Genetic Variation I:

MUTATION
The genetic substrate for natural selection

AND the Raw Material for Evolution

Dr. Carol E. Lee, University of Wisconsin

Copyright ©2020; do not upload without permission
OUTLINE of Next Three Lectures: The Substrate for Natural Selection

(1) Reduction of Variation: Genetic Drift in finite populations → (1) reduction in allelic diversity (2) reduction in heterozygosity (Inbreeding) (Last Time)

(2) Sources of Allelic Variation: Mutations

(3) Sources of Genotypic Variation: Sex (Meiosis)

(4) Heritable variation changes in gene expression without changes in the genetic code: Epigenetic Inheritance
Genetic Variation

• If there is **no genetic variation** neither genetic drift nor natural selection would be able to change allele frequencies, because there would be nothing to change.

• Natural Selection requires **genetic variation** upon which it could act.

• So, I’m going to talk about **genetic variation** today, to prepare you for the lectures on Natural Selection.
Sources of Genetic Variation

• **Mutations (change in the genetic code)** → new alleles and/or new genes:
  - Nucleotide substitutions, insertions, deletions → new alleles
  - Gene duplications or deletions → new genes
  - Exon Shuffling → new genes
  - **Horizontal gene transfer** (not always considered “mutation”) → new genes
  - Chromosomal duplications or deletions
  - Deletions of large chromosomal regions
  - Chromosomal inversions
  - Whole Genome Duplications

• **Sex:** No novel alleles, only **novel genotypes:**
  - **Genetic Recombination:** Shuffling of combinations of alleles along a chromosome
  - **Random Mating:** Shuffling of combinations of haploid chromosomes into new genotypes
This Lecture: Mutation
Types of Mutations
Types of Mutations

- **At the Nucleotide Level (Point mutations):**
  - Single nucleotide substitutions (transitions, transversions)
  - Insertion (nucleotide insertion)
  - Deletion (nucleotide deletion)

- **At the “Gene” Level:**
  - Gene Insertions (Gene Duplications, transposons, horizontal gene transfer)
  - Gene Deletions (pseudogenization, transposons)
  - Exon Shuffling

- **At the Chromosome Level:**
  - Chromosome duplications, deletions, inversions, fusions

- **At the Genome Level:**
  - autopolyplaidization
  - allopolyplaidization
Eukaryotic gene.

DNA
- Promoter
- Terminator

Pre-mRNA
- Cap
- Leader
- Exon
- Intron
- Exon
- Intron
- Exon
- Poly(A) tail

mRNA
- Protein-coding sequence

RNA-coding sequence
- Transcription
- RNA processing: introns removed

Translation

Polypeptide
Within functional coding regions of the genome, mutations could have very different effects depending on where they occur

**STRUCTURAL**: changes in the actual coding region of the gene

- Primary: Amino Acid composition (Amino Acid substitutions)
- Secondary, Tertiary, Quaternary structure

**REGULATORY**: changes in gene regulation

- Gene expression (transcription, RNA processing, translation, etc)
EXAMPLES: Sources of Genetic Variation (types of mutations)

- **Point mutations**
  = single nucleotide change
  - Single nucleotide substitutions (transitions, transversions)
  - Insertion (nucleotide insertion)
  - Deletion (nucleotide deletion)
• **Point mutations**

- **DNA replication error during Mitosis or Meiosis** (e.g. DNA, RNA polymerases, reverse transcriptase)
- **Error in repair of sites damaged by mutagens** (e.g. UV light, chemicals)
A Nucleotide

Phosphate group

Nitrogenous base

5-Carbon sugar
Transitions are more common than transversions; that is, purines are more likely to mutate to purines, and pyrimidines to pyrimidines (transitions).

The leading hypothesis is that because transitions are mutations between nucleotides of similar structure, they cause less disruption of the DNA helical structure and are less detectable by DNA polymerase or mismatch repair enzymes.
Complementary Base Pairs

Sugar-phosphate backbone

Adenine (A) and Thymine (T) form hydrogen bonds.

Guanine (G) and Cytosine (C) form hydrogen bonds.

Figure 5-1cd Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.
“The Central Dogma” of Molecular Biology

Information flow

<table>
<thead>
<tr>
<th>DNA</th>
<th>C</th>
<th>A</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>mRNA</td>
<td>G</td>
<td>U</td>
<td>U</td>
<td>G</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>Protein</td>
<td>Valine</td>
<td>Alanine</td>
<td>Glycine</td>
<td>Cysteine</td>
<td>Serine</td>
<td></td>
</tr>
</tbody>
</table>

Francis Crick (1958)
<table>
<thead>
<tr>
<th>First base</th>
<th>Second base</th>
<th>Amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU</td>
<td>U</td>
<td>Phenylalanine</td>
</tr>
<tr>
<td>UUC</td>
<td>C</td>
<td>Serine</td>
</tr>
<tr>
<td>UUA</td>
<td>U</td>
<td>Tyrosine</td>
</tr>
<tr>
<td>UUG</td>
<td>A</td>
<td>Stop</td>
</tr>
<tr>
<td>CUU</td>
<td>C</td>
<td>Proline</td>
</tr>
<tr>
<td>CUC</td>
<td>C</td>
<td>Histidine</td>
</tr>
<tr>
<td>CUA</td>
<td>C</td>
<td>Glutamine</td>
</tr>
<tr>
<td>CUG</td>
<td>C</td>
<td>Start (Methionine)</td>
</tr>
<tr>
<td>AUU</td>
<td>C</td>
<td>Isoleucine</td>
</tr>
<tr>
<td>AUC</td>
<td>C</td>
<td>Threonine</td>
</tr>
<tr>
<td>AUA</td>
<td>C</td>
<td>Asparagine</td>
</tr>
<tr>
<td>AUG</td>
<td>CA</td>
<td>Valine</td>
</tr>
<tr>
<td>GUU</td>
<td>GC</td>
<td>Alanine</td>
</tr>
<tr>
<td>GUC</td>
<td>GA</td>
<td>Aspartic Acid</td>
</tr>
<tr>
<td>GUA</td>
<td>GC</td>
<td>Valine</td>
</tr>
<tr>
<td>GUG</td>
<td>GC</td>
<td>Glutamic Acid</td>
</tr>
</tbody>
</table>

**Figure 5-3b** Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.
RNA Codons

- In the case of amino acids
- Mutations in Position 1, 2 lead to Amino Acid change
- Mutations in Position 3 often don’t matter
RNA Codons

- In the case of amino acids
- Mutations that lead to Amino Acid changes (Position 1, 2): “Nonsynonymous Substitutions”
- Mutations that do not lead to Amino Acid change (often Position 3): “Synonymous Substitutions”
Examples: Sources of Genetic Variation (types of mutations)

- **Gene Duplications**
- Often followed by differentiation between the duplicates
- These are common sources of new genes
- End up with “gene family”: different opsin genes, hemoglobin, ATPases, etc.
Gene Duplications

Could happen either due to (1) “Slippage” during DNA replication (gene copied twice), or (2) unequal crossing over during genetic recombination during meiosis

Lynch and Connery (2000)
• 0.01 duplications per gene per million years
• Half life for a gene is 3-8 million years
Gene Duplications

- **Duplication** = chromosome segment present in multiple copies
- **Tandem duplications** = repeated segments are adjacent
- **Tandem duplications** often result from *unequal crossing-over* due to mispairing of homologous chromosomes during meiotic recombination

Above: unequal crossing over during Recombination
Unequal cross-over and the origin of gene duplications

The chromosomes on the left have synapsed, but cross-over has occurred at nonhomologous points.

As a result, one of the cross-over products (chromosome #2) lacks gene C and one (chromosome #3) has a duplication of gene C.

Example: https://www.youtube.com/watch?v=tNZrhWhFVQw
Unequal crossing over can happen during meiosis

So, UNEQUAL CROSSING OVER is the cause of mutation (gene duplication), not meiosis itself
Gene Duplications

• Duplicate genes in Eukaryotes are continuously created, tested, and discarded

• Duplicated genes either degenerate into pseudogenes (no function), become new genes, or subfunctionalize with an existing gene
Fate of Duplicated Genes

- **Full function**
- **Dead function**
- **New function**

Nonfunctionalization

Neofunctionalization

Subfunctionalization

Loss of function of extra gene copy

New function of extra gene copy

Partition of function between the gene copies
Examples: Gene Families resulting from gene duplications

- Olfactory receptors
- Steroid hormone receptors
- Heat shock proteins
- Ion uptake enzymes
- Hemoglobin
- Opsins
- Melanins
- Detoxification enzymes (cytochrome P450s)
- Hox genes
- **Exon Shuffling:** different exons either within a gene or between two nonallelic genes are mixed (end up with new protein)
Chromosomal Alterations (chromosome duplications, deletions, inversions, fusions, etc)

Types of mutations
Chromosome Inversion
Types of Mutations

• **Transposable Elements (Transposons)**

  • A DNA sequence that can change its relative position (self-transpose) within the genome

  • Barbara McClintock's discovery of these **jumping genes** earned her a Nobel prize in 1983.

  • The mechanism of transposition can be either "copy and paste" (retrotransposons) or "cut and paste" (DNA transposons).

  • TEs make up a large fraction of the genome of eukaryotic cells, and are often considered "junk DNA" (85% of the maize genome; 44% of the human genome).

http://en.wikipedia.org/wiki/Transposable_element
Polyploidization is the generation of more than two pairs of homologous chromosomes due to failure of reduction of chromosomes during cell division (mitosis or meiosis). This is most important and common mechanism of speciation in plants: will discuss this topic more in lecture on Speciation.
Rates of Mutations and Evolution of Mutation Rate
Rate of Mutations

• Mutation rates vary among species and can even vary among populations within a species
  – DNA polymerase (or reverse transcriptase for RNA genomes) can vary in accuracy
  – DNA repair mechanisms can vary in efficiency
Rate of Mutations

**High in HIV (point mutations)**

- 1 error/10^4-10^5 bp/cycle, or 1 error per genome /replication cycle
- Replication rate is high, 10^9 cells infected/day
- Every possible point mutation occurs in an AIDS patient about 10^4-10^5 times/day (Coffin, 1995 *Science*)
- The virus goes through ~1000 generations before reaching the next person
In most species, mutation rate is lower.

**Bacteria**: 1 error/10^8-10^{10} bp/cycle
or 0.0001-0.0002 mutations per genome/generation

**Drosophila**: 8 x 10^{-11}/bp/cycle
or 0.93/genome/generation

**Human**: 2 x 10^{-8}/bp/generation
~120 new mutations/genome/generation (Crow 1993)
# RATE OF MUTATIONS

In most species, mutation rate is relatively low

<table>
<thead>
<tr>
<th>Organism</th>
<th>Base pairs in haploid genome</th>
<th>Base pairs in effective genome</th>
<th>Mutation rate per base pair per replication</th>
<th>Mutation rate per replication per haploid genome</th>
<th>Mutation rate per replication per effective genome</th>
<th>Mutation rate per sexual generation per effective genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2, T4 phage</td>
<td>1.7*10^5</td>
<td>-</td>
<td>2.4*10^-8</td>
<td>0.0041</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli</td>
<td>4.6*10^6</td>
<td>-</td>
<td>5.4*10^-10</td>
<td>0.0025</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>1.2*10^7</td>
<td>-</td>
<td>2.2*10^-10</td>
<td>0.0026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. elegans</td>
<td>8.0*10^7</td>
<td>1.8*10^7</td>
<td>2.3*10^-10</td>
<td>0.0184</td>
<td>0.0041</td>
<td>0.036</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>1.7*10^8</td>
<td>1.6*10^7</td>
<td>3.4*10^-10</td>
<td>0.0578</td>
<td>0.0054</td>
<td>0.140</td>
</tr>
<tr>
<td>Mouse</td>
<td>2.7*10^9</td>
<td>8.0*10^7</td>
<td>1.8*10^-10</td>
<td>0.4860</td>
<td>0.0144</td>
<td>0.900</td>
</tr>
<tr>
<td>Human</td>
<td>3.2*10^9</td>
<td>8.0*10^7</td>
<td>5.0810^-11</td>
<td>0.1600</td>
<td>0.0040</td>
<td>1.600</td>
</tr>
</tbody>
</table>
Mutation Rate

• Mutation rate varies among species, populations, and even individuals. Why?

• Under what conditions would high vs. low mutation rate be advantageous or disadvantageous?
Mutation rates can vary among populations within a species (and even among individuals)

• **Mistakes are Made**: Each day a human cell (or any cell) is estimated to suffer hundreds of thousands of occurrences of base damage and single-strand DNA breaks. Also errors occur in DNA replication (the wrong nucleotide added), meiosis, etc. (DNA damage + mistakes)

• **Repairs are Made**: DNA repair pathways, such as direct reversal of base damage by enzymes such as photolyase, to the repair of double-strand DNA breaks by recombination repair and non-homologous end-joining pathways, counter the massive load of DNA damage experienced by the genome.

• Variation in mutation rate among species and populations arises from: differences in (1) the accuracy of DNA replication and/or (2) their abilities to recognize and repair DNA damage
Mutation Rates can Evolve

- Elevated mutation rates are advantageous when faced with novel or stressful environments (especially in Bacteria): provides new genetic variation upon which natural selection can act to respond to the environment.

- In the face of environmental stress, some “mutator” strains in some bacterial species will evolve, that have an elevated mutation rate.

- Sometimes selection will favor less accurate DNA replication systems to elevate the mutation rate (e.g. HIV).
Variation within the Genome

- Mutation rate is much higher in organelle genomes (mitochondria, chloroplasts) relative to nuclear genomes – due to lack of DNA repair enzymes

- Mutation rate is elevated in some parts of the genome (mutational “hot spots”)

• What are evolutionary causes of differences in mutation rate?
Evolutionary causes of mutation rate variation

Hypotheses on mutation rate variation among lineages (not mutually exclusive):

• **Generation-time hypothesis.** Groups with shorter generations evolve faster because they experience more rounds of germ-cell divisions during an arbitrary unit of time. More rounds of germ-line divisions mean additional DNA synthesis and extra opportunities for mutations that are due to DNA replication errors.

• **Metabolic-rate hypothesis.** Mutation rate that is due to endogenous or exogenous mutagens, such as oxygen radicals. This hypothesis argues that groups with higher metabolic rates produce more free radicals, which leads to greater DNA damage and faster mutation and evolutionary rates.

• **DNA repair hypothesis.** In groups with better DNA repair systems, more mutations are corrected before transmission, which reduces mutational output.

• **Genetic drift interfere with selection.** In smaller populations, selection is less efficient, so fewer deleterious mutations are removed from the genome. The end result is an increased presence of deleterious mutations in smaller populations.
Effects of Mutations
Mutations: Double-Edged Sword

- Most mutations in multicellular eukaryotes are ‘neutral’ with no effect on fitness, as most of the genome is nonfunctional.
- Most mutations that affect functional genes are harmful.
- Mildly deleterious mutations persist longer in a population because it takes longer to select them out.
- Recessive mutations remain longer in the population, because they are eliminated when homozygous, not when heterozygous; when they are heterozygotes, they are “masked” from selection.
- Selection for favorable mutations leads to adaptation.
% of individuals that survived to adulthood through time in populations that were allowed to accumulate all mutations versus control lines where natural selection eliminated most deleterious mutations.
Mutations: Double-Edged Sword

- Occasionally, a very small number of mutations are favorable, due to chance:

Selection for these favorable mutations leads to adaptation.
Most Mutations have no Effect

• 3.12 billion nucleotides in the human genome

• Most of the genome is non-coding sequence and has no function (up to 95%):
  – mutations here are “Neutral”

• Mutations that affect function are what matter (within genes, or within regulatory sequences that affect the expression of genes)
RNA Codons

- In the case of amino acids
  - Mutations in Position 1, 2 lead to Amino Acid change ("nonsynonymous")
  - Mutations in Position 3 often don’t matter ("synonymous")
Mutations

Mutations that matter, in an evolutionary sense, are those that get passed on to the next generation: i.e., those that occur in the cells that produce gametes (the “germ line”)

Mutations that occur in somatic cells do not get passed on to the next generation
Mutations in “germ line” get passed on to next generation

In Humans:

- ~100+ new mutations per individual
- ~1.6 new deleterious mutations/generation in protein-coding sequences
- More harmful, dominant mutations get selected out quickly
- Recessive mutations stick around longer (when masked in Heterozygote form, not exposed to selection)
Question:

• Would you expect sex differences in mutation rate in the germ line?

• Why?
The generation-time hypothesis. Groups with shorter generations evolve faster because they experience more rounds of germ-cell divisions during an arbitrary unit of time. More rounds of germ-line divisions mean additional DNA synthesis and extra opportunities for mutations that are due to DNA replication errors.

One prediction of this hypothesis is that the mutation rate for males should be greater than for females because of their greater number of germ-line divisions per generation.

Such male mutation bias (or male-driven evolution) has been reported in many mammalian groups and other vertebrates (including birds where females are homogametic), and even in plants.
Sex differences in Mutation Rate

The Male germ line accumulates more mutations, and thus males are more likely to pass on genetic diseases, especially with increasing age.
James Crow
University of Wisconsin, Madison

Discussed the exponential growth of mutations in the male germ line

http://www.genetics.wisc.edu/CATG/crow/index.html
Mutations

- **Females:** only one set of eggs are made

- **Males:** sperm production ongoing
  Continuous cell division... mutations accumulate exponentially

- Male germline is much more prone to replication error
• Females: mutation rate is constant with age
• Males: mutation rate increases exponentially with age

From James Crow
# Cell divisions and Mutation Rate

- Males: mutation rate increases exponentially with age

---

**Figure 3** | Relative frequency of *de novo* achondroplasia and Apert syndrome for different paternal ages. The ordinate is the ratio of the observed number of mutations (O) to the number expected (E), if all paternal ages are associated with the same frequency of mutation. The blue line gives the actual data; the red line is the best-fitting exponential curve. (Figure adapted from REF. 4.)

James Crow, 2000, *Nature Genetics Reviews*

The characterization of mutational processes that generate sequence diversity in the human genome is of paramount importance both to medical genetics and to evolutionary studies. To understand how the age and sex of transmitting parents affect de novo mutations, here we sequence 1,548 Icelanders, their parents, and, for a subset of 225, at least one child, to 35 × genome-wide coverage. We find 108,778 de novo mutations, both single nucleotide polymorphisms and indels, and determine the parent of origin of 42,961. The number of de novo mutations from mothers increases by 0.37 per year of age (95% CI 0.32–0.43), a quarter of the 1.51 per year from fathers (95% CI 1.45–1.57). The number of clustered mutations increases faster with the mother’s age than with the father’s, and the genomic span of maternal de novo mutation clusters is greater than that of paternal ones. The types of de novo mutation from mothers change substantially with age, with a 0.26% (95% CI 0.19–0.33%) decrease in cytosine–phosphate–guanine to thymine–phosphate–guanine (CpG>TpG) de novo mutations and a 0.33% (95% CI 0.28–0.38%) increase in C>G de novo mutations per year, respectively. Remarkably, these age-related changes are not distributed uniformly across the genome. A striking example is a 20 megabase region on chromosome 8p, with a maternal C>G mutation rate that is up to 50-fold greater than the rest of the genome. The age-related accumulation of maternal non-crossover gene conversions also mostly occurs within these regions. Increased sequence diversity and linkage disequilibrium of C>G variants within regions affected by excess maternal mutations indicate that the underlying mutational process has persisted in humans for thousands of years. Moreover, the regional excess of C>G variation in humans is largely shared by chimpanzees, less by gorillas, and is almost absent from orangutans. This demonstrates that sequence diversity in humans results from evolving interactions between age, sex, mutation type, and genomic location.
“From an evolutionary point of view, the greatest mutational health hazard in the human population is present in older males”

Professor James Crow
Male Mutation Rate

Diseases with a strong paternal age effect:

acrodysostosis, achondroplasia, Apert syndrome, basal cell nevus, cleidocranial dysostosis, Crouzon syndrome, fibrodysplasia ossificans progressiva, Marfan syndrome, oculodentodigital syndrome, Pfeiffer syndrome, Progeria, Waardenburg syndrome
Asymmetry of Sex

• Higher mutation rate in male germ line

• Greater sexual selection of males (driven by female choice)

• Male-driven evolution?
How might the following affect allele frequencies in a population? Genotype frequencies?

- Selection
- Genetic Drift
- Inbreeding
- Recombination
- Random Mating
- Mutations
- Migration (Gene flow)
- Epigenetic Inheritance
Concepts

Mutation
Recombination
Inbreeding
Genetic Drift
Natural Selection
Codon Bias
Questions

(1) What are the sources of genetic variation?

(2) What are mutations and are they harmful or beneficial?

(3) Why are there sex differences in mutation rate in the germ line?

(4) What is sex and why did it evolve?

(5) What are the costs and benefits of Sex?

(6) What is the relationship between Genetic Variation and Natural Selection?

(7) Was Lamarck wrong? Or not? In what way?
1. Which of the following is most FALSE regarding the genetic substrate (variation) on which selection acts?

(A) Sex creates new combinations of genotypes  
(B) Genetic drift could reduce the levels of allelic and genotypic variation  
(C) Inbreeding, caused by genetic drift, increases levels of homozygosity in a population  
(D) Mutations are a source of allelic variation  
(E) Epigenetic modifications give rise to allelic diversity
2. Which of the following alleles would tend to be removed MOST quickly from a population through natural selection? (Hint: play with the Allele A1 software and think about the results)

(A) Dominant highly deleterious allele
(B) Dominant slightly deleterious allele
(C) Recessive slightly deleterious allele
(D) Recessive highly deleterious allele
3. Which of the following is FALSE regarding inbreeding?

(A) Inbreeding results from genetic drift

(B) Populations with lower allelic diversity tend to have lower genotypic diversity (more homozygous)

(C) Selection acts more slowly in inbred populations to remove deleterious recessive alleles

(D) One way to reduce inbreeding in a population is to bring in migrants from another population
4. Which is NOT a consequence of Sex?

(A) The increase in allelic diversity
(B) Death (in an evolutionary sense)
(C) The creation of many new genotypes across the genome (Evolution of individuality)
(D) Reduction in population growth rate relative to clonal reproduction (1/2 of the population does not bear offspring)
5. Which of the following are INCORRECT regarding mutations?

(A) Mutations can be harmful
(B) Mutations can be beneficial
(C) Mutations generate allelic variation
(D) Most mutations have significant effects on fitness
(E) Mutations accumulate to a much greater degree in the male germline (sperm) than in the female germline (eggs) with age
6. Which of the following mechanisms would LEAST likely contribute to the creation of novel genes?

(a) Slippage during DNA replication
(b) Unequal crossing over
(c) Different mutations in adjacent copies of genes
(d) Histone deacetylation
(e) Exon shuffling
7. Which of the following mechanisms would \textbf{LEAST likely} contribute to the creation of novel genes?

(a) Slippage during DNA replication (DNA replication error)
(b) Gene Duplication
(c) DNA Methylation
(d) Polyploidization
(e) Exon shuffling
• Answers:

1E
2A
3C
4A
5D
6D
7C
WHY IS THIS SNOWMAN LOOKING AT A SNOWBALL?

HE'S CONTEMPLATING SNOWMAN EVOLUTION.

OBVIOUSLY, IF HE EVOLVED FROM A SNOWBALL, IT RAISES TOUGH THEOLOGICAL QUESTIONS FOR HIM.

LIKE THE MORALITY OF THROWING ONE'S PRECURSORS AT SOMEONE?

SURE, AND WHAT ABOUT SHOVELING ONE'S GENETIC MATERIAL OFF THE WALK?
• Optional Slide:

• Mutation Rates can Evolve
Mutation rate variation in multicellular eukaryotes: causes and consequences

Charles F. Baer, Michael M. Miyamoto & Dee R. Denver
Nature Reviews Genetics 8, 619-631 (August 2007)