

BIOL 239 Course Notes

Genetics

Michael Socha

4A Software Engineering
University of Waterloo
Spring 2018

Contents

1	Course Overview	1
1.1	Logistics	1
1.2	Overview of Topics	1
2	Mendelian Genetics	2
2.1	Artificial Selection	2
2.2	Key Definitions	2
2.3	Mendel's Experiments	2
2.4	Genetics Prior to Mendel	3
2.5	Mendel's Laws	3
2.5.1	Law of Segregation	3
2.5.2	Law of Independent Assortment	3
2.5.3	Law of Dominance	3
2.6	Types of Crosses	3
2.6.1	Monohybrid crosses	3
2.6.2	Dihybrid crosses	4
2.6.3	Multihybrid crosses	4
2.7	Analysis Techniques	4
2.7.1	Punnett Squares	4
2.7.2	Branched-line Diagrams	4
2.7.3	Probability Rules	4
3	Extensions to Mendel for Single-Gene Inheritance	5
3.1	Incomplete Dominance	5
3.2	Codominance	5
3.3	Genes with more than 2 Alleles	5
3.4	Gene Mutations	5
3.5	Allele Frequencies	5
3.6	Pleiotropy	5
3.7	Recessive Lethal Alleles	6
4	Extensions to Mendel for Gene Interactions	7
4.1	Complementary Gene Action	7
4.2	Epistasis	7
4.3	Redundant Genes	7
4.4	Heterogeneous Traits and Complementation	8
4.5	Penetrance and Expressivity	8
4.6	Modifier Genes	8
5	Pedigrees	9
5.1	Single-Gene Traits in Humans	9
5.2	Studying Hereditary Traits in Humans	9

5.3	Pedigree Notation	9
5.4	Patterns of Inheritance	9
6	Chromosomes	10
6.1	Experimentation History	10
6.2	What are Chromosomes?	10
6.3	Sex Chromosomes and Sex Determination	10
7	Mitosis	11
7.1	Interphase	11
7.1.1	G1	11
7.1.2	S	11
7.1.3	G2	11
7.2	Phases of Mitosis	11
7.2.1	Prophase	11
7.2.2	Prometaphase	12
7.2.3	Metaphase	12
7.2.4	Anaphase	12
7.2.5	Telophase	12
7.2.6	Cytokinesis	12
8	Meiosis	13
8.1	Meiosis I	13
8.1.1	Prophase I	13
8.1.2	Metaphase I	13
8.1.3	Anaphase I	13
8.1.4	Telophase I	14
8.2	Meiosis II	14
8.3	Mistakes in Meiosis	14
8.4	Meiosis and Genetic Diversity	14
9	Gametogenesis	15
9.1	Oogenesis	15
9.2	Spermatogenesis	15
10	Validation of Chromosome Theory	16
10.1	Key Claims	16
10.2	Overview of Evidence	16
10.3	Experimental Results	16
11	Aneuploidy	17
11.1	Autosomal Aneuploidy	17
11.2	Sex Chromosome Aneuploidy	17
11.3	Meiotic Nondisjunction	17
11.4	Mitotic Nondisjunction	17

12 Euploidy	19
12.1 Monoploidy	19
12.2 Triploidy	19
12.3 Tetraploidy	19
13 Gene Linkage and Recombination	20
13.1 Testcrosses for Linkage Detection	20
13.2 Observing Recombination	20
13.3 Recombination Frequencies	20
13.4 Linkage Analysis	20
13.4.1 Chi-Square Test	21
14 Gene Mapping	22
14.1 Two-Point Crosses	22
14.2 Three-Point Crosses	22
15 DNA	23
15.1 Evidence DNA is Genetic Material	23
15.2 Structure of DNA	23
15.2.1 Alternative Forms of DNA	23
15.3 DNA Replication	24
15.3.1 Models of Replication	24
15.3.2 Molecular Mechanism of Replication	24
15.3.3 Ensuring Integrity of Genetic Information	25
15.3.4 Replication of Circular Chromosome	25
15.3.5 Replication of Long Chromosomes	26
15.3.6 Replication of Linear Chromosomes	26
16 Genetic Code	27
16.1 “Cracking” the Genetic Code	27
16.1.1 Yanofsky’s Experiments	27
16.1.2 Triplet Codons	27
16.1.3 Mapping Codons to Amino Acids	27
16.2 Nonsense Codons	28
16.3 Universality	28
17 Transcription	29
17.1 Steps of Transcription	29
17.1.1 Initiation	29
17.1.2 Elongation	29
17.1.3 Termination	29
17.2 Addition of Cap and Tail	30
17.3 Exons and Introns	30
17.4 RNA Splicing	30
17.5 Purpose of Introns	30

18 Translation	31
18.1 Transfer RNA	31
18.1.1 Structure	31
18.1.2 Wobble	31
18.2 Ribosomes	31
18.2.1 Structure	31
18.3 Translation Process	32
18.3.1 Initiation	32
18.3.2 Elongation	32
18.3.3 Termination	32
18.4 Prokaryotic vs Eukaryotic Differences	32
18.5 Post-translational Processing	33
19 Prokaryotic Gene Expression	34
19.1 Catabolic and Anabolic Pathways	34
19.2 E. Coli's Utilization of Sugar Lactose	34
19.2.1 Catabolic Proteins	34
19.2.2 Operon Theory Overview	34
19.2.3 Evidence for Repressor Protein	35
19.2.4 Repressor Mutants	35
19.2.5 Trans vs Cis Action	35
20 Eukaryotic Gene Expression	36
20.1 Regulation Through Transcription	36
20.1.1 RNA Polymerase	36
20.1.2 Regulatory Regions	36
20.1.3 Basal Factors	36
20.1.4 Activators	37
20.1.5 Repressors	37
20.2 Chromatin Structure	37
20.2.1 Chromatin Remodeling	37
20.2.2 Histone Modification	38
20.2.3 DNA Methylation	38
21 Restriction Enzymes	39
21.1 Use as a Defense Mechanism	39
21.2 Cutting	39
21.3 Restriction Mapping	39
22 Molecular Cloning	40
22.1 Cloning Vectors	40
22.2 Host Cells	40
23 Population Genetics	41
23.1 Key Terms	41

23.2	Hardy-Weinberg Principle	41
23.2.1	Hardy-Weinberg Assumptions	41
23.2.2	Hardy-Weinberg Proportions	42
23.2.3	Reaching Equilibrium	42
23.3	Changes from Equilibrium	42
23.3.1	Genetic Drift	42
23.3.2	Natural Selection	42
24	Genetic Analysis	43
24.1	Genomic and cDNA Libraries	43
24.2	DNA Probes	43
24.3	Polymerase Chain Reaction	43
24.4	DNA Sequencing	43
25	Mutation	45
25.1	Classes of Mutations	45
25.2	Rates of Spontaneous Mutation	45
25.3	Causes of Mutations	45
25.4	Molecular Mechanisms of Mutation	46
25.5	Repairing Damaged DNA	46
25.5.1	Base Excision Repair	46
25.5.2	Nucleotide Excision Repair	46
25.5.3	Repairing Mistakes During DNA Replication	46
26	Cancer Genetics	48
26.1	Characteristics of Cancer Cells	48
26.2	Genetic Basis of Cancer	48
26.3	Tumor-Suppressor Genes	48
26.3.1	TP53	48
26.3.2	RB	49
26.3.3	BRCA1 and BRCA2	49
26.4	Oncogenes	49

Course Overview

Logistics

- **Professor:** Christine Dupont

Overview of Topics

Topics covered in this course include:

- Transmission of traits
- Gene structure, function, and transmission
- Genes as carriers of information
- Gene regulation
- Population genetics

Mendelian Genetics

Artificial Selection

Artificial selection is the process in which humans select plants and animals to breed based on particular traits. Artificial selection has been performed for tens of thousands of years, long before humans had a solid understanding of genetics. Examples of plants and animals domesticated through selective breeding include dogs, goats, sheep, cattle, cabbage, and maize.

Key Definitions

- **Phenotype:** An observable characteristic in an individual, which can include an appearance or behaviour
- **Genotype:** The genetic composition of an individual, which in turn controls phenotypes
- **Gene:** A discrete unit of heredity
- **Allele:** An alternative form of a single gene
- **Dominant Allele:** An allele that is manifested in a phenotype regardless of the other allele. The phenotype in which this allele is manifested is known as the dominant phenotype.
- **Recessive Allele:** An allele that does not have an effect on a phenotype in when a dominant allele is present. The phenotype in which this allele is manifested is known as the recessive phenotype.
- **Parental Types:** Phenotypes that reflect a previously existing parental combination
- **Recombination Types:** Phenotypes that reflect a new combination of genes that occurs during gamete formation
- **Filial Generation:** Subsequent offspring in a sequence of breeding, starting with the parental (P) generation

Mendel's Experiments

Mendel was a 19th century Augustinian monk that ran a series of experiments regarding trait inheritance. Much of the value of Mendel's experiments was due to how they featured:

- Using garden peas, which are easy to cross-fertilize, produce a large number of offspring, and have a short growing season (reproduce frequently). Garden peas also have clear-cut forms of the traits to test (e.g. yellow vs green, round vs wrinkled).

- Establishment of pure-breeding lines, which are lines are only breed with other members of the same genotype
- Carefully controlled breeding (use of reciprocal crosses and self-fertilization)

Genetics Prior to Mendel

Multiple incorrect theories existed prior to Mendel's laws of inheritance. These included:

- One parent contributing more to an offspring the other - disproved through reciprocal crosses
- Parental traits become blended (i.e. no discrete traits) - disproved through reappearance of reciprocal traits

Mendel's Laws

Through his experimentation, Mendel devised the following laws of inheritance:

Law of Segregation

This law states that each trait is controlled by two alleles which separate upon gamete formation and recombine when a zygote is formed.

Law of Independent Assortment

This law states that alleles for separate phenotypic traits are transmitted to offspring independent of one another. This can be demonstrated through dihybrid experiments, where individuals differing in multiple traits are crossed with one another.

Law of Dominance

This law introduces the concept of dominant and recessive alleles, and states that recessive alleles do not affect the phenotype unless both alleles are recessive.

Types of Crosses

Monohybrid crosses

Monohybrid crosses are mating between homozygous (i.e. both alleles the same) individuals that only differ in one trait. The F1 filial features only the dominant phenotype, while the

F2 filial features the dominant phenotype and recessive phenotype in a 3:1 ratio.

Dihybrid crosses

Dihybrid crosses are mating between homozygous individuals that only differ in two unrelated traits. The F1 filial features both dominant phenotypes, while the F2 filial features a 9:3:3:1 ratio of traits 1 and 2 both being dominant, trait 1 being dominant and trait 2 being recessive, trait 1 being recessive and trait 2 being dominant, and both traits being recessive.

Multihybrid crosses

Dihybrid crosses are mating between individuals that only differ in three or more unrelated traits.

Analysis Techniques

Punnett Squares

A Punnett Square is a chart that lists the possible gametes that may combine in a cross (one gamete list forms 1 dimension), and lists all possible resulting combinations.

Branched-line Diagrams

A branched-line diagram features one column for each gene in a cross, with the resulting phenotypes shown at the end. This type of diagram is useful for multihybrid crosses, where Punnett Squares become hard to read.

Probability Rules

Probability rules (i.e. product rule and sum rule) can be also be used to analyze the results of crosses, often in combination with the above techniques.

Extensions to Mendel for Single-Gene Inheritance

Incomplete Dominance

Incomplete dominance describes the situation where F1 offspring have a phenotype that resembles neither of their parents, but rather appears as a blend of the two. An example is crossing red and white flowers to generate the F1 generation, and the F1 generation's flowers being pink.

Codominance

Codominance describes the situation where F1 offspring have a phenotype that combines features from both parents. An example would be crossing dotted and spotted lentils to generate the F1 generation, and the F1 generation being both spotted and dotted.

Genes with more than 2 Alleles

Some genes can have more than 2 alleles. In this case, filial genotype and phenotype ratios can be much different from those experiments with only 2 alleles.

Gene Mutations

Mutations of genes occur in nature at a fairly low frequency; the frequency of gametes with a mutation for a particular gene is typically between 1 in 10000 to 1 in 1000000, depending on the gene. Such mutations allow for the creation of new alleles.

Allele Frequencies

The percentage of a particular allele in a gene's entire population is known as its allele frequency. Alleles that have a relatively high frequency (typically over 1%) are known as wild-type alleles, while other alleles are known as mutant alleles. A gene with only one wild-type allele is known as monomorphic, while genes with multiple wild-type alleles are known as polymorphic.

Pleiotropy

Pleiotropy is the concept of a single gene determining multiple (often seemingly unrelated) phenotypic traits. Mendel himself observed this phenomenon during his experiments, where pea seed coat color matched to flower color, implying that there was a common control for

both traits. Note that a single allele may be dominant with respect to some traits and recessive with respect to others, so dominant and recessive for such alleles must be defined with the context of a specific phenotype.

Recessive Lethal Alleles

Some alleles that are harmless when combined with a different allele can render an individual inviable when there are two copies of that allele. This allele is recessive in terms of lethality, but may be dominant with respect to some other phenotype. In this case, the phenotypic ratios of this dominant trait in the F1 generation would be 2:1 instead of 3:1, since individuals with two copies of this trait's dominant allele cannot survive. Note that some recessive lethal alleles may cause delayed lethality, in which case the phenotypic ratios would not be 2:1, since affected individuals may survive for some time (and may even reproduce).

Extensions to Mendel for Gene Interactions

Traits in an organism that arise from the actions of multiple genes are known as polygenic, while traits that are controlled by multiple genes and the environment are known as multifactorial. These two types of traits are known as complex traits, and account for the majority of traits in organisms. Genotypic classes are groupings of related genotypes that produce a particular phenotype.

Complementary Gene Action

Complementary gene action refers to multiple genes working together to produce a particular trait. As an example, consider a combination of 2 genes with alleles Aa and Bb, and assume that a certain trait only surfaces in the A-B- phenotype. In the F₂ population, this will result in a 9:7 dominant to non-dominant trait ratio.

Epistasis

Epistasis refers to one allele masking the effect of another allele. The allele performing the masking is known as the epistatic allele.

A situation where two recessive alleles are required to mask the effect of some other allele is known as recessive epistasis. In a recessive epistasis example where the b allele may be epistatic and a BBEE generation is crossed with a bbee generation, the F₂ ratios are 9:3:4 (9 parts B-E-, 3 parts bbE-, and 4 parts -ee).

A situation where one dominant allele is sufficient to mask the effect of some other allele is known as dominant epistasis. In a dominant epistasis example where the B allele may be epistatic and a BBEE generation is crossed with a bbee generation, the F₂ ratios are 12:3:1 (12 parts B—, 3 parts bbE-, and 1 part bbee). Should a bbee phenotype be the same as a B— phenotype, this ratio can be simplified to 13:3 - this case is known as dominant suppression.

Redundant Genes

Redundant genes are two genes that control a very similar phenotype. As an example, if genes A and B are redundant and a AABB generation is crossed with a aabb generation, the F₂ ratios are 15:1.

Heterogeneous Traits and Complementation

Phenotypic traits that can arise as a result of multiple different genes are known as heterogeneous. Thus, it is possible that two individuals sharing the same traits have differing genetic causes of those traits. If these individuals have offspring which exhibit a wild-type phenotype, then complementation has occurred, meaning that different genes controlled the mutant phenotype for both parents, and that they were both recessive. Thus, each parent could apply a dominant allele to complement the other parent's recessive mutant allele.

Penetrance and Expressivity

A phenotype may depend on more factors than just the underlying genotype. These factors may include environmental factors, modifier genes, and random chance. Penetrance is used to describe how many members of a population with a particular genotype show the expected phenotype. Penetrance can be complete (i.e. 100%) or incomplete. Expressivity refers to the intensity with which a particular genotype is expressed in a phenotype, which may be variable or unvarying.

Modifier Genes

Genes that alter the phenotype produced by alleles of other genes are known as modifier genes.

Pedigrees

Single-Gene Traits in Humans

Most human traits are controlled by the interaction of many genes. In fact, most single-gene traits in humans evoke life-threatening diseases, such as cystic fibrosis and Tay-Sachs disease.

Studying Hereditary Traits in Humans

Humans make poor genetic testing subjects; on top of many ethical issues, humans tend to not have many offspring, do not interbreed much (i.e. no F2 generations), and rarely choose mates on purely genetic grounds. Human genetic lineage is often analyzed using pedigrees, which are diagrams detailing a family's genetic features over several generations.

Pedigree Notation

- **Sex:**
 - **Male:** Square
 - **Female:** Circle
 - **Unspecified:** Diamond
- **Trait infliction:**
 - **Trait inflicted:** Filled shape
 - **Trait not inflicted:** Unfilled shape
- **Mating:**
 - **Consanguineous (i.e. from same ancestor) mating:** Double horizontal line
 - **Non-Consanguineous mating:** Single horizontal line

Patterns of Inheritance

Dominant traits tend to follow a vertical pattern of inheritance, being transferred over multiple generations. On the other hand, recessive traits tend to follow a horizontal pattern of inheritance, where in order for offspring to exhibit a recessive phenotype, both parents must carry the allele that causes that phenotype.

Chromosomes

Experimentation History

In the 17th century, semen was determined to contain spermatozoa, which was hypothesized to be able to enter an egg and achieve fertilization. This hypothesis was confirmed by further experimentation in the 19th century. Around 1880, threadlike structures later confirmed to be chromosomes were observed during cell division. In the early 1900s, a link was established between chromosomes and the Mendelian principles of segregation and independent assortment, giving rise to the chromosomal theory of inheritance.

What are Chromosomes?

Chromosomes are organelles that package DNA. Most cells that compose an individual, as well as zygotes, carry two matching sets of chromosomes. These cells are known as diploids. Chromosomes in a pair that contain the same genes are called homologous chromosomes (or homologues). Since one chromosome in each pair comes from each parent, homologues may contain different alleles. Cells that only contain one set of chromosomes, such as gametes, are known as haploids.

A karyotype is a visual description of the complete set of chromosomes found in the cell of an organism. In a karyotype, homologous pairs are shown together ordered by decreasing size. The matching pairs of chromosomes are known as autosomes, which are distinct from sex chromosomes.

Sex Chromosomes and Sex Determination

Sex chromosomes determine the sex of an organism. In humans, the 23rd pair of chromosomes are sex chromosomes, where females have an XX pair and males have a non-homologous XY pair. The Y chromosome's SRY (sex-determining region of Y) gene is specifically what determines maleness. Sex-determining mechanisms can vary significantly between different species.

Mitosis

Mitosis is a process of cell division in which a cell produces two genetically identical daughter cells. In the cell cycle, mitosis is known as the M phase.

Interphase

Interphase is the part of the cell cycle that occurs between divisions. Interphase consists of the gap 1 (G1), synthesis (S), and gap 2 (G2) phases.

G1

G1 lasts from the birth of a new cell until chromosome replication. In this phase, most of a cell's growth occurs. The duration of G1 can vary greatly, from between a few hours in a newly developing embryo, to the remaining of an individual's lifetime for mature brain cells.

S

In the S phase, the cell duplicates its genetic material by synthesizing DNA. Each chromosome doubles to produce identical sister chromatids, which become visible once chromosomes condense at the beginning of mitosis. These sister chromatids remain joined to each other at the centromere until mitosis begins.

G2

During G2, the cell may grow, though typically less than during G1. The cell also synthesizes proteins necessary for mitosis.

Also, during interphase, fine microtubules known as centrioles radiate out into a cell's cytoplasm from a structure known as a centrosome.

Phases of Mitosis

Prophase

During interphase, the cell nucleus remains intact, and chromosomes are masses of chromatin, which look like a mass of fine tangled string. At prophase, chromosomes condense, and move apart toward opposite poles of a cell. Also, nucleoli begin to disappear, and microtubules disappear and are replaced by a set of dynamic microtubules that grow from and shrink toward their centrosomal organizing centers. The centrosomes also move apart to opposite poles of the cell.

Prometaphase

In prometaphase, the nuclear envelope breaks down, and microtubules from the centrosomes invade the nucleus. Sister chromatids attach to microtubules from opposite centrosomes through the kinetochore, a structure in the centromere region of each chromatid specialized for transport.

Metaphase

In metaphase, chromosomes align on the metaphase plate, which is an equator halfway between the cell's two poles. In this phase, forces pulling sister chromatids toward opposite poles are in balance with tension across the chromosomes.

Anaphase

During anaphase, the centromeric connections are severed, and separated sister chromatids begin to move to opposite cell poles to which they are connected through kinetochore microtubules. This separation of sister chromatids to pass genetically identical material onto both daughter cells is known as disjunction.

Telophase

During telophase, the nuclear membrane and nucleoli reform, the spindle fibres disappear, and chromosomes untangle to become a mass of chromatin. This phase can be seen as somewhat of a reverse of prophase.

Cytokinesis

In cytokinesis, the nuclei emerging at each end of the elongated parent cell are packaged into two separate daughter cells. Processes of cytokinesis may actually begin during anaphase, though do not complete until after telophase.

Meiosis

The majority of cells that make up an individual are somatic cells, which either mitotically divide or are arrested in a G phase of the cell cycle. On the other hand, Germ cells are incorporated into reproductive organs, where they eventually undergo meiosis, a type of cell division that yields 4 gamete haploid cells from a single diploid cell. Meiosis contains two separate cell divisions, known as meiosis I and meiosis II.

Meiosis I

Prophase I

Leptotene is the first substage of prophase I, and is the phase in which chromosomes start to thicken. These chromosomes have already been priorly duplicated.

In Zygotene, each chromosomes seeks out its homologous partner, which become zipped together in a process known as synapsis. Each synapses chromosome pair is known as a bivalent (contains two chromosomes), or a tetrad (contains four chromatids). One side of a bivalent has maternal genetic material, and the other has paternal genetic material.

Pachytene begins after synapsis, and consists of structured called recombination nodules exchanging parts of maternal and paternal chromatids of homologous chromosomes. This exchange is known as recombination or crossing-over.

In diplotene, the synaptonemal zipper complex begins to disappear, and the tetrads of four chromatids become visible and start to move apart. However, homologous chromosomes remain tightly connected at areas called chiasmata, which represent sites where recombination occurred.

In diakinesis, the chromatids condense and shorten even further. Also, the nuclear membrane begins to break down, and spindles begin to attach to kinetochere structures.

Metaphase I

During metaphase I, the tetrads formed in prophase I line up along the metaphase plate. These tetrads are attached to spindles on opposite poles, and begin to separate.

Anaphase I

In anaphase I, the chiasmata joining homologues are dissolved, allowing the chromosomes to separate and move toward opposite poles.

Telophase I

Telophase I begins when nuclear membranes start to form around the chromosomes that have moved to each pole. Each daughter nuclei contains half the number of chromosomes in the original parent (meiosis I is known as reductional division), but there are two sister chromatids for each chromosome.

Meiosis II

Following a brief interphase following meiosis I known as interkinesis, Meiosis II is performed on both daughter cells. Meiosis II is a very similar process to mitosis, and yields four haploid cells in total.

Mistakes in Meiosis

Segregational errors during meiotic divisions can lead to aberrations. An example cause is when homologues fail to separate during meiosis I, causing a trisomy through a process known as nondisjunction. Many trisomies are lethal in utero, though others, such as trisomy 21 (i.e. Down syndrome), are not.

On a separate note, some hybrid animals carry non-homologous chromosomes that cannot pair up and segregate during meiosis. Viable gametes cannot form in such animals, so they are sterile.

Meiosis and Genetic Diversity

Two main factors contribute to genetic diversity in meiosis. The first is that the pole maternal or paternal homologues migrate to is random, so the homologue pairs are able to assort independently of one another. The second factor is the reshuffling of genetic material that occurs during recombination.

Gametogenesis

Embryonic germ cells (collectively known as a germ line) are specialized diploid cells that can give rise to haploid cells through meiosis. These meiotic divisions lead to the production of gametes, with either the resulting haploids being gametes, or some of the resulting haploids undergoing some sort of differentiation to become gametes.

Oogenesis

Oogenesis is the process that produces egg cells, also known as ovums. Oocytes are cells that undergo meiotic divisions to ultimately generate one ovum. About half a million primary oocytes are stored in the ovaries, and are arrested in the diplotene substage of prophase I. During ovulation, a primary oocyte is released and produces a secondary oocyte and a far smaller polar body. The first polar body usually arrests its development, while the secondary oocyte proceeds as far as the metaphase of meiosis II. Upon fertilization, meiosis II completes, with the secondary oocyte producing a mature ovum and a second polar body. The long interval between the beginning and completion of meiosis for oocytes may lead to segregational errors such as trisomies for older females.

Spermatogenesis

Spermatogenesis is the process that produces sperm cells. Spermatocytes are cells that undergo meiotic divisions to produce sperm cells. Unlike oocytes, spermatocytes undergo a symmetric meiosis I and II, yielding four spermatids for each spermatocyte. Spermatogenesis can continue throughout a male's lifetime.

Validation of Chromosome Theory

Key Claims

The key claims of the chromosomal theory of inheritance are that:

1. Hereditary information is stored on genes, which are located in chromosomes.
2. The egg and sperm contribute equally to the genetic composition of the resulting zygote.

Overview of Evidence

Based on already covered material, the following points support the chromosomal theory of inheritance:

1. The number of chromosomes in somatic cells is constant.
2. Sexual identity is associated with particular chromosomes.

The following evidence would strengthen the theory's claims:

1. The inheritance of genes corresponding to the inheritance of chromosomes.
2. The transmission of particular chromosomes corresponding with the transmission of particular traits (i.e. traits other than sex).

Experimental Results

Key experiments were run by American biologist Thomas Hunt Morgan on a type of fly called *Drosophila*. These experiments ultimately confirmed that a gene controlling eye color resides on an X chromosome, which is also used for sex determination (this makes the eye color an X-linked trait). These results were obtained by observing the phenotypic ratios of eye color between males and females, as well as analyzing individuals that had undergone non-disjunction to confirm that eye color was ultimately controlled by genes on X chromosomes rather than an individual's sex.

Aneuploidy

Aneuploidy refers to an abnormal number of chromosomes in a cell (i.e. not an exact multiple of haploid number n). Individuals lacking a chromosome in a pair are known as monosomic, while individuals with an extra chromosome are known as trisomic.

Autosomal Aneuploidy

Autosomal aneuploidy tends to be lethal for most chromosomes. However, there are some exceptions, such as trisomy 21 in humans, which causes Down syndrome. As a general rule, autosomal trisomies can sometimes be viable, while autosomal monosomies cannot.

Sex Chromosome Aneuploidy

Aneuploidies in sex chromosomes tend to be more viable than in autosomal chromosomes. For aneuploidies of the X chromosome in particular, X-chromosome inactivation represses the expression of most genes on all but one X chromosome. However, X-chromosome aneuploidies commonly cause infertility, since X-chromosome reactivation occurs in the oogonia, which is normally necessary in order to ensure that all of the produced ovums can support development after fertilization (as opposed to half with only X-chromosome inactivation). X-chromosome reactivation also occurs in males with multiple X chromosomes, typically preventing them from producing sperm.

Meiotic Nondisjunction

Mistakes in chromosome segregation during meiosis can cause aneuploidies. Nondisjunction during meiosis I produces four cells with aneuploidies, while nondisjunction during meiosis II produces only two such cells. Studies suggest that meiotic nondisjunction may result from problems during recombination. In particular, a lack of chiasmata formation removes the mechanism for ensuring that maternal and paternal chromosomes go to opposite poles during anaphase I.

Mitotic Nondisjunction

Aneuploidies may also result from the failure of sister chromatids to separate during mitosis, producing trisomic and monosomic daughter cells. Other mistakes may also occur, such as chromosome loss, which produces one monosomic and one diploid cell. Aneuploid cells may survive and undergo further rounds of mitosis, potentially producing a mosaic organism with differing chromosome numbers in across different tissue. One such example is with gynandromorphic organisms, which are organisms that contain both male and female characteristics.

Another example is with Turner syndrome (XO); most affected individuals start off as XX zygotes, but lose an X chromosome early during embryonic development.

Euploidy

Euploid cells contain only complete sets of chromosomes, but might be monoploids (containing only 1 set of chromosomes) or polyploids (containing 3 or more sets of chromosomes).

Monoploidy

Monoploid organisms tend to develop by parthenogenesis, where an ovum develops without having been fertilized. Due to the mechanics of meiosis, monoploids are typically infertile. One potential benefit of monoploidy, often used by plant breeders, is that desirable recessive traits appear since there is no matching (potentially dominant) allele.

Triploidy

Triploids, which have three sets of chromosomes, can result from the union of gamete and diploid cells, and are almost always sterile. This sterility is often used by plant breeders, who may want to deliberately grow seedless fruit. This characteristic of sterility can be extended in general to polyploids with odd numbers of chromosome sets, since they cannot produce balanced gametes.

Tetraploidy

Tetraploids have four complete sets of chromosomes. Tetraploids can arise from the nondisjunction of all chromosomes during mitosis. Unlike triploids, tetraploids can be fertile, since they can produce balanced (diploid) gametes. This possibility of fertility can be extended in general to polyploids with even numbers of chromosome sets, since they may be able to produce balanced gametes.

Gene Linkage and Recombination

Some genes do not follow Mendel's law of independent assortment, but tend to travel together, exhibiting a behavior known as gene linkage. In such genes, parental classes of alleles tend to appear in offspring much more frequently than recombinant classes.

A common cause for gene linkage is that linked genes are on the same chromosome, in which case they are considered syntenic. These genes do not always travel together due to recombination, but tend to travel together in most cases, especially when they are close together on the same chromosome.

Testcrosses for Linkage Detection

Two-point (i.e. "two gene") testcrosses can be carried out to detect gene linkage. These tests work by breeding a parent that only contributes recessive alleles with a parent that is dihybrid for the traits to test. Should the resulting phenotypes not be in a 1:1:1:1 ratio, then gene linkage is likely occurring.

Observing Recombination

Recombination is a key reason why genes on the same chromosome do not always travel together. This can be confirmed by experiments by observing chiasmata, which form where gene recombination occurs.

Recombination Frequencies

Recombination Frequencies (RFs) concerning two genes can be used as an indicator for the distance between two genes on a chromosome. Units of measure along a chromosome include RF percentage points, centimorgans (cM) and map units (m.u.), where $1\% \text{ RF} = 1\text{cM} = 1\text{m.u.}$ Recombination frequencies tend to be 50% for genes on different chromosomes and can approach 50% for genes on the far apart on the same chromosome, but never exceed 50%.

Linkage Analysis

Determining whether two genes are linked required some statistical analysis. It is not sufficient to say that the genes did not assort in a perfect 1:1:1:1 ratio, since real-world experimentation, especially on small sample sizes, can produce deviations from theoretically expected values.

Chi-Square Test

A chi-square test can be used to determine the probability that a certain gene arrangement could have occurred by chance, assuming an expected 1:1:1:1 ratio (i.e. assume the linkage hypothesis is false (null hypothesis), and then determine whether this assumption matches the data).

$$\chi^2 = \sum \frac{(n_{obs} - n_{exp})^2}{n_{exp}}$$

The degrees of freedom (df) must also be determined, which is 1 less than the number of observed classes. For example, if the experiment contains parental and recombinant types, then the df is 1. Based on the df and the χ^2 value, the p-value (probability) of the null hypothesis can be determined. Generally speaking, p-values below 0.05 are considered a boundary between significance and non-significance; null hypotheses below 0.05 are typically rejected.

Gene Mapping

Gene mapping can be used to find the positions of genes on locations (i.e. loci) on chromosomes. These maps can be used to predict inheritance patterns by estimating recombination frequencies.

Two-Point Crosses

Two-point crosses involve determining the relative positions of two genes on a chromosome. While fairly simple to analyze, two-point crosses have many limitations, including that it may be hard to determine gene order for genes close together, and that actual map distances do not always add up accurately for genes far apart. For genes far apart, it is better to perform a series of two-point crosses using genes that lie between the two genes being analyzed, or to perform a three-point cross.

Three-Point Crosses

Three-point crosses involve determining the relative positions of three genes on a chromosome. The main advantage of three-point crosses over two-point crosses are that they allow for correcting mapping distance discrepancies caused by double crossovers. In the resulting progeny, parental classes will be the most frequent resulting genotypes. The next most frequent genotypes will result from single crossover recombinations between the two most distant genes (known as region I). The most frequent after that will result from single crossover recombinations between the two closest adjacent genes (known as region II). The two genotypes with the lowest frequency result from double crossovers. These double crossovers must be counted twice when calculating distances between the two outside genes.

Note that although three-point crosses are able to accurately estimate mapping distances between the second (i.e. middle) gene with the first and third gene, mapping distances between the first and third gene tend to be incorrect.

DNA

Evidence DNA is Genetic Material

In 1869, a substance later confirmed to be DNA (deoxyribonucleic acid) was extracted from human cells. A further experiment carried out in 1923, which used a Schiff reagent to stain DNA red, showed that DNA is stored almost exclusively in chromosomes.

Even though DNA was confirmed to exist in chromosomes, this was not strong evidence that it had anything to do with genes. Certain proteins stored in chromosomes, which can be much more complex than DNA, were considered by many scientists to be the building blocks of genes.

In 1923, Frederick Griffith carried out an experiment that showed that genetic material from dead bacteria cells could somehow be transmitted to live cells. This ability of a substance to change the genetic characteristics of another organism is known as transformation. These results were published in 1928, and Oswald Avery subsequently led a team that published findings in 1944 that this transforming principle was DNA.

Further evidence that DNA is genetic material was found in 1952, where Alfred Hershey and Martha Chase determined that viruses called phages infect bacteria by injecting DNA into cells.

Structure of DNA

DNA is a large polymer built from subunits known as nucleotides. Each nucleotide is composed of a deoxyribose sugar, a phosphate, and one of four nitrogenous bases (adenine (A), guanine (G), thymine (T), and cytosine (C)). A nucleoside consists of a base with a covalent attachment to the deoxyribose sugar, and the addition of a phosphate forms a nucleotide.

The double helix structure of DNA was first proposed by James Watson and Francis Crick in a 1953 paper. A piece of evidence for this theory is X-ray diffraction data, which showed that DNA is helical. Another piece of evidence is that A:T and C:G exist in nearly perfect 1:1 ratios, which occurs because two anti-parallel strands of DNA are held together by hydrogen bonding between these base pairs.

The DNA helix completes a full turn every 34 angstroms (an angstrom is 10^{-10}m), the diameter of the molecule is 20 angstroms, and the spacing between repeating units is 4.3 angstroms.

Alternative Forms of DNA

DNA molecules in a standard double-helix right spiral formation are known as B-form DNA. An alternative form of DNA is known as Z-form, where the helix spirals to the left and the

backbone has a zigzagging shape. While there is evidence that Z-form DNA may exist in living organisms, no clear biological role has yet been determined.

Eukaryotic nuclear chromosomes are all long, linear double helices. However, in some instances, such as with chromosomes of prokaryotic bacteria or the chromosomes of some organelles such as mitochondria, chromosomes can be circular.

Some viruses carry only a single stand of DNA, which can serve as a template for making a second strand once inside a host. Some viruses use RNA as their genetic material, which contains the sugar ribose instead of deoxyribose, uracil (U) as a base instead of thymine (T), and tends to be single-stranded and short.

DNA Replication

Models of Replication

The three main models of replication are:

- **Conservative:** One of two resulting helices consists only of original DNA strands, while the other consists of newly synthesized strands.
- **Semiconservative:** One strand of each resulting helix comes from the parent, other is synthesized.
- **Dispersive:** Both resulting helices carry strands of original DNA interspersed with newly synthesized DNA.

The Watson-Crick model of DNA replication involves semiconservative replication, which has been experimentally verified. One relevant experiment was carried out by Matthew Meselson and Franklin Stahl, which dealt with controlling the isotopic composition of nucleotides incorporated into daughter DNA strands. These isotopes varied in density, so the composition of the resulting DNA could be studied by spinning the DNA in a centrifuge, where the results were consistent with semiconservative replication.

Molecular Mechanism of Replication

DNA replication is a complex process that occurs at a specific stage in the cell cycle (S phase for eukaryotes). The process consists of two stages, namely initiation and elongation.

Initiation consists of preparing the double-helix for complementary base pairing. At first, a small region of the double helix is unwound, exposing short sequences of nucleotides in what is known as the origin of replication.

An initiator protein binds to the origin of replication, which attracts an enzyme known as DNA helicase. This enzyme catalyzes the unwinding of the double helix, forming what is known as a replication bubble, which is bounded by two unwound DNA strands called replication forks.

Formation of new DNA stands depends on an enzyme called DNA polymerase III. This enzyme acts according to the following rules:

1. Only unwound and single-stranded DNA can be copied.
2. Nucleotides can only be added to the end of a chain.
3. The DNA strand can only be built in a single direction (i.e. from 5' to 3').

To start off the chain, a short strand of RNA consisting of a few nucleotides is used as a primer, off of which a DNA strand can be built. This primer is synthesized by an enzyme known as primase.

Elongation consists of connecting nucleotides into a continuous strand of DNA. Once base-pairing nucleotides are formed from the unwound strands, they form a chain by the formation of phosphodiester bonds in a process known as polymerization. This polymerization is catalyzed by DNA polymerase III.

Since polymerization only takes place in a 5' to 3' direction, the two strands forming the replication bubble will be replicated through a slightly different mechanism. One strand is built in the direction the helix unwinds, which is known as the leading strand. The other strand's 5'-to-3' direction runs in the opposite direction of the strand unwinding, and is known as the lagging strand. The lagging strand is replicated using around 1000 small strands known as Okazaki fragments, which are later merged by an enzyme known as DNA polymerase I that replaces the many RNA primers with DNA.

Ensuring Integrity of Genetic Information

Since DNA is the sole repository for a vast amount of information concerning how a cell functions, the integrity of its information is of utmost importance. Mechanisms for ensuring integrity include:

- **Redundancy:** One strand of double-stranded DNA specifies the sequence of the other, which provides a mechanism for detecting errors.
- **Precision of cellular replication machinery:** Errors during replication are very rare in the first place.
- **Enzymes that repair chemical damage to DNA:** Cells contain enzymes that can repair various types of damage to DNA.

Replication of Circular Chromosome

A replication bubble in circular chromosomes may cause excessive torsion between intertwined strands as it expands. Topoisomerase mitigates this issue by unwinding DNA strands and separating the generated molecules.

Replication of Long Chromosomes

DNA replication in eukaryotes would take too long if it only involved a single replication bubble. Instead, DNA replication occurs at many points along a chromosome, forming replicons that are later joined.

Replication of Linear Chromosomes

DNA replication only occurs in a single direction (i.e. from 5' to 3'). Thus, there can be issues in replicating DNA at the 5' ends of chromosomes. This issue is mitigated by telomeres, which consist of repeating sequences of DNA on the end of DNA molecules that do not encode proteins. A ribonucleoprotein called telomerase can add additional telomere sequences to the end of a DNA molecule, delaying eventual cell death. However, most somatic cells do not express telomerase, so their number of replications is limited. Increased levels of telomerase in somatic cells are known to increase the risk of cancer, so adding telomerase is not a simple “fountain of youth”.

Genetic Code

Gene expression refers to the flow of genetic information from DNA or RNA to polypeptides (chains of amino acids). Like nucleotides are the building blocks of DNA, amino acids are the building blocks of proteins. There are 20 amino acids that are encoded by DNA or RNA; rarer amino acids undergo further modification after synthesis. Amino acids are encoded by nucleotide triplets called codons. There are 64 possible codons in total, 61 of which represent the 20 amino acids, and 3 of which signify stops.

“Cracking” the Genetic Code

Yanofsky’s Experiments

In the 1960s, Charles Yanofsky determined the following:

- **Nucleotide sequences are colinear with amino acid sequences.** The evidence is that changing the order of DNA mutations mapped to positions of amino acid substitutions.
- **Codons are composed of multiple nucleotides.** The evidence is that altering different nucleotide pairs may affect the same amino acid.
- **Each nucleotide is only part of one codon.** The evidence is that point mutations of a single nucleotide pair affect only a single amino acid.

Triplet Codons

In research published in 1961, Francis Crick and Sydney Brenner provided evidence that DNA was composed of trios of nucleotides with a single starting point called reading frames. Changes that alter the grouping of nucleotides into codons are known as frame-shift mutations, and almost always abolish the function of the generated polypeptide. Crick and Brenner demonstrated that combining three such insertion or deletion mutations might not result in such mutations. These mutations can perform what is known as intragenic suppression, which is the restoration of gene function by one mutation canceling the other in the same gene. The relative commonality of intragenic suppression implies that most amino acids are specified by multiple codons.

Mapping Codons to Amino Acids

In the 1950s, researchers detected that although most DNA is stored in the cell nucleus, protein synthesis takes place in a cell’s cytoplasm. This implies that some intermediate DNA molecule must transport the genetic information into the cytoplasm, later shown to be messenger RNA (mRNA). Knowledge of mRNA helped lead to two breakthroughs that

helped crack the genetic code. The first was that mRNA combined with certain other substances could be used to synthesize polypeptides in test tubes (in vitro). The second was development of techniques enabling the synthesis of some strands of artificial mRNA.

Nonsense Codons

There are three different triplets (UAA, UAG, and UGA) that serve as stop (or nonsense) codons. A mutation that changes an amino-acid producing codon into a stop codon is known as a nonsense mutation.

Universality

The genetic code is nearly universal, with genetic information often interchangeable between different species. Although there are a handful of exceptions, codon encoding of polypeptides is almost always the same.

Transcription

Transcription is the process by which parts of a single DNA strand are copied into an RNA strand. This is done by polymerization of ribonucleotides guided by complementary base pairing, producing a single strand that is complementary to the template strand in the copied gene. In eukaryotes, this primary transcript undergoes further processing in the nucleus before being used for protein synthesis.

Steps of Transcription

Transcription follows the steps below:

Initiation

An enzyme called RNA polymerase binds to promoters, which are specialized DNA sequences that represent the beginning of a gene where transcription will start. After binding, the RNA polymerase starts to unwind part of the double helix, exposing unpaired bases to form what is known as an open promoter complex.

Elongation

The RNA polymerase separates from the promoter and moves along the chromosome, unwinding part of the double helix to produce what is called a transcription bubble. The RNA chain builds up ribonucleotides as the RNA polymerase moves along the chromosome, which has complementary bases to the template strand except for uracil being used instead of thymine. These base pairs are built up in the 5'-to-3' direction. Some genes can undergo transcription by several RNA polymerases simultaneously.

Termination

Terminators are sequences of RNA that signal the end of transcription. These terminating sequences are transcribed from specific DNA sequences. Terminators can be intrinsic, in which case transcription terminates on its own, or extrinsic, in which case additional proteins (in particular a polypeptide known as rho) is needed to bring about termination.

Terminators often form hairpin loops in which nucleotides within the RNA strand pair with complementary nucleotides.

Addition of Cap and Tail

In eukaryotes, after the primary strand is copied, a special capping enzyme adds a backwards guanine triphosphate to the 5' end of the RNA strand. An enzyme known as methyl transferase then adds methyl groups to the backwards G and one or more of the succeeding nucleotides. This forms what is known as a methylated cap, and is essential for the efficient translation of RNA.

A tail is also added in most eukaryotes, which is known as a poly-A-tail. This usually consists of 100-200 adenosines. In eukaryotes, the interaction between the cap and tail forms the RNA strand into a loop, which helps stabilize the RNA strand (prevents degradation) and aids in the efficiency of translation.

Exons and Introns

Eukaryotes often feature far longer primary transcripts than mRNA strands. The reason for this is that not all sequences of a primary strand encode protein product. Exons are sequences that encode protein product, and are found in mature mRNA. Introns do not encode protein product, and are removed from the primary transcript.

RNA Splicing

RNA splicing is the process by which introns are removed from a primary transcript and successive exons are joined. RNA splicing involves three short sequences within each intron, namely a splice donor, splice acceptor, and branch site. In each intron, a cut occurs at its splice donor, and the newly formed end attaches to an A nucleotide within the branch site. The splice acceptor is also cut, after which the separated intron is degraded. Splicing is usually carried out with the help of a complex known as a spliceosome. However, some RNA transcripts are self-splicing.

Purpose of Introns

The exact reason why introns exist has not been determined. One hypothesis is that introns allow for exons to be reshuffled to produce new genes, known as alternative splicing. This may occur, for instance, if splicing occurs between the splice donor site of one intron and the splice acceptor site of another. Trans-splicing is a special form of alternative splicing in which exons from two different transcripts are joined.

Translation

Translation is the process in which nucleotide sequences in mRNA direct the assembly amino acids to produce proteins. Translation takes place on ribosomes, which coordinate the movement of transfer RNAs (tRNAs), which can carry specific amino acids, with mRNAs, which carry genetic instructions.

Transfer RNA

Structure

tRNAs serve as adapter molecules that mediate the transfer of information from nucleic acid to protein. tRNAs are short, single-stranded RNA molecules 75-94 nucleotides in length. Several of the nucleotides in tRNAs contain modified bases produced by chemical alternations of the principal A, G, C, and U nucleotides. Each tRNA encodes one particular amino acid, so cells must have at least 1 tRNA for each of the amino acids specified by the genetic code.

tRNA carries an anticodon, which consists of three nucleotides complementary to an mRNA codon specifying the tRNA's amino acid. Base pairing between an mRNA codon and a tRNA anticodon determines where an amino acid becomes incorporated into a growing polypeptide chain. At the other end of the tRNA molecule, enzymes known as aminoacyl-tRNA synthases catalyze the attachment of a tRNA to its amino acid. A tRNA coupled to its amino acid is known as charged tRNA.

Wobble

Each tRNA is associated with a particular amino acid. However, cells do not necessarily carry tRNAs with anticodons complementary to all 61 possible codon triplets. Some tRNAs recognize more than one amino acid for the codon they carry; this flexibility in base-pairing is known as wobble.

Ribosomes

Ribosomes are the sites of protein synthesis, and are complex structures composed of protein and RNA.

Structure

Ribosomes consist of ribosomal RNA (rRNA) and ribosomes proteins. Each ribosome consists of a large and small subunit, which start off as separate entities and come together once translation begins. The larger subunit contributes an enzyme known as peptidyl transferase,

which catalyzes the formation of peptide bonds joining adjacent amino acids. Both subunits contribute to three distinct tRNA-binding sites known as the aminoacyl (A) site, peptidyl (P) site, and exit (E) site.

Translation Process

Ribosomes and charged tRNAs collaborate to translate mRNAs into polypeptides. This follows a process with the following stages:

Initiation

A special signal is used to indicate where along mRNA translation should begin. In prokaryotes, this is known as the ribosome-binding site, and consists of a purine-rich (usually AGGAGG) sequence known as the Shine-Dalgarno sequence, and an AUG initiation codon. This AUG codon signals the initiation of translation by adding a tRNA carrying formyl-methionine (fMet), which is a derivative of the amino acid methionine. The fMet tRNA is placed in the P site of the ribosome.

In eukaryotes, the small ribosomal subunit binds the to mRNA and then migrates to the initiation site. The initiation sequence is usually AUG, but in some mammals, also depends on the surrounding sequences. Also, initiator tRNA carries methionine (Met) instead of fMet.

Elongation

Elongation involves adding amino acids to a growing polypeptide. Proteins known as elongation factors place the appropriate tRNA into the A site of the ribosome. Peptidyl transferase catalyzes the formation of peptide bonds between the carboxyl terminus of fMet and amino terminus of the second amino acid. Free tRNA is moved to the E site and released.

Termination

When a stop codon is encountered in the mRNA, proteins called release factors recognize these codons and bring polypeptide synthesis to a halt. The polypeptide is released, and the rRNA, tRNA, and ribosomal subunits dissociate from one another.

Prokaryotic vs Eukaryotic Differences

A few other differences in translation for prokaryotes and eukaryotes include:

- Prokaryotes undergo replication, transcription, and translation in cytoplasm (no nucleus), while eukaryotes undergo replication and transcription in the nucleus and then undergo translation in cytoplasm.
- In prokaryotes, mRNA is short-lived, and translation may begin before transcription has completely finished. In eukaryotes, mRNA must be complete before it can travel out of the nucleus, and tends to be much more long-lived.

Post-translational Processing

Post-translational processing refers to changes to proteins after translation. Examples of post-translational processing include the removal of amino acids, generating several smaller polypeptides from a single chain, or adding chemical groups to an existing polypeptide.

Prokaryotic Gene Expression

Catabolic and Anabolic Pathways

Catabolic pathways are metabolic pathways in which complex molecules are broken down into simpler molecules for use in a cell (e.g. breaking down sugars). Anabolic pathways involve the production of end-product molecules (e.g. amino acids).

Catabolic pathways require inducible regulation, meaning that they should only be turned on when the molecules to be broken down are present. Anabolic pathways require repressible regulation, meaning that they should only be turned on when the cell lacks enough of the end product produced.

E. Coli's Utilization of Sugar Lactose

As a case study, this section covers the catabolic pathway that allows E. Coli to utilize sugar lactose as a source of carbon and energy.

Catabolic Proteins

Two key proteins involved in this catabolic pathway are lactose permease and beta-galactosidase. Lactose permease transports lactose into a cell, while beta-galactosidase splits the lactose into galactose and glucose. The presence of lactose induces the expression of genes required for lactose utilization; the addition of lactose induces a 1000-fold increase in the production of these two proteins.

Operon Theory Overview

Much of the early work done on the gene expression of E. Coli was done in the 1950s by Jacques Monod and Francois Jacob. They introduced the operon theory, which suggests that a single signal can regulate the expression of several genes clustered together on a chromosome. The reasoning behind this is that the genes clustered together would be transcribed together (forming polycistronic mRNA, which is one mRNA strand with multiple transcribed genes), so anything that regulates transcription affects all the genes in this cluster.

In E. Coli, there are three main genes encoding proteins for lactose utilization:

- **lac Z**: Encodes beta-galactosidase.
- **lac Y**: Encodes permease.
- **lac A**: Encodes transacetylase (adds acetyl group to lactose, not required for lactose breakdown).

Operons have a single operator located before the first gene. Molecules that interact with the operon are the repressor, which binds to the operon's operator, and the inducer, which binds to the repressor to prevent it from binding to the operator (i.e. negative control). In the operon under study, beta-galactosidase also serves to produce the inducer allolactose from lactose.

Evidence for Repressor Protein

Mutations in another gene known as *lacI* result in mutants that synthesize permease and beta-galactosidase even in the absence of lactose. These mutations that result in the production of a protein regardless of environmental conditions are known as constitutive mutations. This information led to the hypothesis that *lacI* encodes a negative regulator (i.e. repressor) that binds at the operator site.

Repressor Mutants

The *lac* repressor induces a loop in the DNA that prevents RNA polymerase from binding. This loop is not formed through covalent bonding, but rather due to the laws of thermodynamics. Mutations that result in an inactive repressor typically alter the repressor's shape to prevent it from binding to the operator.

Some mutations of the *lacI* gene can erase repressor activity (e.g. *lacI*⁻ allele). There are also super-repressor mutations, where an operator is permanently repressed even when lactose is present, since a mutation in the repressor prevents it from interacting with the inducer (e.g. *lacI*^S allele).

Trans vs Cis Action

Elements acting in trans can diffuse through cytoplasm and act on any DNA molecule in a cell. Cis elements only influence their own DNA molecule (e.g. operators).

Proteins act in trans, so repressor proteins can influence other operons. Because of this behavior, operators are rarely truly on or off, but are expressed in a level dependent on the repressor and inducer concentrations.

Eukaryotic Gene Expression

Regulation Through Transcription

Like in prokaryotes, eukaryotes can have their genes regulated through transcription. However, these genes are not organized in operons. Additional levels of complexity are introduced by some eukaryote-specific traits, including:

- Chromatin structure can make DNA unavailable to transcription machinery.
- Additional RNA processing events may occur.
- Transcription occurs in the nucleus, while translation occurs in the cytoplasm.
- Gene regulation controls cellular differentiation into many kinds of specialized cells.

RNA Polymerase

In eukaryotes, three kinds of RNA polymerase transcribe genes. These are:

- **RNA polymerase I:** Transcribes genes that encode the major RNA components of ribosomes.
- **RNA polymerase II:** Transcribes protein-encoding genes.
- **RNA polymerase III:** Transcribes genes that encode tRNA and some other small RNA molecules.

Regulatory Regions

The kinds of DNA sequences essential to cell regulation in eukaryotes are promoters and enhancers. The core promoter is located just upstream of where transcription begins, and usually contains a so-called “TATA box” (TATA(A or T)A(A or T) sequence. RNA polymerase II attaching to the TATA box allows for a basal (i.e. low) level of transcription.

Enhancers can be far away upstream from the core promoter, and vary greatly in size. The binding of proteins to enhancers can augment or repress basal levels of transcription, depending on whether an activator or repressor binds to the enhancer.

Basal Factors

Basal factors assist in the binding of RNA polymerase II to the promoter. The key component of the basal factor complex is the TATA-box-binding-protein, also known as the TBP. TBPs in turn attract other proteins known as TBP-association factors, also known as TAFs, which ultimately allow for RNA polymerase II to initiate transcription.

Activators

Activators bind to enhancer sequences, and can increase transcription far above (i.e. hundreds of times) the basal level. Activators increase transcription by doing one of the following:

- Recruiting basal factors and RNA polymerase II to core promoter sequences through direct interaction with the promoter sequence. In order for direct interaction between a distant promoter and enhancer, the DNA must form loops known as topologically associating domains (TADs).
- Recruiting coactivators, which in turn open the local chromatin structure to allow for transcription.

Most eukaryotic activators must form structures called dimers to function. Dimers are structures where the activators form a bond with another molecule. In homodimers, activators form dimers with one another. In heterodimers, activators form dimers with a non-identical molecule.

Repressors

Repressors can diminish the transcriptional activity of a gene. Most repressors work by attracting corepressor proteins to enhancers, where the corepressor can do one of the following:

- Interact directly with the promoter to prevent it from attaching to the basal complex.
- Act as enzymes that close chromatin to prevent transcription.

Chromatin Structure

DNA in eukaryotes is packaged into chromatin. The basic structural unit of chromatin is a nucleosome, which consists of a ball of histone proteins. Changes in chromatin structure can result in altered gene expression. For example, heterochromatin is a dense form of chromatin that prevents gene transcription.

Chromatin Remodeling

Chromatin remodeling involves repositioning nucleosomes to expose certain DNA sequences. Proteins that carry out this process are known as chromatin remodeling complexes. As mentioned above, certain coactivators function by unwinding promoter sequences to make them accessible.

Histone Modification

Some histone tails extend beyond the nucleosome, and are able to affect the overall chromatin structure. An example is acetylation, which reduces the attraction between adjacent nucleosomes, making the DNA easier to unwind.

DNA Methylation

DNA methylation involves methyl groups joining to CpG sequences on a DNA strand (p stands for phosphate). Promoters are often rich in methylated sequences, and methylated promoters are effectively turned off, preventing transcription.

DNA methylation patterns may be inherited from a parent. In this case, the expression of an allele may depend on which parent it originates from, which is a phenomenon known as gene imprinting. In general, inheritable changes to gene expression that do not involve sequence changes are known as epigenetic changes.

Restriction Enzymes

Restriction enzymes (also known as restriction endonucleases) are enzymes that remove portions of DNA by severing the phosphate bonds between nucleotides at a specific sequence of bases. The fragments generated by restrictive enzymes are known as restrictive fragments, and the act of cutting is known as digestion.

The DNA sequences where cutting occurs are known as recognition sequences, and each restriction enzyme cuts a specific recognition sequence. These sequences are palindromic, meaning that they are the same from the 5' to 3' end as the 3' to 5' end.

Use as a Defense Mechanism

Restriction enzymes originate in bacteria, which use them as a defense mechanism against viruses. These enzymes can restrict what type of DNA can exist in their host by digesting foreign sequences, after which they are further degraded. Bacteria can shield their own recognition sites from digestion through the addition of methyl groups.

Cutting

Restriction enzymes cut through recognition sequences in two main ways:

- Cutting through both strands of DNA in the same location, producing what are known as blunt ends.
- Cutting through opposite strands of DNA in slightly different locations, producing overhanging strands that are known as sticky ends.

Restriction Mapping

Restriction maps describe where certain recognition sequences occur in a sequence of DNA. They can be formed by introducing a restriction enzyme to a DNA sequence and observing where cuts occur. To determine cutting locations, the restriction fragments are ordered by their length, which is done using a process known as electrophoresis. Electrophoresis involves charged particles in an electric field moving through a thick gel matrix; large fragments are less likely to find a pore small enough to squeeze through, which slows their movement and separates them from smaller fragments.

Molecular Cloning

Molecular cloning is the process of using living cells to make many replicas of a fragment of DNA. Molecular cloning is useful for amplifying a sequence of DNA (i.e. making enough copies of it so that it becomes easier to study).

Cloning Vectors

The end goal of molecular cloning is to have a host cell replicate a DNA fragment. However, simply inserting an isolated DNA fragment into a cell would result in the fragment's degradation.

Instead, the DNA fragment to be replicated is first added to a cloning vector, which is a piece of DNA that can be stably maintained in an organism. Plasmids are often used as cloning vectors. A restriction enzyme can be used to make a cut in the vector, which can attract the DNA fragment if the “sticky ends” of the cut match.

Host Cells

Under certain conditions, cells can take up a cloning vector, after which they undergo transformation and replicate the vector. However, most cells exposed to a vector do not absorb it. A common technique for isolating cells that took up the vector from ones that did not is placing them in an environment where some gene contained in the plasmid is necessary for survival.

Population Genetics

Key Terms

A population is a group of individuals of the same species that live in the same geographical area and are capable of interbreeding. A gene pool is the collection of alleles carried by individuals within a population. A few common types of frequencies analyzed are:

- **Phenotype frequency:** The proportion of individuals in a population of a particular phenotype.
- **Genotype frequency:** The proportion of individuals in a population of a particular genotype.
- **Allele frequency:** The proportion of genes in a gene pool of a particular allele.

Hardy-Weinberg Principle

The Hardy-Weinberg principle describes the relationship between genotype and allele frequencies within a population. The Hardy-Weinberg principle states that if certain assumptions are met, these allele and genotype frequencies remain constant between generations.

Hardy-Weinberg Assumptions

Some simplifying assumptions underlying the Hardy-Weinberg principle include:

1. The population is a very large number (i.e. can be approximated as infinite) of diploid individuals with equal access to mating.
2. Individuals mate at random; genotypes do not play a role in mate selection.
3. No new mutations occur in the gene pool.
4. No migration of individuals into or out of the population takes place.
5. Natural selection does not occur; there are no genotype-dependent differences in rates of surviving to reproductive age and reproducing.

Populations that satisfy these assumptions are said to be at Hardy-Weinberg equilibrium. In practice, populations do not fit these assumptions perfectly; however, the Hardy-Weinberg principle is still effective for providing estimates of genotype and allele frequencies over a limited number of generations.

Hardy-Weinberg Proportions

Let A_1, A_2, \dots, A_N represent the frequencies of alleles of a gene A . The genotype frequencies that result from a cross between two random individuals can be found by squaring the sum of the alleles. Since $A_1 + A_2 + \dots + A_N = 1$, $(A_1 + A_2 + \dots + A_N)^2 = 1$.

As an example, assume that a gene has an allele frequency of 70% for a p allele and a 30% frequency for a q allele. $(p + q)^2 = p^2 + 2pq + q^2$. Thus, the chance of an individual being homozygous for the p allele is $0.7^2 = 0.49$, the chance of an individual being heterozygous is $2(0.7)(0.3) = 0.42$, and the chance of an individual being homozygous for the q allele is $(0.3)^2 = 0.09$.

Reaching Equilibrium

If a population's genotypic ratios are not currently at Hardy-Weinberg equilibrium (e.g. there are only homozygotes), a single generation of breeding is sufficient for autosomal genes to attain this equilibrium. The situation is more complex for X-linked genes, where several generations of breeding are necessary to approach equilibrium.

Changes from Equilibrium

Genetic Drift

Genetic drift is a change in the frequency of an allele as a result of random sampling of individuals. Small populations are more susceptible to genetic drift than larger ones.

In practice, genetic drift is commonly observed through the founder effect, where a small group of individuals establishes a new population. Genetic drift is also commonly observed in population bottlenecks, which occur when a population size is rapidly reduced.

Natural Selection

In practice, the genotype of an individual often does affect its ability to survive to reproductive age and reproduce, contradicting one of the Hardy-Weinberg assumptions. The relative ability of an individual to survive and transmit its genes to the next generation is known as fitness. Natural selection is the process that progressively eliminates individuals of low fitness and chooses individuals of high fitness to survive and become parents of the next generation.

Note that fitness is often dependent on the environment; a genotype that is relatively fit in one environment may be unfit in a different one.

Genetic Analysis

Genomic and cDNA Libraries

Genomic libraries consist of a single copy of every DNA fragment in an entire genome, each of which is placed into a suitable vector for storage. The number of fragments is known as the genetic equivalent of the library.

cDNA libraries only consist of transcribed sections of DNA. These DNA sequences are produced through reverse transcription of mRNA sequences.

DNA Probes

DNA probes are purified single-stranded sections of DNA that can be used to identify complementary DNA sequences by hybridizing (annealing) with them. For successful hybridization, the two DNA strands being combined need to be of a sufficient length (i.e. around 50-100 bp) and at least 80% of their sequence should be complementary. For visualization, DNA probes are typically labeled with fluorescent dyes or radioactive isotopes.

DNA probes can be constructed through chemical synthesis. These DNA probes can be built nucleotide-by-nucleotide to produce a short DNA sequence known as an oligonucleotide. Alternatively, a previously cloned fragment can be used as a probe.

Polymerase Chain Reaction

Polymerase chain reaction (PCR) is a technique used to rapidly make copies of DNA segments. First introduced by Kary Mullis in 1985, PCR is widely used as a faster and more flexible alternative to cloning.

To make copies of a particular sequence of DNA, two primers are used that are complementary to that region's ends. Then, these primers, DNA polymerase, and the template DNA are mixed together. The temperature is increased to unwind the template DNA into single strands. After this, the temperature is lowered, allowing the primers to base pair with their complementary regions. The temperature is then raised again to allow DNA polymerase to replicate the DNA of the region of interest. This effectively doubles the amount of copies of the region of interest; running this reaction iteratively allows for an exponential growth in the total number of copies.

DNA Sequencing

DNA sequencing involves determining the exact order of nucleotides in a DNA molecule. The main method for DNA sequencing still employed today was developed in the mid-1970s

by Fred Sanger.

Sanger sequencing involves combining a template to be sequenced with deoxyribonucleotide triphosphates necessary for adding base pairs (dATP, dCTP, dGTP, dTTP), a primer that is complementary to part of the template, and DNA polymerase. Along with the deoxyribonucleotide triphosphates, smaller concentrations of dideoxyribonucleotide triphosphates are added. These function similarly to deoxyribonucleotide triphosphates, but do not have a 3' end on which the constructed strand can further expand.

The result is a set of DNA strands of varying length with a dideoxyribonucleotide triphosphate at their end. This dideoxyribonucleotide triphosphate can be identified by dyeing the triphosphate for different base pairs different colors. Electrophoresis can be used to sort the generated strands by length, and their color can be used to determine their last nucleotide. Running this process many times on a section of DNA can produce a complete sequencing.

Mutation

Mutations are heritable changes in base sequences that modify the content of DNA.

Classes of Mutations

Mutations may be classified by whether they involve wild-type or mutant alleles:

- **Forward mutations** change a wild-type (i.e. occurring at greater than 1%) allele to a mutant form.
- **Reverse mutations** revert a mutant allele back to a wild-type allele.

Mutations may also be classified by how they alter DNA on a molecular level.

- **Substitutions** occur when a base pair is replaced with one of the three other bases. In transitions, purines are swapped with purines and pyrimidines are swapped with pyrimidines. In transversions, purines are swapped with pyrimidines, or vice versa.
- **Deletions** occur when a block of nucleotide pairs is lost.
- **Insertions** occur when a block of nucleotide pairs is added.

Rates of Spontaneous Mutation

Spontaneous mutations tend to occur at a very low rate. This rate varies by gene and species, with the average being around $2 - 12 \cdot 10^6$ per gene per generation. The majority of mutations do not occur in genes, and have no phenotypic effect.

Multicellular organisms tend to have higher rates of mutation than unicellular ones. This is because multicellular organisms pass on hereditary information through germ cells, which typically take multiple rounds of cell division to form, and there is a chance of a mutation at each round of division. Also, since sperm require more rounds of cell division to form than eggs, they tend to carry more mutations.

Causes of Mutations

Spontaneous mutations arise as a result of random events. This was demonstrated in 1943 by the Luria-Delbruck experiment, where bacteria colonies were allowed to grow and then contaminated with a phage. The result was that the number of surviving bacteria between colonies varied greatly. This led to the conclusion that resistance to the phages is not a result of active adaptation once the phages attack a colony, but rather a result of random mutations that occur before the phages are introduced.

Molecular Mechanisms of Mutation

Changes to DNA occur quite frequently, but are usually corrected by enzymatic systems for DNA repair. Mutagens are physical or chemical agents such as radiation that raise the frequency of mutations above their spontaneous rate.

Molecular processes that can change the information stored in DNA include:

- **Depurination:** Involves the removal of a purine base, and can result in a random base being added during DNA replication.
- **Deamination:** Involves the removal of an amino group than can change cytosine to uracil, and can result in transition mutations.
- **Oxidation:** Involves oxidative damage, and can result in transversion mutations.

Other processes that can result in DNA mutation include:

- X-rays causing deletions by breaking parts of the DNA sugar-phosphate backbone.
- UV rays causing adjacent T bases to form dimers, which can prevent replication.
- Unequal crossing over, resulting in one chromosome with duplicate information and another with deleted information.

Repairing Damaged DNA

Base Excision Repair

Base excision repair involves adding enzymes called DNA glycosylases to cleave an altered base from the sugar of its nucleotide. Different DNA glycosylases recognize different bases. DNA polymerase can fill in this gap based on the opposite strand, and DNA ligase can seal the formed gap.

Nucleotide Excision Repair

Nucleotide excision repair is used when there is no DNA glycosylase to recognize the problem base sequence. Nucleotide exclusion repair relies on enzyme complexes to detect and release damaged fragments, which are then filled in by DNA polymerase and sealed with DNA ligase.

Repairing Mistakes During DNA Replication

Mistakes during DNA replication are extremely rare, occurring once per around 10^9 base pairs. This is largely because of the following repair mechanisms.

- **DNA polymerase mismatch repair.** DNA polymerase is able to recognize and remove mispaired bases.

- **Methyl-directed mismatch repair.** Some bacterial recognition and repair systems rely on tagging parental strands with methyl groups. This allows for the correction of some errors by forming a nick between the error and the not-yet methylated tag just after replication, after which the errored DNA on the non-methylated strand can be removed. Eukaryotic cells follow a similar repair process, but their “tagging” mechanism is not well understood.

Cancer Genetics

Cancer refers to a variety of diseases characterized by uncontrolled cell growth and division. While cancer is not inherited, some inherited mutations can increase the risk of cancer.

Characteristics of Cancer Cells

A few common controls for cell division include:

- Division only induced by growth factors (e.g. hormones).
- Division stopping when cells come into contact with other cells (known as contact inhibition).
- Division being controlled by gap junctions, which are small pores in the membranes of adjacent cells that allow for the transfer of molecules controlling cell growth.
- Apoptosis (i.e. programmed cell death), which is activated by the expression of genes in a cell starved of growth factor or a cell with heavily damaged DNA.

Cancer cells sometimes create their own growth stimuli (autocrine stimulation), or sometimes lose gap junctions or gap junctions. Also, cancer cells often become “immortal” through the generation of large amounts of telomerase, which can prevent apoptosis.

Genetic Basis of Cancer

The formation of a cancer involves normal cells being transformed into cancer cells, which is a process known as oncogenesis. The tumor that develops is a mass of clones of these cells. Generally speaking, multiple mutations must occur in a cell to make it cancerous. Some environmental mutagens, such as arsenic, can increase the risk of cancer. Since cancer is usually caused by multiple mutations accumulated over time, cancer rates also increase with age.

Tumor-Suppressor Genes

Tumor-suppressor genes implement a G1-to-S checkpoint that prevents cells from going into the S phase with damaged DNA. Cells can become cancerous when both copies of a tumor-suppressor gene are mutated and fail to function (i.e. mutant alleles acts recessively).

TP53

The TP53 gene produces the p53 protein, which in turn turns on transcription for a cyclin-dependent kinase (CDK) inhibitor known as p21. Active CDK commits a cell to the S phase

of the cell cycle by inducing promoters of genes needed for DNA synthesis. p53 also turns on the expression of some genes encoding DNA repair enzymes. If a cell is too damaged for repair, the p53 proteins initiates apoptosis. A large (50%) percentage of cancers are associated with mutations of the TP53 gene.

RB

The retinoblastoma (RB) gene produces a protein called Rb that inhibits a transcription factor known as E2F. Some CDK complexes, which can be inhibited by p21, phosphorylate Rb to prevent it from inhibiting E2F. Thus, Rb is also necessary, along with p53, to control cell growth.

BRCA1 and BRCA2

The BRCA1 and BRCA2 genes produce proteins that are part of a surveillance system for repairing DNA breaks. Mutations to these genes may result in damaged DNA being replicated.

Oncogenes

Cells can become cancerous when mutations change proto-oncogenes to oncogenes. Some oncogenes can also be transmitted by viruses. Oncogenes act dominantly, such as by inactivating tumor-suppressing proteins.