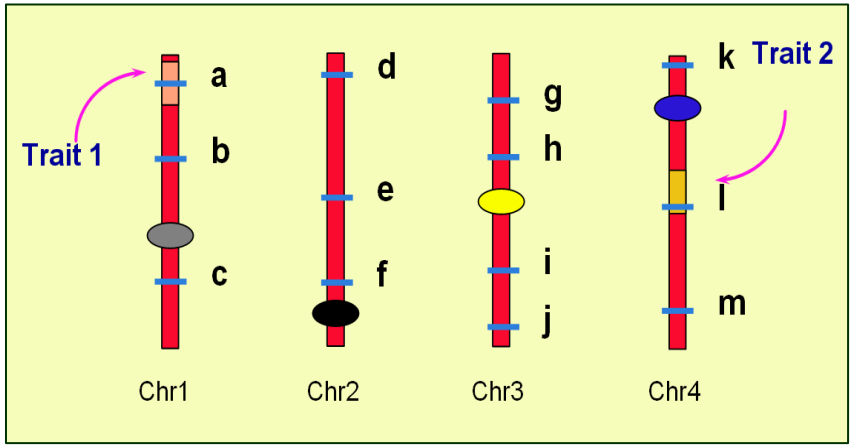
Parental and recombinant chromosomes



Gardner, E. J., M. J. Simmons, and D. P. Snustal. 1991. Principles of Genetics. 8th Edition. Reprinted with permission of John Wiley and Sons, Inc.

Markers track inheritance





- Markers allow for tracking of inheritance
- Linkage between markers allows for the creation of genetic maps
- Association between markers and phenotypic traits, observed within crosses, allows for mapping of QTL

Figure Credit: David Harry, Oregon State University

Linkage disequilibrium

- The correlation of alleles at different loci
- LD is a measure of non-random association among alleles – it describes the extent to which the presence of an allele at one locus predicts the presence of a specific allele at a second locus
- Though partly a function of physical distance, LD has several other causes, including recombination frequency, inbreeding, and selection
- LD is the basis upon which association genetics functions

Genotype and phenotype

- Genotype refers to the particular gene or genes an individual carries
- Phenotype refers to an individual's observable traits
- Only rarely can we determine genotype by observing phenotype
- Genomics offers tools to better understand the relationship between genotype and phenotype (genetic markers in abundance)

Single-gene traits in trees are rare… Here's one in alder (*F. pinnatisecta*)

P  x 

F1 

F1 x F1 (crossed or selfed)

F2  3 : 1 

Mendel's predictions hold up

Image Credit: David Harry, Oregon State University

Traits run in families but are not typically simply inherited

Sib 1 →

Family 1

Sib 2 →

← Sib 3

Family 2

← Sib 4

Image Credit: David Harry, Oregon State University

Landscape genomics

Image Credit: Nicholas Wheeler, Oregon State University

References cited

- Gardner, E. J., M. J. Simmons, and D. P. Snustal. 1991. Principles of Genetics. 8th Ed. John Wiley and Sons, Hoboken, NJ.
- Gibson, G., and S. V. Muse. 2004. A primer of genome science. Sinauer Associates, Sunderland, MA.
- Lewin, B. 2000. Genes VII. Oxford University Press, New York.
- Strickberger, M. W. 1976. Genetics 2nd edition. Macmillan Publishing, New York.
- White, T. L, W. T. Adams, and D. B. Neale. 2007. Forest genetics. CAB International, Wallingford, United Kingdom. (Available online at: http://bookshop.cabi.org/?page=2633&pid=2043&site=191) (verified 27 Apr 2011).

External Links

- The quest for genes [Online]. Microbial Genomics Workshop, DOE Joint Genome Institute, U.S. Department of Energy, Office of Science. The Regents of the University of California. Available at: www.jgi.doe.gov/education/microbialworkshop/The_Quest_for_Genes.ppt (verified 31 May 2011).

- Wikipedia contributors. 2010. File: Plant cell structure. Wikipedia, The Free Encyclopedia. Available at: http://commons.wikimedia.org/w/index.php?title=File:Plant_cell_structure.png&oldid=45093602 (verified 31 May 2011).

Thank You.

Conifer Translational Genomics Network
Coordinated Agricultural Project

CTGN CAP

UCDAVIS

USDA

United States
Department of
Agriculture

National Institute
of Food and
Agriculture