The Origins of Variation
The Cell Cycle

Interphase
  G1 – gap or growth 1
  S – synthesis
  G2 – gap or growth 2

Mitosis – karyokinesis and cytokinesis
  Prophase – condensation
  Metaphase – chromosomes at equatorial plane
  Anaphase – chromosomes separate
  Telophase – decondensation
cytokinesis
Meiosis

defines “sexual reproduction” (even in parthenogenic species)
two consecutive reduction divisions

chromosome replication already occurred during S phase of interphase

differences in meiosis and mitosis occur during in prophase

increases genetic variation through recombination
Prophase I

Leptotene – condensation
Zygotene – homologous *chromosomes* pair (*synapsis*) to form tetrad of 4 *chromatids* (two *diploid* chromosomes)
Pachytene – crossing over occurs, complement is now *haploid* because each chromatid is now unique even though still joined to another at the centromere
Diplotene – paired chromosomes begin to separate, cross-overs are visible as *chiasmata*
Diakinesis – separation of maternal and paternal homologs complete in places, producing *dyads* but the segregation and assortment of maternal and paternal dyads is random

Prophase II

*dyads* split as individual chromatids
chromatids again segregate and assort independently
two mechanisms of recombination in meiosis

**crossing over** - the exchange of genetic material between chromosomes of different parental origin, during synapsis (in Prophase I) of meiosis; a minimum of at least one exchange of chromosomal arms between non-sister chromosomes in each tetrad

**independent assortment** - a unique mixture of chromatids of maternal, paternal, and recombined chromatids in each gamete
Recombination: Crossing Over (Meiosis I)

2 parental chromatids → 4 gametic chromatids
Recombination: Independent Assortment

Chromosome 1

Chromosome 2

16 gametic combinations
1a/2a
1a/2b
1a/2c
1a/2d
1b/2a
1b/2b
1b/2c
1b/2d
1c/2a
1c/2b
1c/2c
1c/2d
1d/2a
1d/2b
1d/2c
1d/2d
Recombination: Independent Assortment

4(N)^2 possible gametic combinations

N Chromosomes

e.g., 23 chromosomes = 8,464 combinations
Origins of *Genetic Variation*

1) mutation - an alteration in DNA sequence, various types

2) intragenic recombination - results in entirely new associations of genes not present in either parental genome

   Two forms of intragenic recombination:
   a) crossing over
   b) independent assortment

3) reticulation – acquisition of genetic material from unrelated or relatively unrelated sources (e.g., hybrid species, horizontal gene transfer)
General Types of Mutations

1) point mutations

2) chromosomal rearrangements and chromosomal number variation
The “Central Dogma”

Transcription – polymerization of mRNA from ssDNA template

Translation – tRNA- and ribosome-mediated polymerization of amino acids from mRNA template
Point Mutation

redundancy of genetic code - 64 combinations of nucleotide triplets (codons) but only 22 amino acids in living organisms (selenocysteine is a post-transcriptionally modified from UGA codon; pyrrolysine genetically coded in methanogenic Archaea)

synonymous vs nonsynonymous substitution

missense mutation – change in amino acid translation

nonsense mutation – results in premature stop codon

Redundancy of genetic code

1st position - intermediate translational effect
2nd position - large translational effect
3rd position - almost never a translational effect
Vertebrate Mitochondrial Genetic Code (others exist)

<table>
<thead>
<tr>
<th>Second letter</th>
<th>U</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>UUU, UUC, UUA, UUG {Phe}</td>
<td>UCU, UCC, UCA, UCG {Ser}</td>
<td>UAU, UAC {Tyr}</td>
<td>UGU, UGC {Cys}, UGA {Trp}</td>
</tr>
<tr>
<td>C</td>
<td>CUU, CUC, CUA, CUG {Leu}</td>
<td>CCU, CCC, CCA, CCG {Pro}</td>
<td>CAU, CAC {His}</td>
<td>CGU, CGC, CGA, CGG {Arg}</td>
</tr>
<tr>
<td>A</td>
<td>AUA, AUG {Ile, Met}</td>
<td>ACU, ACC, ACA, ACG {Thr}</td>
<td>AAU, AAC {Asn}</td>
<td>AGU, AGC {Ser}, AGA {Stop}</td>
</tr>
<tr>
<td>G</td>
<td>GUU, GUC, GUA, GUG {Val}</td>
<td>GCU, GCC, GCA, GCG {Ala}</td>
<td>GAU, GAC {Asp}, GAG {Glu}</td>
<td>GGU, GGC, GGA, GGG {Gly}</td>
</tr>
</tbody>
</table>

First letter: C
Third letter: G
Point Mutation = nucleotide substitution or single nucleotide polymorphisms (SNPs)

1.3% of human genome (36.8 million total)

transition C ↔ T, A ↔ G little effect on translation, common

transversion purine ↔ pyrimidine, big effect on translation, uncommon

both spontaneous deamination and error in replication of methylated cytosine → thymine (the most common nucleotide substitution, especially in hypermutation of sex chromosomes)
Mutation Rate = \( \mu \) (substitution rate)

\( \mu = 10^{-8}/\text{locus/generation} \) in human nuclear genome
\( \mu = 10^{-5}/\text{locus/generation} \) in human mitochondrial genome
\( \mu = \text{up to } 10^{-3}/\text{locus/generation} \) in some viruses

in humans, at least 1 new phenotypically distinct mutant per gamete per generation

>120 point mutations per human

A typical genome differs from the reference human genome at 4.1 to 5.0 million sites

99.9% of variants consist of SNPs and short insertions/deletions
implications of mutation rate on molecular clock

apparent mutation rate is based only on substitutions that persist in the genome

mutations that occur at nucleotide positions that affect phenotype (nonsynonymous) may be eliminated by selection

selection, speciation, population size, or other factors may accelerate or retard the molecular clock

to the extent that the molecular clock is a good time-keeper, recent lungfishes are hypothetically as different genetically from Devonian lungfishes as humans are from Devonian lungfishes
other types of mutations - Chromosomal Rearrangements

*fusion and fission* - so large that they can be appreciated under a microscope

chromosomal rearrangements may include centromeres, which may result in their loss or doubling on rearranged chromosomes and ultimately to loss or destruction of the chromatid through *nondisjunction* (failure of chromatids to segregate during diakinesis)

*Metacentric* chromosome - symmetrical arms
*Acrocentric* chromosome - long and short arms

*Robertsonian Rearrangements* - conversion from metacentric to/from acrocentric by fusion or fission

Chromosomal Duplication
e.g., *autopolyploidy*, Down’s syndrome (trisomy 21), Klinefelter’s syndrome (XXY)
Insertions and Deletions ("Indels")

the addition or removal of DNA sequence to a chromosome range in size from one nucleotide to thousands

mechanisms of insertion/deletion:

a) replication slippage – usually mono-, di- tri-nucleotide repeats that causes DNA polymerase enzyme to ‘slip’ and add or delete a repeat; form hotspots of length variation - microsatellite alleles (also known as STRs [single tandem repeats] and SSRs)

b) mispairing during synapsis followed by unequal crossing over or repair

c) mobile genetic elements
Role of Indels

*frameshift mutation* - change in open reading frame (ORF) of translation due to insertion or deletion (that is not a multiple of 3 nucleotides) within a protein coding gene; results in missense mutation of all downstream codons

non-frameshift mutant allele causative agents of disease
e.g., cystic fibrosis, Huntington’s disease, fragile X syndrome

reduced recombination in heterogametic sex chromosomes

in viruses and bacteria - roles in immunoevasion and cytoadherance, fusion protein expression
Translocation and Transposition

Translocation - movement of DNA sequences to a non-homologous chromosome
Transposition - movement of DNA sequences to new position

Mobile Genetic Elements

Insertion sequences - small, include only transposase gene (mostly prokaryotes only)

Transposons

also:

Integrating Plasmids or “Episomes”
Bacteriophages
Group II Introns – self-catalyzing ribozymes
Transposons

Class I ("Retrotransposons" or "retroelements")
include reverse transcriptase gene
“copy and paste” DNA → RNA → DNA
very large and may include additional genes or parts thereof
comprise large percentages of the genome of many organisms
~42% of human genome, ~90 of wheat genome

Types:
Viral, with long terminal repeats (LTRs)
Long Interspersed Nuclear Elements (LINEs) or non-LTR viral
retrotransposons, at least 21% of human genome
e.g., LINE-1 active in human genome, 6kb, implicated in epithelial
cell carcinoma, schizophrenia
Short Interspersed Nuclear Elements (SINEs), nonviral
e.g., Alu I repeat ~300 bp, ~1,500,000 copies, ~10.7-13% of human genome, classically referred to as ‘selfish DNA’
Transposons

Class II ("DNA Transposons")
include transposase gene
“cut and paste” short direct repeats followed by inverted repeats
large and may include additional genes

Role of Transposons (Classes I and II)
Duplicative or Replicative Transposition
  e.g., believed to have been the vehicle of gene duplication
  producing three color receptors in primate retina
in DNA Transposons – may result in gene duplication during S-
  phase of interphase when donor has been replicated but target
  site has not
in Retrotransposons – simply by repeated replication
Role of Transposons, continued
Disrupt gene function thru frameshift of open reading frame (ORF) of protein coding gene or translocation of gene to another chromosome
e.g., colorectal cancer, breast cancer, Ewing’s sarcoma,
hypercholesterolemia, hemophilia, neurofibromatosis, diabetes mellitus type 2

hypothetically regulatory
e.g., Alzheimer’s syndrome, lung cancer, gastric cancer

Permanent Translocation Heterozygosity or “meiotic chains” - non-independent segregation of chromosomes in meiosis mediated by pairing of homologous segments in different chromosomes, mostly in plants, some invertebrates, and monotreme mammals

vehicle of Horizontal Gene Transfer (HGT)
Still more chromosomal rearrangements

Inversion - the flipping over (reverse orientation) of a DNA sequence
one example known to have produced increased fertility in women

*Pericentric* inversion – includes centromere
*Paracentric* inversion – does not include centromere

Unequal crossing over involving pericentric inversions result in centromere loss in one chromosome, but gain in the other which in turn may result in:

*nondisjunction* (failure of homologous chromosomes to segregate) and *aneuploidy* (incorrect number of chromosomes per gamete, and/or chromosomal fission
Still more chromosomal rearrangements

Duplication or *Copy Number Variation* (CNV) - mispairing during meiotic synapsis followed by unequal crossing over 12%-19% of human genome (possibly as much as 50%) – much more than SNPs (1.3% of genome)

*May have strong phenotypic effects!* (at least 40 diseases)

Potential fates of gene duplications

- multicopy genes
- *defunctionalization* - pseudogenes
- *neofunctionalization* - multigene families
  - e.g., globin gene family
- addition or deletion of repeats

Humans have >1,750 duplicated genes compared with chimps
Orthology – strictly homologous single-copy genes

Paralogy – duplicated genes; between species these share common ancestry that predates population or species divergence
Still more chromosomal rearrangements

Conversion - mispairing during meiotic synapsis followed by repair, two mechanisms:

- Double Holliday Junction (DHJ) – dsDNA exchange during synapsis
- Synthesis Dependent Strand Annealing (SDSA) – information exchange only

may result in concerted evolution of multicopy genes or repeated elements
Holliday Junction

a secondary structure formed between 4 DNA strands

its role in recombination
Reticulation

acquisition of genetic material from unrelated or relatively unrelated sources

Hybridization

Horizontal or Lateral Gene Transfer (HGT or LGT)
Hybridization
Potential outcomes
species fusion - the formation of *hybrid species*, usually *allopolyploids* but sometimes *homoploids*; after a few generations the genomes of the parental species become thoroughly mixed by crossing over and fusions and fissions e.g., many angiosperms

e.g., red wolf

*introgression* (introgressive hybridization) - gene flow or "leaking" of genes of one species into another by way of backcrossing hybrid offspring to parental types, i.e., movement of genetic material from one population to another by sexual reproduction

e.g., movement of mitochondrial genome of one species into another
e.g., acquisition of genetically engineered genes (e.g., pest or pesticide resistance, frost tolerance) from crop plants into weed species
HGT (LGT)
the insertion of retroelements and exogenous transposons into a genome
the translocation of genetic material between endosymbionts and their hosts or by bacteriophage vectors

e.g., mitochondria - endosymbiotic origin, evidence from cell membranes, gene structure, origin of replication, the fact that mitochondrial rRNAs are more similar to endosymbiotic bacterial (Rickettsia) rRNAs than to nuclear rRNAs of eukaryotes, and ongoing movement of genetic material from mitochondria (and chloroplast) genomes to the nuclear genome (numt's) resulting in obligate endosymbiosis

e.g., plasmids in bacteria - genetic piracy of advantageous genes (e.g., antibiotic resistance) by restriction endonucleases
HGT (LGT)

e.g., genome of house mouse - more than 50% retroviral in origin

e.g., Endogenous retroviruses (ERVs) - expressed during implantation in placental mammals, act as immunodepressors and create a syncytium around the developing embryo

e.g., obligate insect bacterial endosymbionts – e.g., protists in termites involved in digestion or nutrient production

e.g., photosynthetic zooxanthellae in corals
Epilogue: A typical genome differs from the reference human genome at 4.1 to 5.0 million sites.

99.9% of variants consist of SNPs and short indels (insertions/deletions).

2,100 to 2,500 structural variants, including:
- 1,000 large deletions
- 160 copy-number variants (CNVs)
- 915 Alu insertions (non-autonomous retroelements)
- 128 L1 insertions (LINE-1 autonomous retroelements)
- 51 SVA insertions (SINE-VNTR-Alu autonomous retroelements)
- 4 NUMTs (nuclear mitochondrial DNA fragments)
- 10 inversions
  - affecting 20 million bases of sequence