

## Welcome to BMB 400

- "Molecular Biology of Genes and Genomes"
- Biochemical basis for genetic phenomena:
  - structure of genes and chromosomes
  - replication and maintenance of DNA
  - pathway of gene expression
  - regulation of gene expression
- Instructor: Ross Hardison  
TA: Cathy Vrentas

## Text

- Customized textbook, 2 volumes
  - Text (built from lecture notes)
  - Problems
  - Answers
- You can supplement it with other texts for broader coverage  
E.g. Lewin's Genes VII

## Required work and points

- **4 scheduled EXAMS**
  - Midterm 1: 80 points
  - Midterm 2: 80 points
  - Midterm 3: 80 points
  - Final exam: 160 points
- Required project report
  - 100 points
- Total points for required material:
  - 500 points

## Grading Policy

- Point cut-offs for letter grade assignments determined by distribution of the sums of the scores on the required material (4 exams + project report)
- In addition, you will have extra credit opportunities
  - About 4-5 unannounced "quizzes", which are more like in-class exercises. Total about 20-25 pts
  - Additional projects to explore Internet-based resources and servers in biochemistry, molecular genetics and genomics
- E.g. cut-off for A may be 400, you have 380 for exams+project, and 30 points for extra credit. Your 410 total points gets an A.

## Make-up Exam Policies

- Make-ups for **EXAMS**
  - can be scheduled for students who must miss the exam for an acceptable excuse
    - E.g. illness, death in the immediate family.
  - will be problem-solving/essay
  - may be written or oral at the discretion of the instructor.
- No make-ups will be offered for any "quizzes"

## Academic Integrity Policy

- Academic integrity is the pursuit of scholarly activity free from fraud and deception.
- Dishonesty includes, but is not limited to, cheating, plagiarizing, facilitating acts of academic dishonesty by others, submitting work of another person, or tampering with the academic work of other students.
- Cite the source for any material or ideas obtained from others.
- All exam answers must be your own, and you must not provide any assistance to other students during exams.
- Academic dishonesty can result in assignment of "F" by the course instructors or "XF" by Judicial Affairs as the final grade for the student.

## Fundamental properties of genes

- **Genes are heritable units, arranged linearly along chromosomes.**
- Complementation analysis of a large number of mutants defines genes that determine a function.
  - E.g., biosynthetic pathway or DNA replication.
- Genetic techniques in microorganisms were used to determine the fine structure of a gene.
- Genes encode polypeptides
  - Codons are triplets of nucleotides that encode an amino acid.

## What are genes?

## Characteristics of Genes

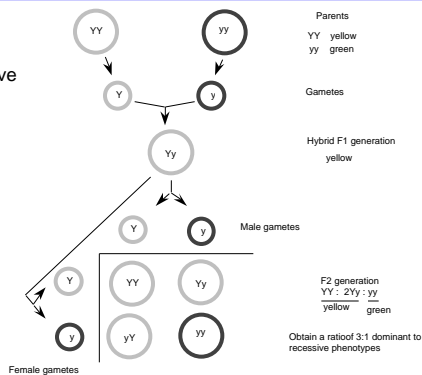
- Determine heritable phenotypes
- Are mutable: allelic variants
- Units of heredity
- Are on chromosomes
  - Behavior of genes mimics movement of chromosomes
  - Allelic variants segregate equally (1st Law)
  - Different genes usually sort independently (Mendel's 2nd Law)
- Linked on chromosomes in a linear array

## Types of alleles

- Wild type: normal, functional product
- Loss-of-function: usually recessive
  - Null: No product
  - Hypomorph: Less product
- Gain-of-function: usually dominant
  - New function
  - Hypermorph: More product
- Dominant negative: mutant product interferes with function of wild-type product
- Some allelic variants have no observable effects

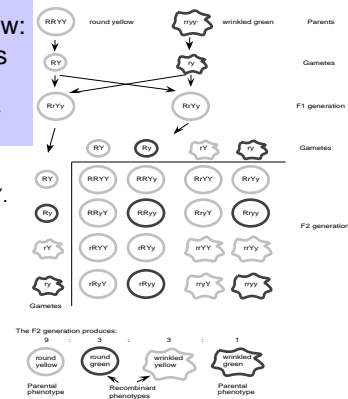
## Mendel's 1st Law: Alleles segregate equally

Genes behave as units:  
Discrete phenotypes



## Mendel's 2nd Law: Different genes assort independently

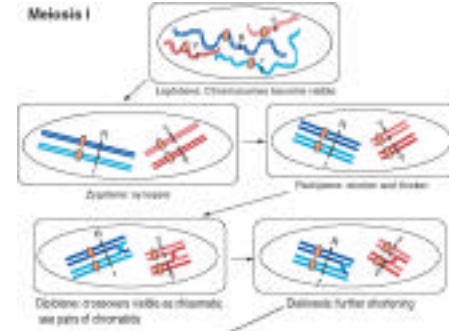
R does not stay with Y.  
r does not stay with y.  
Get nonparental phenotypes.



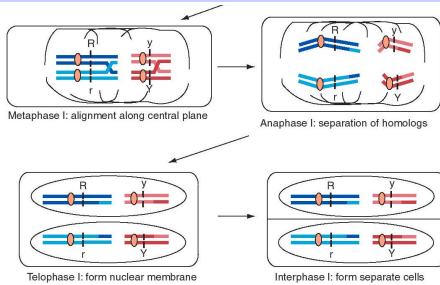
### Behavior of alleles mimics movement of chromosomes during meiosis

- Alleles correlate with homologous pairs of chromosomes
- Equal segregation of alleles: separation of homologous chromosomes at anaphase I of meiosis
- Independent assortment of different genes: independent separation of homologs of different chromosomes during meiosis
- Chromosomal theory of inheritance (Sutton and Boveri)

### Meiosis I



### Meiosis I (continued)



1st Law: R goes to precursor to 1 germ cell, r goes to another.  
2nd Law: R can assort with y or Y.

### Meiosis II



### Linked genes lie along chromosomes in a linear array

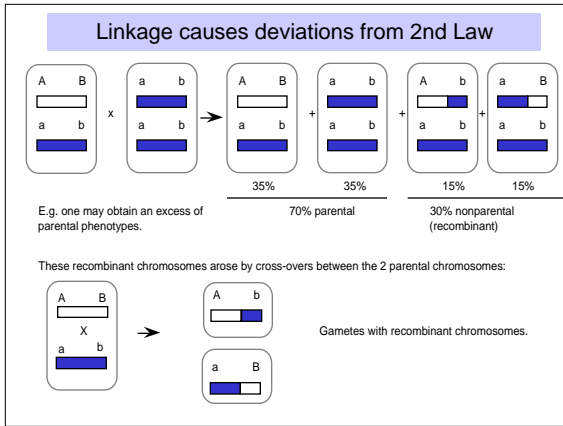
- Number of genes > number of chromosomes
- Some pairs of genes show substantial deviation from the predictions of Mendel's 2nd Law.
- Propensity of two genes to stay together rather than assorting independently is **linkage**.
  - Most easily seen in a **backcross** between an F1 heterozygote and a recessive homozygote.
- Genes on the same chromosome can be separated by recombination between homologous chromosomes.
  - Chiasmata formed between chromosomes in meiosis
- Recombination maps are linear.

### Expectation for unlinked genes in a backcross

Parents: AABB x aabb  
Generate F1 AaBb  
Backcross between F1 AaBb x aabb

A. If no linkage, expect 50% parental and 50% nonparental phenotypes.

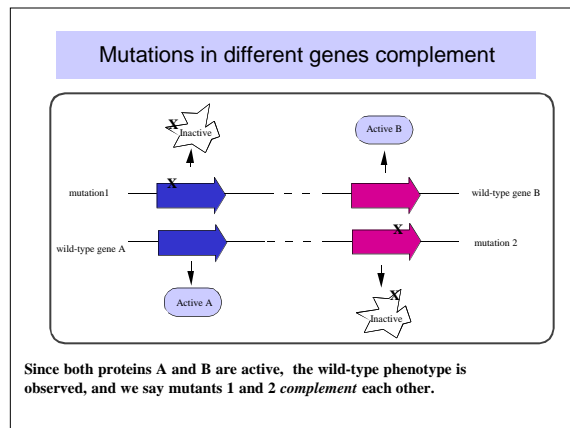
		Gametes from heterozygote			
		AB	Ab	aB	ab
Gametes from recessive double heterozygote	ab	AaBb	Aabb	aaBb	aabb
	Phenotypes:	Parental	Nonparental	Nonparental	Parental



- ### Fundamental properties of genes
- Genes are heritable units, arranged linearly along chromosomes.
  - Complementation analysis of a large number of mutants defines genes that determine a function.**
    - E.g., biosynthetic pathway or DNA replication.
  - Genetic techniques in microorganisms were used to determine the fine structure of a gene.
  - Genes encode polypeptides
    - Codons are triplets of nucleotides that encode an amino acid.

- ### Complementation
- The ability of **two mutants in combination** to **restore** a normal phenotype
  - A and B are different genes, allele 1 is wild-type, allele 2 is LOF mutant
 
$$A2A2 \ B1B1 \times A1A1 \ B2B2 \quad \text{parents}$$

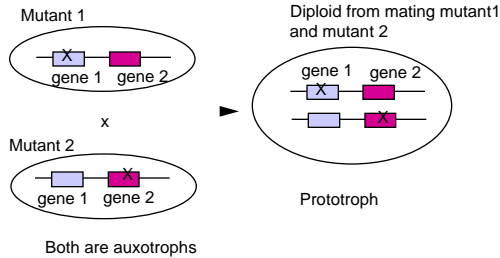
$$A2A1 \ B1B2 \quad \text{F1 progeny}$$
  - The function missing in each parent is restored in the progeny. The mutants **complement** each other.



- ### Use of complementation analysis in deducing number of genes in a pathway
- Start with many mutants that generate the same phenotype
  - Test all pairwise combinations of the mutants for complementation
  - Those pairs of mutations that **complement** are in **different** genes.
  - Those pairs that **fail to complement** are in the **same** gene.

- ### Growth factor requirements
- Auxotrophs** =
    - “increased growth requirements”
    - cells that require some additional nutrient (growth factor) to grow (e.g. “Arg auxotroph”).
  - Prototrophs**
    - wild type cells
    - do not have the need for the additional factor ; grow on minimal medium (e.g. they still make their own Arg)

### Complementation restores prototrophy



### Mutants that fail to complement constitute a complementation group

- Non-complementing strains carry different mutant alleles of the **same gene**.
- Thus a complementation group comprises a set of mutant alleles of the same gene, and it is an operational description of a gene (also called a **cistron**).
- Complementation distinguishes between mutations in the same gene or in different genes.

### Growth of diploids in the absence of arginine

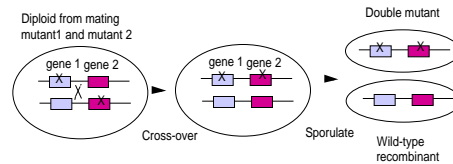
Mutant number	Mutant number					
	1	2	3	4	5	6
1	-	+	+	-	+	+
2		-	-	+	+	+
3			-	+	+	+
4				-	+	+
5					-	+
6						-

How many different “complementation groups” (genes)?

- 4 complementation groups:
- Gene 1: mutant strains 1 and 4
  - Gene 2: mutant strains 2 and 3
  - Gene 3: mutant strain 5
  - Gene 4: mutant strain 6

### Recombination

A physical exchange of DNA between chromosomes



### Fundamental properties of genes

- Genes are heritable units, arranged linearly along chromosomes.
- Complementation analysis of a large number of mutants defines genes that determine a function.
  - E.g., biosynthetic pathway or DNA replication.
- **Genetic techniques in microorganisms were used to determine the fine structure of a gene.**
- Genes encode polypeptides
  - Codons are triplets of nucleotides that encode an amino acid.

### Advantages of microorganisms for genetic analysis

- **have a haploid genome**
  - recessive phenotypes easily detected.
- **can be partially diploid (merodiploid)**
  - test whether alleles are dominant or recessive
- **increase cell number very rapidly**
  - can obtain large quantities of mutant organisms for biochemical fractionation
- **capable of sexual transfer of genetic information**

## Bacteriophage

- Bacteriophage have been a powerful model genetic system, because they
  - have small genomes
  - have a short life cycle
  - produce many progeny from one infected cell.
- They provide a very efficient means for transfer of DNA into or between cells.
- The large number of progeny make it possible to measure very rare recombination events.

## Fundamental properties of genes

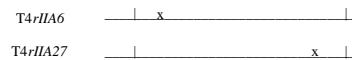
- Genes are heritable units, arranged linearly along chromosomes.
- Complementation analysis of a large number of mutants defines genes that determine a function.
  - E.g., biosynthetic pathway or DNA replication.
- Genetic techniques in microorganisms were used to determine the fine structure of a gene.
  - **Genes encode polypeptides**
    - **Codons** are triplets of nucleotides that encode an amino acid.

## How genes encode proteins

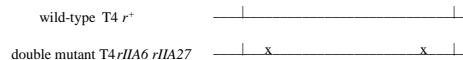
- Genes are composed of a series of mutable sites that are also sites for recombination (now recognized as nucleotides).
- Many genes encode at least one polypeptide (“One gene encodes one polypeptide”).
- The gene and the polypeptide are colinear.
- Single amino acids are specified by a set of three adjacent mutable sites; this set is called a codon.

## Recombination within genes allows construction of a gene map

Consider the results of infection of a bacterial culture with two mutant alleles of the T4 gene *rIIA* (causes rapid cell lysis but phage do not grow on E. coli K12)



Progeny from this infection include the parental phage (in the great majority) and, at a much lower frequency, two types of recombinants :



## Conclusions from recombination mapping of *rII*

- A large number of mutable sites occur within a gene: these are nucleotides.
- The genetic maps are clearly linear.
- Most mutations change a single mutable site (they are point mutations).
- Other mutations cause the deletion of a string of mutable sites.

## “One gene encodes one polypeptide”

Intermediates:

	M	--->	N	--->	O	--->	P	--->	Arg
Enzyme	1		2		3		4		
Gene	1		2		3		4		

Mutation in “Gene 2” results in loss of enzymatic activity 2 and accumulation of intermediate N.

“Gene 2” encodes enzyme 2.

More generally: **Many genes encode at least one polypeptide.**

## Alternative models for gene and codon structure

Model 1: The coding units = codons within genes could specify both composition and address of amino acids.

Encode:

Ser at 256	Ala at 144	Thr at 2	Met at 97	Cys at 187	Gly at 211	Glu at 11	etc.
---------------	---------------	-------------	--------------	---------------	---------------	--------------	------

The codons in this "gene" could be scrambled with no effect on the encoded polypeptide. The position of codons in the gene does not correspond to the position of amino acids in the polypeptide; i.e. the gene and polypeptide are not colinear.

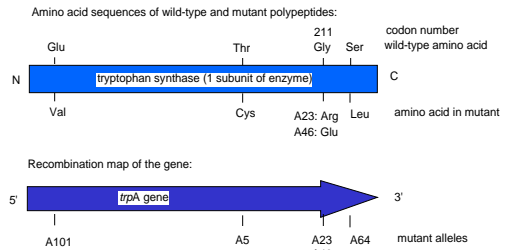
Model 2: The codons could specify only composition of an amino acid, and the address be deduced from the position of the codon within the gene.

Encode:

Ala	Ser	Thr	Gly	Arg	Gly	Cys	etc.
-----	-----	-----	-----	-----	-----	-----	------

e.g. Arg is inserted at position 5 of the polypeptide only because it is the 5th codon in the gene.

## The gene and its polypeptide product are colinear



## Implications of colinearity

- This correspondence between the positions of the mutations in each allele and the positions of the consequent changes in the polypeptide contradict the predictions of Model 1.
- Coding units (codons) do **not** provide information about the **address** of the amino acid.
- Model 2 is supported: the codon conveys information only about the **composition** of the amino acid.

## Characteristics of codons

- Single amino acids are specified by a set of three **adjacent mutable sites** (nucleotides)
- The set of three adjacent nucleotides is called a **codon**.
- The codons for a gene do **not overlap**.
- No punctuation** separates codons.

### 1. Amino acids are specified by adjacent mutable sites

- This was shown by recombination between different mutations in amino acid 211 of Trp synthase.
 

GGA	(Gly 211) -->	AGA	(Arg 211) mutant allele A23
		X	
GGA	(Gly 211) -->	GAA	(Glu 211) mutant allele A46

GGA (wt Gly 211)  
 in 2 out of 100,000 progeny
- Recombination to yield wild type occurs, albeit at a very low frequency. If mutations involved the **same** mutable site, one would *never* see the wild-type recombinant.

### 2. The genetic code is NOT overlapping

A: Overlapping code:

```

GCCGAC
GCC-Ala
CCG-Ser
CGA-Thr
GAC-Gly
    
```

A mutation at a single nucleotide would result in the alteration of more than one amino acid (e.g. changing the 2nd C would change Ala, Ser and Thr).

- ➔ However, alterations of a single nucleotide change only one amino acid, thus the code is non-overlapping.

### 3. Effect of frameshift mutations rule out a punctuated code

B: Punctuated code :

GCC UGAC UACG UGGC UAGA  
Ala Ser Thr Gly Arg

In this example, U means "end of codon." Insertions or deletions would affect only the codon with the insertion or deletion, not any of the other codons.

C: Non-overlapping, non-punctuated code :

ATGGCUUCUACGGGCAGA  
Met Ala Ser Thr Gly Arg

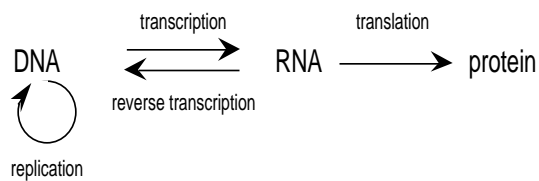
Insertions or deletions will affect the codon with the insertion or deletion **plus** all codons that follow. The reading frame will be changed.

### Restoration of function by indels of 3n nucleotides show that the code is read in triplets from a fixed starting point

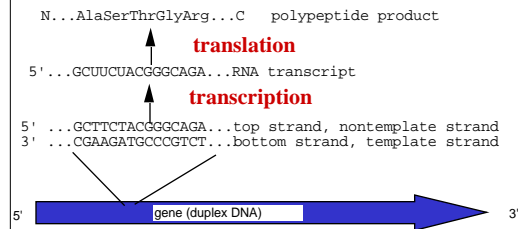
Wild-type	GCUUCUACGGGCAGA AlaSerThrGlyArg
Insertion (+)	v Insert A GCU <u>A</u> UCUACGGGCAGA Ala <u>Ile</u> ThrGlyGln
Deletion (-)	v Delete U GCUUCACGGGCAGAC Ala <u>Leu</u> ArgAlaAsp
Double mutant (+)	v Insert A and delete U GCU <u>A</u> CUACGGGCAGA Ala <u>Thr</u> ThrGlyArg
Triple mutant (+++)	v v v Insert A at 3 positions GCU <u>AUC</u> UACGGGCAGA Ala <u>IleIle</u> ThrGlyArg

Underlined amino acids or nucleotides differ from the wild-type.

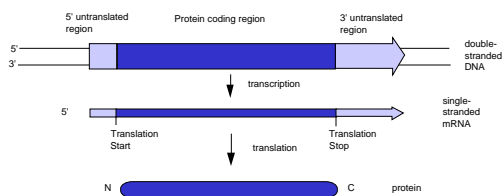
### Central Dogma of Molecular Biology



### Only one strand of duplex DNA codes for a product



### Untranslated sequences are at the ends of mRNA



### Regulatory signals are parts of genes

- Signals to start transcription, e.g. **promoters**
- Signals for regulating the amount of transcription
- Signals to stop transcription, e.g. **terminators**
- The gene includes the *transcription unit*, which is the segment of DNA copied into RNA in the primary transcript.

### Finding the function of genes

- Genes were originally defined in terms of phenotypes of mutants
- Now we have sequences of lots of DNA from a variety of organisms, so ...
- Which portions of DNA actually do something?
  - What do they do?
    - code for protein or some other product?
    - regulate expression?
    - used in replication, etc?