GENOMIC DATA ANALYSIS

LVA-Nr. 320.301 and 320.304

Irene Tiemann-Boege
Theresa Schwarz

JOHANNES KEPLER
UNIVERSITÄT LINZ

OUTLINE FOR TODAY-PART 1

- Genomic variation
- From SNPs to copy number variants
- Evolution of polymorphisms
- Human diversity HapMap project / Simons project

GENOMIC VARIATION

- Sequence among humans is 0.5% different
  - ACGGTC G/A CCATTTT
  - Some people may have an A at a particular locus while others have a G.
  - Such a site is known as a single nucleotide polymorphism (SNP)
  - Allele = each of the two possibilities

JKU

GENOMIC VARIATION

- Differences in the DNA:
- Some examples:
  - Single nucleotide polymorphisms, insertion-deletions
  - Repeats
    - Micro- and minisatellites
    - Transposons
  - Structural variation
    - Copy number variants (CNVs)
    - Segmental duplications or low-copy repeats (LCR)

JKU

TYPES OF GENETIC VARIATION

- Large differences in sizes
  - Single nucleotide polymorphisms to chromosomal rearrangements

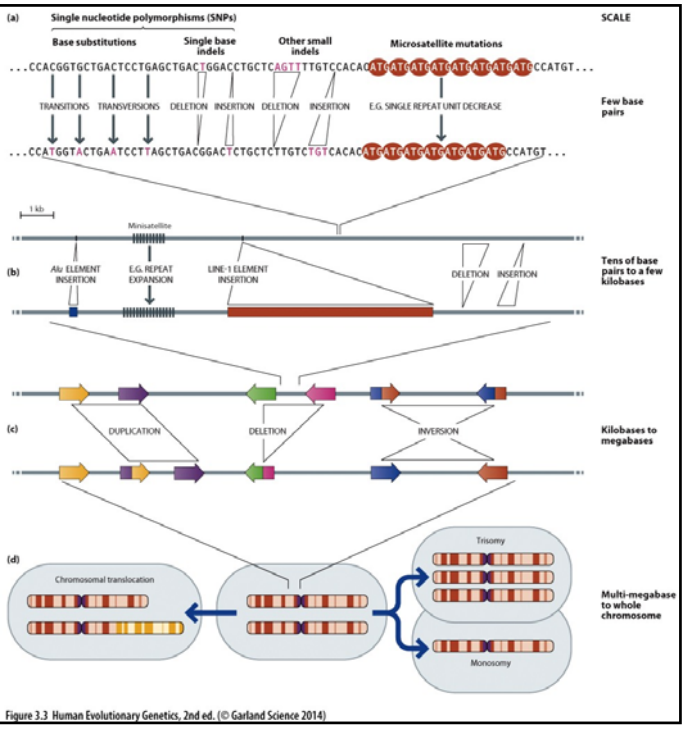
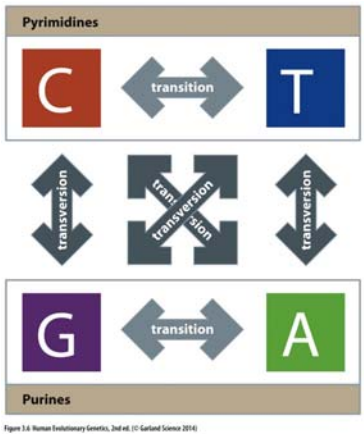


Figure 3.3 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

JKU

SINGLE NUCLEOTIDE POLYMORPHISMS

- base substitution, in which one base is exchanged for another
  - Transitions: pyrimidine base is exchanged for another pyrimidine
  - Transversions: a pyrimidine is exchanged for a purine, or vice versa
  - Observation: there are many more transitions than transversions in our genome; is this expected?



JKU

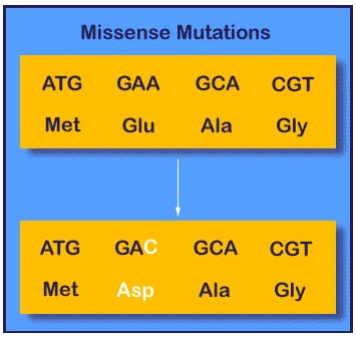
EFFECT OF DNA CHANGES ON GENES

- Transitions are more prevalent
- Synonymous vs. non-synonymous (silent mutations vs. missense mutations)
- Frameshift mutations
- Nonsense mutations

JKU

EFFECT OF DNA CHANGES ON GENES

- Synonymous vs. non-synonymous (silent mutations vs. missense mutations)



JKU

EFFECT OF DNA CHANGES ON GENES

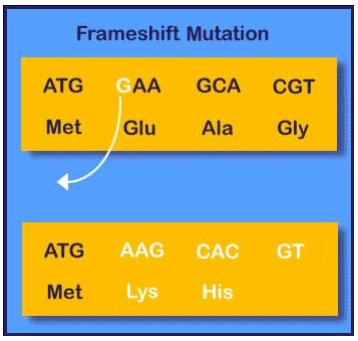
- DNA changes (substitutions) are at the third codon position are usually silent—redundancy of genetic code

1st base	2nd base: T		2nd base: C		2nd base: A		2nd base: G	
T	TTT	(Phe/F) Phenylalanine	TCT	(Ser/S) Serine	TAT	(Tyr/Y) Tyrosine	TGT	(Cys/C) Cysteine
	TTC	(Phe/F) Phenylalanine	TCC	(Ser/S) Serine	TAC	(Tyr/Y) Tyrosine	TGC	(Cys/C) Cysteine
	TTA	(Leu/L) Leucine	TCA	(Ser/S) Serine	TAA	Stop (*Ochre*)	TGA	Stop (*Opal*)
	TTG	(Leu/L) Leucine	TCG	(Ser/S) Serine	TAG	Stop (*Amber*)	TGG	(Trp/W) Tryptophan
C	CTT	(Leu/L) Leucine	CCT	(Pro/P) Proline	CAT	(His/H) Histidine	CGT	(Arg/R) Arginine
	CTC	(Leu/L) Leucine	CCC	(Pro/P) Proline	CAC	(His/H) Histidine	CGC	(Arg/R) Arginine
	CTA	(Leu/L) Leucine	CCA	(Pro/P) Proline	CAA	(Gln/Q) Glutamine	CGA	(Arg/R) Arginine
	CTG	(Leu/L) Leucine	CCG	(Pro/P) Proline	CAG	(Gln/Q) Glutamine	CGG	(Arg/R) Arginine
A	ATT	(Ile/I) Isoleucine	ACT	(Thr/T) Threonine	AAT	(Asn/N) Asparagine	AGT	(Ser/S) Serine
	ATC	(Ile/I) Isoleucine	ACC	(Thr/T) Threonine	AAC	(Asn/N) Asparagine	AGC	(Ser/S) Serine
	ATA	(Ile/I) Isoleucine	ACA	(Thr/T) Threonine	AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine
	ATG[1]	(Met/M) Methionine	ACG	(Thr/T) Threonine	AAG	(Lys/K) Lysine	AGG	(Arg/R) Arginine
G	GTT	(Val/V) Valine	GCT	(Ala/A) Alanine	GAT	(Asp/D) Aspartic acid	GGT	(Gly/G) Glycine
	GTC	(Val/V) Valine	GCC	(Ala/A) Alanine	GAC	(Asp/D) Aspartic acid	GGC	(Gly/G) Glycine
	GTA	(Val/V) Valine	GCA	(Ala/A) Alanine	GAA	(Glu/E) Glutamic acid	GGA	(Gly/G) Glycine
	GTG	(Val/V) Valine	GCG	(Ala/A) Alanine	GAG	(Glu/E) Glutamic acid	GGG	(Gly/G) Glycine

JKU

EFFECT OF DNA CHANGES ON GENES

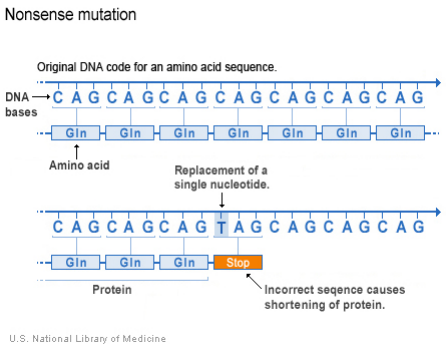
- Frameshift mutation
- Caused by insertion/and deletions



JKU

EFFECT OF DNA CHANGES ON GENES

- Nonsense mutation
- Caused by a change into a stop codon shortening the protein



JKU

OTHER DNA CHANGES IN THE GENOME

- Changes in non-coding regions of the genome are usually neutral (no effect)
- Changes outside the open reading frame might have an effect

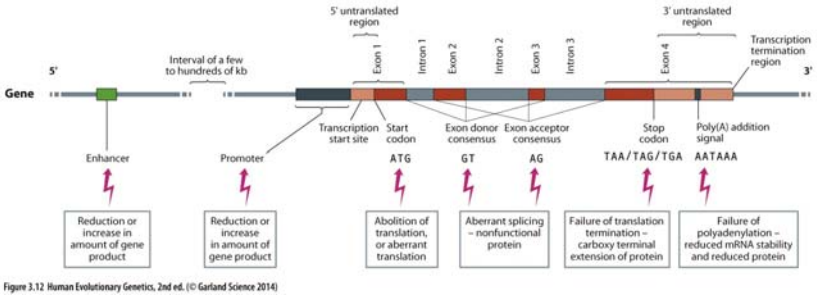


Figure 3.12 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

JKU

IN THE HUMAN GENOME: OF 55 MILLION SNPS, ONLY A FRACTION OCCURS IN ORF

TABLE 3.2:
ESTIMATED NUMBERS OF POTENTIAL CODING AND LOSS-OF-FUNCTION VARIANTS WITHIN PROTEIN-CODING GENES

Type of variant	Average number per genome
Synonymous	10,572–12,126[a]
Nonsynonymous (missense)	9966–10,819[a]
Generation of stop codon (nonsense)	26.2 (5.2)[b]
Splice site variant	11.2 (1.9)[b]
Small indel causing frameshift	38.2 (9.2)[b]
Large deletion	28.3 (6.2)[b]
Total number of LoF variants	103.9 (22.5)[b]

Data from the low-coverage dataset of 1000 Genomes Project Consortium (2010) *Nature* 467, 1061.
[a] Interquartile range of the number of variants per individual across the CEU, CHB, JPT, and YRI HapMap samples (see Box 3.6 for the three-letter abbreviations of the populations).
[b] Average number of variants in the CEU sample, with average number in homozygous state in parentheses; from MacArthur DG et al. (2012) *Science* 335, 823.
LoF, loss of function.

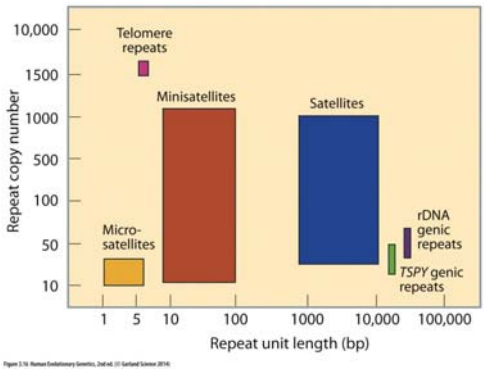
OUTLINE FOR TODAY-PART 1

- Genomic variation
- From SNPs to copy number variants
- Evolution of polymorphisms
- HapMap project

JKU

TANDEMLY REPEATED DNA SEQUENCES (HIGH COPY NUMBER)

- Changes in the numbers of repeated DNA sequences arranged adjacently in tandem arrays
  - Also known as: variable number of tandem repeats: VNTRs
  - Highly variable; multiallelic loci
  - Cover vast range of different scales



MICROSATELLITES

- Known as short tandem repeats (STRs) or as simple sequence repeats (SSRs)
  - tandem arrays of repeat units 1–7 bp in length
  - typical copy number of 10–30.
  - Cover vast range of different scales
  - Some repeats cause diseases with dramatic expansion throughout life

JKU

EXAMPLES OF MICROSATELLITE STRUCTURES

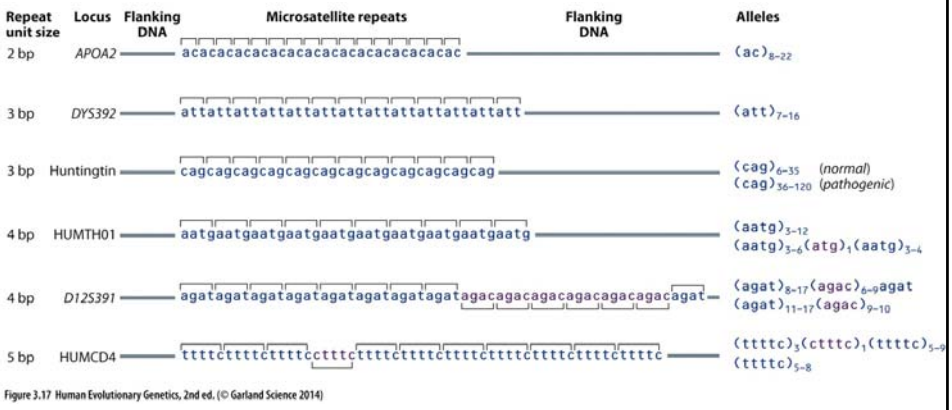


Figure 3.17 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

How are microsatellites annotated?

JKU

PHENOTYPES ASSOCIATED WITH MICROSATELLITES

- CGG, CCG, CTG, GAA expansion
  - Expansions from 5-50 to 50-4000 in the promoter, UTR, or an intron can abolish transcription, eg. Fragile X syndrome (OMIM 309550) or Friedriech's ataxia
  - Expansion of a trinucleotide repeat in the 3' UTR of a gene leads to accumulation of RNAs that interfere with the normal balance of RNA-binding proteins and disrupts splicing;
    - Eg. disease myotonic dystrophy (OMIM 160900), with expansions from 5–37 CTG repeats, to 50–thousands.

JKU

USE OF MICROSATELLITES AS MARKERS

TABLE 3.4:
PROPERTIES OF MICROSATELLITES BY REPEAT UNIT SIZE

Repeated unit/bp	Properties and distribution	Utility
1	Mostly poly(A)/poly(T), associated with *Alu*, LINE, and other retroelements	Not used, due to small differences in allele size and problem of allele-calling due to **PCR stutter**, resulting from slippage synthesis errors by the PCR polymerase
2	$(AC)_n/(GT)_n$ most common, representing 0.5% of genome; $(GC)_n$ extremely rare	Widely used in early studies because of ease of discovery; stutter a problem
3	Wide range of different repeat units; some arrays are within or close to genes and can cause diseases through expansion. $(AAT)_n$ and $(AAC)_n$ most common	Widely used. Alleles easily discriminated, and little stutter
4	Wide range of different repeat units. $(AAAC)_n$ and $(AAAT)_n$ most common; $(GATA)_n/(GACA)_n$ frequent, and clustered near centromeres	Widely used. Alleles easily discriminated, and little stutter; form basis of most forensic microsatellite profiling (Chapter 18)
5, 6, 7	Range of different repeat units	Not widely used because of relative scarcity

Table 3.4 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

MINISATELLITES

- repeat units from about 8 to 100 bp in length,
- with copy numbers from as low as 5 to over 1000
- GC-rich minisatellites tend to be clustered toward the ends of chromosomes and might be associated with recombination hotspots
- mutate through recombination mechanisms (eg. unequal crossing over)

JKU

TRANSPOSONS

- Selfish DNA
  - Can change its position within the genome
  - Causes mutations and genome rearrangements

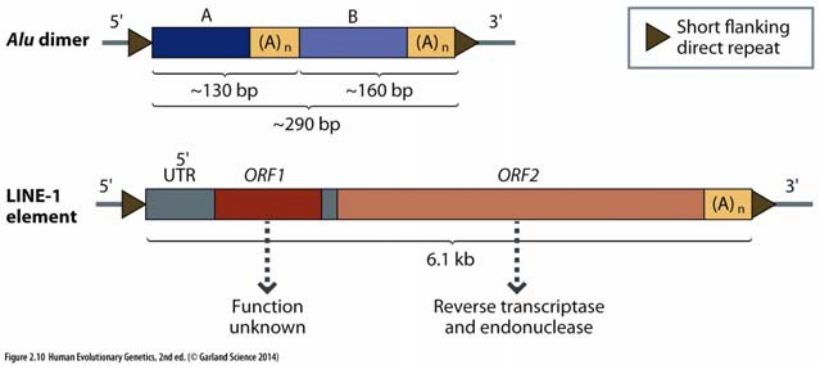
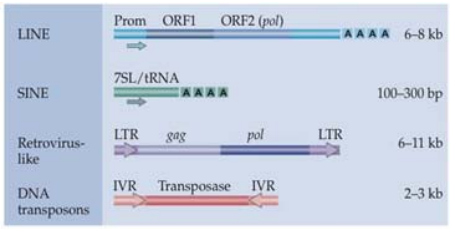


Figure 2.10 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

JKU

RETROTRANSPOSONS

- >850,000 LINEs (long interspersed nuclear elements)
  - Eg. L1 in humans
  - (with reverse transcriptase to integrate a DNA copy of the RNA genome into the host genome.
- >1,500,000 SINEs (short interspersed nuclear elements)
  - *Alu* in humans



JKU

Greg Gibson. A Primer of Genome Science, 3rd Ed.

TRANSPOSONS

TABLE 2.2:
CLASSES OF DISPERSED REPEATS IN THE HUMAN GENOME

Class	Copy no. per haploid genome	Fraction of genome	Autonomous transposition or retrotransposition?	Length of complete copies
LINEs	850,000	21%	yes	up to 6–8 kb
SINEs	1,500,000	13%	no	up to 100–300 bp
Retrovirus-like elements	450,000	8%	complete copies, yes	6–11 kb
DNA transposon copies	300,000	3%	complete copies, yes	2–3 kb

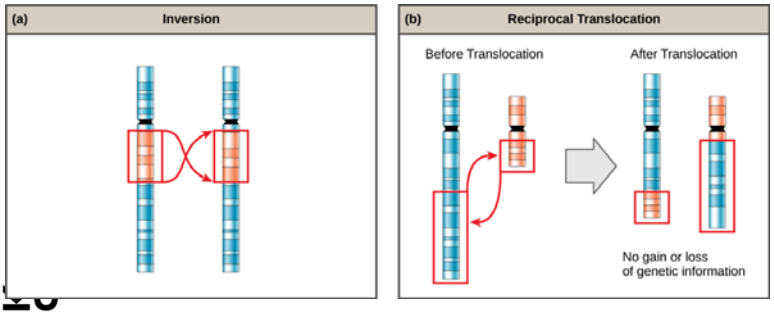
Incomplete elements, incapable of autonomous transposition, are common (see **Section 3.5**). [Data from International Human Genome Sequencing Consortium (2001) *Nature* 409, 860.]

Table 2.2 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

JKU

STRUCTURAL VARIATION

- changes in chromosome structure
- Balanced:
  - no alteration of copy number
    - Inversions or translocations between a pair of chromosomes.



JKU

STRUCTURAL VARIATION (LOW COPY VARIATION)

- Copy number variation (CNV)
  - differences in the numbers or copies of particular sequences (>1kb)

- Segmental duplications or low-copy repeats (LCR)
  - duplicated sequences (paralogs); >10kb in size

JKU

STRUCTURAL VARIATION

- Copy number variation (CNV)
  - involve differences in the numbers or copies of particular sequences (>1kb)
  - Segmental duplications or low-copy repeats (LCR)
    - duplicated sequences (paralogs) that occurred within <40 million years
    - >90% similar in sequence; >10 kb in size
    - Profound implications for the evolution of our genome and for disease because such paralogous repeats promote non-allelic homologous recombination (NAHR).

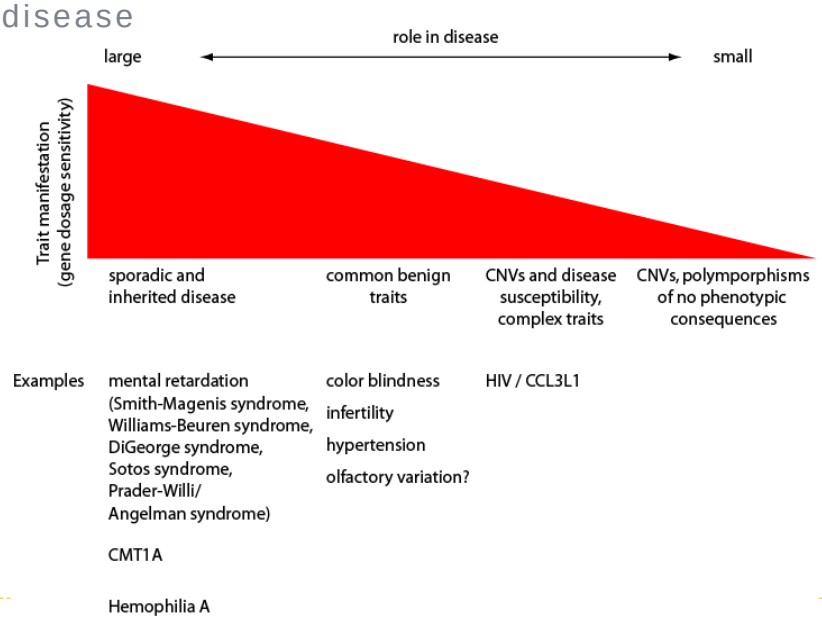
JKU

COPY NUMBER VARIATION (CNV)

- Low copy repeats (LCRs) susceptible to genomic rearrangements result in CNVs.
- CNVs are widespread in the human genome
- CNV traditionally refers to a DNA segment at least 1 kb in size, for which differences in copy number have been observed when comparing two or more genomes.
- Also known as copy-number polymorphism (CNP) for CNV that exists at >1% frequency in a population.

JKU

Copy number variation (CNVs) can lead to disease



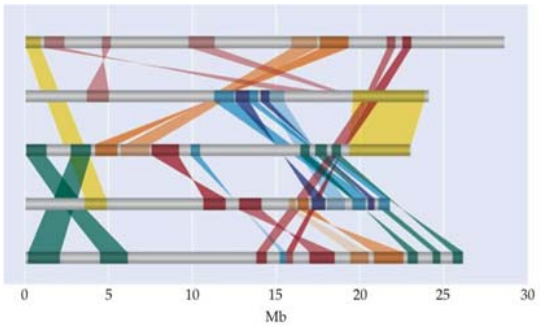
SEGMENTAL DUPLICATIONS

- Very common
  - 3% of our sequence matches sequences at different genomic locations by 90% identity
- Intra-chromosmal and interchromosomal
- 300kb-50Mb
- Paralog genes
  - What is the effect of segmental duplications?
  - Potential for evolutionary divergence
  - Allow for gene modifications with novel functions

JKU

SEGMENTAL DUPLICATIONS

- Segmental duplications in Arabidopsis



JKU

OUTLINE FOR TODAY-PART 1

- Genomic variation
- From SNPs to copy number variants
- Evolution of polymorphisms
- HapMap project

JKU

HOW ARE MUTATIONS INTRODUCED

- Environment
  - chemical and physical processes that alter bases or damage the physical structure of DNA (UV rays, ionizing radiation, chemical mutagens)
- Spontaneous events (time dependent)
  - spontaneous, endogenous chemical processes going on in all cells that lead to base modification or loss:
    - Deamination (loss of an amine-group, $-NH_2$)
    - Depurination
- Replication linked

JKU

SPONTANEOUS EVENTS

- Bases are susceptible to change with time
- Depurination
  - Each DNA genome molecule looses 5000 purine bases per day
  - N-glycosyl linkage to deoxyribose hydrolyzes releasing guanine or adenine
- Deamination
  - Major type is cytosine deamination
  - Converts cytosine into uracil
  - Deamination is 1000x more frequent in ssDNA
  - 400 cytosines are deaminated daily in each human cell

JKU

SPONTANEOUS EVENTS-DEAMINATION

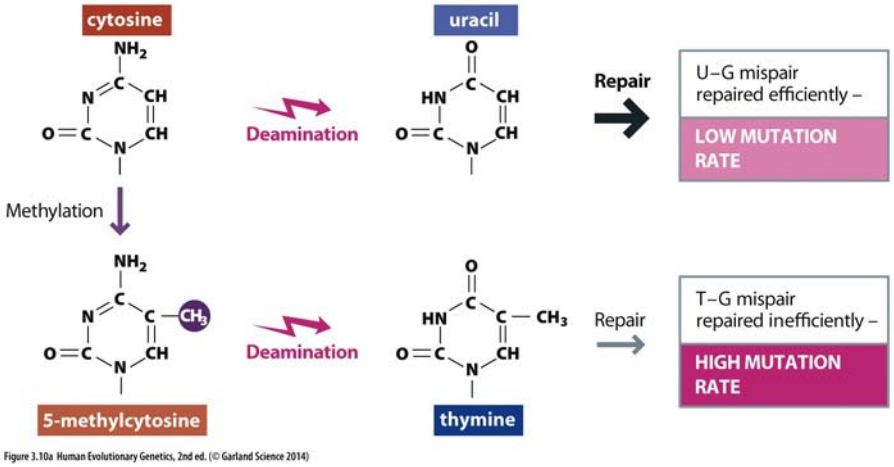
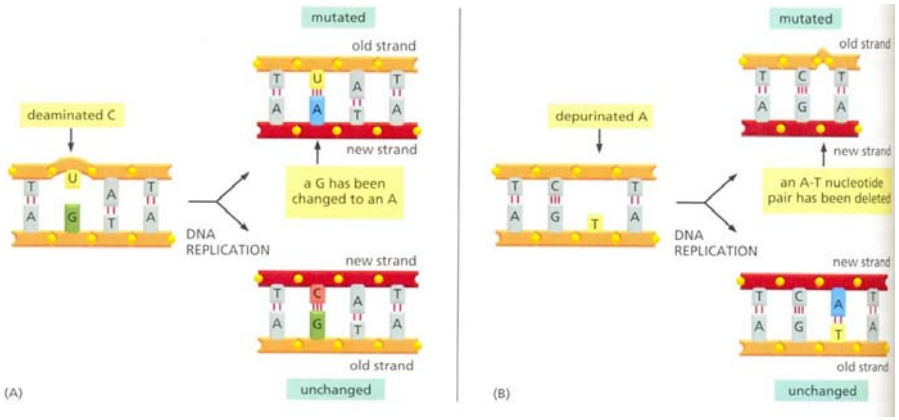


Figure 3.10a Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

THE FATE OF DAMAGES, IF UNREPAIRED

- Depurinations: Indels
- Deamination: base substitutions S (CG) >W (TA

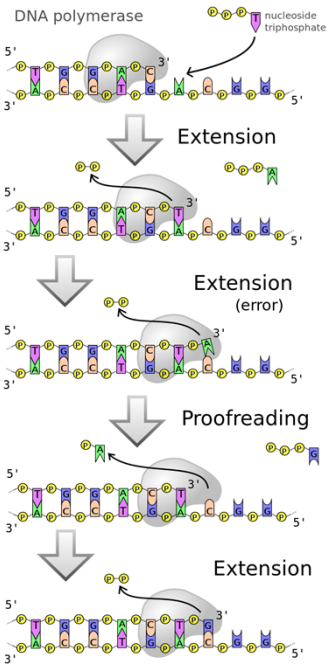


JKU

Replication derived events

☐ Replication derived events

- misincorporation of nucleotides during replication
- Polymerases (replisome) has proofreading activity
- replication errors occur with a frequency of only about 10^{-9}–10^{-11} per nucleotide per replication event
- Replication errors lead mainly to C>T substitutions



JKU

REPLICATION SLIPPAGE OF STR

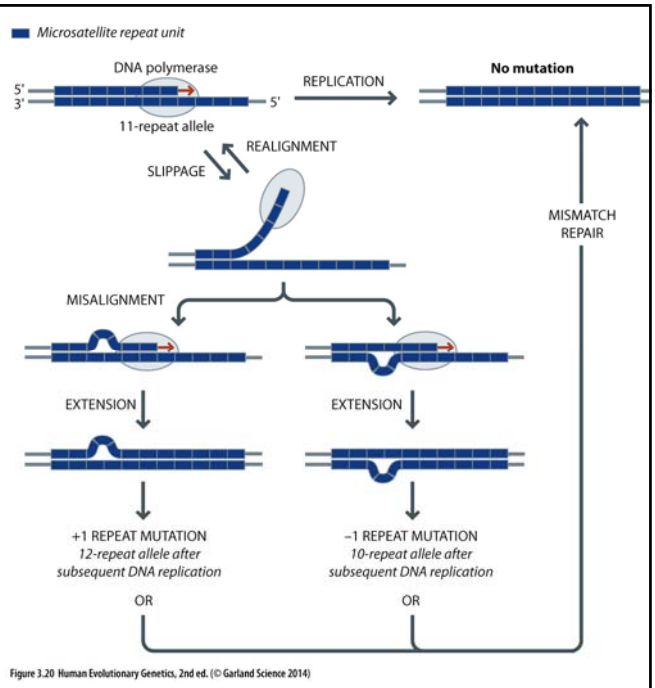
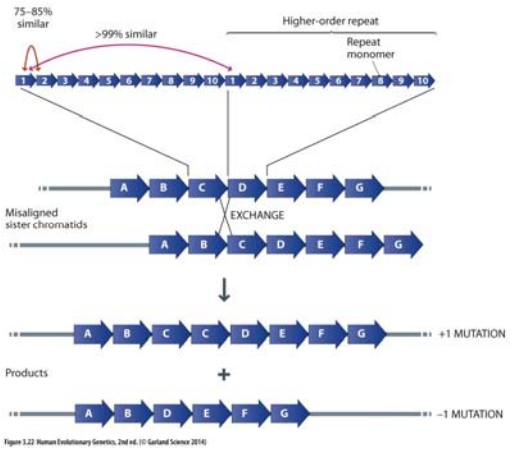


Figure 3.20 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

JKU

MINISATALLITES

- Instability (gain or loss of repeats) by recombination (mitosis or meiosis)



JKU

STRUCTURAL VARIATION

- LCR promote non-allelic homologous recombination (NAHR)

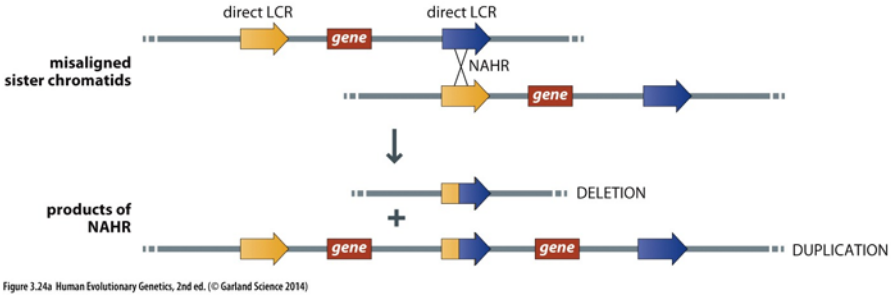


Figure 3.24a Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

JKU

RATES OF MUTATIONS BETWEEN GENETIC ELEMENTS ARE VERY DIFFERENT

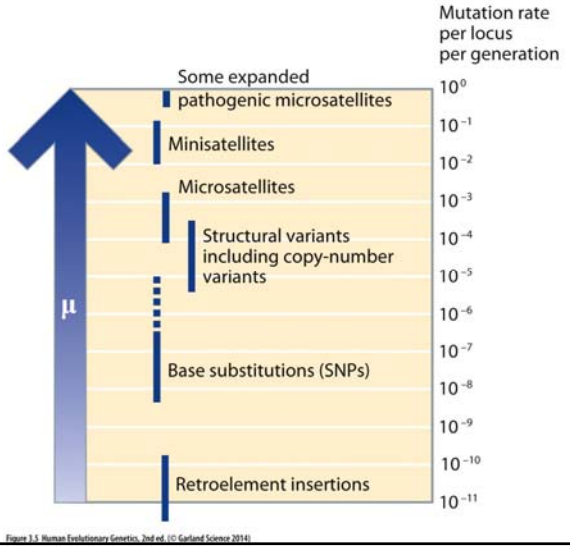


Figure 3.5 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

JKU

DEFINITION OF ALLELES/GENOTYPES/HAPLOTYPES

- Genotype
  - Nucleotide at the polymorphic site
    - Homozygote or heterozygote
    - SNPs = 2 alleles in the population
- Haplotype
  - Combination of single nucleotide types on a single chromosome
  - Many different haplotypes
- Allele frequency
  - Two alleles; sum of alleles equals 1

JKU

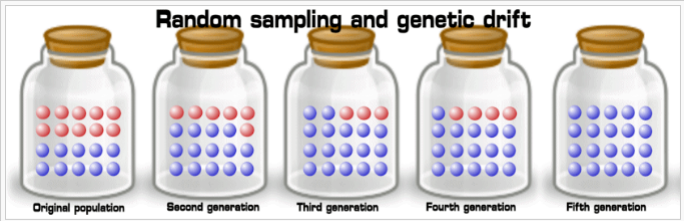
SNP FREQUENCIES VARY

- Genotypes/Allele frequency change
  - genetic drift
  - bottlenecks
  - selection
- Haplotype change
  - Recombination

JKU

ALLELE FREQUENCY DISTRIBUTION

- Neutral theory of molecular evolution
  - SNPs are introduced via mutation
  - SNP are lost or fixed due to random sampling effects (genetic drift)

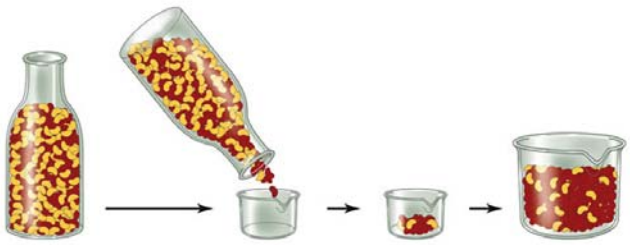


In this simulation, there is fixation in the blue "allele" within five generations.

JKU

PROCESSES THAT CHANGE ALLELE FREQUENCIES

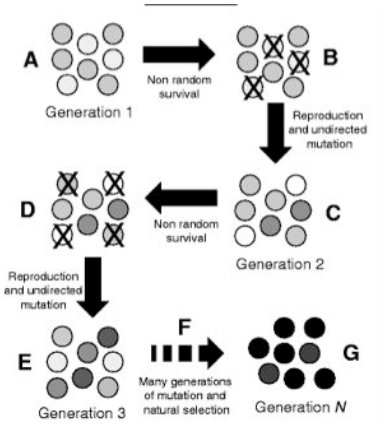
- Population bottleneck
- Migrations, islands are strong bottlenecks



JKU

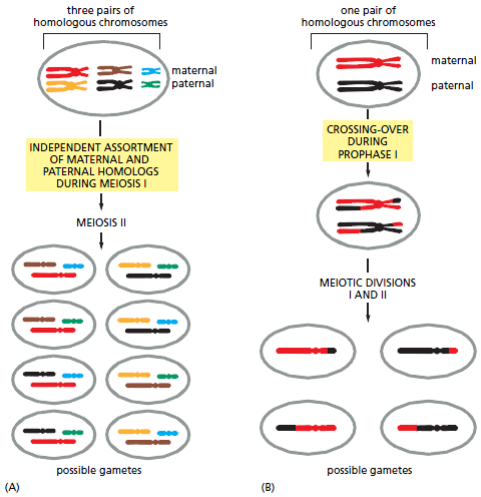
NATURAL SELECTION

- Selection (non-random process)



JKU

HAPLOTYPES CHANGE: REASSORTMENT OF SNPS OCCURS DURING MEIOSIS



JKU

HARDY-WEINBERG EQUILIBRIUM (HWE)

- p+q=1
- $p^2+2pq+q^2=1$ (HW allele frequencies)

- In the Hardy-Weinberg model, the mathematical relation between the allele frequencies and the genotype frequencies is given by:
  - AA: p^2 Aa: 2pq aa: q^2
  - AA, Aa, and aa are the frequencies of the genotypes and p and q are the allele frequencies of A and a

- Homozygotes: p^2 or q^2
- Heterozygotes: 2pq

JKU

HARDY-WEINBERG EQUILIBRIUM

- What is the allele frequency of the A-blood allele in a population with the following genotypes

Genotype	A (AA)	A (AO)	O (OO)
number	9%	42%	49%

- Answer: p(A)= $\sqrt{0.09}$ = 0.3; q=0.7

- What is the frequency of the heterozygotes in a population if the allele frequency p(A)=0.4
- Answer: 2pq (q=1-0.4) = 2x0.4x0.6 = 0.48 = 48%

JKU

OUTLINE FOR TODAY-PART 1

- Genomic variation
- From SNPs to copy number variants
- Evolution of polymorphisms
- HapMap project

JKU

HAPMAP PROJECT

- Finding common SNPs with frequency of at least 1%
- Genotyping of 269 individuals
  - 4 populations:
    - 30 adult-and-both-parents trios from Ibadan, Nigeria (YRI)
    - 30 trios of U.S. residents of northern and western European ancestry (CEU)
    - 44 unrelated individuals from Tokyo, Japan (JPT)
    - 45 unrelated Han Chinese individuals from Beijing, China (CHB).

JKU

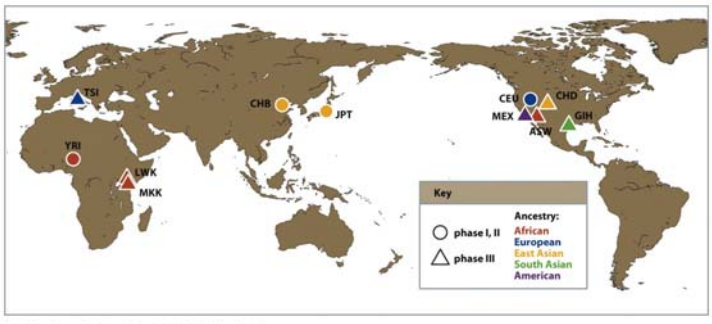
SNP DISCOVERY

- How many individuals should be typed?
- It depends on the allele frequency
- $Pr = 1- (1-p)^{2n}$
  - Pr: probability of detecting the SNP
  - p = frequency of rare allele
  - n= sample size
- Pr=94% to detect an allele with frequency of 0.5 in two individuals
- For an allele frequency of 0.1, 13.35 individuals need to be typed for Pr=94%

JKU

HAPMAP PROJECT

- Finding common SNPs with frequency of at least 1%
- Genotyping of individuals from different populations across the world



JKU

Box 3.6 Figure 1 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

THE HAPMAP PROJECT- CHARACTERIZE HUMAN VARIATION

TABLE 1:
DETAILS OF POPULATION SAMPLES USED IN HAPMAP

Standard abbreviation	Origin	Sample size and composition
Phase I and II		
YRI	Yorubans from Ibadan, Nigeria	90 (30 parent–child trios)
CEU	Utah (US) residents of N and W European ancestry	90 (30 parent–child trios)
CHB	Han Chinese from Beijing, China	45 unrelated individuals
JPT	Japanese from Tokyo, Japan	44 unrelated individuals
Phase III		
ASW	individuals of African ancestry from Southwest USA	90 (11 parent–child trios, 24 parent–child duos + 9 unrelated individuals)
CHD	Chinese from Metropolitan Denver, Colorado, USA	100 unrelated individuals
GIH	Gujarati Indians from Houston, Texas, USA	100 unrelated individuals
LWK	Luhya from Webuye, Kenya	100 unrelated individuals
MEX	individuals of Mexican ancestry from Los Angeles, California, USA	90 (30 parent–child trios)
MKK	Maasai from Kinyawa, Kenya	180 (30 parent–child trios + 90 unrelated individuals
TSI	Tuscans from Italy	100 unrelated individuals

Box 3.6 Table 1 Human Evolutionary Genetics, 2nd ed. (© Garland Science 2014)

JKU

HAPMAP PROJECT

- Phase I--1 SNP per 5000bp
  - 1mill SNPs total
- Re-sequencing for SNP discovery
  - 10million SNPs with 40% being polymorphic
- Phase II--1 SNP per 1000bp
  - ~3mill SNPs

JKU

1000 GENOMES PROJECT

- 2008-2015
- largest public catalogue of human variation and genotype data
- genetic variants with frequencies of at least 1%

1000 GENOMES PROJECT

Three pilot studies provided data to inform the design of the full-scale project:

Pilot	Purpose	Coverage	Strategy	Status
1 - low coverage	Assess strategy of sharing data across samples	2-4X	Whole-genome sequencing of 180 samples	Sequencing completed October 2008
2 - trios	Assess coverage and platforms and centres	20-60X	Whole-genome sequencing of 2 mother-father-adult child trios	Sequencing completed October 2008
3 - gene regions	Assess methods for gene-region-capture	50X	1000 gene regions in 900 samples	Sequencing completed June 2009

1000 GENOMES PROJECT

- 26 populations of European, East Asian, South Asian, American, and sub-Saharan African ancestry
- focused on demographically large populations
- Sequence of ~2500 individuals at 4-6x coverage

1000 GENOMES PROJECT

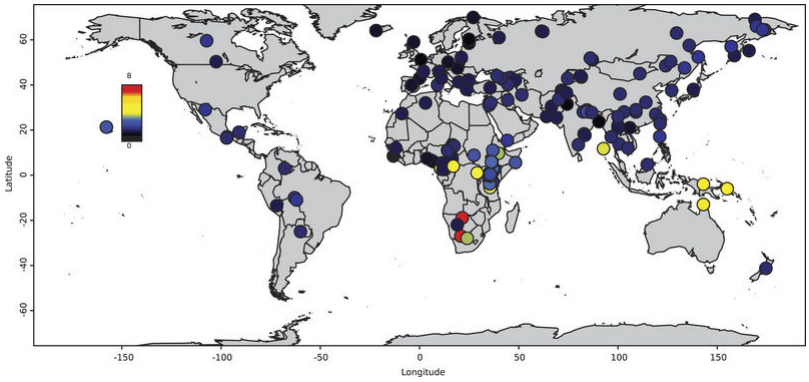
JKU

SIMONS GENOME DIVERSITY PANEL

- 300 individuals from 142 diverse populations
- To obtain a complete picture of human diversity:
  - genetic, linguistic, and cultural variation
  - 43x coverage (34-83x coverage)
  - PCR-free library
  - New variation discovered: 11% in the KhoeSan and 5% in New Guineans and Australians

JKU

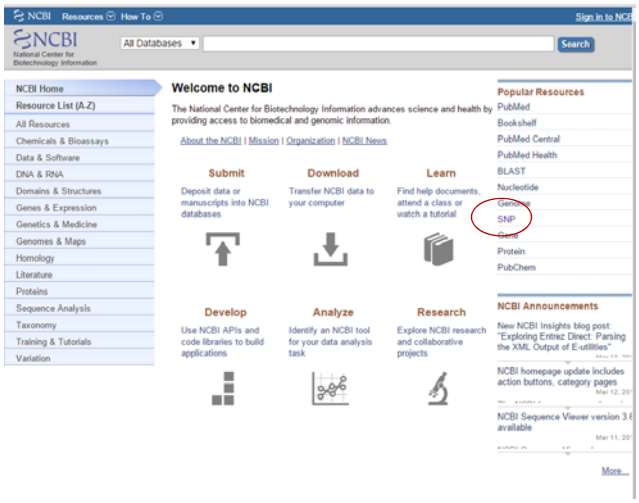
SIMONS GENOME DIVERSITY PANEL



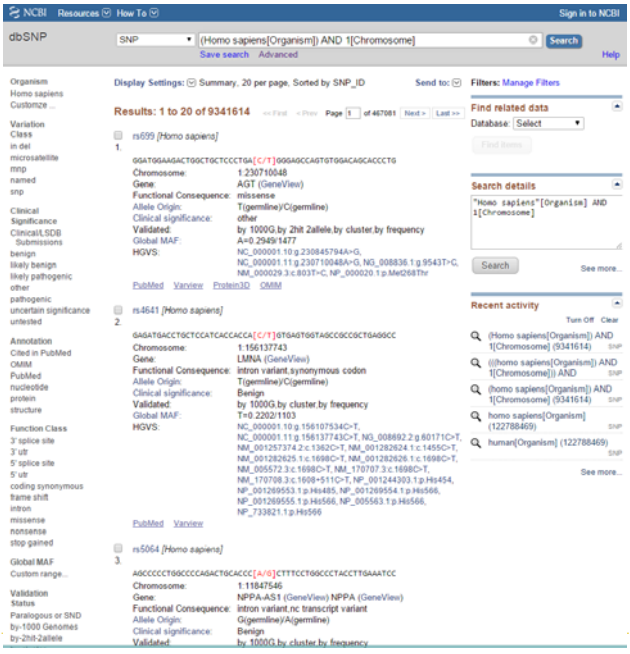
Heat map of fraction of heterozygous sites missed in the 1000 Genomes project but detected in the Simons panel

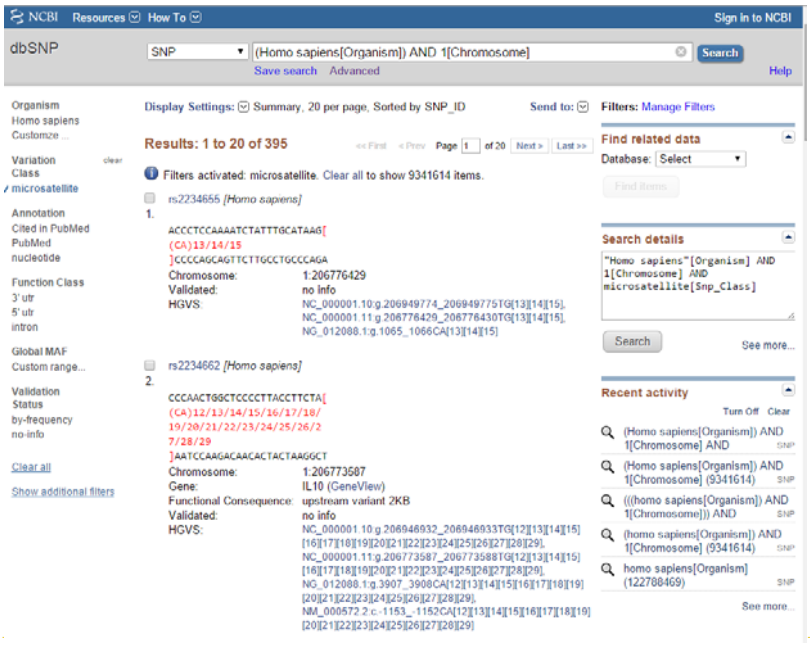
JKU

DATABASES OF POLYMORPHISMS

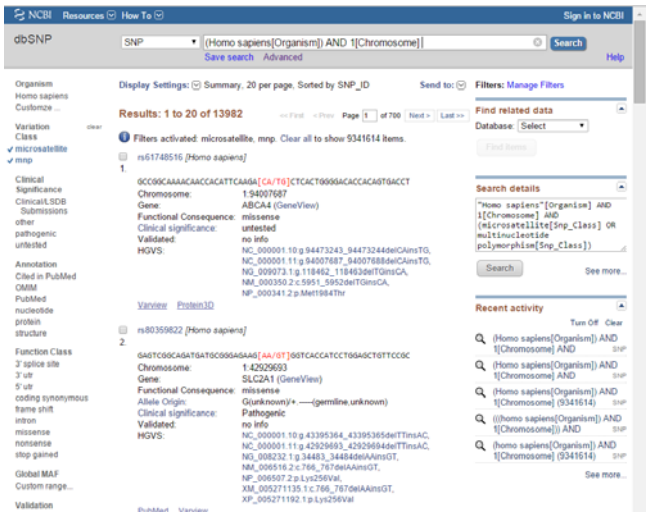


JKU





MNP-Multiple Nucleotide polymorphism



WHY ARE SNPs USEFUL?

- Association studies (genetic marker)
  - Relationship between SNPs and phenotype
  - quantitative trait loci (phenotype caused by the effect of many genes. E.g height, diabetes, heart disease, cancer)
  - Genome wide association studies (GWAS)
- Identity and identification
- Evolution
  - Human migration
  - Population stratification

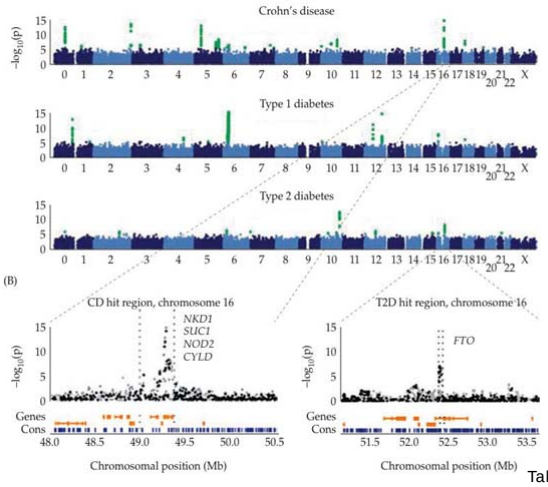
JKU

Uses of SNPs: High density Genetic Map

- High density Genetic Map is necessary to identify Quantitative trait loci (QTL)
- Phenotypic variance: height, skin color, body mass, cancer, diabetes, hypertension
- Attributed to the interactions between two or more genes and their environment
- How can we identify QTLs (genes that underlie the trait in question?
- QTLs can be identified by association studies that link a phenotype with a specific genetic region.

JKU

FINDING QUANTITATIVE TRAIT LOCI (QTL) IN GWAS



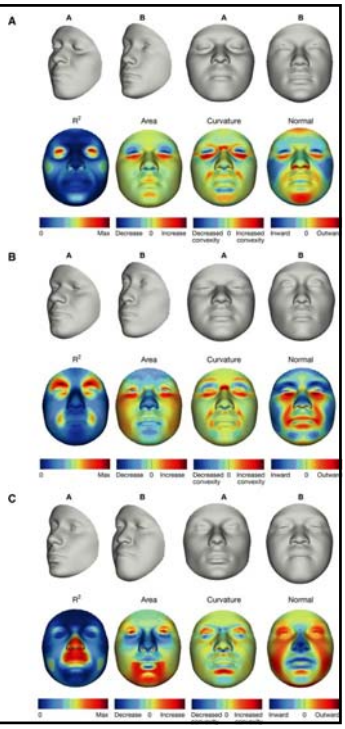
Taken from Gibson. A Primer of Genome Science, 3rd ed.)

Uses of SNPs-Identity and identification

- Some phenotypic characteristics are predictable from DNA
  - Eg. Prediction of eye color
  - IrisPlex assay: six most informative SNPs, from six pigmentation genes (HERC2, OCA2, SLC24A4, SLC45A2, TYR, and IRF4),
  - Predictive accuracy of 0.93 and 0.91 for brown and blue eye color respectively,
  - validated for forensic use.

Uses of SNPs-Identity and identification

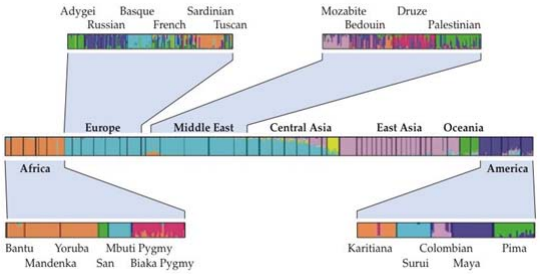
- DNA-based phenotyping is helping police derive visual information from crime scene samples to aid in the hunt for suspects
- http://embor.embopress.org/content/16/7/782?etoc



INFORMATION INFERRED FROM SNPS

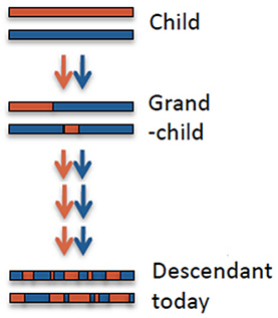
☐ Population stratification and demographic history

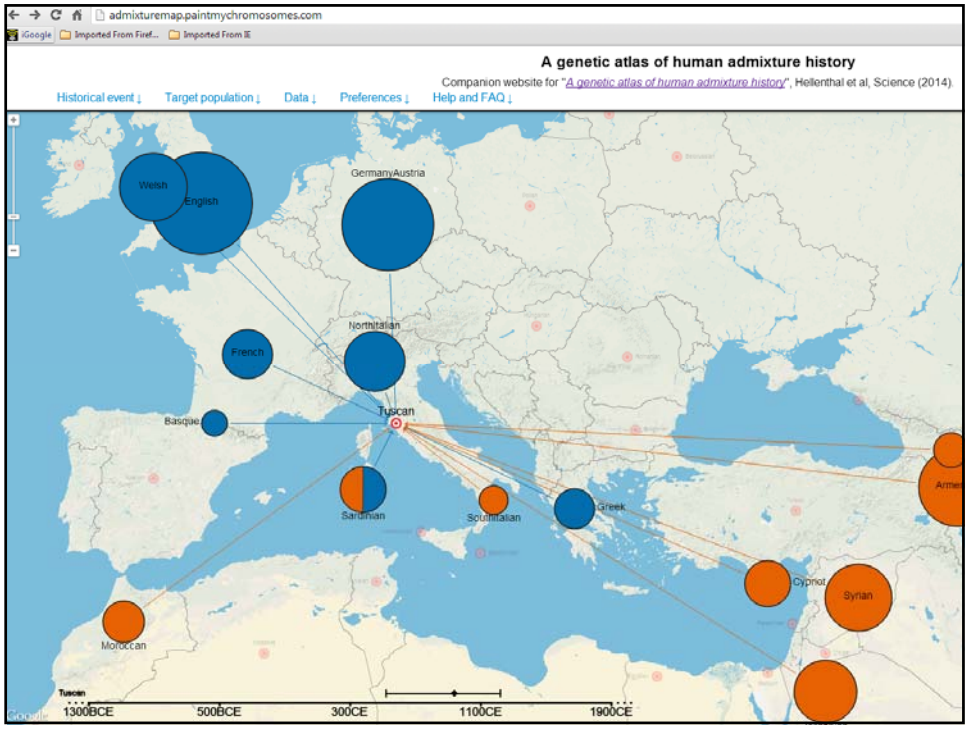
- Assuming six different human population groups, in the central bar, each individual within a population is represented by a thin vertical line. Colors identify what proportion of the microsatellite alleles is attributable to each population group (Rosenberg 2002)



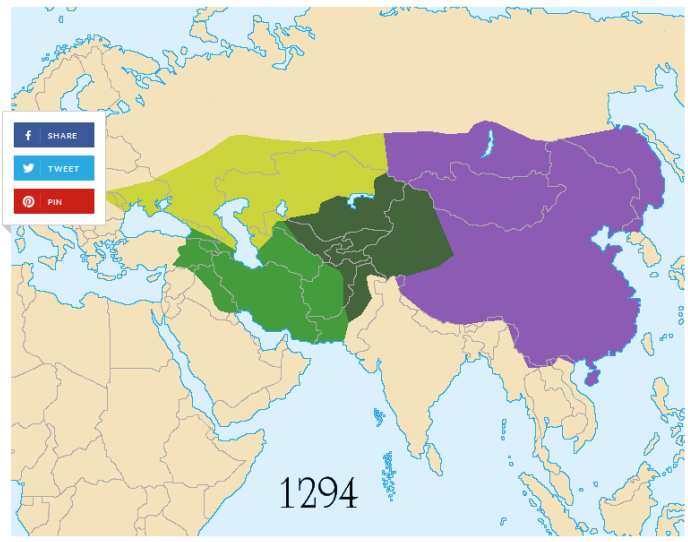
Information inferred from SNPs

- Pieces of DNA from modern day genomes
  - Pieces are passed down through generations and get smaller
  - These segments tell us when and where the mixing occurred and which groups mixed
  - The longer the pieces of DNA, the more recent the mixing.





5. The Spread of the Mongol Empire (1206-1368)



The expansion of the Mongol Empire was a famously rapid. These studies confirm that there was an abrupt transfer of DNA across Asia during this period. DNA evidence from the Hazara people of Pakistan agrees with historical records showing that they descended from Mongol warriors. Six other populations, from as far west as Turkey, demonstrated similar evidence of genetic mixing with Mongols around the same time.

JOHANNES KEPLER
UNIVERSITÄT LINZ

QUESTIONS?

JOHANNES KEPLER
UNIVERSITÄT LINZ
Altenberger Straße 69
4040 Linz, Österreich
www.jku.at