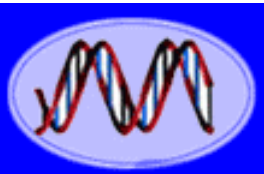
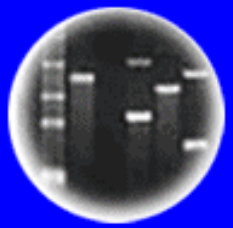


DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

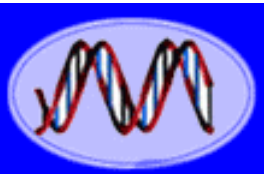
HC70A & SAS70A Winter 2010
Genetic Engineering in Medicine, Agriculture, and Law

Professors Bob Goldberg & John Harada

Lecture 5
The Age of Genomics:Your Personal Genome & Tracing Your Ancestry

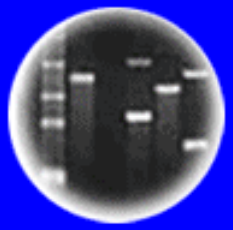
UCLA

UCDAVIS
UNIVERSITY OF CALIFORNIA



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

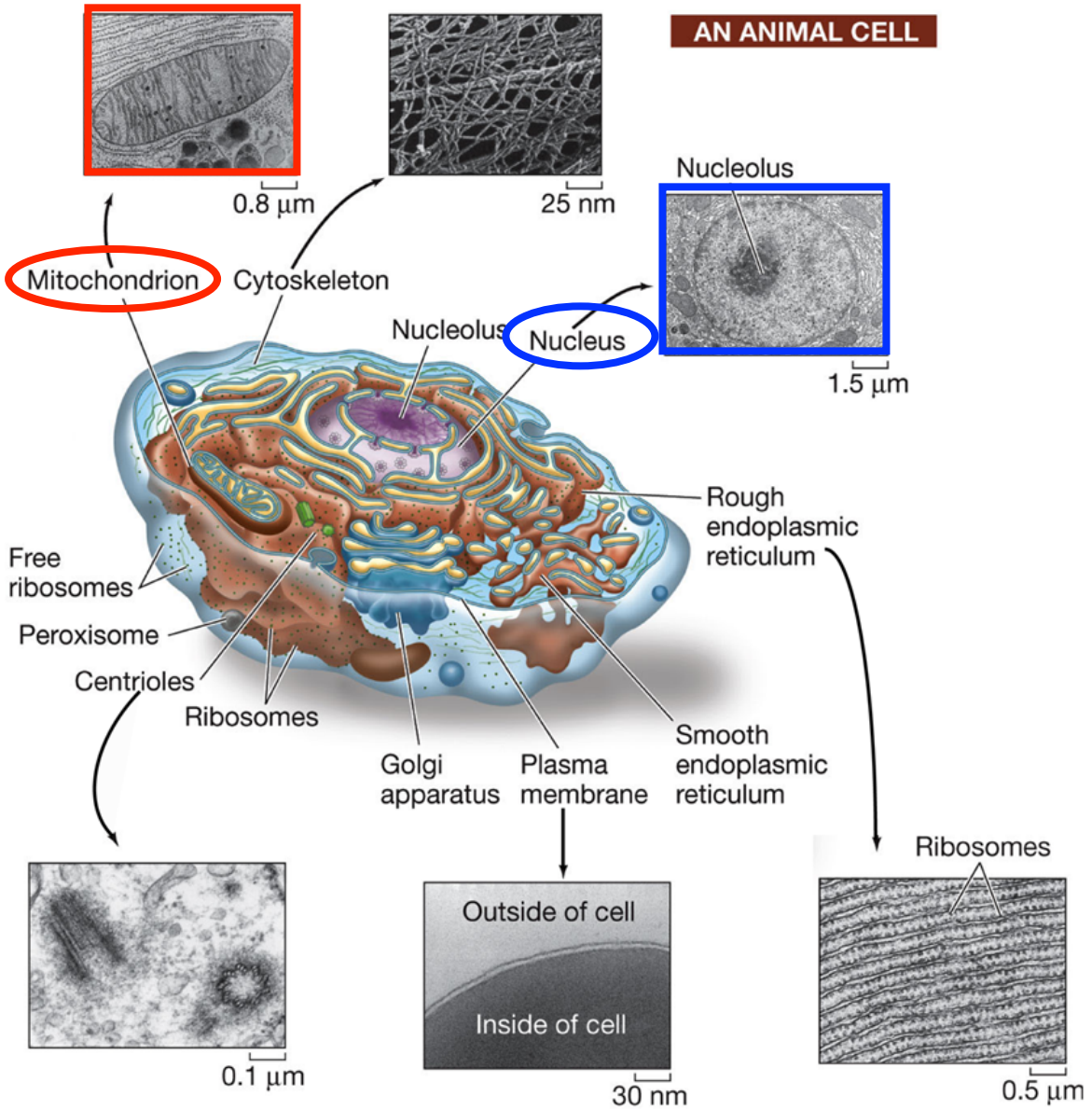
Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

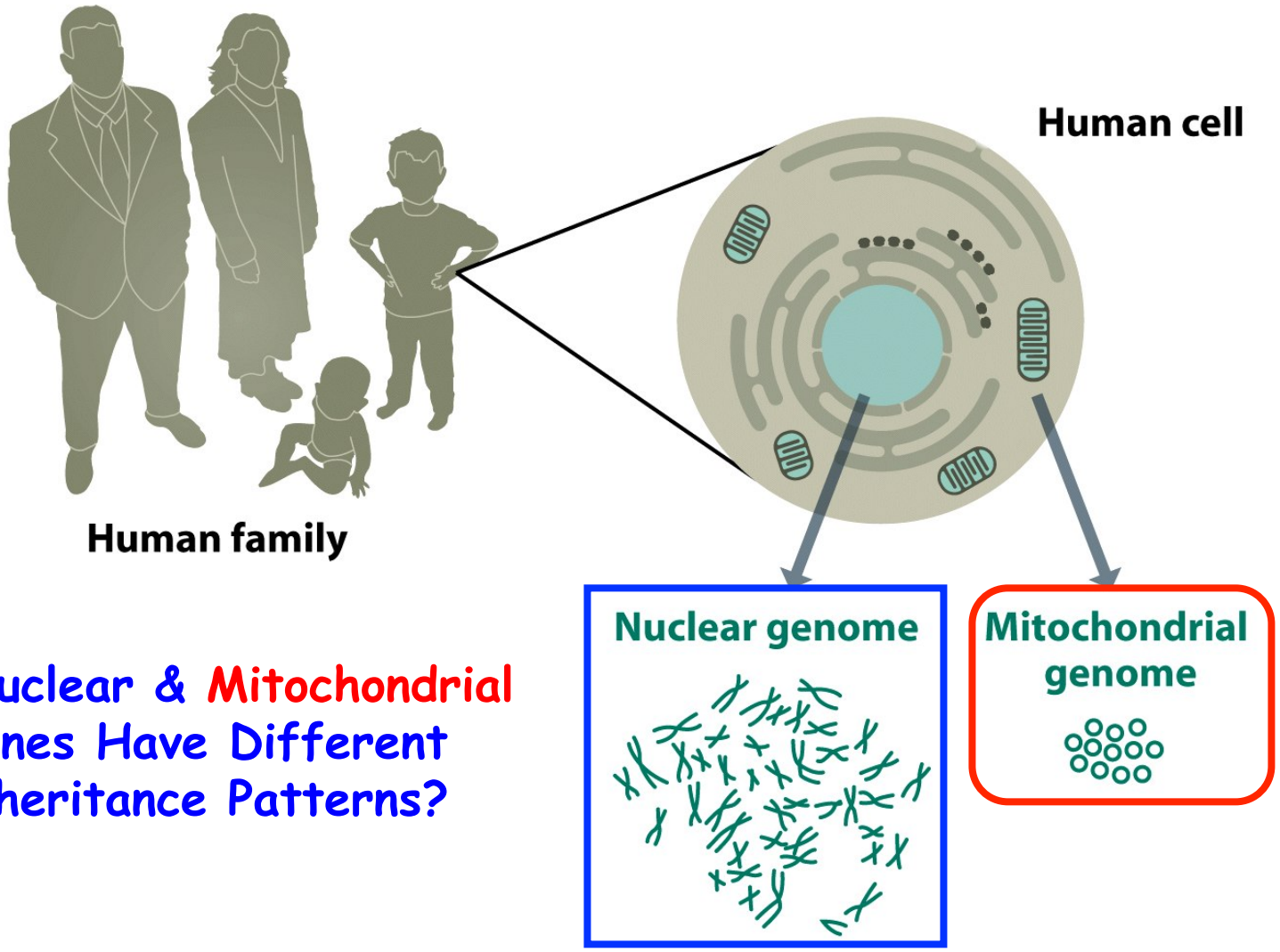
THEMES

1. Two Genomes in a Cell!
2. What is the Mitochondrial Genome and How is it Inherited?
3. What Are the Characteristics of the Human Genome?
4. The Age of the Personal Genome Has Arrived!
5. How Many Mammalian Genomes Have Been Sequenced and What Can We Learn From Comparative Genomics?
6. How Does Genetic Variation Arise in the Human Genome?
7. How to Use DNA Markers to Find Human Disease Gene Alleles?
8. How to Detect DNA Sequence Variation: SNPs and VNTRs?
9. What Can SNPs Be Used For?
10. Tracing Human Ancestry Using SNPs
11. Are Their Human Races?
12. Knowledge or Certainty?

Human Cells Have Two Genomes



……One in the Nucleus and One in the Mitochondria



Do Nuclear & Mitochondrial Genes Have Different Inheritance Patterns?

Figure 1-1 Genomes 3 (© Garland Science 2007)

The Nuclear and Mitochondrial Genomes in Size & Shape

Nuclear

3.2 Mb
25,000 Genes
24 Linear Pieces

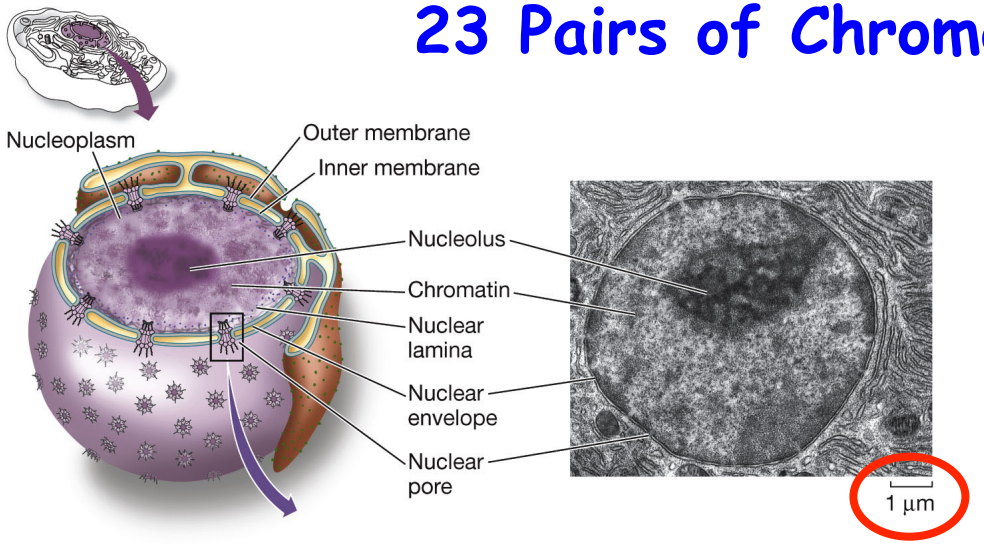
Mitochondrial

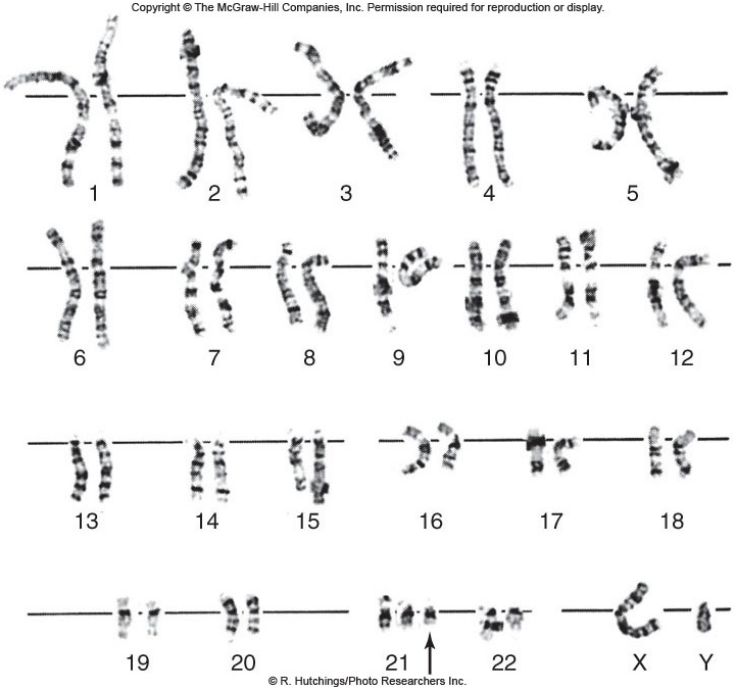
17 kb
30 Genes
1 Circle

Table 9.1: The human nuclear and mitochondrial genomes

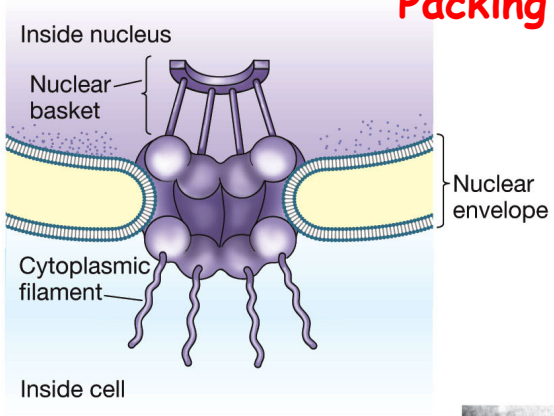
	Nuclear genome	Mitochondrial genome
Size	3200 Mb	16.6 kb
No. of different DNA molecules	23 (in XX cells) or 24 (in XY cells); all linear	One circular DNA molecule
Total no. of DNA molecules per cell	46 in diploid cells, but varies according to ploidy	Often several thousands (but variable – see *Box 9.1*)
Associated protein	Several classes of histone and nonhistone protein	Largely free of protein
No. of genes	~ 30 000–35 000	37
Gene density	~ 1/100 kb	1/0.45 kb
Repetitive DNA	Over 50% of genome, see *Figure 9.1*	Very little
Transcription	The great bulk of genes are transcribed individually (*monocistronic transcription units*)	Co-transcription of multiple genes from both the heavy and the light strands (*polycistronic transcription units*)
Introns	Found in most genes	Absent
% of coding DNA	~ 1.5%	~ 93%
Codon usage	See *Figure 1.22*	See *Figure 1.22*
Recombination	At least once for each pair of homologs at meiosis	Not evident
Inheritance	Mendelian for sequences on X and autosomes; paternal for sequences on Y	Exclusively maternal

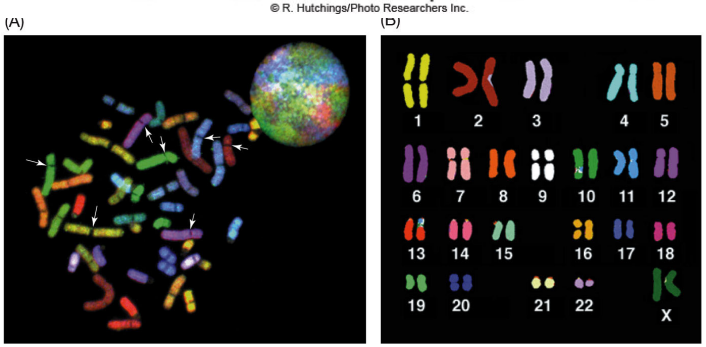
The Nucleus Is A Complex Organelle With 23 Pairs of Chromosomes (Humans)

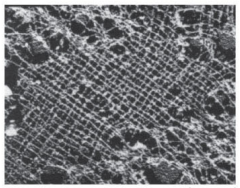




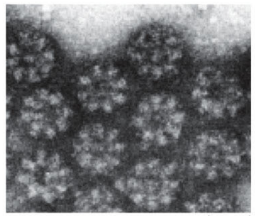
Packing Problem?







RNA & Protein Transport

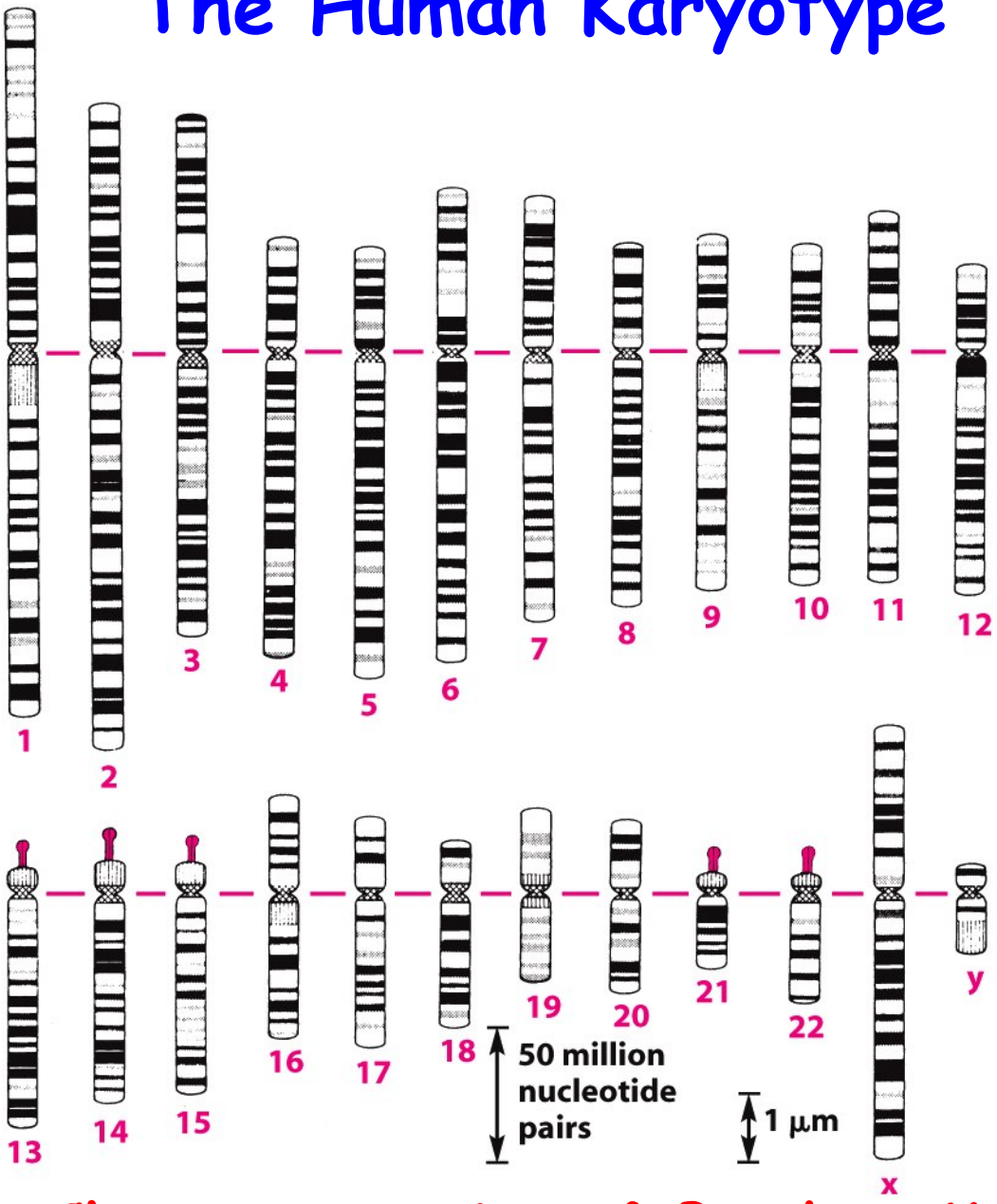


The Human Genome End to End is 1.1 Meters in Length!!!!!!

LIFE 8e, Figure 4.8 (Part 2)

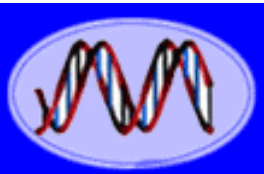
Note: Chromosome Sizes & Bands = Markers

The Human Karyotype



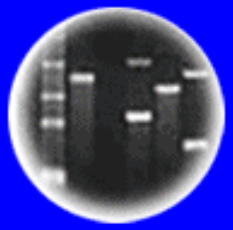
3.2×10^9 bp
(3.2 Mb)

Note: Chromosome Sizes & Bands = Markers



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria

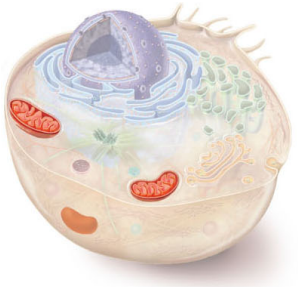


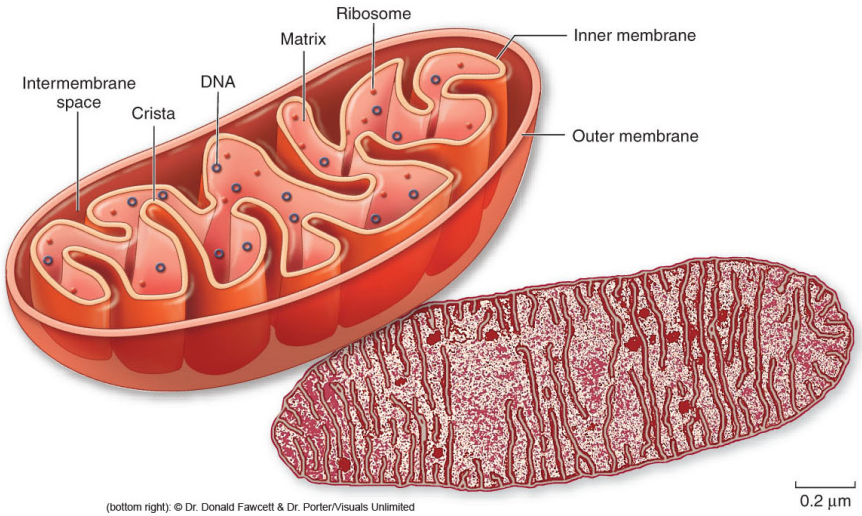
DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

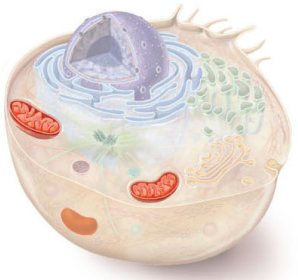
Plants of Tomorrow

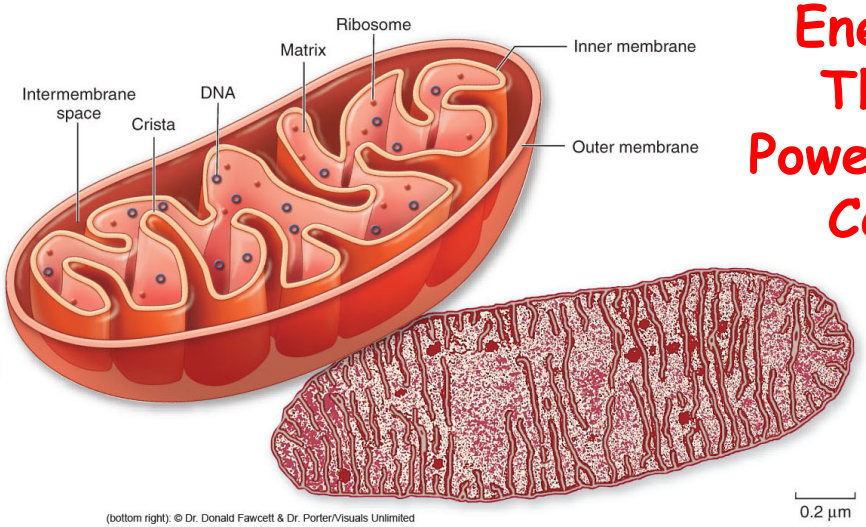
The Mitochondrial Genome





Mitochondria Power Human Cells and Contain a Circular Genome





(bottom right): © Dr. Donald Fawcett & Dr. Porter/Visuals Unlimited

Makes ATP Energy That Powers All Cells!

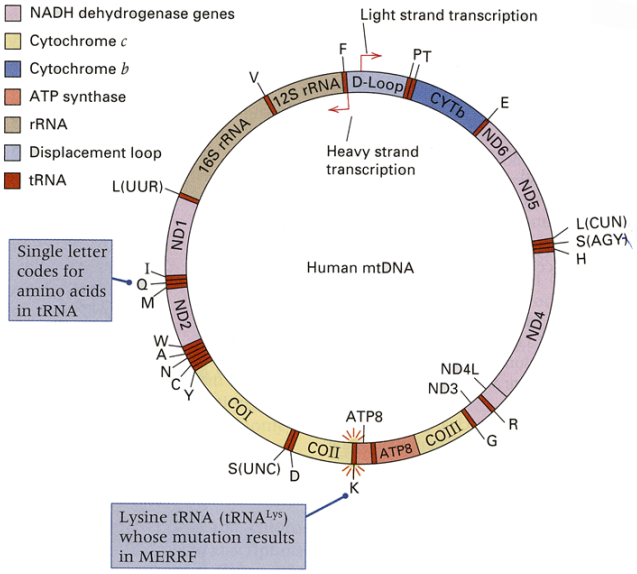


Figure 16.3 Genes in human mitochondrial DNA. The tRNA genes are indicated by the one-letter amino acid symbols; hence tRNA^{Lys} is denoted K. The positions of these and other genes in the mitochondrial DNA are indicated by color according to the key at the upper left. The arrows indicate the promoters for transcription of the heavy and light strands. [Courtesy of N-G. Larsson and D. A. Clayton. With permission, from the *Annual Review of Genetics* 29: 151. Copyright 1995 by Annual Reviews, www.AnnualReviews.org.]

Semi-Autonomous Genome

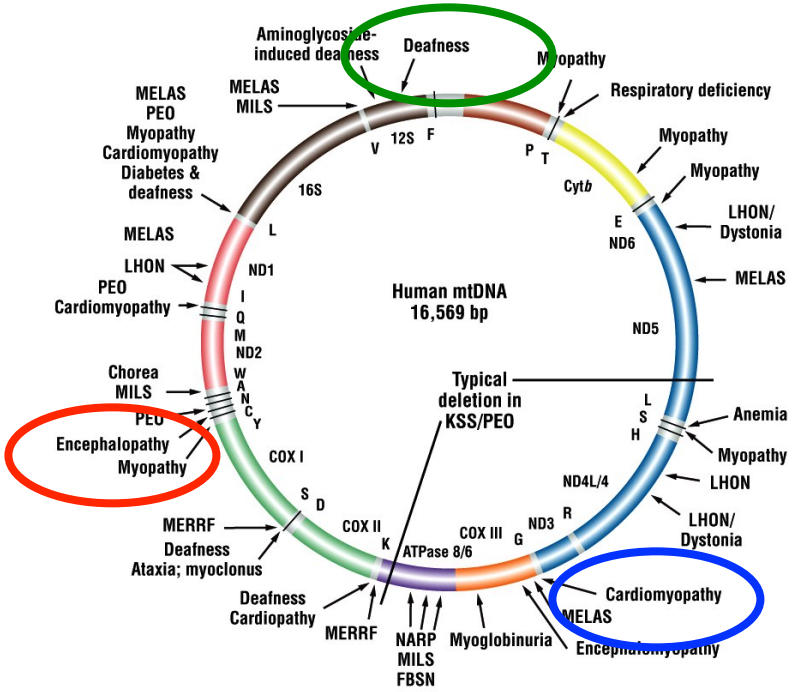
DNA Divides
Transcription
Translation

→ **Mitochondrial Proteins**

Mitochondrial Genes Are Inherited:

a. Paternally
b. Maternally

The Circular Mitochondrial Genome is Inherited Maternally



Disease Genes Present on the Mitochondrial Genome

Many Affect Muscles Because Mitochondria Produce Energy Needed For Muscle Activity

Diseases:

MERRF	Myoclonic epilepsy and ragged red fiber disease	MMC	Maternally inherited myopathy and cardiomyopathy
LHON	Leber hereditary optic neuropathy	PEO	Progressive external opthalmoplegia
NARP	Neurogenic muscle weakness, ataxia, and retinitis pigmentosum	KSS	Kearns–Sayre syndrome
MELAS	Mitochondrial encephalomyopathy, lactic acidosis, and strokelike symptoms	MILS	Maternally inherited Leigh syndrome

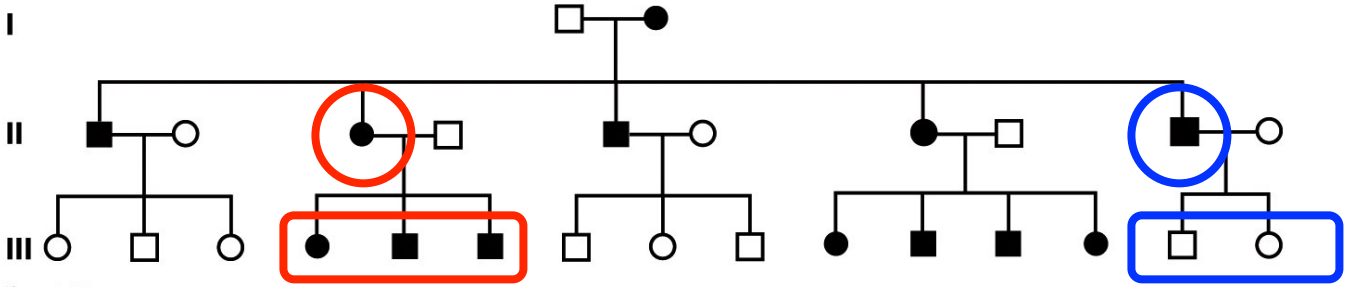
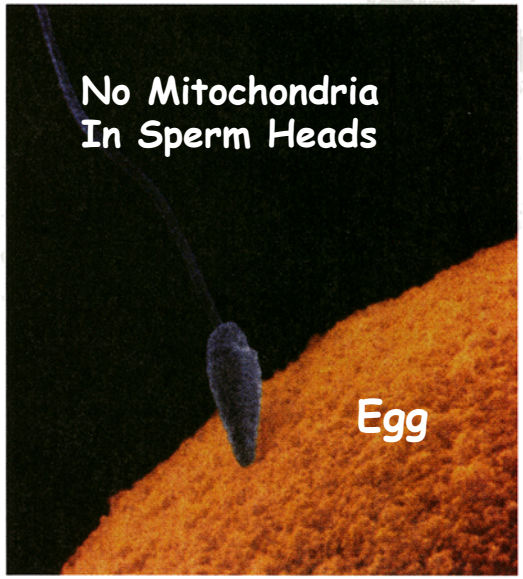


Figure 3-25
Introduction to Genetic Analysis, Ninth Edition

How Are Mitochondrial Gene Defects Inherited?



A human sperm and egg. The volume of the egg cell is about 5000 times the volume of the sperm head and contributes virtually all of the cytoplasm to the zygote, including the mitochondria. [D. W. Fawcett/Photo Researchers, Inc.]

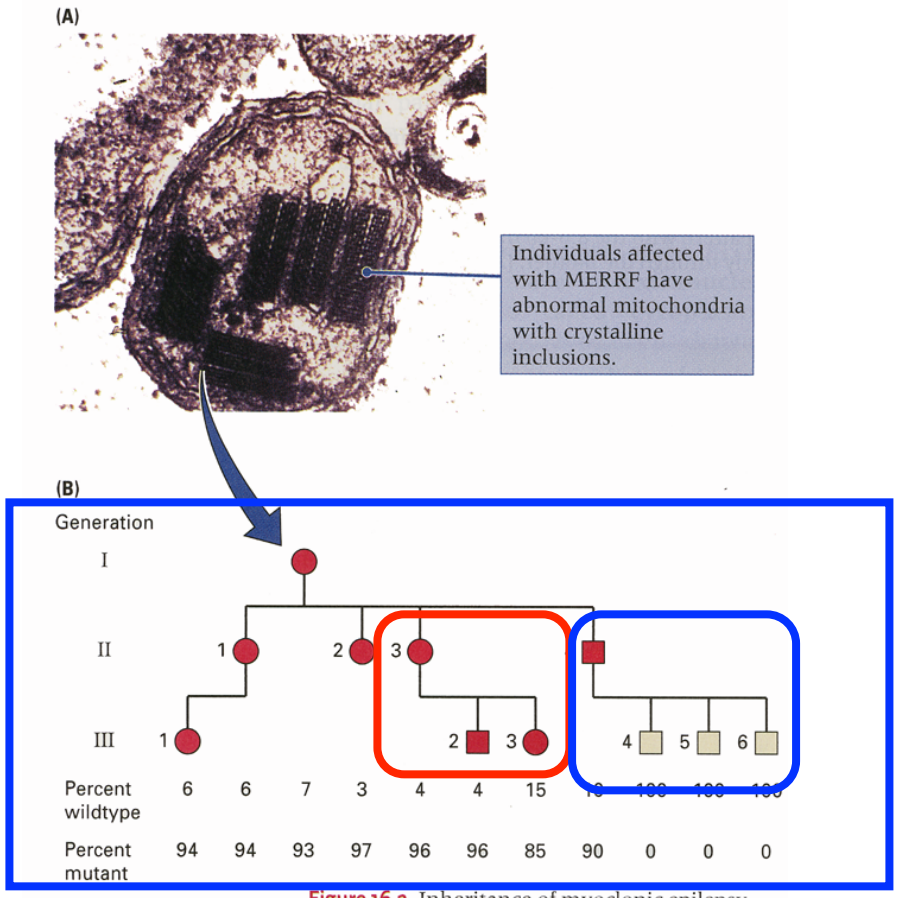


Figure 16.2 Inheritance of myoclonic epilepsy with ragged-red fiber disease (MERRF) in humans. (A) Electron micrograph of an abnormal MERRF mitochondrion containing paracrystalline inclusions. (B)The pedigree shows inheritance of MERRF in one family and the percentage of the mitochondria in each person found to be wildtype or mutant. [Micrograph courtesy of D. C. Wallace, from J. M. Shoffner, M. T. Lott, A. M. S. Lezza, P. Seibel, S. W. Ballinger, and D. C. Wallace. 1990. *Cell* 61: 931.]

Note Maternal Inheritance:
Disease Passed From Mother to All of Her Children and Not Passed on By Father

RFLPs Can Be Used to Identify Individuals Using Mitochondrial DNAs

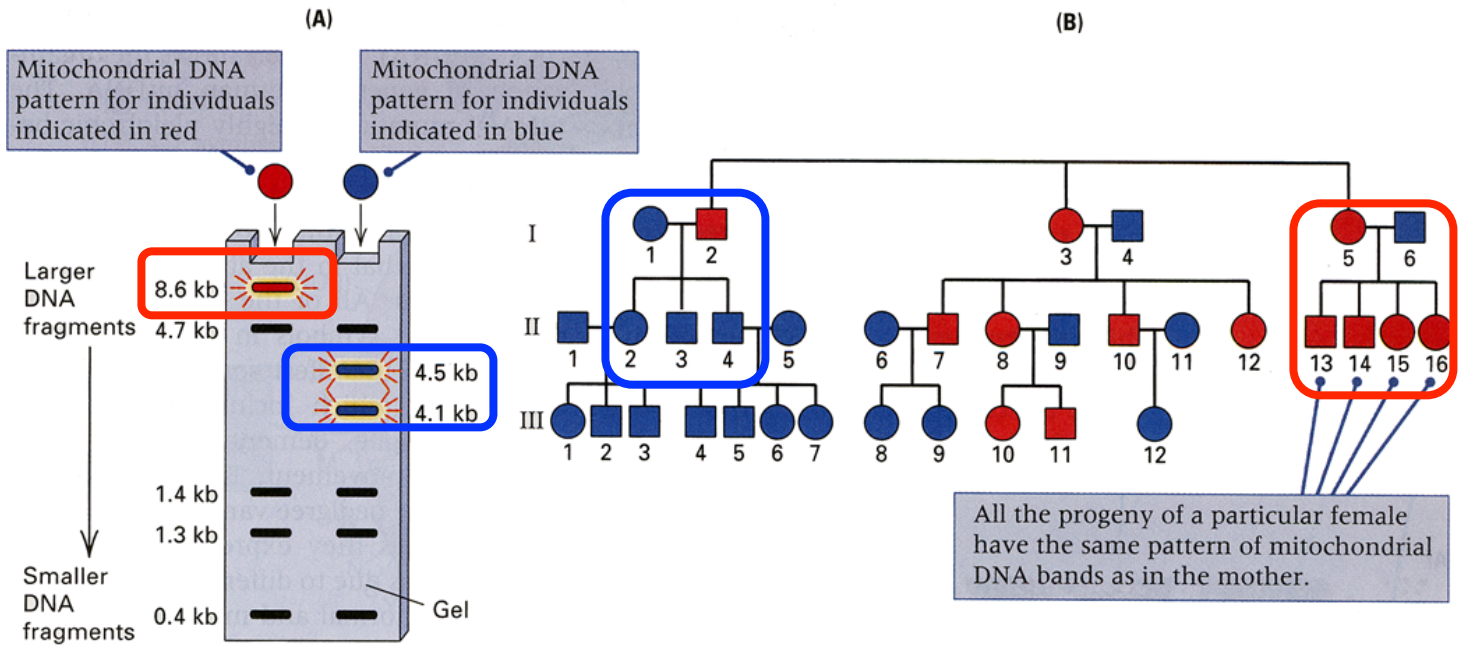


Figure 16.1 Maternal inheritance of human mitochondrial DNA. (A) Pattern of DNA fragments obtained when mitochondrial DNA is digested with the restriction enzyme *Hae*II. The DNA type at the left includes a fragment of 8.6 kb (red). The DNA type at the right contains a cleavage site for *Hae*II within the 8.6-kb fragment, which results in smaller fragments of 4.5 kb and 4.1 kb (blue). (B) Pedigree showing maternal inheritance of the DNA pattern with the 8.6-kb fragment (red symbols). The mitochondrial DNA type is transmitted only through the mother. [After D. C. Wallace. 1989. *Trends in Genetics* 5: 9.]

Note How Mitochondrial RFLP Markers Are Inherited !!

NUCLEAR TRANSPLANTATION

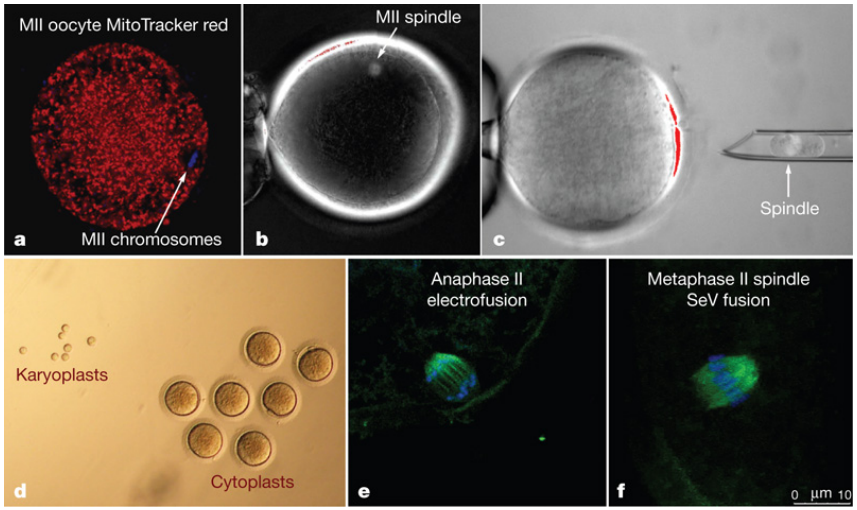
Researchers Prevent Inheritance of Faulty Mitochondria in Monkeys

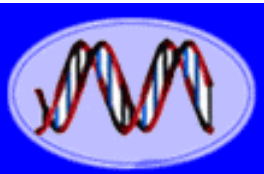
Vol 461|17 September 2009|doi:10.1038/nature08368 nature

Nature 461, September 17, 2009

ARTICLES

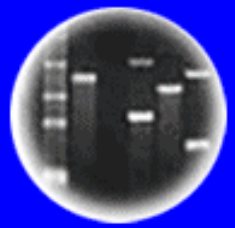
Mitochondrial gene replacement in primate offspring and embryonic stem cells





DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria

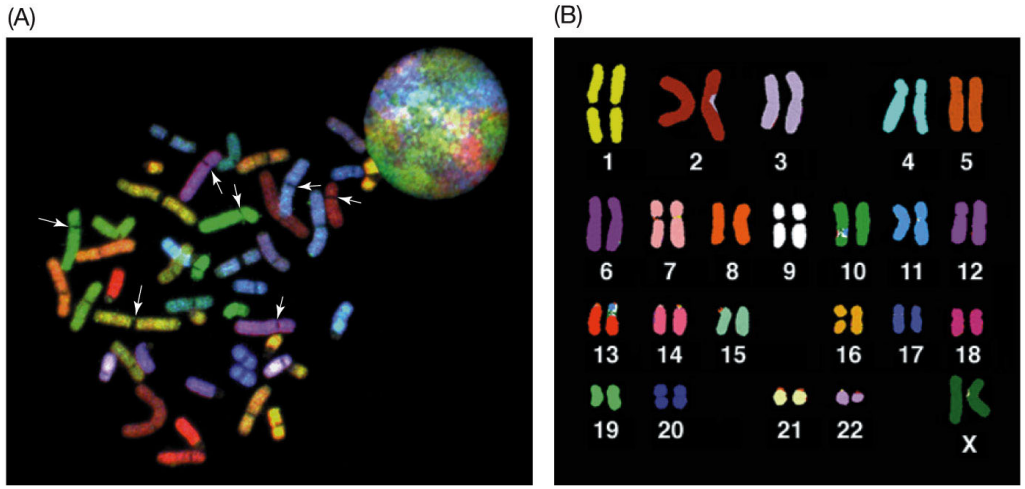


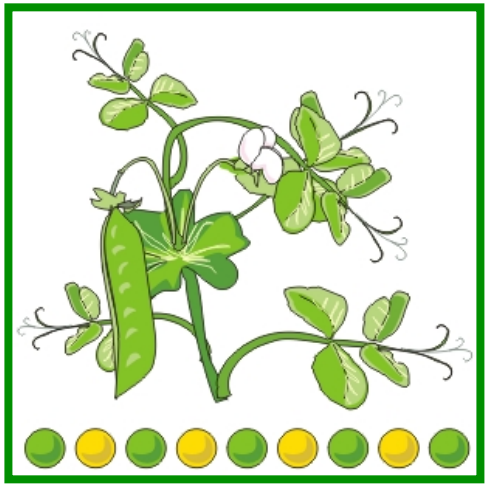
DNA Fingerprinting

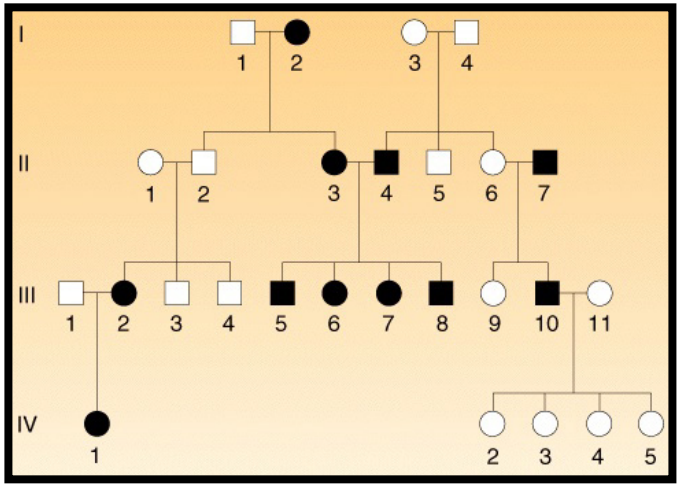
Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

The Nuclear Genome







The Human Genome Was Sequenced Ten Years Ago!

The Human Genome Project

ws
Print"

The New York Times

National Edition

Arizona and New Mexico... h
cloudy in New Mexico, thunderst
in the mountains. Partly sunny
where. Highs 80 mountains, ove
deserts. Weather map is on Pag

. No. 51,432 TUESDAY, JUNE 27, 2000 ONE DOLL

tic Code of Human Life Is Cracked by Scientist

become part that Congress was entitled to the last word because Miranda's presump-

The Book of Life

The 3 billion base pairs ... of the interweaving double helix of DNA ... that make up the set of chromosomes in our cells, have been sequenced.

BASE PAIRS Rungs between the strands of the double helix

BASES
A adenine
C cytosine
G guanine
T thymine

By ordering the base units, scientists hope to locate the genes and determine their functions.

The New York Times

A SHARED SUCCI

2 Rivals' Announcem
Marks New Medic
Era, Risks and All

By NICHOLAS WADE

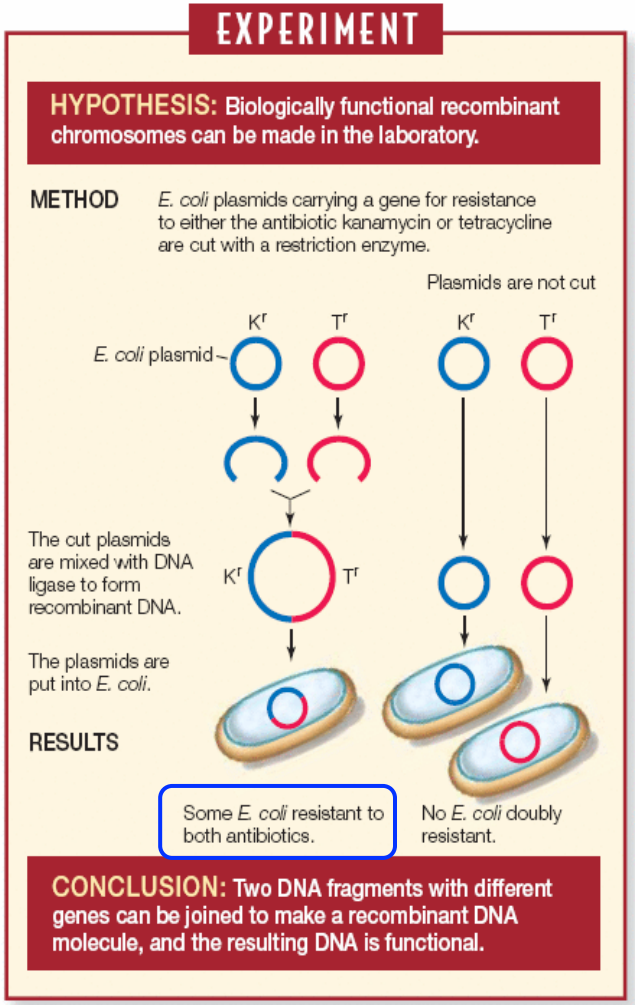
WASHINGTON, June 26 — I
achievement that represents
nacle of human self-knowledge
rival groups of scientists said
that they had deciphered the h
tary script, the set of instruc
that defines the human organis

Public & Private Effort Using Different Strategies - A Race!

3 Billion Dollars & Took 15 Years

The Human Genome Could Not Have Been Sequenced Without The Invention of Genetic Engineering

Cohen & Boyer Experiment That "Invented" Genetic Engineering



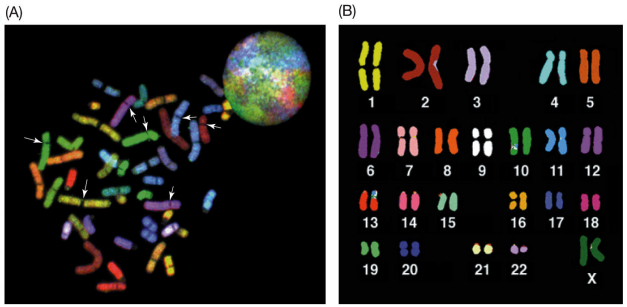
Genes Need to Be Cloned Before They Can Be Sequenced!!!!!!!

The Age of Genomics is a Result of the Age of Genetic Engineering

Only A Small Fraction of the Human Genome Encodes Proteins

TABLE 18.1 Classes of DNA Sequences Found in the Human Genome

Class	Frequency (%)	Description
Protein-encoding genes	1.5	Translated portions of the 25,000 genes scattered about the chromosomes
Introns	24	Noncoding DNA that constitutes the great majority of each human gene
Segmental duplications	5	Regions of the genome that have been duplicated
Pseudogenes (inactive genes)	2	Sequence that has characteristics of a gene but is not a functional gene
Structural DNA	20	Constitutive heterochromatin, localized near centromeres and telomeres
Simple sequence repeats	3	Stuttering repeats of a few nucleotides such as CGG, repeated thousands of times
Transposable elements	45	21%: Long interspersed elements (LINEs), which are active transposons 13%: Short interspersed elements (SINEs), which are active transposons 8%: Retrotransposons, which contain long terminal repeats (LTRs) at each end 3%: DNA transposon fossils



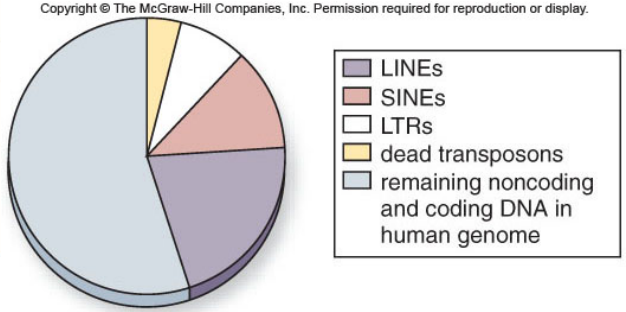
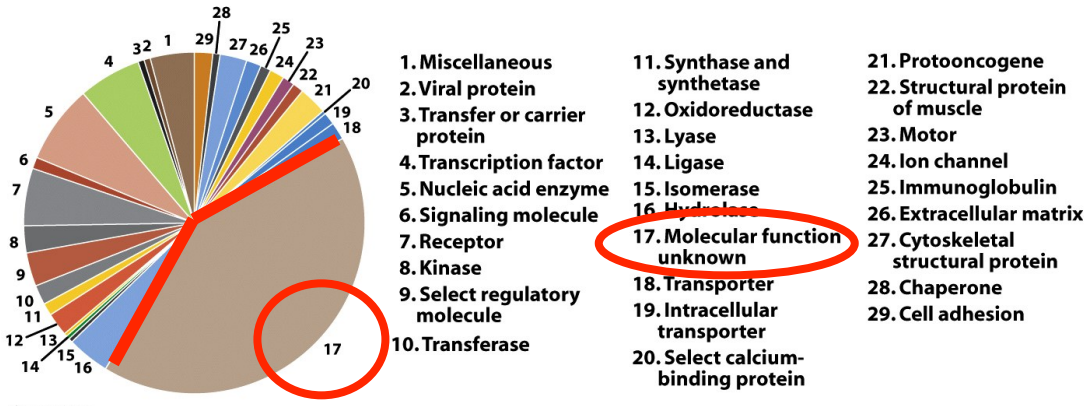


Table 20.6 Average characteristics of genes in the human genome

Characteristic	Average
Number of exons	8.8
Size of internal exon	145 bp
Size of intron	3,365 bp
Size of 5′ untranslated region	300 bp
Size of 3′ untranslated region	770 bp
Size of coding region	1,340 bp
Total length of gene	27,000 bp

Table 20-6

The Human Genome Has DNA Sequences Present Once As Well as Repeated Many Times



Human Genes are Large but Contain Mostly Introns

Characteristics of the Human Genome

Table 4–1 Some Vital Statistics for the Human Genome

	HUMAN GENOME
DNA length	3.2×10^9 nucleotide pairs*
Number of genes	approximately 25,000
Largest gene	2.4×10^6 nucleotide pairs
Mean gene size	27,000 nucleotide pairs
Smallest number of exons per gene	1
Largest number of exons per gene	178
Mean number of exons per gene	10.4
Largest exon size	17,106 nucleotide pairs
Mean exon size	145 nucleotide pairs
Number of pseudogenes**	more than 20,000
Percentage of DNA sequence in exons (protein coding sequences)	1.5%
Percentage of DNA in other highly conserved sequences***	3.5%
Percentage of DNA in high-copy repetitive elements	approximately 50%

* The sequence of 2.85 billion nucleotides is known precisely (error rate of only about one in 100,000 nucleotides). The remaining DNA primarily consists of short highly repeated sequences that are tandemly repeated, with repeat numbers differing from one individual to the next.

** A pseudogene is a nucleotide sequence of DNA closely resembling that of a functional gene, but containing numerous mutations that prevent its proper expression. Most pseudogenes arise from the duplication of a functional gene followed by the accumulation of damaging mutations in one copy.

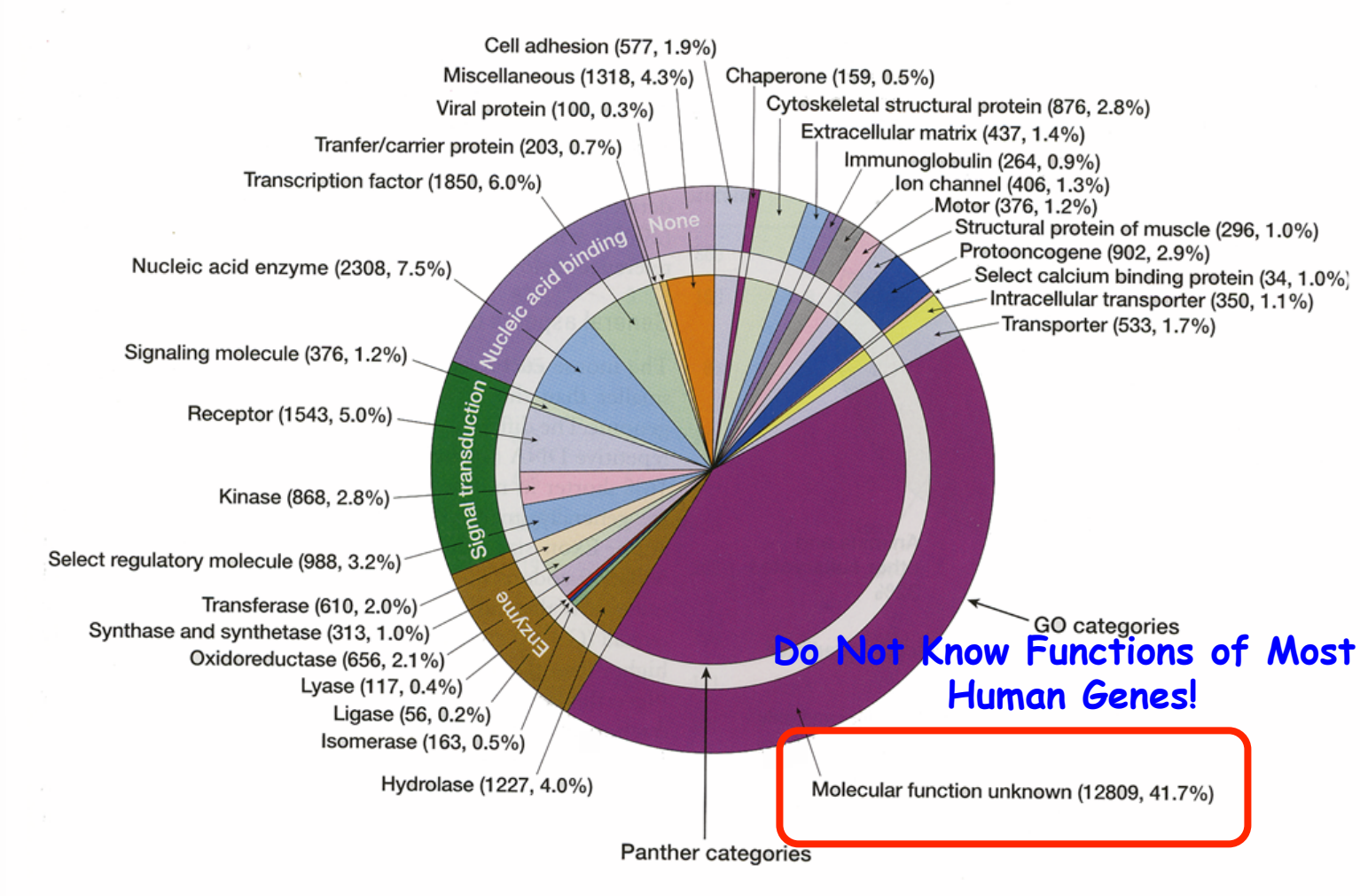
*** Preserved functional regions; these include DNA encoding 5′ and 3′ UTRs (untranslated regions), structural and functional RNAs, and conserved protein-binding sites on the DNA.

Duchenne Muscular Dystophy

Smallest Gene is 252 bp & Ecodes an Insulin-like Growth factor

Table 4-1 *Molecular Biology of the Cell* (© Garland Science 2008)

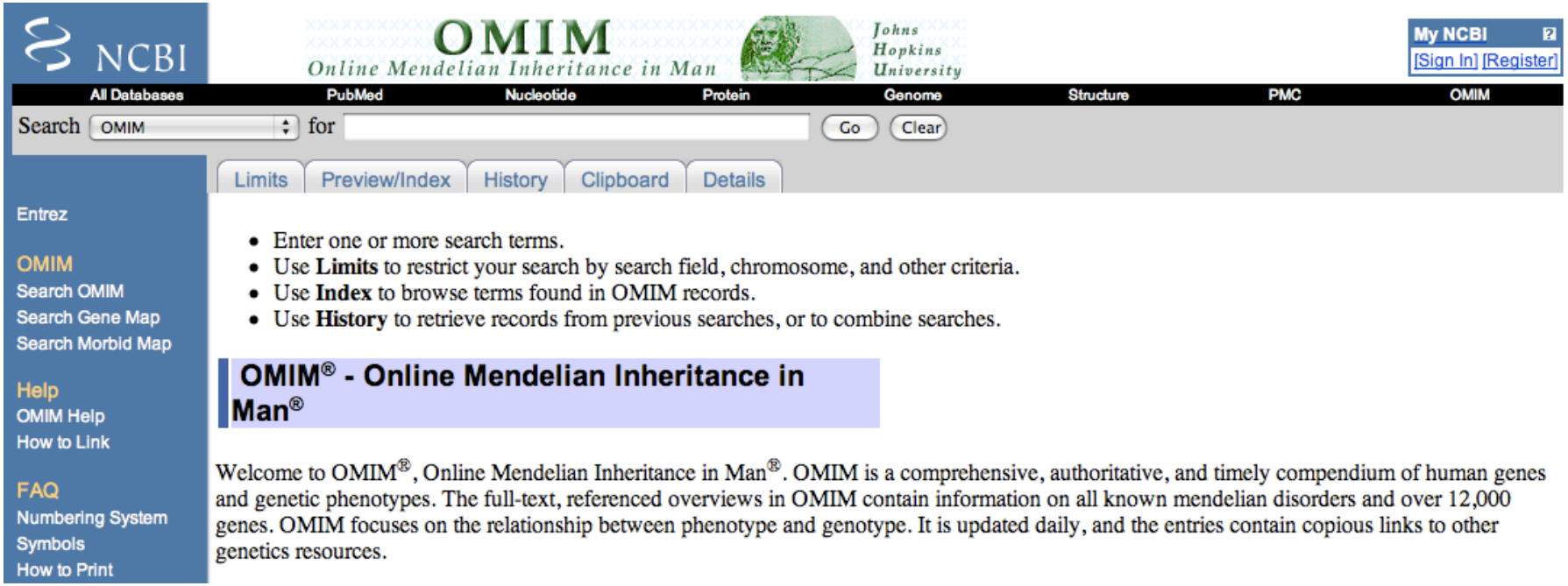
The Human Genome Contains ~25,000 Different Genes



Do Not Know Functions of Most Human Genes!

How Many Encoded Proteins? Alternative Splicing?

How Many Human Disease Genes Have Been Identified?



There are ~25,000 Genes in The Human Genome

1. ~2,700 Genes Correlate With a Disease Phenotype
2. The Molecular Basis of 90% of These Genetic Diseases Are Known (e.g., Sickle Cell Anemia, Hemophilia A)

Nature Education 1(1), (2008)

http://www.nature.com/scitable

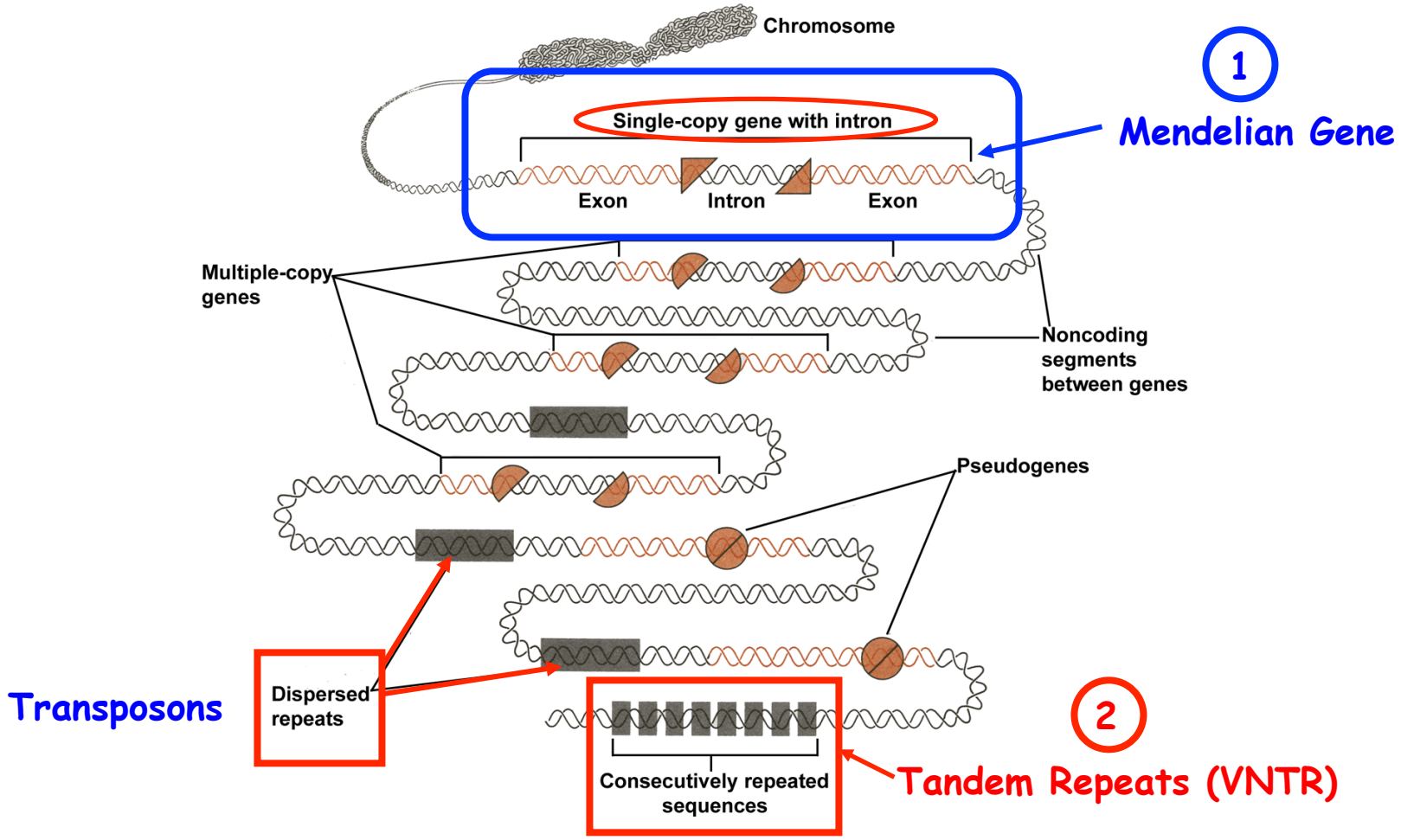
Examples of Human Disease Genes That Are Known

Table 1: Examples of Human Diseases, Modes of Inheritance, and Associated Genes

Disease	Type of Inheritance	Gene Responsible
Phenylketonuria (PKU)	Autosomal recessive	Phenylalanine hydroxylase (*PAH*)
Cystic fibrosis	Autosomal recessive	Cystic fibrosis conductance transmembrane regulator (*CFTR*)
Sickle-cell anemia	Autosomal recessive	Beta hemoglobin (*HBB*)
Albinism, oculocutaneous, type II	Autosomal recessive	Oculocutaneous albinism II (*OCA2*)
Huntington's disease	Autosomal dominant	Huntingtin (*HTT*)
Myotonic dystrophy type 1	Autosomal dominant	Dystrophia myotonica-protein kinase (*DMPK*)
Hypercholesterolemia, autosomal dominant, type B	Autosomal dominant	Low-density lipoprotein receptor (*LDLR*); apolipoprotein B (*APOB*)
Neurofibromatosis, type 1	Autosomal dominant	Neurofibromin 1 (NF1)
Polycystic kidney disease 1 and 2	Autosomal dominant	Polycystic kidney disease 1 (*PKD1*) and polycystic kidney disease 2 (*PKD2*), respectively
Hemophilia A	X-linked recessive	Coagulation factor VIII (*F8*)
Muscular dystrophy, Duchenne type	X-linked recessive	Dystrophin (*DMD*)
Hypophosphatemic rickets, X-linked dominant	X-linked dominant	Phosphate-regulating endopeptidase homologue, X-linked (*PHEX*)
Rett's syndrome	X-linked dominant	Methyl-CpG-binding protein 2 (*MECP2*)
Spermatogenic failure, nonobstructive, Y-linked	Y-linked	Ubiquitin-specific peptidase 9Y, Y-linked (*USP9Y*)

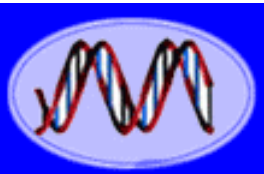
Genetic Tests Exist For These Disease Genes

The Human Genome Landscape



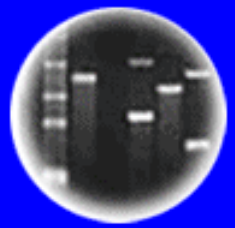
Tandem Repeats or VNTRs are Useful for DNA Fingerprinting Studies!

e.g., DIS80 Locus For Class DNA Fingerprint on Chromosome 4 Core = 16bp



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

Comparative Genomics

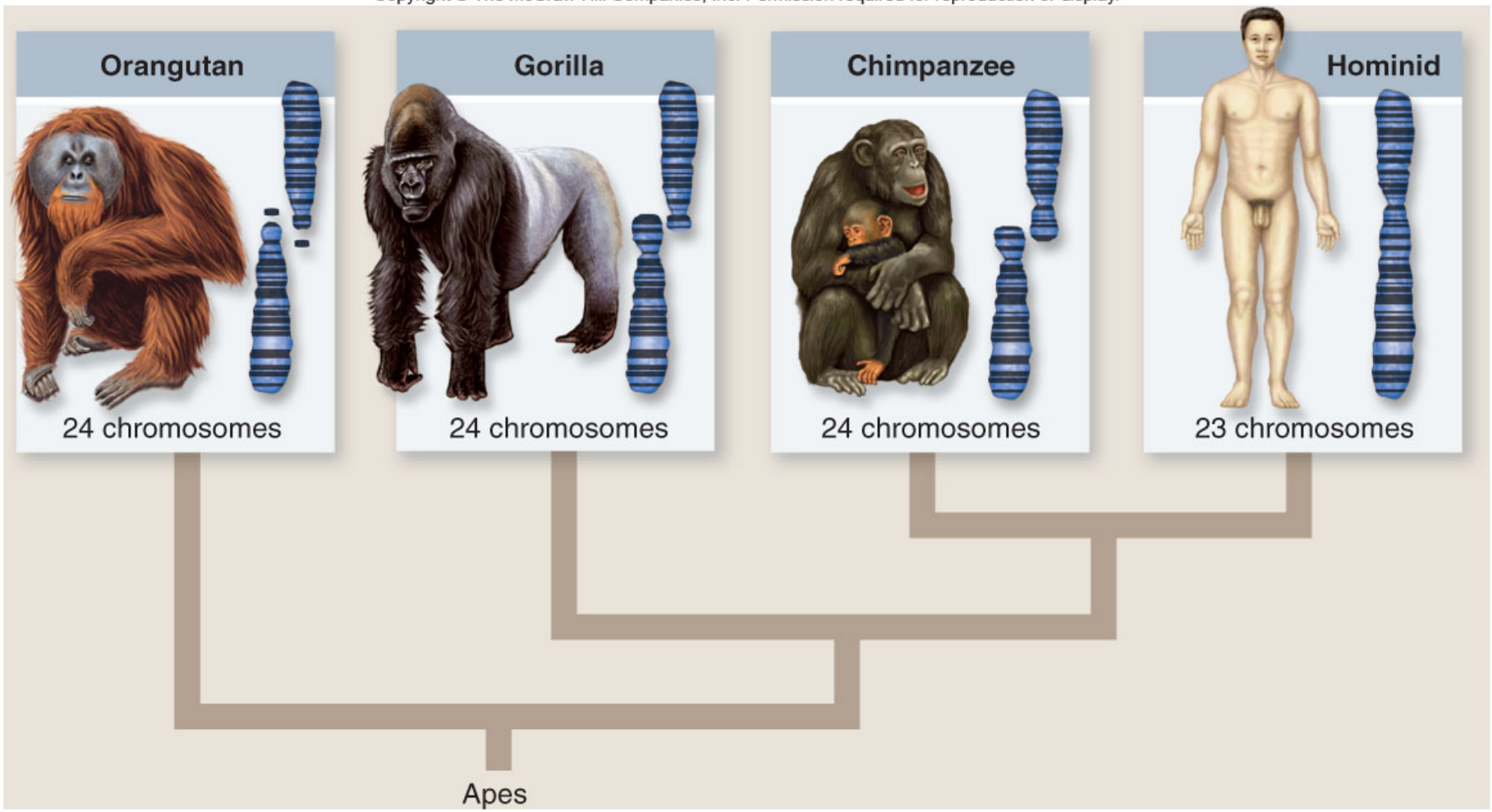
Many Mammalian Genomes Have Been Sequenced And More Are Being Sequenced

Human	Rabbit
Mouse	Rat
Dog	Ground Squirrel
Cow	Tree Shrew
Guinea Pig	Dolphin
Sloth	Chimpanzee
Armadillo	Gorilla
Kangaroo Rat	Orangutan
Horse	Rhesus Monkey
Cat	Wallaby

+ Fifty Individual Human Genomes Including James Watson
Because of Major Breakthroughs in Sequencing Technology

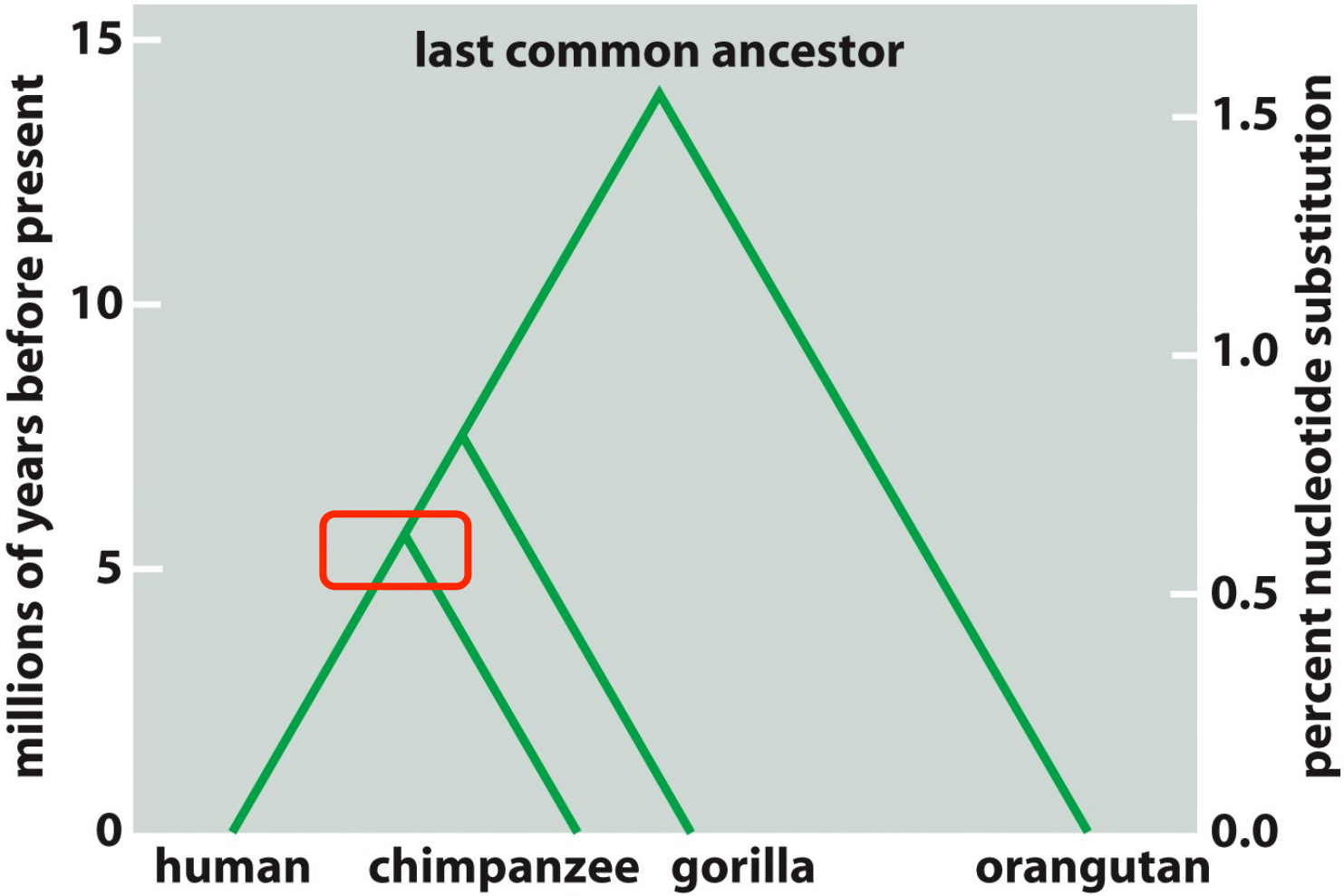
2010

Comparison of Mammalian Genomes Attempts To Determine "What Makes a Man, a Man and a Mouse a Mouse"



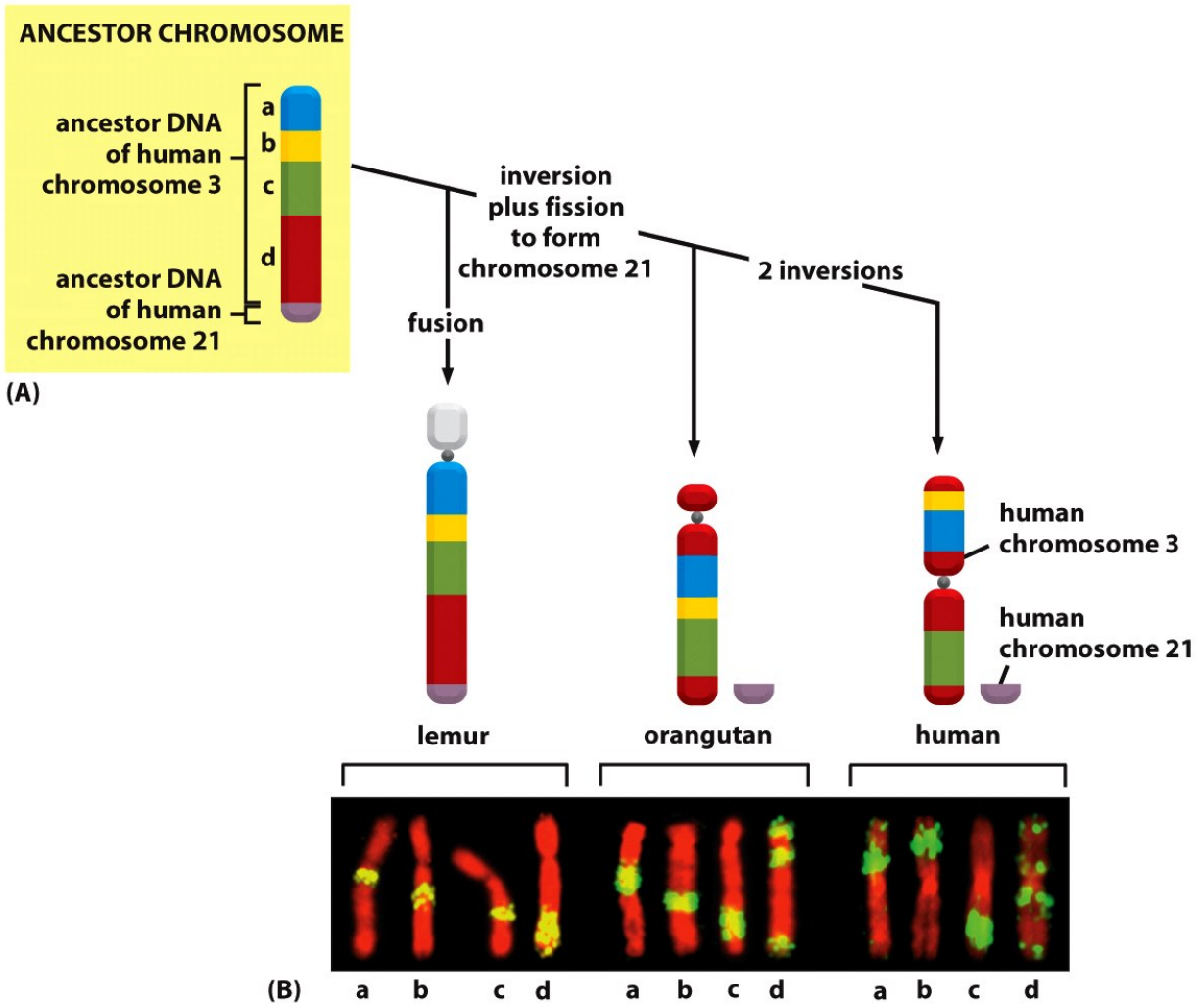
There is <1% Difference Between Human & Chimpanzee DNAs!

Comparison Between Primate Genomes

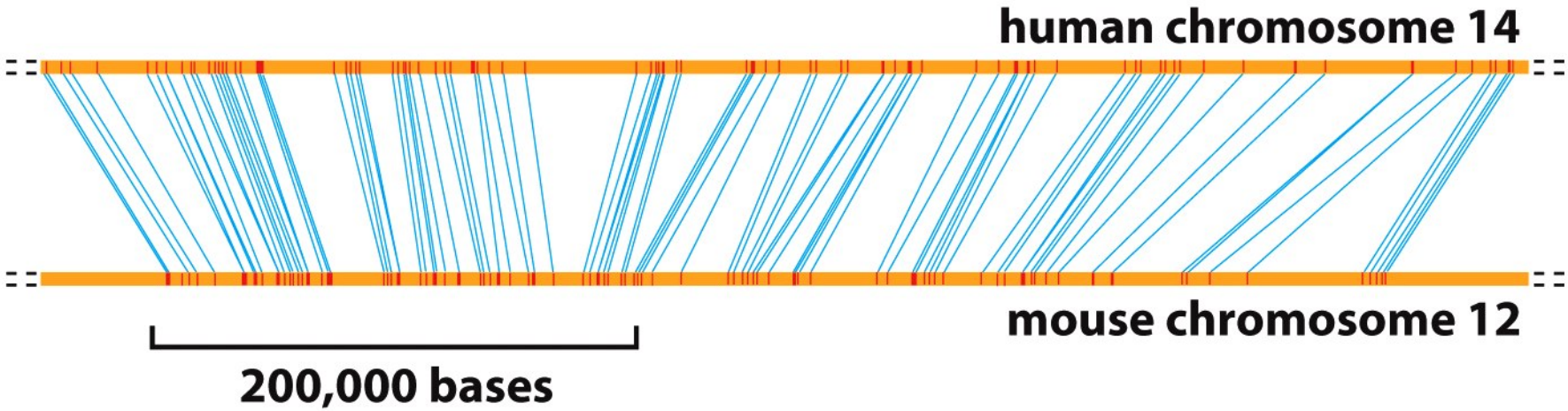


Note the Small Sequence Differences in These Genomes-
What Makes a "Human a Human?"

Comparative Genomics Can Uncover the Origin of Human Chromosomes and Relationship to Other Mammalian Chromosomes

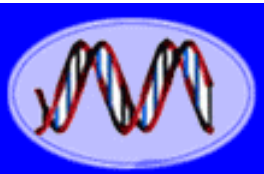


Comparative Genomics Can Align Related Genes in Two Different Genomes



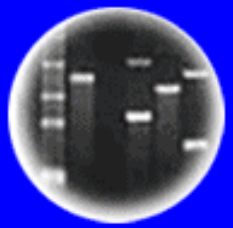
Note "Synteny," or Alignment, of Related Genes Between Human and Mouse Chromosome Regions

What Does This Say About Genome Evolution?



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria

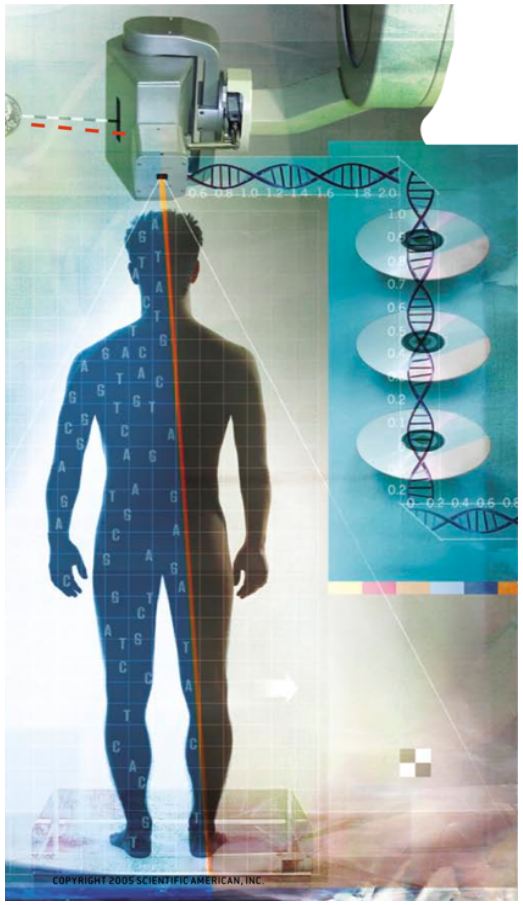


DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

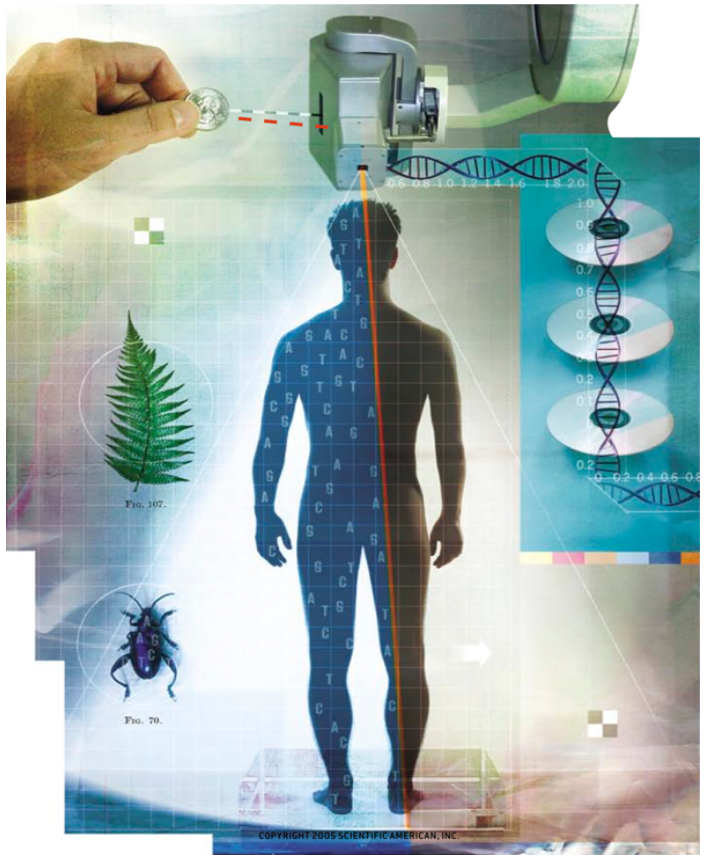
The Personalized Genome



The Ultimate Measure of Individuality
Personal Genome Sequence & Comparing Individual Human Genomes

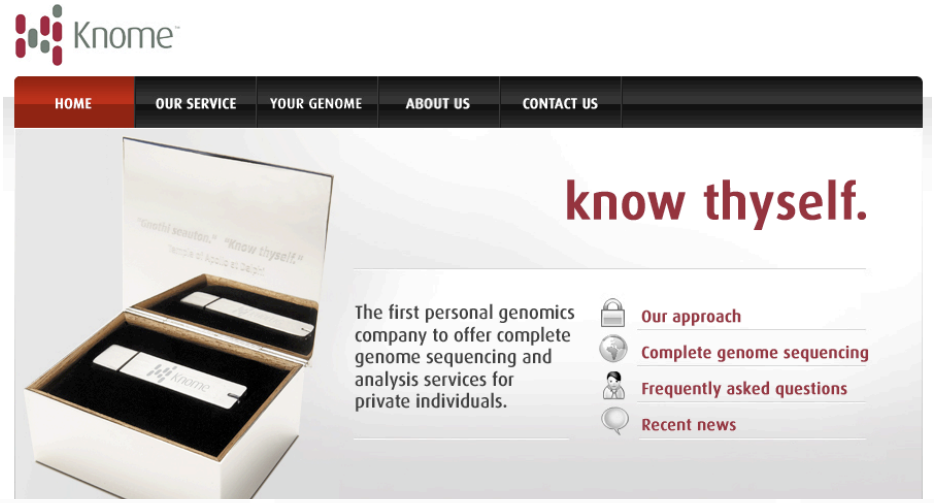
Genomes for ALL

Next-generation technologies that make
reading DNA fast, cheap and widely accessible
are coming in less than a decade.
Their potential to revolutionize research
and bring about the era of truly personalized medicine
means the time to start preparing is now



Find DNA Variability in
<u>All</u> Genes & Associate
with Specific Traits!

The Age of Personal Genomics Has Begun!



$50,000-Soon Down to $5,000

Problems?

Genome of DNA Discoverer Is Deciphered

By NICHOLAS WADE
Published: June 1, 2007

The full genome of James D. Watson, who jointly discovered the structure of DNA in 1953, has been deciphered, marking what some scientists believe is the gateway to an impending era of personalized genomic medicine.

LETTERS

Nature April 17, 2008

The complete genome of an individual by massively parallel DNA sequencing

February 11, 2010

Whole Genome of Ancient Human Is Decoded

Vol 463 | 11 February 2010 | doi:10.1038/nature08835 nature

ARTICLES

Ancient human genome sequence of an extinct Palaeo-Eskimo

Morten Rasmussen[1,2]*, Yingrui Li[2,3]*, Stinus Lindgreen[1,4]*, Jakob Skou Pedersen[4], Anders Albrechtsen[4], Ida Moltke[4], Mait Metspalu[5], Ene Metspalu[5], Toomas Kivisild[5,6], Ramneek Gupta[7], Marcelo Bertalan[7], Kasper Nielsen[7], M. Thomas P. Gilbert[1,2], Yong Wang[8], Maanasa Raghavan[1,9], Paula F. Campos[1], Hanne Munkholm Kamp[1,4], Andrew S. Wilson[10], Andrew Gledhill[10], Silvana Tridico[11,12], Michael Bunce[12], Eline D. Lorenzen[1], Jonas Binladen[1], Xiaosen Guo[2,3], Jing Zhao[2,3], Xiuqing Zhang[2,3], Hao Zhang[2,3], Zhuo Li[2,3], Minfeng Chen[2,3], Ludovic Orlando[13], Karsten Kristiansen[2,3,4], Mads Bak[14], Niels Tommerup[14], Christian Bendixen[15], Tracey L. Pierre[16], Bjarne Grønnow[17], Morten Meldgaard[18], Claus Andreasen[19], Sardana A. Fedorova[5,20], Ludmila P. Osipova[21], Thomas F. G. Higham[9], Christopher Bronk Ramsey[10], Thomas v. O. Hansen[22], Finn C. Nielsen[22], Michael H. Crawford[23], Søren Brunak[7,24], Thomas Sicheritz-Pontén[7], Richard Villems[5], Rasmus Nielsen[4,8], Anders Krogh[2,4], Jun Wang[2,3,4] & Eske Willerslev[1,2]

We report here the genome sequence of an ancient human. Obtained from ~4,000-year-old permafrost-preserved hair, the genome represents a male individual from the first known culture to settle in Greenland. Sequenced to an average depth of 20×, we recover 79% of the diploid genome, an amount close to the practical limit of current sequencing technologies. We identify 353,151 high-confidence single-nucleotide polymorphisms (SNPs), of which 6.8% have not been reported previously. We estimate raw read contamination to be no higher than 0.8%. We use functional SNP assessment to assign possible phenotypic characteristics of the individual that belonged to a culture whose location has yielded only trace human remains. We compare the high-confidence SNPs to those of contemporary populations to find the populations most closely related to the individual. This provides evidence for a migration from Siberia into the New World some 5,500 years ago, independent of that giving rise to the modern Native Americans and Inuit.

From 5,000 Year-Old Hair!

How Determine Phenotype?!

nature Vol 444|16 November 2006|doi:10.1038/nature05336

ARTICLES

Nature, November, 2006

Analysis of one million base pairs of Neanderthal DNA

From a 45,000 Year-Old Bone

Richard E. Green[1], Johannes Krause[1], Susan E. Ptak[1], Adrian W. Briggs[1], Michael T. Ronan[2], Jan F. Simons[2], Lei Du[2], Michael Egholm[2], Jonathan M. Rothberg[2], Maja Paunovic[3]‡ & Svante Pääbo[1]

What About the Future?

The 1,000 Genomes Project Will Provide Novel Insight in Human Genomes, Ancestry, & Disease Genes

Only Possible Using New Sequencing Methods

1,000 Genomes

Gene-sequencing projects keep getting bigger.
Tuesday, January 22, 2008
By Emily Singer

In a testament to the steady plummet in sequencing costs, today the National Human Genome Research Institute (NHGRI) announced a massive international collaboration to sequence the genomes of 1,000 people from around the world.

According to the NHGRI statement,

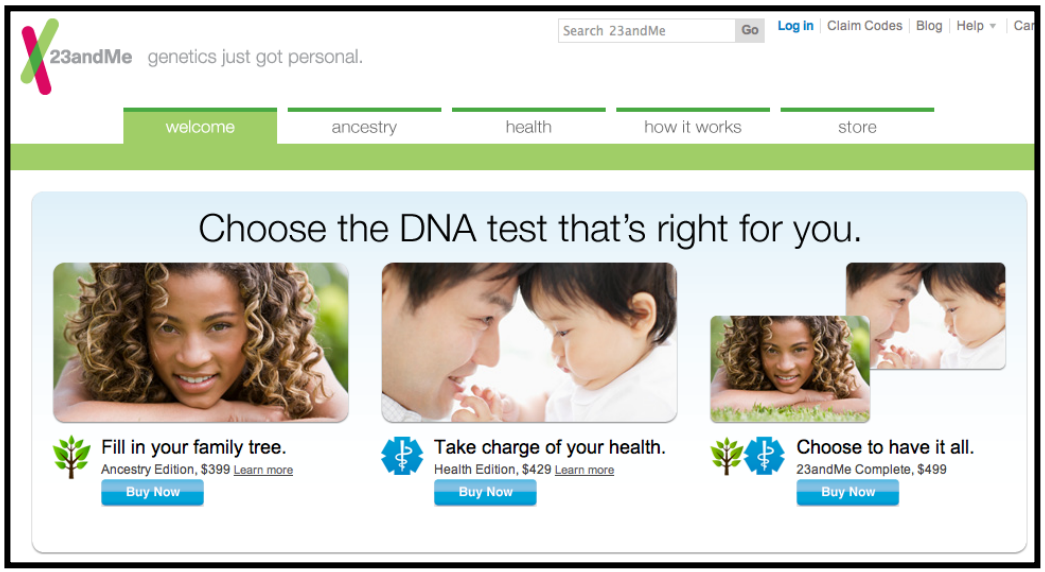
> "The 1000 Genomes Project will examine the human genome at a level of detail that no one has done before," said Richard Durbin, Ph.D., of the Wellcome Trust Sanger Institute, who is co-chair of the consortium. "Such a project would have been unthinkable only two years ago. Today, thanks to amazing strides in sequencing technology, bioinformatics and population genomics, it is now within our grasp. So we are moving forward to build a tool that will greatly expand and further accelerate efforts to find more of the genetic factors involved in human health and disease."
>
> During its two-year production phase, the 1000 Genomes Project will deliver sequence data at an average rate of about 8.2 billion bases per day, the equivalent of more than two human genomes every 24 hours. The volume of data--and the interpretation of those data--will pose a major challenge for leading experts in the fields of bioinformatics and statistical genetics.

2 Human Genomes Every 24 hrs !

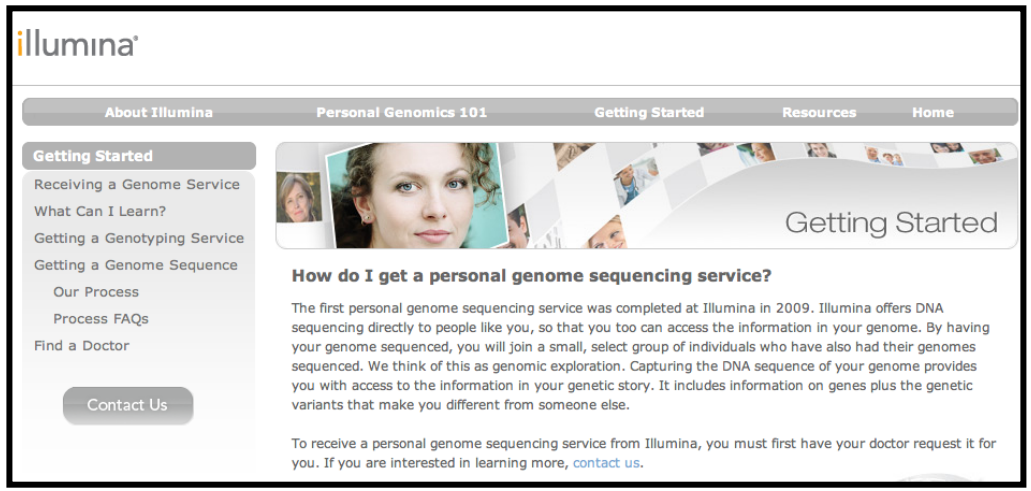
The 1,000 volunteers will be selected from those who participated in the HapMap project, a map of common genetic variation (see "A New Map for Health"), and will include:

Yoruba in Ibadan, Nigeria; Japanese in Tokyo; Chinese in Beijing; Utah residents with ancestry from northern and western Europe; Luhya in Webuye, Kenya; Maasai in Kinyawa, Kenya; Toscani in Italy; Gujarati Indians in Houston; Chinese in metropolitan Denver; people of Mexican ancestry in Los Angeles; and people of African ancestry in the southwestern United States.

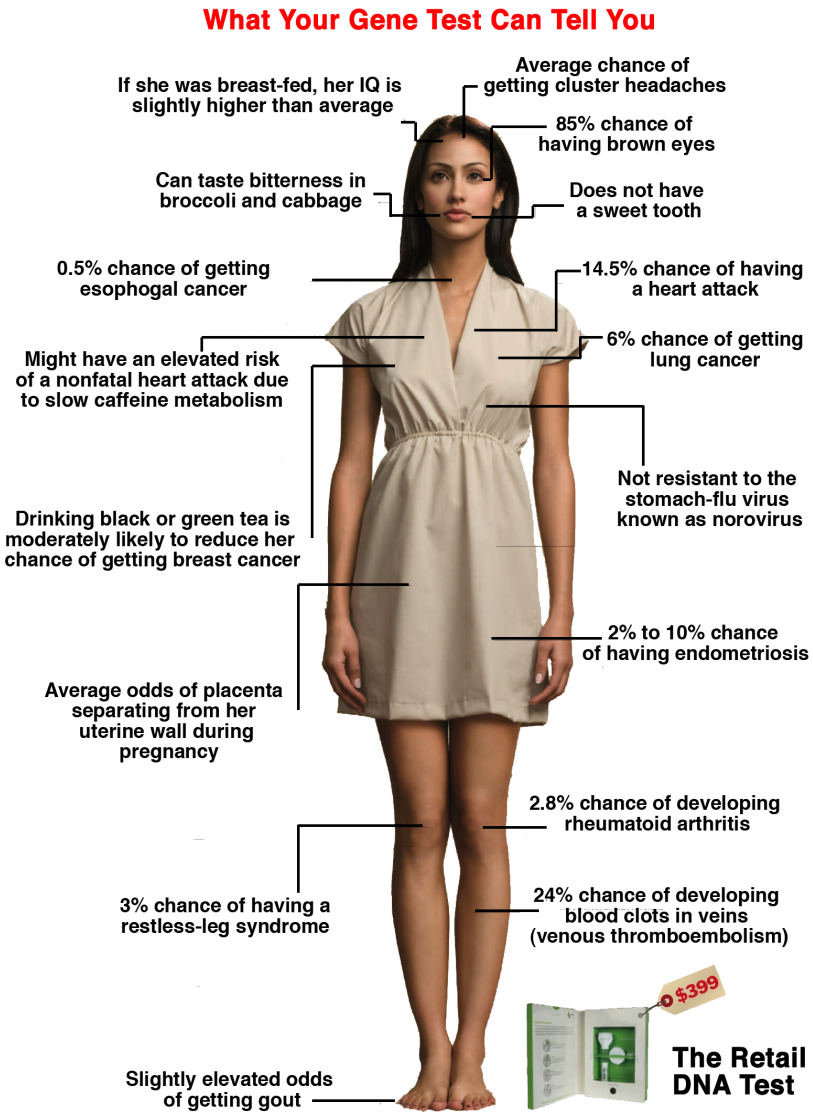
Personal Genome Sequencing Companies



$5,000 Genome!



Time Magazine 2008 - Invention of the Year
Your Personal Genome - 23andMe®



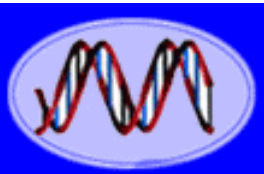
Invention Of the Year

And Before Birth!!!

https://www.23andme.com/

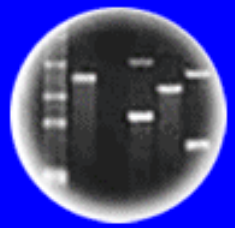
The Problems With Human Genome Sequencing Companies Are

a. Reliability of Results?
b. Privacy?
c. What To Do With Information Obtained?
d. Regulatory Oversight?
e. All of Above?



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria

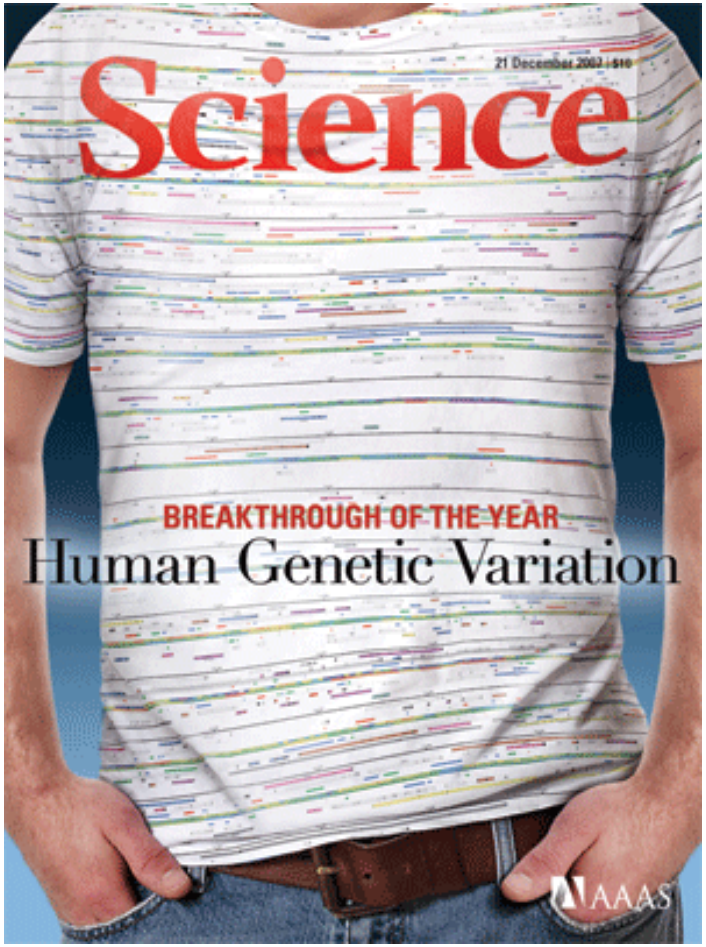


DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

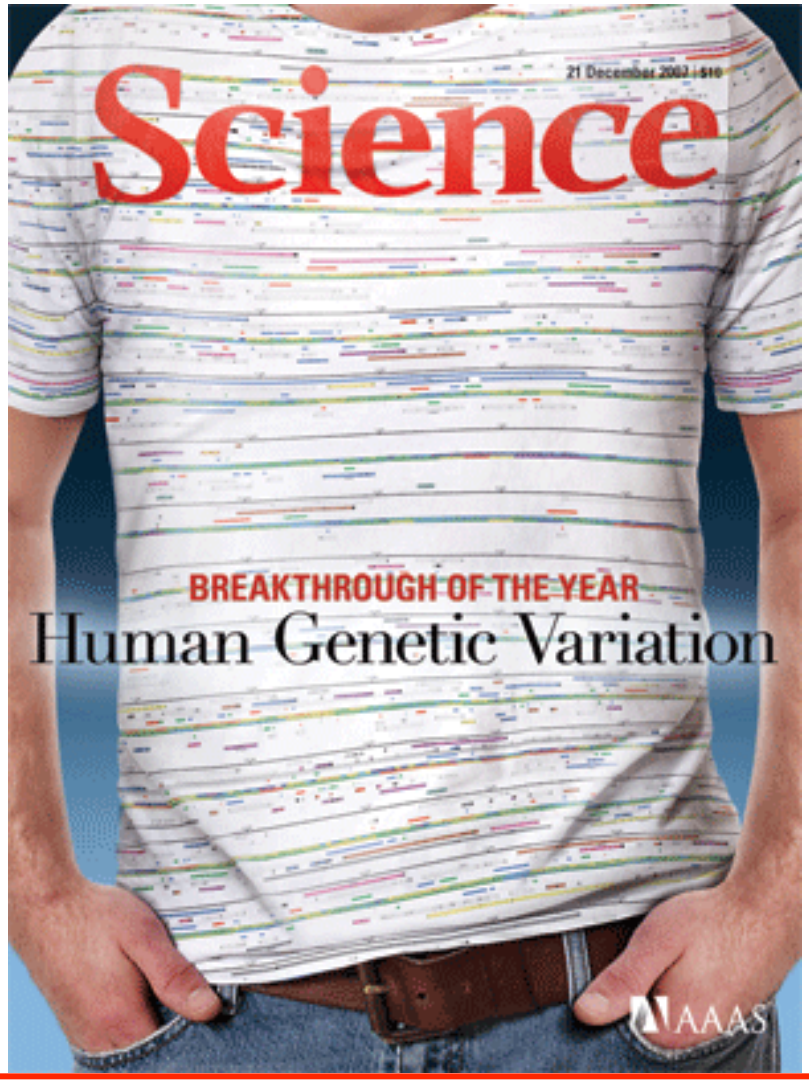
Plants of Tomorrow

Individual DNA Sequence Variability



There Are Large DNA Sequence Variations in Human Populations

Variation in Genes (e.g., Disease Genes) Accounts For Only a Small Amount of Human DNA Variation



DNA Sequence Variation Makes us Individuals! Genetic Variability-Allelic Differences

There Are Large DNA Sequence Variations in Animal Populations

Vol 438|8 December 2005 nature

Nature, December 2005

NEWS & VIEWS

C. COLLINS/CORBIS

GENOMICS

The dog has its day

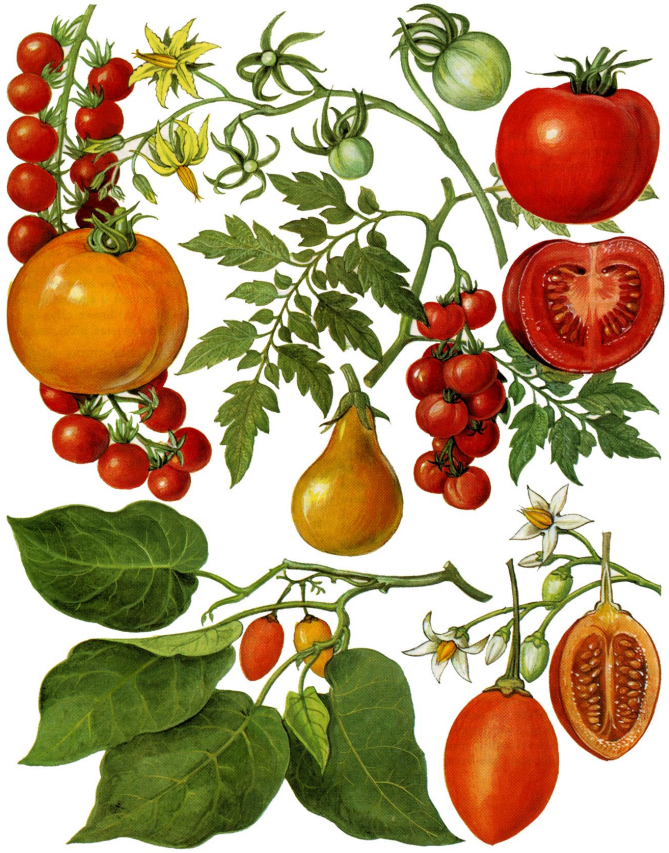
Hans Ellegren

Domestication and selective breeding have transformed wolves into the diversity of dogs we see today. The sequence of the genome of one breed adds to our understanding of mammalian biology and genome evolution.

The Dog Genome Has Been Sequenced!

Breeding Uses Natural DNA Sequence Variability of Genes As Raw Material - Variability Generated by Mutations

Tomato Genetic Diversity



Mutations in a Gene That Change Its Chemical Sequence & Slightly Alters Its Function (e.g., fruit size, color)

Alleles And Homologous Chromosomes-A Reminder

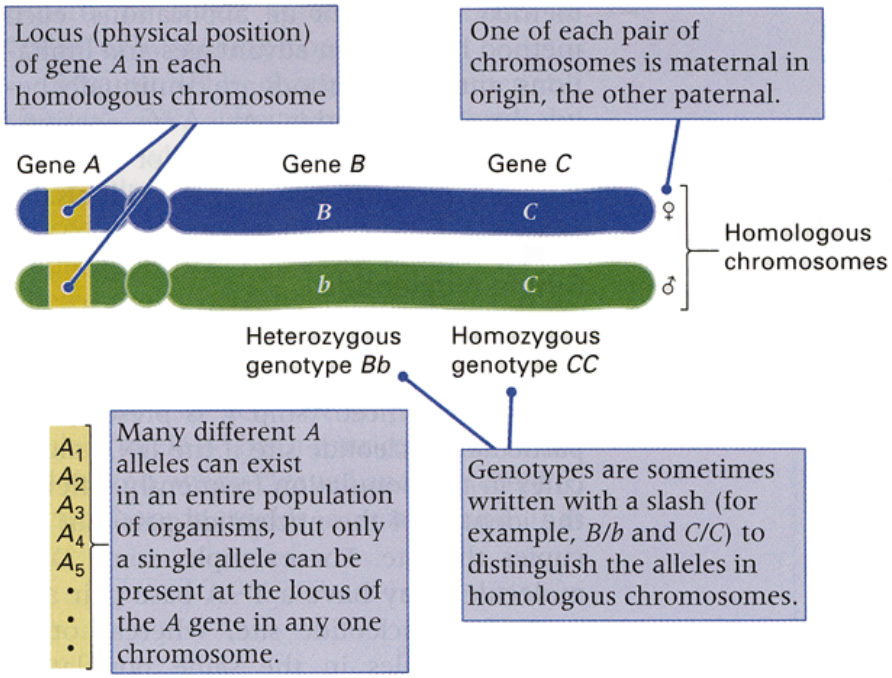


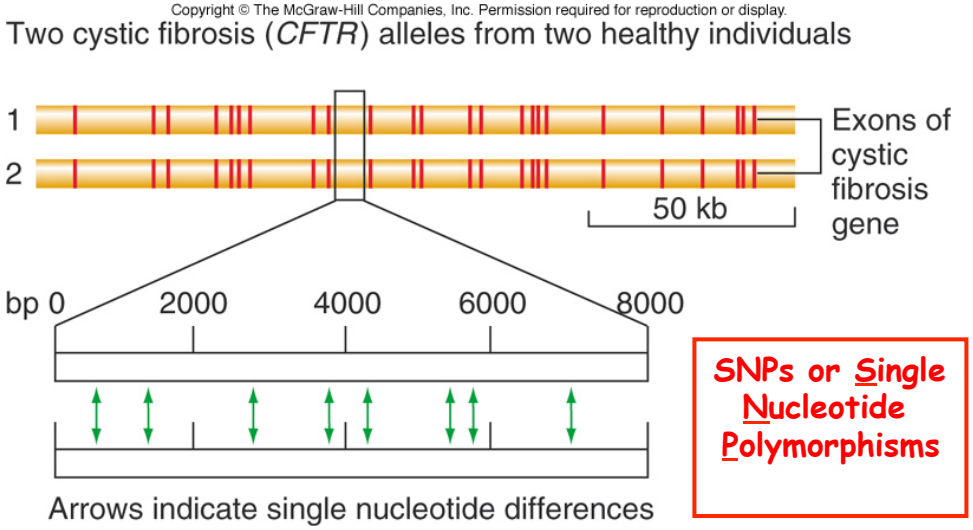
Figure 2.22 Key concepts and terms used in modern genetics. Note that a single gene can have any number of alleles in the population as a whole, but no more than two alleles can be present in any one individual.

Marker

Individuals May Contain Two Different Alleles at any DNA Location

There can be an Infinite # of Alleles for any Gene (or DNA sequence in a Population

Most DNA Variations Between Individuals Occur Because of Base-Pair Changes in Non-Coding Regions of the Genome



SNPs or Single Nucleotide Polymorphisms

To Be on the safe side, suppose you assume that only 80% (0.8) of the 3 billion base pairs in the genome are noncoding, and on average only 1 base pair in 700 is polymorphic. With these assumptions, you can determine the frequency of polymorphism within a single individual by multiplying 3 billion by 0.8 and then multiplying that amount by 1/700:

(3x109) x 0.8 = 2.4 x 109, (2.4 x 109) x 1/700 = 3.4 million.

The result of 3.4 million is astonishing: It means that there are millions of differences between any two haploid sets of human chromosomes. Combined with differences in coding and regulatory sequences (which occur much less frequently), the millions of polymorphisms at anonymous loci contribute to an enormous pool of potential DNA markers.

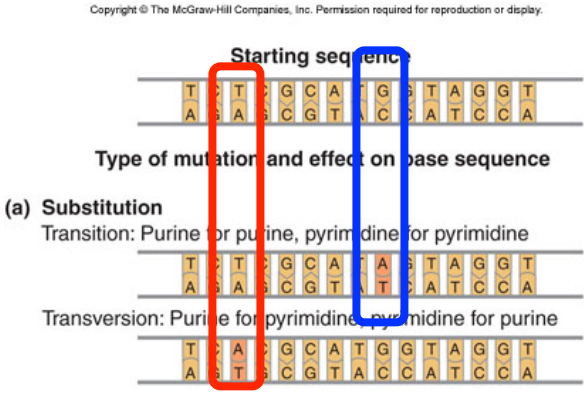
TABLE 11.1 Classes of DNA Polymorphisms

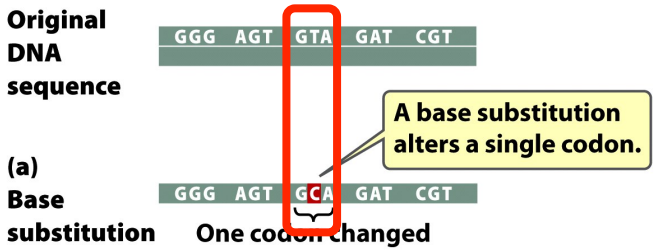
Class	Size of Locus	Number of Alleles	Number of Loci in Population	Rate of Mutation	Use	Method of Detection
SNP	Single base pair	2	100 million	10^{-9}	Linkage and association mapping	PCR followed by ASO hybridization or primer extension
Microsatellite	30–300 bp	2–10	200,000	10^{-3}	Linkage and association mapping	PCR and gel electrophoresis
Multilocus minisatellite	1–20 kb	2–10	30,000	10^{-3}	DNA fingerprinting	Southern blot and hybridization
Indels (deletions and duplications)	1–100 bp	2	N/A	$<10^{-9}$	Linkage and association mapping	PCR and gel electrophoresis

This is What Makes Us Unique Individuals!

There is ~1bp Change per 700bp in Human Genomes or ~3.4 Million bp Differences Between Individuals ~0.1% of Genome

How Do SNPs Arise in the Human Genome During DNA Replication?





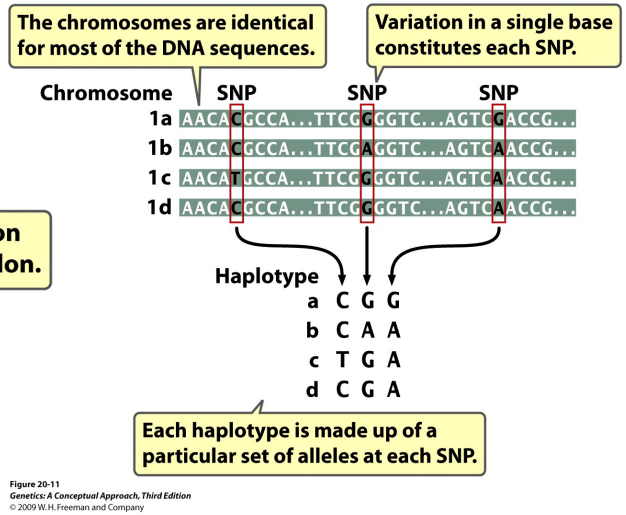


Figure 20-11
Genetics: A Conceptual Approach, Third Edition

Most SNPs are Single Nucleotide Changes that Have No Effect on the Phenotype or Gene Function! They Are Outside Coding Sequence of Genes -- Between Genes or in Introns

Different "Forms" of the Same SNP = Allele!

DNA Sequence Changes in the Genome Are Rare

TABLE 9.1 Five Classes of DNA Polymorphism

Class	Cause	Rate of Mutation per Locus per Gamete	Frequency in Genome	Number per Human Genome (on average)
Single base (SNP)	Mutagens or replication errors	10^{-8}–10^{-9}	1/700 bp	3 million
Microsatellite (VNTR or SSR)	Slippage during replication	10^{-3}	1/30,000 bp	100,000
Minisatellite	Unequal crossovers	10^{-3}	Unknown; discovered by chance	Fewer than 100 families known, yielding 1000 copies in all
Deletions	Mutagens; unequal crossovers	Extremely rare	Very low	0 – a few
Duplications	Mutagens; unequal crossovers	Extremely rare	Very low	0 – a few
Other insertions (excluding those resulting from micro- or minisatellite recombination)	Transposable elements	Extremely rare	Very low	0 – a few
Complex haplotype (any locus of 5 kb or more)	Any of the above	Combination of the above	Not applicable	Not applicable

Only a Few Affect Gene Function & Lead to a Visible Mutation!

Identifying DNA Variations Between Individuals Has Many Uses

1. Epidemiology and Food Safety Science
2. Human Population History and Origins*
3. Improvement of Domesticated Plants and Animals
4. History of Animal & Plant Domestication
5. DNA Polymorphisms as Ecological Indicators
6. Evolutionary Genetics
7. Forensics*
8. Wildlife Identifications (Poachers)
9. Breeding
10. Paternity, Clone Identification, Individual Identification*
11. Marking and Identifying Disease Genes*
12. Marking Drug Efficacy Genes (Pharmacogenomics)*

Identifying SNPs in the Human Genome

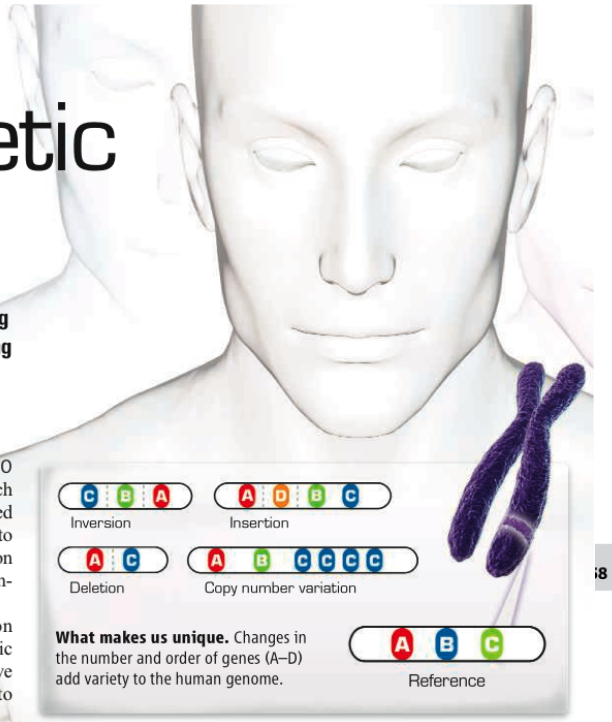
BREAKTHROUGH OF THE YEAR

Human Genetic Variation

Equipped with faster, cheaper technologies for sequencing DNA and assessing variation in genomes on scales ranging from one to millions of bases, researchers are finding out how truly different we are from one another

THE UNVEILING OF THE HUMAN GENOME ALMOST 7 YEARS AGO cast the first faint light on our complete genetic makeup. Since then, each new genome sequenced and each new individual studied has illuminated our genomic landscape in ever more detail. In 2007, researchers came to appreciate the extent to which our genomes differ from person to person and the implications of this variation for deciphering the genetics of complex diseases and personal traits.

Less than a year ago, the big news was triangulating variation between us and our primate cousins to get a better handle on genetic changes along the evolutionary tree that led to humans. Now, we have moved from asking what in our DNA makes us human to striving to know what in my DNA makes me me.



Remember: Most SNPs Are Not in Gene Coding Regions

nature

ARTICLES

Identify From Sequencing the Genome Regions (and soon Genomes) of Individuals From Different Groups

A second generation human haplotype map of over 3.1 million SNPs

The International HapMap Consortium*

We describe the Phase II HapMap, which characterizes over 3.1 million human single nucleotide polymorphisms (SNPs) genotyped in 270 individuals from four geographically diverse populations and includes 25–35% of common SNP variation in the populations surveyed. The map is estimated to capture untyped common variation with an average maximum r^2 of between 0.9 and 0.96 depending on population. We demonstrate that the current generation of commercial genome-wide genotyping products captures common Phase II SNPs with an average maximum r^2 of up to 0.8 in African and up to 0.95 in non-African populations, and that potential gains in power in association studies can be obtained through imputation. These data also reveal novel aspects of the structure of linkage disequilibrium. We show that 10–30% of pairs of individuals within a population share at least one region of extended genetic identity arising from recent ancestry and that up to 1% of all common variants are untaggable, primarily because they lie within recombination hotspots. We show that recombination rates vary systematically around genes and between genes of different function. Finally, we demonstrate increased differentiation at non-synonymous, compared to synonymous, SNPs, resulting from systematic differences in the strength or efficacy of natural selection between populations.

SNPs Become Personal!

It's All About Me

Along with the flood of discoveries in human genetics, 2007 saw the birth of a new industry: personal genomics. Depending on your budget, you can either buy a rough scan of your genome or have the whole thing sequenced. The companies say the information will help customers learn about themselves and improve their health. But researchers worry that these services open up a Pandora's box of ethical issues.

At $300,000 to $1 million per genome, sequencing all 3 billion base pairs is still too costly for all but a few. Although dozens more personal genomes will probably be sequenced in the coming year, most will be done by public and private research organizations—including the institute run by genome maverick J. Craig Venter, whose personal genome was one of three completed in 2007 in the United States and China. In a lower-budget effort, Harvard's George Church this month will deliver initial DNA sequences for the protein-coding sections (1% of the genome) to the first 10 volunteers for his Personal Genome Project. Meanwhile, a new company called Knome is offering full-genome sequencing to 20 customers willing to pay $350,000.

A glimpse of one's genome is already within the reach of ordinary people, thanks to several companies. They include 23andMe, which has financing from Google and may let users link to others with shared traits; Navigenics, which will screen for about 20 medical conditions; and deCODE Genetics in Iceland, a pioneer in disease gene hunting. For $1000 to $2500, these companies will have consumers send in a saliva sample or cheek swab, then use "SNP chips" to scan their DNA for as many as 1 million markers. The companies will then match the results with the latest publications on traits, common diseases, and ancestry.

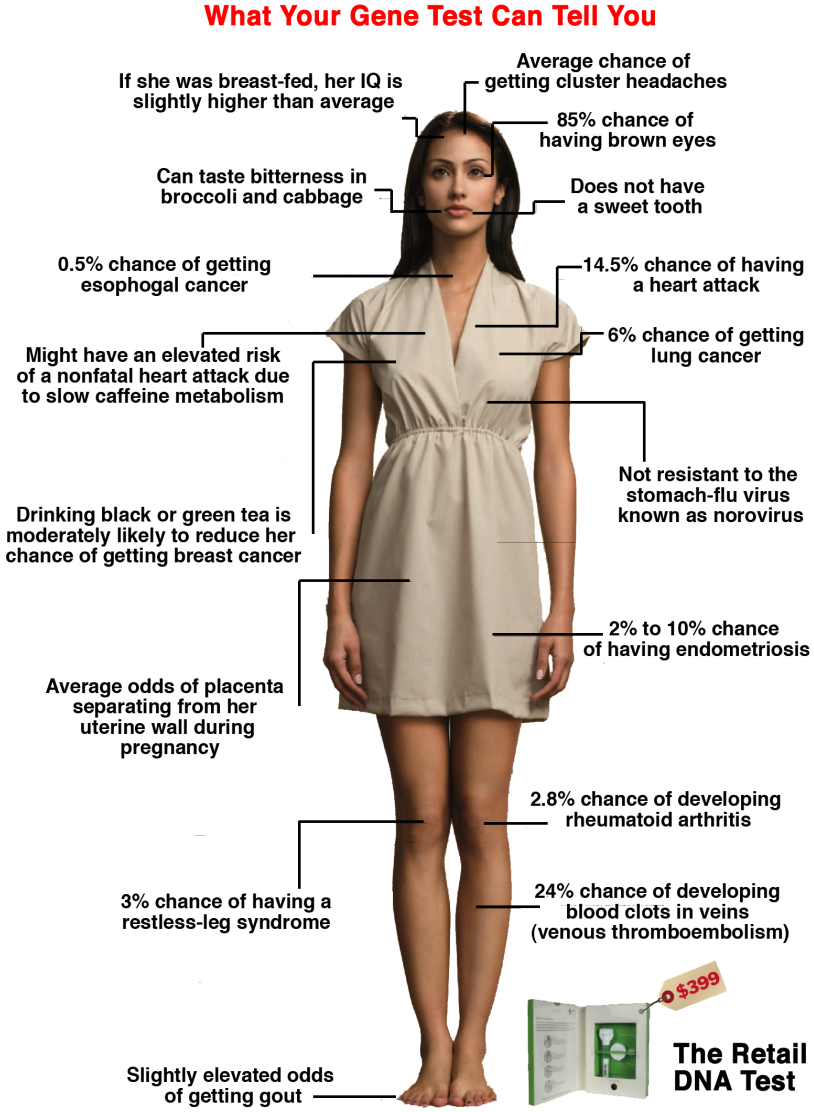
Although many customers may view this exercise as a way to learn fun facts about themselves—recreational genomics, some call it—bioethicists are wary. Most common disease markers identified so far raise risks only slightly, but they could cause needless worry. At the same time, some people may be terrified to learn they have a relatively high risk for an incurable disease such as Alzheimer's.

Pandora's box? This cheek-swab kit could reveal your intimate secrets.

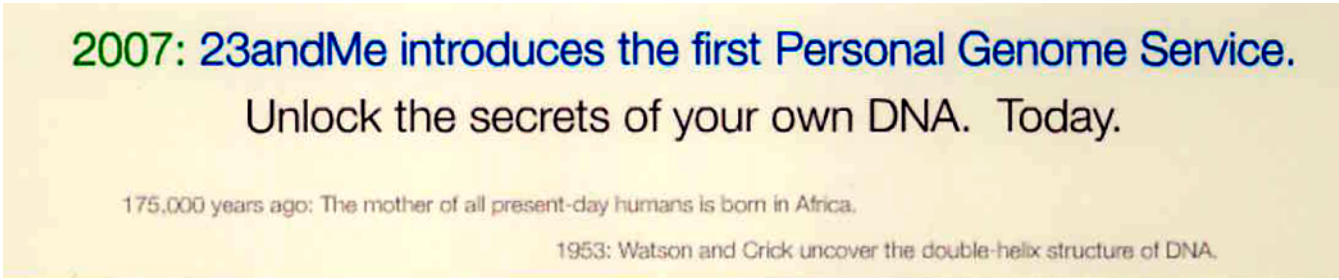
The rush toward personal genome sequences also sharpens long-held worries about discrimination. A bill to prevent insurers and employers from misusing genetic data is stalled in Congress. Complicating matters, your genetic information exposes your relatives' DNA, too.

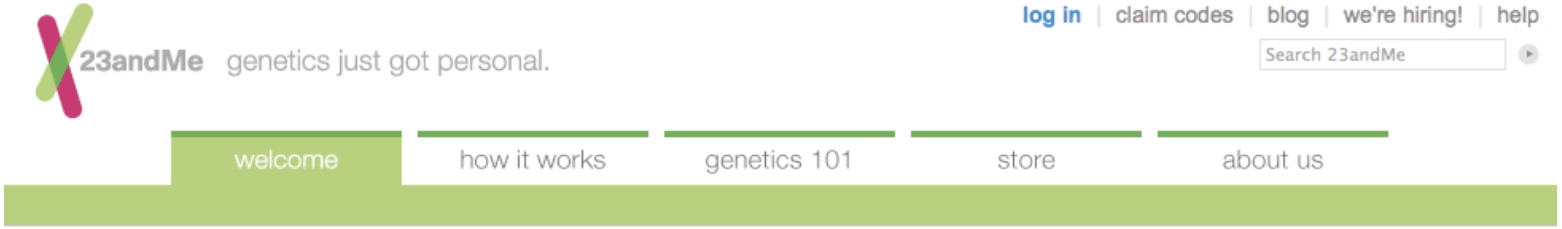
The most profound implications of having one's genome analyzed may not be what it reveals now—which isn't much—but what it may show later on. Perhaps to sidestep such questions, some companies will limit which markers to disclose. Others, however, will hand customers their entire genetic identity, along with all the secrets it may hold.

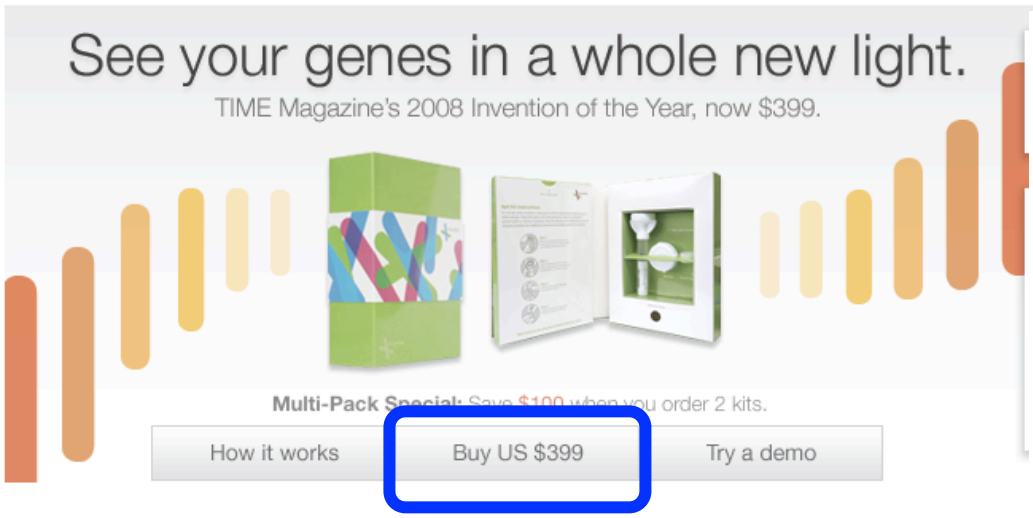
–JOCELYN KAISER

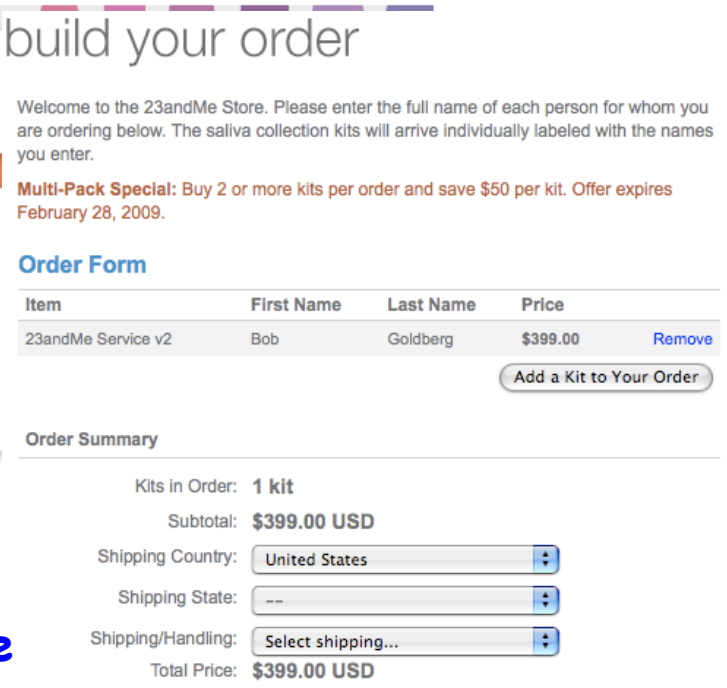


Invention Of the Year



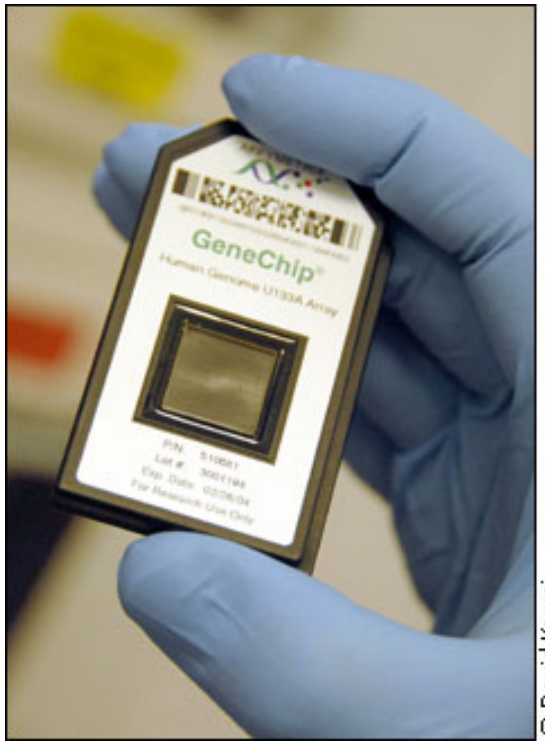


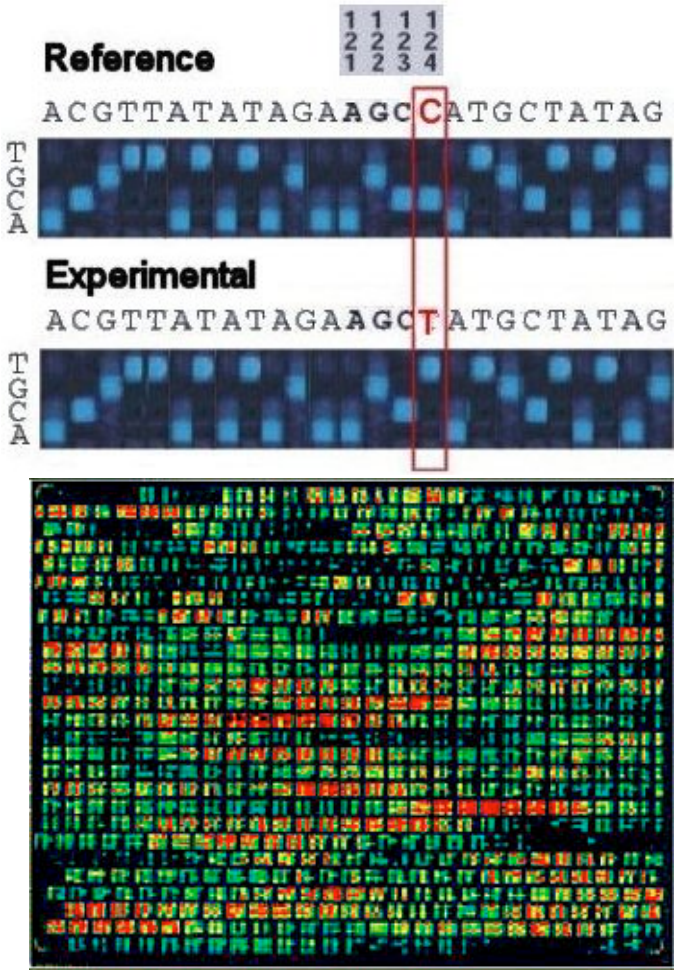




What Are the Problems With This Service and Approach to Personal Genomics?

DNA Chips Can Detect SNP Genotypes (Or Haplotypes) Across An Individuals Genome

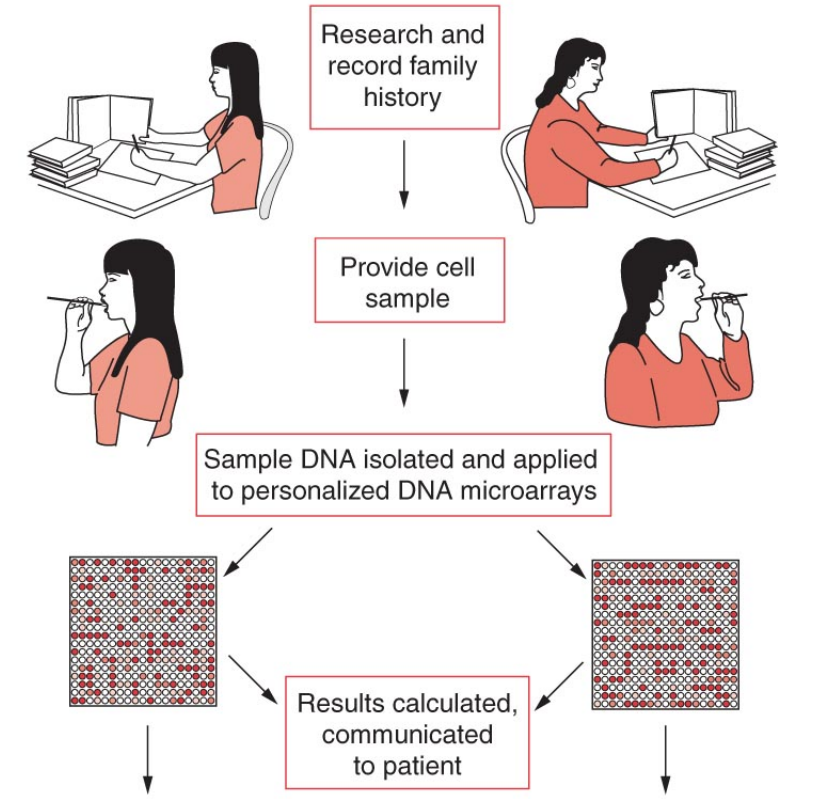




This Can Then Be Correlated With Diseases &/or Geographical Associations

Whole Genome SNP Chips

Associate SNP With Trait From Population Studies (With Trait vs. Without Trait Populations)



Susan's Genetic Profile

Trait	Risk
Addictive behavior	: Greater than general population
Lung cancer	: Greater than general population
Colon cancer	: Less than general population
Alzheimer's disease	: Less than general population

Lisa's Genetic Profile

Trait	Risk
Cystic fibrosis	: 100% diagnosis
Type II diabetes mellitus	: Less than general population
Cardiovascular disease	: Greater than general population

SNPs Can Be Associated/Linked With Specific Traits & Used By Genetic Testing Companies

OCA2

From SNPedia

OCA2, the oculocutaneous albinism gene (also known as the human P protein gene, or, DN10), is a gene associated with albinism and certain pigmentation effects in general such as eye color, skin color, and hair color.

A large (>3,000 individuals) study of Caucasians indicates that the following **OCA2** variants, all located in the first intron of the gene, are preferentially linked to blue eye color inheritance; together, they form haplotypes that (in some cases at least) predict eye color with greater than 50:50 odds. [PMID 17236130; OMIM 203200.0013 (http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=203200&a=203200_AllelicVariant0013)]

- rs7495174
- rs6497268
- rs11855019

The haplotypes are defined in order as listed above for these 3 SNPs, so, for example, the TGT haplotype refers to rs745174(T)-rs6497268(G)-rs11855019(T). The correspondence between diplotypes (the two haplotypes in one individual) and the % of individuals with blue/gray, green/hazel/ and brown eye color , respectively, was reported as follows for the most common diplotypes[PMID 17236130]:

- TGT/TGT: 62.5, 28.0, 9.5
- TGT/***TTC***: 47.1, 20.3, 32.6
- TGT/***CGT***: 28.6, 14.3, 57.1
- TGT/***TGC***: 27.9, 22.1, 50.0
- ***TGC/TTC***: 25.0, 8.3, 66.7
- ***TTT/TGC***: 20.7, 31.0, 48.3
- TGT/***TTT***: 17.6, 38.5, 44.0
- TGT/***CTC***: 7.9, 23.3, 68.8

The haplotypes shown in ***bold italics*** represent the ones reported by the authors of this study to be most associated with brown eye color. Furthermore, the haplotypes shown above are as published, and the associated SNPs - which have since changed # as well - are not in the orientation shown in dbSNP.

More recently, a study of a large Danish family led to associations with 2 SNPs in a different region of **OCA2** as linked to blue or brown eye color:

- rs12913832
- rs1129038

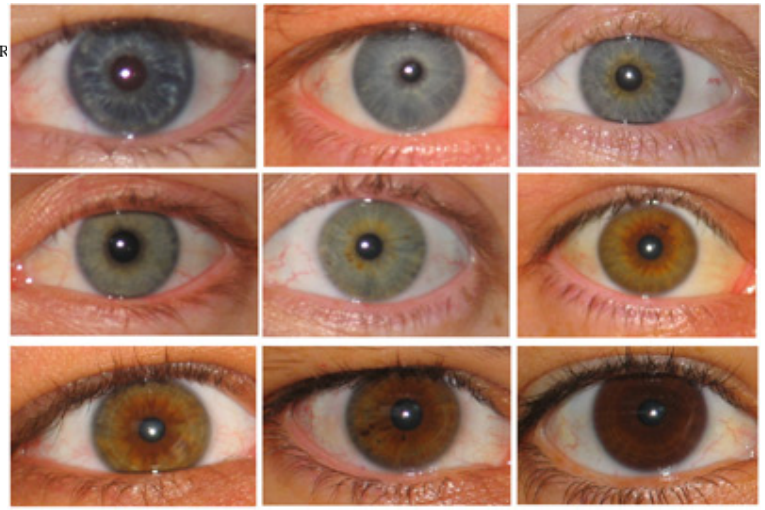
Earlier studies found different regions of the **OCA2** gene to also be predictive of eye color;

- **OCA2** SNP rs1800401 helps predict brown eye color. [PMID 12163334, PMID 15889046; OMIM 203200.0011 (http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=203200&a=203200_AllelicVariant0011)]
- **OCA2** SNP rs1800407 may be associated with green/hazel eye color in some populations, but not others. [PMID 12163334, PMID 15889046; OMIM 203200.0012 (http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=203200&a=203200_AllelicVariant0012)]

is a	gene
is	mentioned by
wikipedia	OCA2 (http://en.wikipedia.org/wiki/OCA2)
google	OCA2 (http://www.google.com/search?hl=en&q=OCA2)
gopubmed	OCA2 (http://www.gopubmed.org/search?q=OCA2)
23andMe	OCA2 (https://www.23andme.com/you/explorer/gene/?gene_name=OCA2)
GeneRIF	4948 (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=4948&ordinalpos=1&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_R
dbSNP	4948 (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=4948&chooseRs=all)
PubMed	4948 (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=Link&LinkName=gene_pubmed&from_uid=4948)
HugeNav	4948 (http://hugenavigator.net/HuGENavigator/huGEPedia.do?firstQuery=OCA2}&geneID=4948&typeSubmit=GO&check=y&typeOption=gene&which=2&pubOrderType=pubD)

	Chromosome position
Rs1129038	26,030,454
Rs11631797	26,175,874
Rs12593929	26,032,853
Rs1800401	25,933,648
Rs1800407	25,903,913
Rs2238289	26,126,810
Rs2240203	26,167,797
Rs28934272	25,903,842
Rs3935591	26,047,607
Rs3940272	26,142,318
Rs4778241	26,012,308
Rs7170852	26,101,581
Rs7183877	26,039,328
Rs7495174	26,017,833
Rs8028689	26,162,483
Rs916977	26,186,959

SNPs in Human P Protein (OCA2) Gene Lead To Different Eye Colors (Physical & Molecular Markers)



Human Eye Color

Vol 447|7 June 2007|doi:10.1038/nature05911 nature

ARTICLES

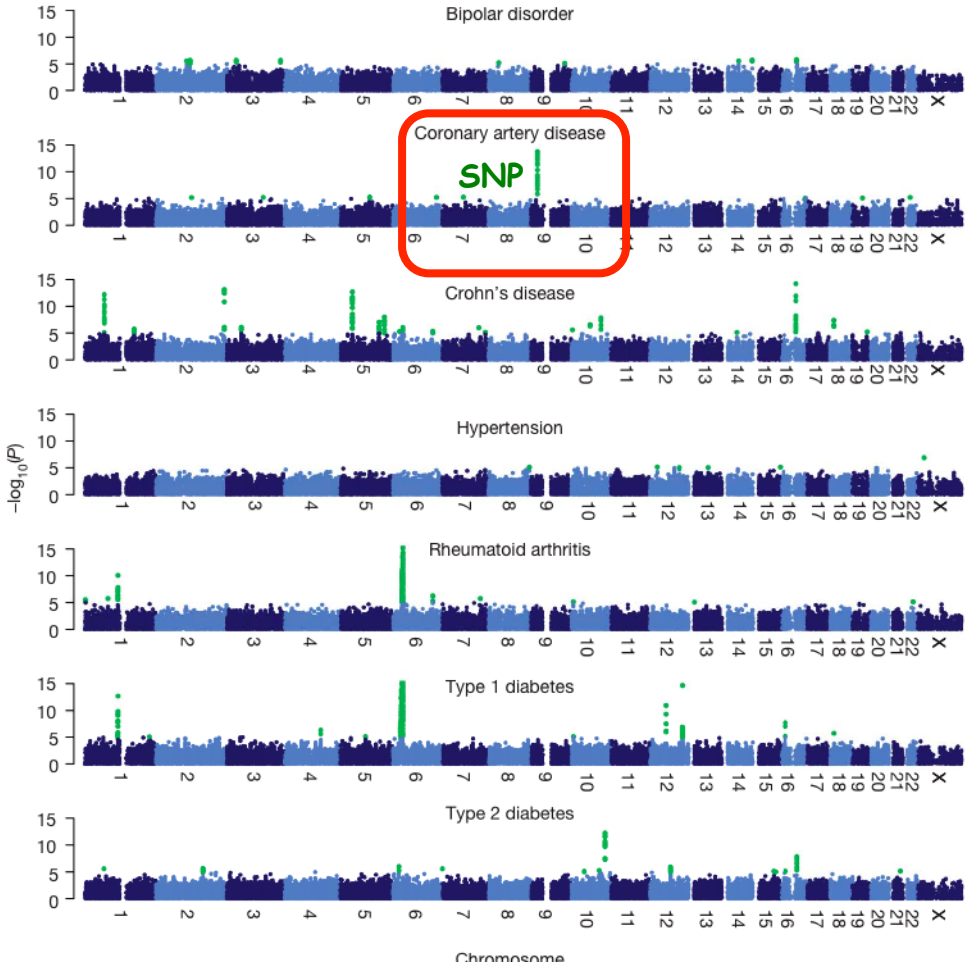
Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

The Wellcome Trust Case Control Consortium*

There is increasing evidence that genome-wide association (GWA) studies represent a powerful approach to the identification of genes involved in common human diseases. We describe a joint GWA study (using the Affymetrix GeneChip 500K Mapping Array Set) undertaken in the British population, which has examined ~2,000 individuals for each of 7 major diseases and a shared set of ~3,000 controls. Case-control comparisons identified 24 independent association signals at $P < 5 \times 10^{-7}$: 1 in bipolar disorder, 1 in coronary artery disease, 9 in Crohn's disease, 3 in rheumatoid arthritis, 7 in type 1 diabetes and 3 in type 2 diabetes. On the basis of prior findings and replication studies thus-far completed, almost all of these signals reflect genuine susceptibility effects. We observed association at many previously identified loci, and found compelling evidence that some loci confer risk for more than one of the diseases studied. Across all diseases, we identified a large number of further signals (including 58 loci with single-point *P* values between 10^{-5} and 5×10^{-7}) likely to yield additional susceptibility loci. The importance of appropriately large samples was confirmed by the modest effect sizes observed at most loci identified. This study thus represents a thorough validation of the GWA approach. It has also demonstrated that careful use of a shared control group represents a safe and effective approach to GWA analyses of multiple disease phenotypes; has generated a genome-wide genotype database for future studies of common diseases in the British population; and shown that, provided individuals with non-European ancestry are excluded, the extent of population stratification in the British population is generally modest. Our findings offer new avenues for exploring the pathophysiology of these important disorders. We anticipate that our data, results and software, which will be widely available to other investigators, will provide a powerful resource for human genetics research.

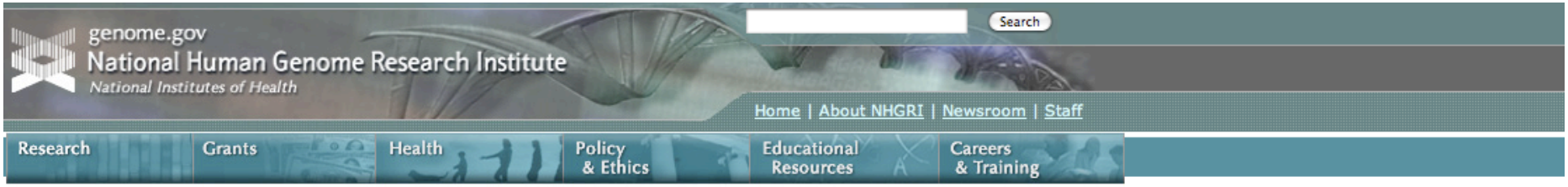
Population Association Studies

Correlating SNPs With Specific Diseases Using SNP Chips & Association Studies



SNPs May Be Near Or In Relevant Genes

All Published Genome-Wide Association Studies Are Listed on the National Human Genome Research Institute Website



Correlate SNPs With Specific Traits And Used By Personal Gene Testing Companies Such as 23andMe®

Using Large Populations SNPs Can Be Used As Markers For Specific Genes/ Traits

SNPedia (http://www.snpedia.com/)

- New model for prostate cancer based on 5 SNPs
- rs1815739 sprinters vs endurance athletes
- rs4420638 and rs429358 can raise the risk of Alzheimer's disease by more than 10x
- rs6152 can prevent baldness
- rs9939609 triggers obesity
- rs662799 prevents weight gain from high fat diets
- rs7495174 green eye color
- rs7903146 in 3% of the population greatly increases the risk of type-2 diabetes
- rs12255372 linked to type-2 diabetes and breast cancer
- rs2395029 asymptomatic HIV viral load set point
- rs324650 influences intelligence and alcohol dependence
- rs1799971 makes alcohol cravings stronger
- rs17822931 determines earwax

Caution

How Will You Use the Information?

How Good Are The Correlations?

What To Do With The Information?

Privacy Issues?

Group Differences? Discrimination?

Using SNPs or DNA Sequence Variation As Markers For Disease Genes

Remember: Only a Small Fraction of Human Genes Are Known To Cause Diseases

Problem: Different Companies-Different Predictions!

TABLE 1: PREDICTIONS FOR DISEASE RELATIVE RISKS FOR FIVE INDIVIDUALS

Disease	Female A	Female B	Female C	Male D	Male E
Breast cancer	↑↑	↑↑	↓↓		
Coeliac disease	↓↓	↓↓	↓↓	↓↓	↓↓
Colon cancer	==	==	=↓	↑↑	=↓
Crohn's disease	↓↑	↓↑	↓↓	↓↓	↓=
Heart attack	↓↓	=↓	=↓	=↓	↑↑
Lupus	↑↓	↓↓	↓↓	↑=	↑=
Macular degeneration	↓↓	↓↓	↑=	↓↓	↓↓
Multiple sclerosis	↑↑		↓↓	↓↓	↓↓
Prostate cancer				↑↑	↓↑
Psoriasis	↓↑		↑↓	↑↑	↓↓
Restless legs syndrome	=↓	↑↑	↓=	↓↑	↑↑
Rheumatoid arthritis	↑↑	↑↑	↓↓	↓↓	↑↑
Type 2 diabetes	↓↓	=↓	↓↓	↑↓	=↓

↑ increased risk (RR > 1.05), ↓ decreased risk (relative risk (RR) < 0.95), = average risk (0.95 ≤ RR ≤ 1.05). First prediction is from 23andMe; second prediction is from Navigenics. Different predictions are highlighted in beige.

DNA Tests Available For Most Known Disease Genes

Table 11.1 GENETIC DISEASE TESTING

Genetic Disease Condition	Genetic Basis for Disease and Symptoms
Cancers (brain tumors; urinary bladder, prostate, ovarian, breast, brain, lung, and colorectal cancers)	A variety of different mutant genes can serve as markers for genetic testing.
Cystic fibrosis	Large number of mutations in the CFTR (cystic fibrosis transmembrane conductance regulator) gene on chromosome 7. Causes lung infections and problems with pancreatic, digestive, and pulmonary functions.
Duchenne muscular dystrophy	Defective gene (dystrophin) on the X chromosome causes muscle weakness and muscle degeneration.
Familial hypercholesterolemia	Mutant gene on chromosome 19 causes extremely high levels of blood cholesterol.
Hemophilia	Defective gene on the X chromosome makes it difficult for blood to clot when bleeding.
Huntington disease	Mutation in gene on chromosome 4 causes neurodegenerative disease in adults.
Phenylketonuria (PKU)	Mutation in gene required for converting the amino acid phenylalanine into the amino acid tyrosine. Causes severe neurological damage, including mental retardation.
Severe combined immunodeficiency (SCID)	Immune system disorder caused by mutation of the adenosine deaminase gene.
Sickle-cell disease	Mutation in β-globin gene on chromosome 11 affects hemoglobin structure and shape of red blood cells, which disrupts oxygen transport in blood and causes joint pain.
Tay-Sachs disease	Rare mutation of a gene on chromosome 5 causes certain types of lipids to accumulate in the brain. Causes paralysis, blindness, retardation, and respiratory infections.

RFLPs or DNA Markers (SNPs) Can Be Used to Follow/ Identify Gene Alleles if Linked

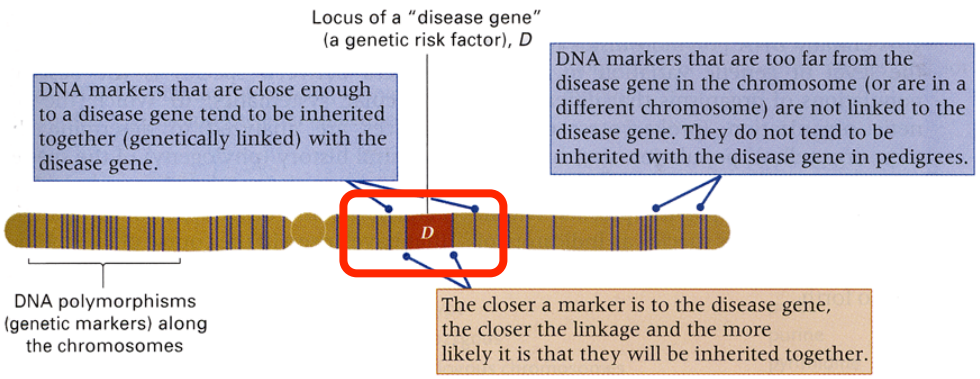


Figure 2.29 Concepts in genetic localization of genetic risk factors for disease. Polymorphic DNA markers (indicated by the vertical lines) that are close to a genetic risk factor (*D*) in the chromosome tend to be inherited together with the disease itself. The genomic location of the risk factor is determined by examining the known genomic locations of the DNA polymorphisms that are linked with it.

Useful for DNA Testing & Genetic Diagnosis!

Markers For Wrinkled Peas

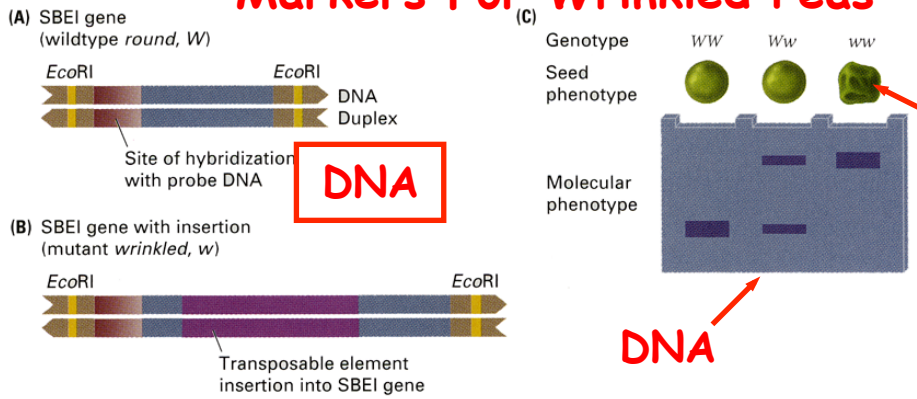
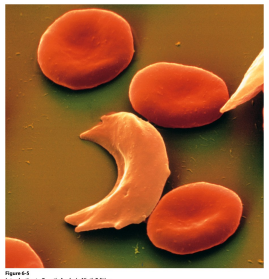
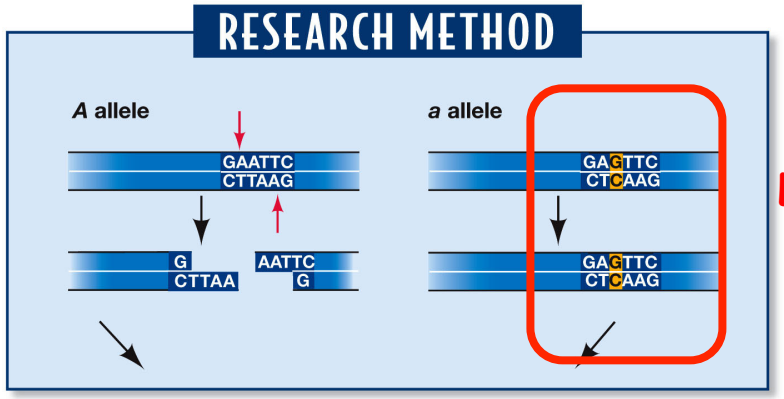


Figure 3.2 (A) *W* (round) is an allele of a gene that specifies the amino acid sequence of starch branching enzyme I (SBEI). (B) *w* (wrinkled) is an allele that encodes an inactive form of the enzyme because its DNA sequence is interrupted by the insertion of a transposable element. (C) At the level of the morphological phenotype, *W* is dominant to *w:* Genotype *WW* and *Ww* have round seeds, whereas genotype *ww* has wrinkled seeds. The molecular difference between the alleles can be detected as a restriction fragment length polymorphism (RFLP) using the enzyme *Eco*RI and a probe that hybridizes at the site shown. At the molecular level, the alleles are codominant: DNA from each genotype yields a different molecular phenotype—a single band differing in size for homozygous *WW* and *ww,* and both bands for heterozygous *Ww.*

Also Useful in Breeding and Obtaining Markers For Specific Traits!

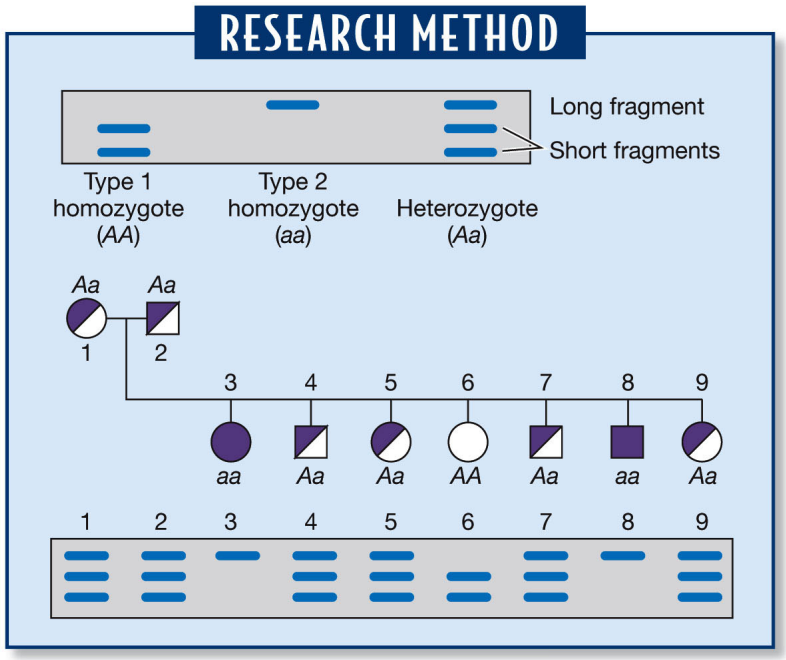


Using RFLPs + Markers to Identify the Sickle Cell Allele (Single Gene Test)



Loss of Restriction Site in a Allele (in gene)

Detected By Blots Or PCR



SNP Leads to RFLP!!!

LIFE 8e, Figure 17.8 (Part 2)

DNA Testing Should Be Carried Out On Every Individual Born in the US:

a. Yes
b. No

DNA Testing Results Should Be Made Widely Available?

a. Yes
b. No

How to Detect
DNA Variation in
Individuals?

Do Not Need SNPs in Coding
Sequences-Can Be Anywhere in
Genome!

Need Cloned Probes and/or DNA
Sequences to Detect

Now Done By Sequencing or Chips on a
Genome-Wide Basis

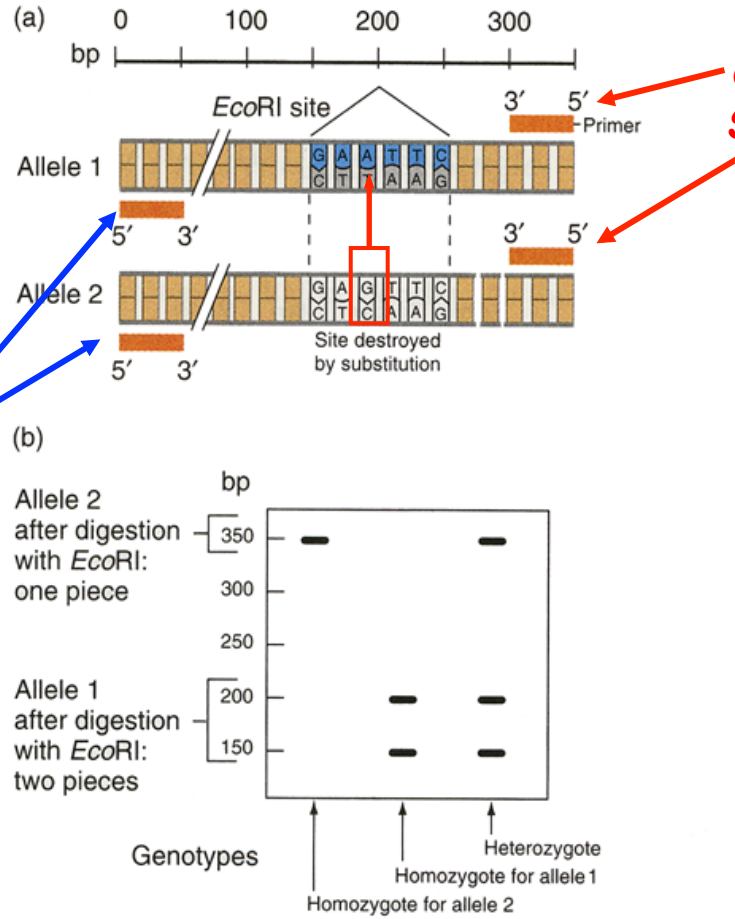
Use PCR/RFLPs For "Simple"
Situations (Paternity, Forensics,
Disease Gene in Family)

Recall: PCR Can Be Used to Identify SNP-Generated RFLPs and DNA Variation

Identify by PCR Using Primers For Surrounding Conserved Sequences

Conserved Sequence 1

Conserved Sequence 2

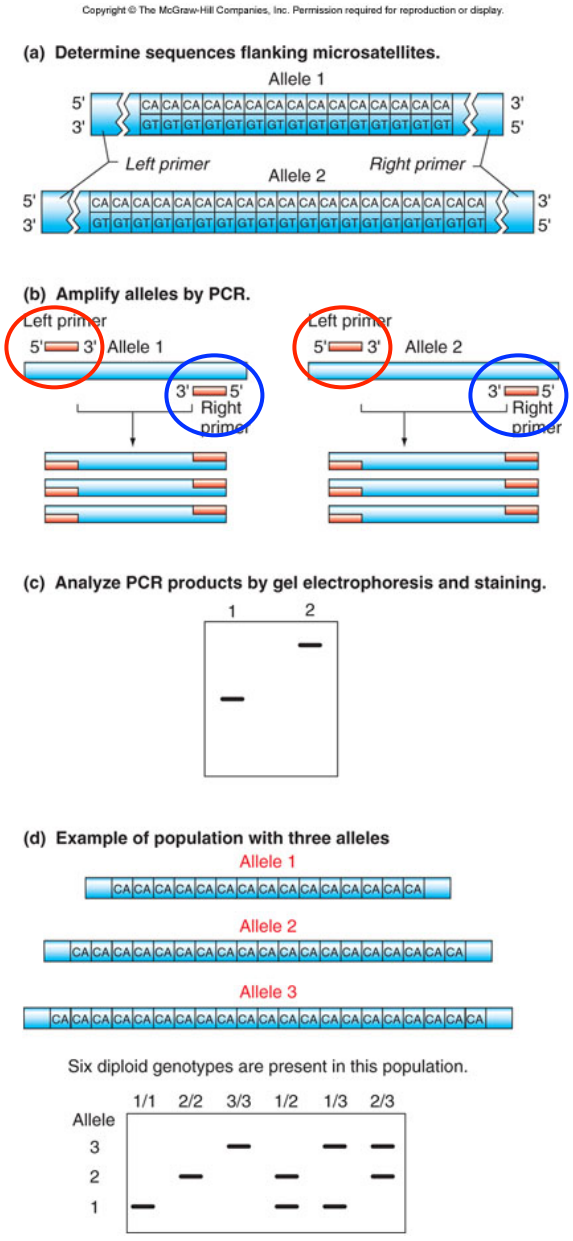


1. PCR
2. Restriction Digest
3. Gel

Figure 9.7 Restriction site polymorphisms can be detected most efficiently with PCR-based protocols. (a) PCR amplification

Recall: VNTRs, STRs, SSRs Can Be Assayed Using PCR

Because VNTRs Vary By Length in Individuals



15 Copies
19 Copies

Use Conserved Neighbor Sequences For PCR Primers

Note: Size Difference on Gel

Used to Identify Individuals

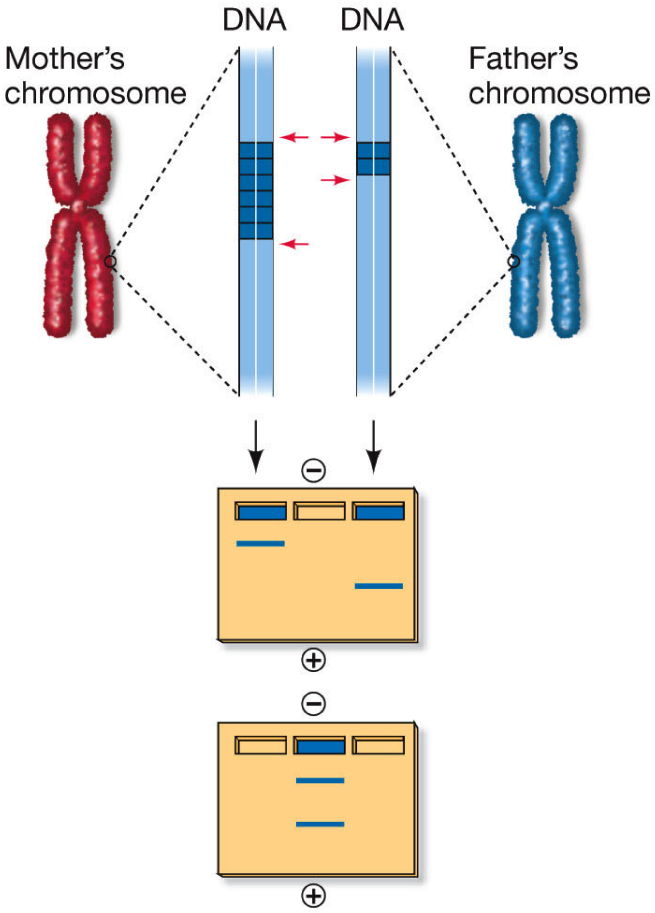
e.g., DIS80 VNTR Class DNA Fingerprinting

STRs Used to Verify Remains of Russian Royal Family

	Number of repeats	
STR-1	15,16	15,16
STR-2	8,8	7,10
STR-3	3,5	7,7
STR-4	12,13	12,12
STR-5	32,36	11,32

Tsarina Alexandra ○ — □ Tsar Nicholas II

STR-1	15,16	15,16	15,16
STR-2	8,10	7,8	8,10
STR-3	5,7	5,7	3,7
STR-4	12,13	12,13	12,13
STR-5	11,32	11,36	32,36



PNAS

Genomic identification in the historical case of the Nicholas II royal family

PNAS, March, 2009

Evgeny I. Rogaev[a,b,c,d,1], Anastasia P. Grigorenko[b,d], Yuri K. Moliaka[b], Gulnaz Faskhutdinova[b], Andrey Goltsov[d], Arlene Lahti[e], Curtis Hildebrandt[e], Ellen L. W. Kittler[f], and Irina Morozova[a]

[a]Department of Genomics and Laboratory of Evolutionary Genomics, Vavilov Institute of General Genetics, Russian Academy of Science, Gubkina Street, 3, Moscow, 119991, Russian Federation; [b]Brudnick Neuropsychiatric Research Institute, University of Massachusetts Medical School, 303 Belmont Street, Worcester, MA 01604; [c]Faculty of Bioinformatics and Bioengineering, Lomonosov Moscow State University, Moscow, 119991, Russian Federation; [d]Research Center of Mental Health, Russian Academy of Medical Science, Zagorodnoe Shosse 2/2, Moscow, 113152, Russia; [e]Molecular World, Inc., Thunder Bay, ON, Canada P7B 2T1; and [f]University of Massachusetts Medical School, Center for AIDS Research, Worcester, MA 01605

Communicated by James D. Watson, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, November 14, 2008 (received for review October 8, 2008)

RESEARCH ARTICLE OPEN ACCESS

Mystery Solved: The Identification of the Two Missing Romanov Children Using DNA Analysis

PLOS, March, 2009

Michael D. Coble[1#*], Odile M. Loreille[1#], Mark J. Wadhams[1], Suni M. Edson[1], Kerry Maynard[1], Carna E. Meyer[1], Harald Niederstätter[2], Cordula Berger[2], Burkhard Berger[2], Anthony B. Falsetti[3], Peter Gill[4,5], Walther Parson[2], Louis N. Finelli[1]

1 Armed Forces DNA Identification Laboratory, Armed Forces Institute of Pathology, Rockville, Maryland, United States of America, 2 Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria, 3 University of Florida, Gainesville, Florida, United States of America, 4 Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, United Kingdom, 5 Institute of Forensic Medicine, University of Oslo, Oslo, Norway

Identifying Victims of 9/11 by DNA Fingerprinting

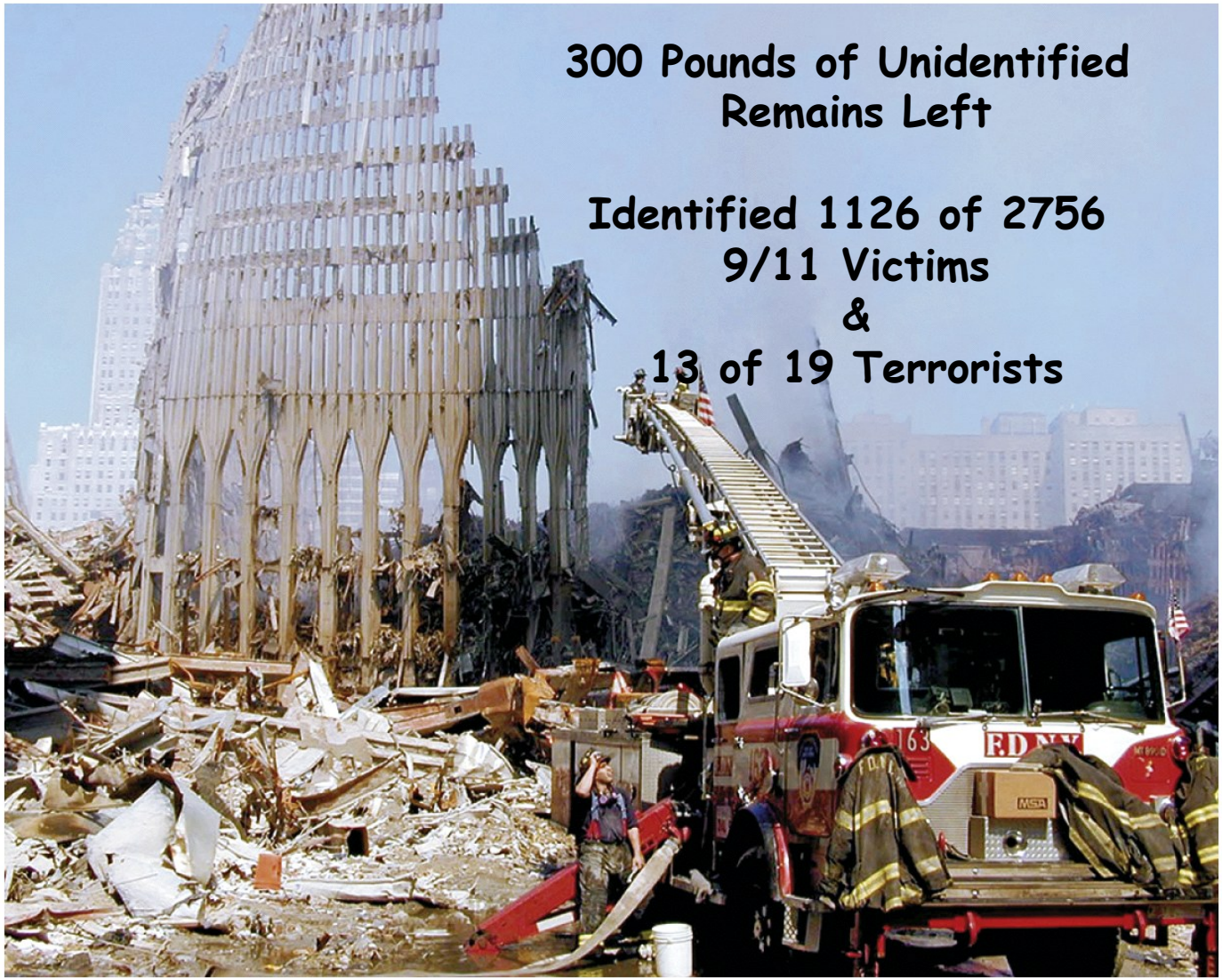
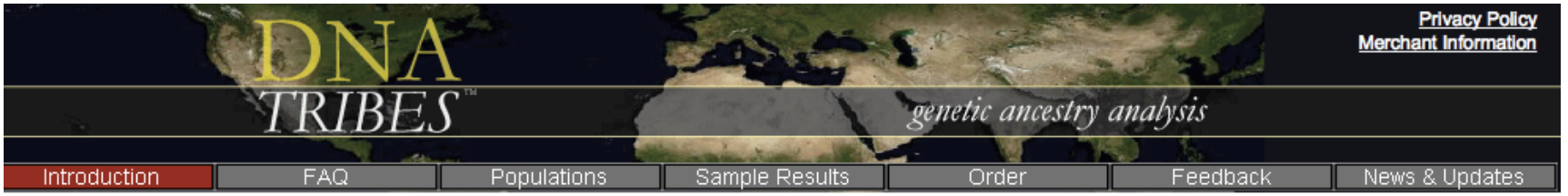
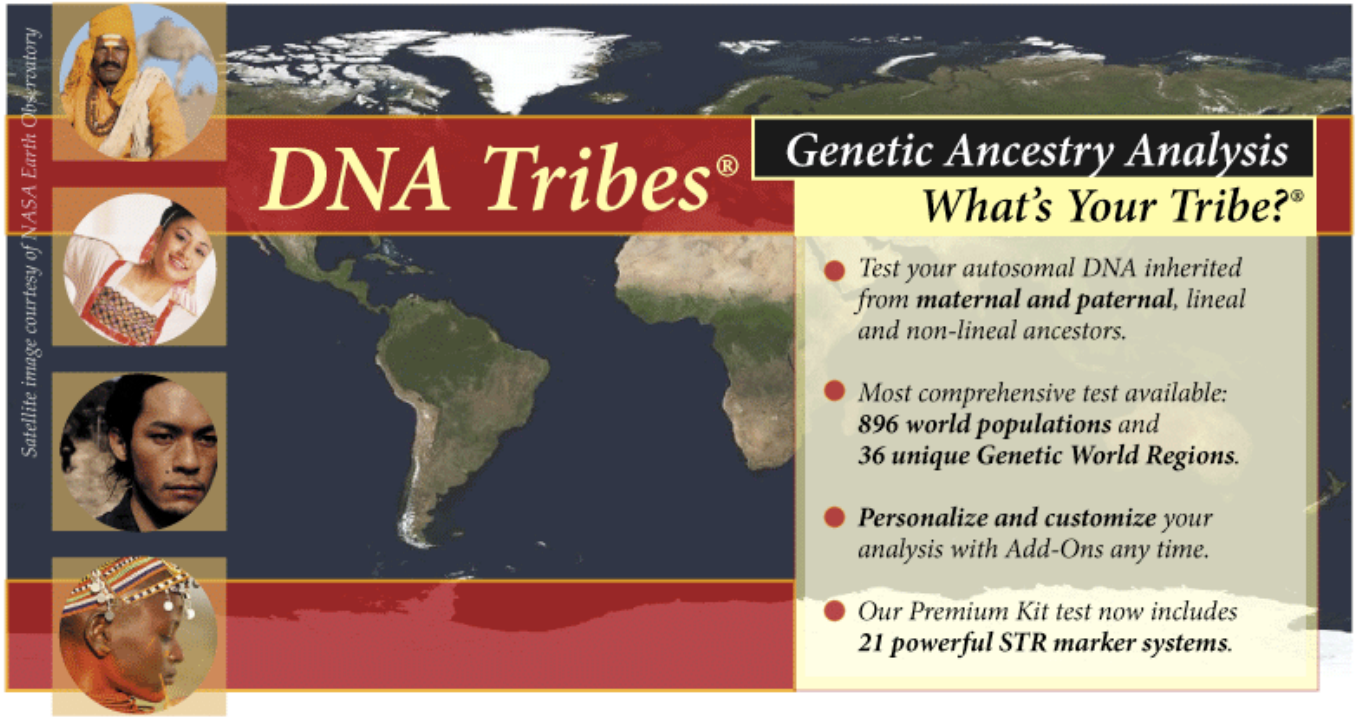


Figure 19-31
Genetics: A Conceptual Approach, Third Edition

Newsweek, January 12, 2009

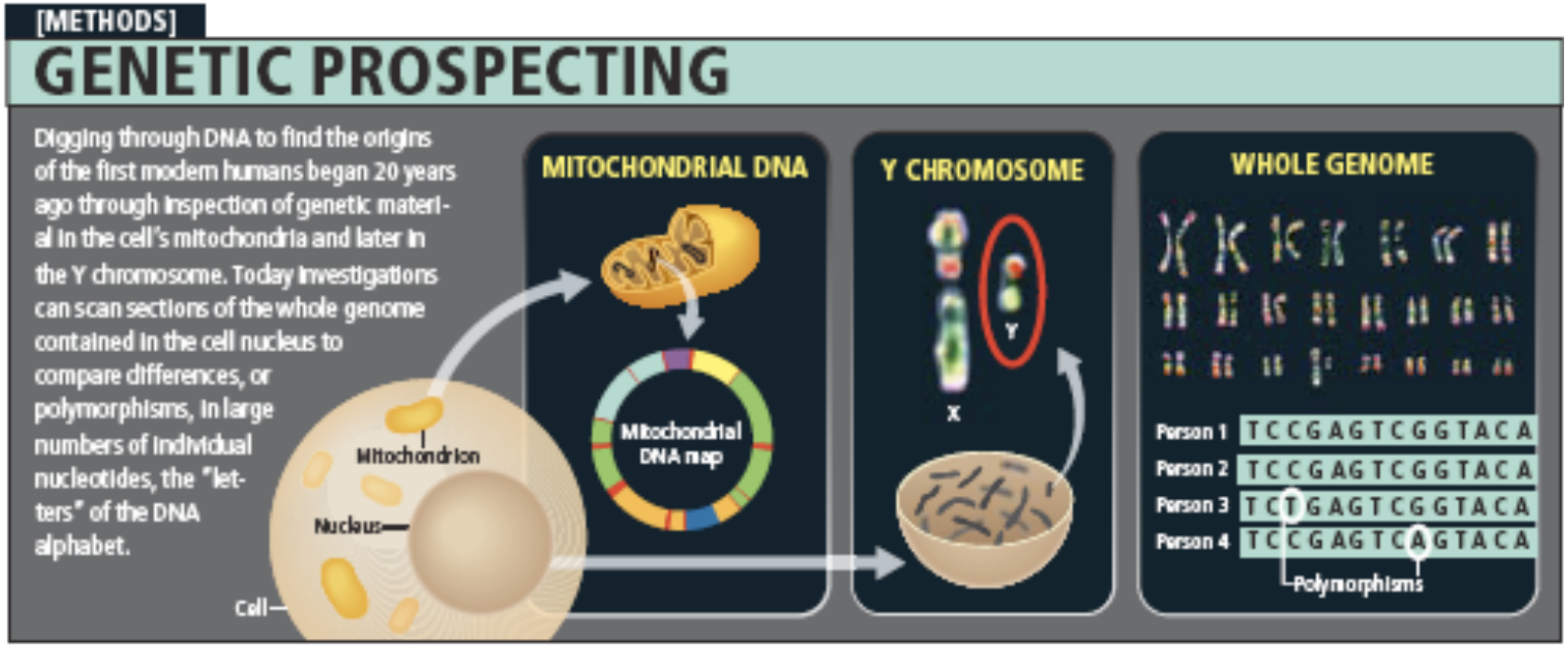
Whole Genome SNP Chips & Personal DNA Sequencing Can Trace Our Ancestry





Most Haplotypes Found In All Human Populations-Some May Be Unique To A Population &/or be Represented At Higher Frequency In A Population (5% of Variation)

Tracing Human Populations Using DNA Polymorphisms



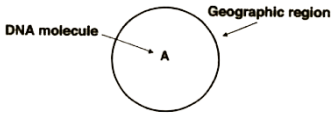
Oldest Populations Contain the Most Diversity

Analysis of human mtDNA led to the Mitochondrial Eve Hypothesis

In the 1980s, Allan Wilson pioneered the use of mtDNA to study human evolution.

In two papers published in 1987 and 1991, he and his colleagues at Cal proposed that we all come from a population of humans that lived in Africa approximately 200,000 years ago.

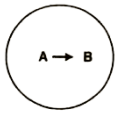
Here's the logic behind the hypothesis.



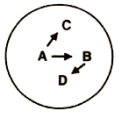
Start with A.

Time

Detected By Using Specific Markers or SNPS



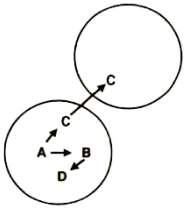
Mutation generates B from A;
now have individuals with both A and B DNAs in population.



Additional mutations generate diversity;
now have individuals with both A, B, C and D DNAs.

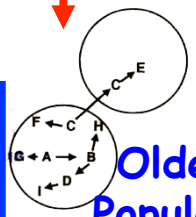
Subset of "Old" Variants + New Variants

Newest Population



C migrates to form separate population.

Lots of "Old" Variants



Oldest Population

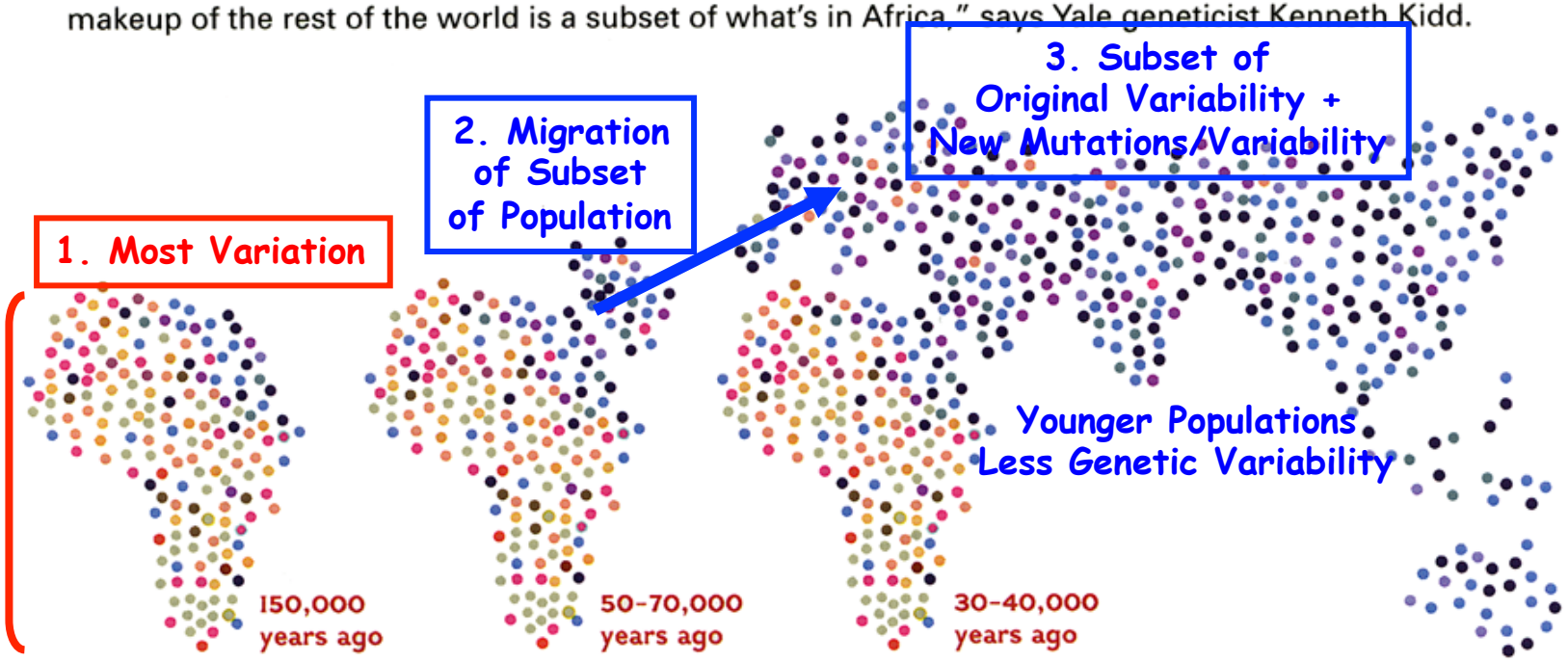
Additional mutations diversify DNAs in populations:
original population more diverse (A, B, C, D, F, G, H, I) than newer population (C, E).

Old Variants Trace Ancient Lineage. New Variants Mark New Populations SPECIFICALLY

Most Genetic Diversity Originated in the Founder Populations to Modern Humans!

Diverse From the Start

The diversity of genetic markers is greatest in Africa (multicolored dots in map), indicating it was the earliest home of modern humans. Only a handful of people, carrying a few of the markers, walked out of Africa (center) and, over tens of thousands of years, seeded other lands (right). "The genetic makeup of the rest of the world is a subset of what's in Africa," says Yale geneticist Kenneth Kidd.



Genetic Variation Proportional to Population Age

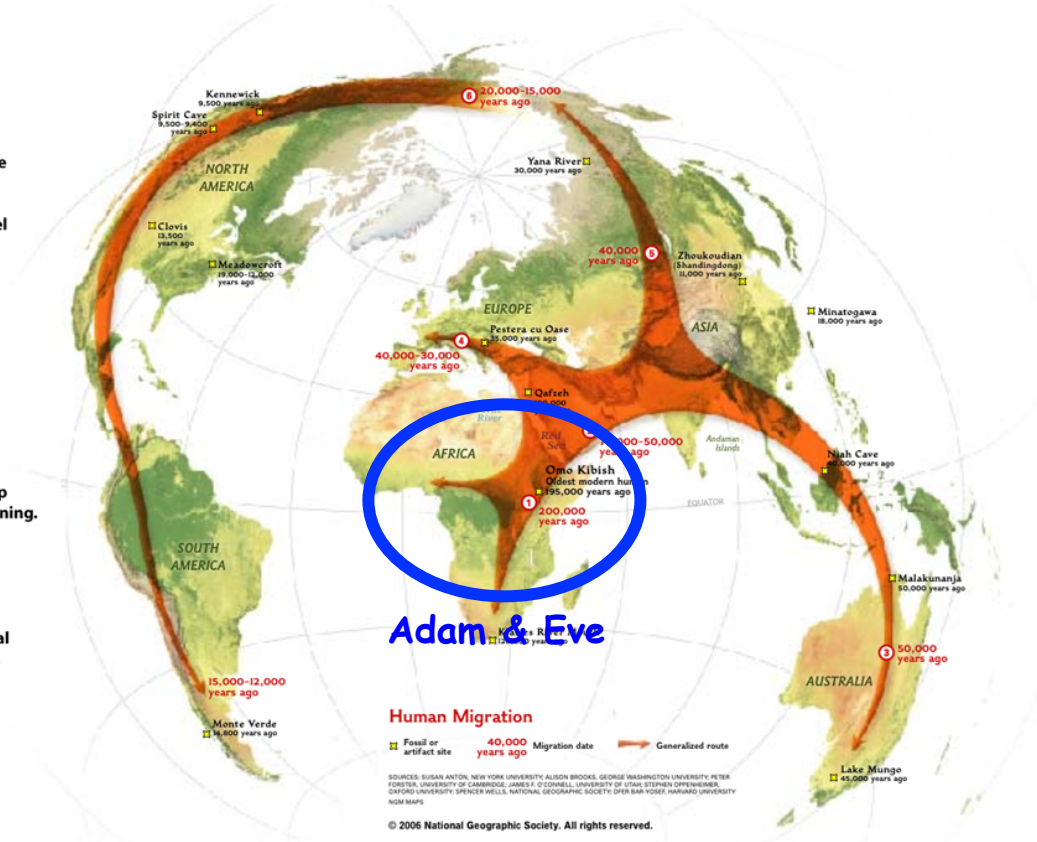
Markers From Original Population + New Markers For "New" Population

Origins of Human Populations From DNA Sequence Comparisons

1. African Cradle
Most paleoanthropologists and geneticists agree that modern humans arose some 200,000 years ago in Africa. The earliest modern human fossils were found in Omo Kibish, Ethiopia. Sites in Israel hold the earliest evidence of modern humans outside Africa, but that group went no farther, dying out about 90,000 years ago.

2. Out of Africa
Genetic data show that a small group of modern humans left Africa for good 70,000 to 50,000 years ago and eventually replaced all earlier types of humans, such as Neandertals. All non-Africans are the descendants of these travelers, who may have migrated around the top of the Red Sea or across its narrow southern opening.

3. The First Australians
Discoveries at two ancient sites—artifacts from Malakunanja and fossils from Lake Mungo—indicated that modern humans followed a coastal route along southern Asia and reached Australia nearly 50,000 years ago. Their descendants, Australian Aborigines, remained genetically isolated on that island continent until recently.

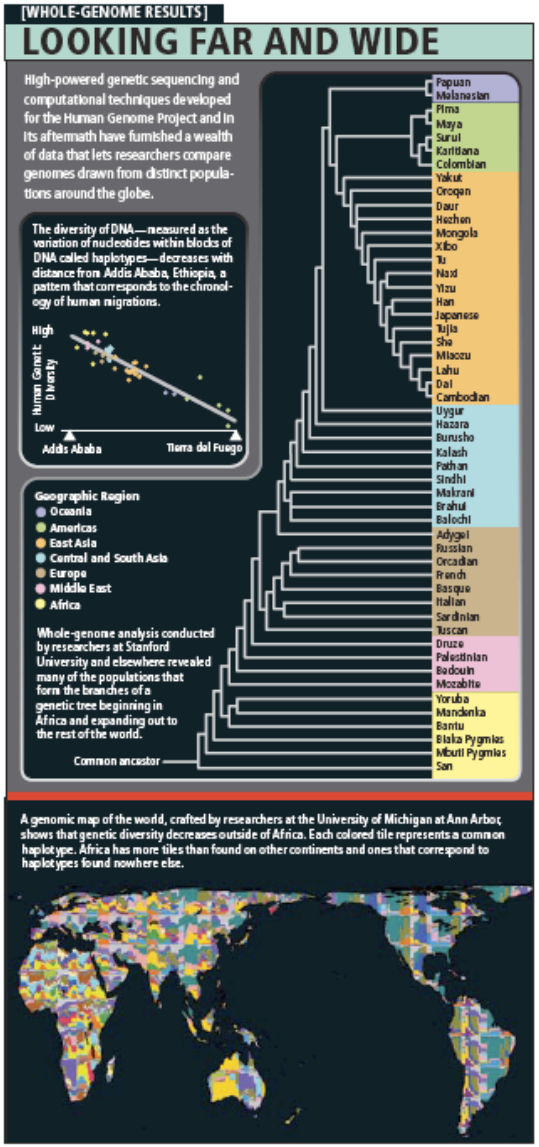


4. Early Europeans
Paleoanthropologists long thought that the peopling of Europe followed a route from North Africa through the Levant. But genetic data show that the DNA of today's western Eurasians resembles that of people in India. It's possible that an inland migration from Asia seeded Europe between 40,000 and 30,000 years ago.

5. Populating Asia
Around 40,000 years ago, humans pushed into Central Asia and arrived on the grassy steppes north of the Himalaya. At the same time, they traveled through Southeast Asia and China, eventually reaching Japan and Siberia. Genetic clues indicate that humans in northern Asia eventually migrated to the Americas.

6. Into the New World
Exactly when the first people arrived in the Americas is still hotly debated. Genetic evidence suggests it was between 20,000 and 15,000 years ago, when sea levels were low and land connected Siberia to Alaska. Ice sheets would have covered the interior of North America, forcing the new arrivals to travel down the west coast

Human Population Relationships Using Whole-Genome Comparisons



Most Genetic Diversity In African Populations

Summary

Mt-DNA, Y-Chromosome, and Whole-Genome Comparisons All Trace Human Origins Back to Africa 100,000-200,000 Years Ago

HUMAN DIVERSITY

Scientific American Library
1982 ISBN 07167-14698

RICHARD LEWONTIN

Human Races Have a Genetic Basis:

a. Yes
b. No

There is More Genetic Diversity Within Populations than Between Populations!! So Much for the Concept of racial "purity"!!!!

More Genetic Diversity Within Any Population Than Between Polulations

Proportion of genetic diversity accounted for within and between populations and races

Gene	Total $H_{species}$	Proportion		
		Within Populations	Within Races between Populations	Between Races
Hp	.994	.893	.051	.056
Ag	.994	.834	—	—
Lp	.639	.939	—	—
Xm	.869	.997	—	—
Ap	.989	.927	.062	.011
6PGD	.327	.875	.058	.067
PGM	.758	.942	.033	.025
Ak	.184	.848	.021	.131
Kidd	.977	.741	.211	.048
Duffy	.938	.636	.105	.259
Lewis	.994	.966	.032	.002
Kell	.189	.901	.073	.026
Lutheran	.153	.694	.214	.092
P	1.000	.949	.029	.022
MNS	1.746	.911	.041	.048
Rh	1.900	.674	.073	.253
ABO	1.241	.907	.063	.030
Mean		.854	.083	.063

Source: R. C. Lewontin, *Genetic Basis of Evolutionary Change* (Columbia University Press, 1974).

1. 85% of Human Genetic Variations Occurs within Populations & Between Individuals in that Populations!
2. Remaining 15% of Human Genetic Variation Split Between Different Populations of Same "race" (8%) & Between Different "races" (6%)
3. Only 6% of Human Genetic Variation are to Differences between races!!! Mostly Geographic. Note: THERE ARE GROUP DIFFERENCES!

Within Population Differences Account For 95% of Human Genetic Variation

Genetic Structure of Human Populations

Noah A. Rosenberg,[1]* Jonathan K. Pritchard,[2] James L. Weber,[3] Howard M. Cann,[4] Kenneth K. Kidd,[5] Lev A. Zhivotovsky,[6] Marcus W. Feldman[7]

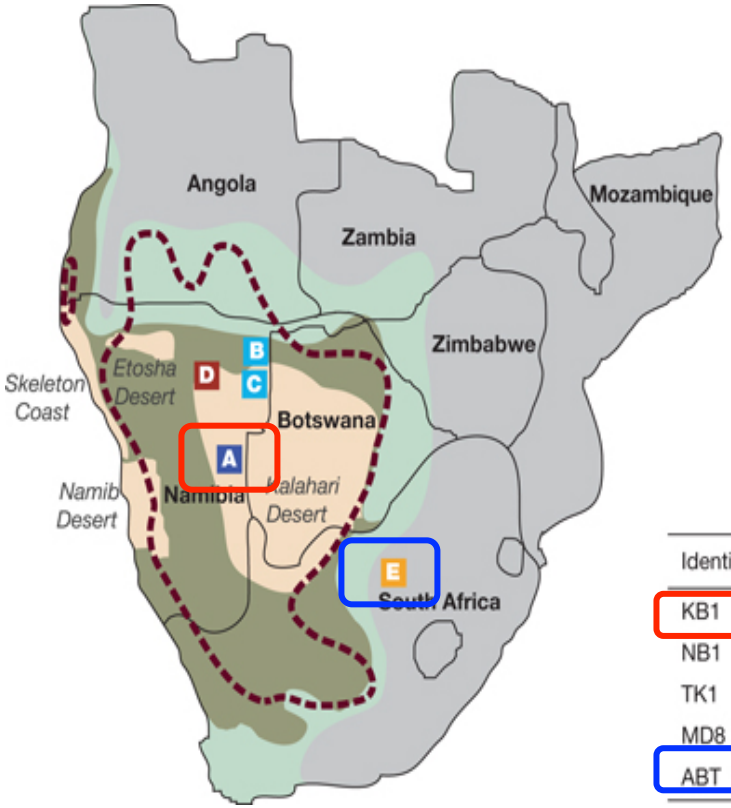
We studied human population structure using genotypes at 377 autosomal microsatellite loci in 1056 individuals from 52 populations. Within-population differences among individuals account for 93 to 95% of genetic variation; differences among major groups constitute only 3 to 5%. Nevertheless, without using prior information about the origins of individuals, we identified six main genetic clusters, five of which correspond to major geographic regions, and subclusters that often correspond to individual populations. General agreement of genetic and predefined populations suggests that self-reported ancestry can facilitate assessments of epidemiological risks but does not obviate the need to use genetic information in genetic association studies.

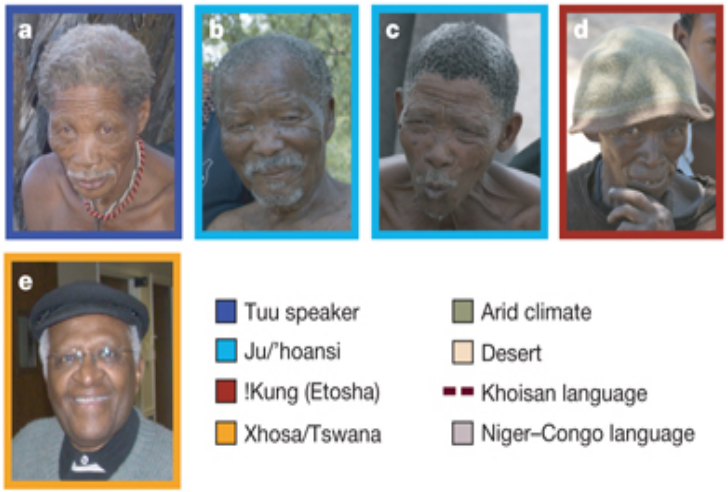
Table 1. Analysis of molecular variance (AMOVA). Eurasia, which encompasses Europe, the Middle East, and Central/South Asia, is treated as one region in the five-region AMOVA but is subdivided in the seven-region design. The World-B97 sample mimics a previous study (*6*).

Sample	Number of regions	Number of populations	Variance components and 95% confidence intervals (%)		
			Within populations	Among populations within regions	Among regions
World	1	52	94.6 (94.3, 94.8)	5.4 (5.2, 5.7)	
World	5	52	93.2 (92.9, 93.5)	2.5 (2.4, 2.6)	4.3 (4.0, 4.7)
World	7	52	94.1 (93.8, 94.3)	2.4 (2.3, 2.5)	3.6 (3.3, 3.9)
World-B97	5	14	89.8 (89.3, 90.2)	5.0 (4.8, 5.3)	5.2 (4.7, 5.7)
Africa	1	6	96.9 (96.7, 97.1)	3.1 (2.9, 3.3)	
Eurasia	1	21	98.5 (98.4, 98.6)	1.5 (1.4, 1.6)	
Eurasia	3	21	98.3 (98.2, 98.4)	1.2 (1.1, 1.3)	0.5 (0.4, 0.6)
Europe	1	8	99.3 (99.1, 99.4)	0.7 (0.6, 0.9)	
Middle East	1	4	98.7 (98.6, 98.8)	1.3 (1.2, 1.4)	
Central/South Asia	1	9	98.6 (98.5, 98.8)	1.4 (1.2, 1.5)	
East Asia	1	18	98.7 (98.6, 98.9)	1.3 (1.1, 1.4)	
Oceania	1	2	93.6 (92.8, 94.3)	6.4 (5.7, 7.2)	
America	1	5	88.4 (87.7, 89.0)	11.6 (11.0, 12.3)	

But - There Are Differences! But…They Fall Into Geographical Groups -- Groups Divided Originally by Geographic Barriers (Ocean, Desert, Mountains). The 5% Difference Allows Us to Mark and Trace Ancestry!

Recent Sequencing of Two African Genomes Reveals Remarkable Genetic Diversity





Identifier	Name	Location of origin	Linguistic group	Y chromosome
KB1	!Gubi	Southern Kalahari	Tuu-speaker	B2b
NB1	G/aq'o	Northern Kalahari	Juu (Ju/'hoansi)	A3b1
TK1	D#kgao	Northern Kalahari	Juu (Ju/'hoansi)	A2
MD8	!Aî	Northern Kalahari	Juu (!Kung)	E1b1b1
ABT	Tutu	South Africa	Bantu (Xhosa/Tswana)	E1b1a8a *

SC Schuster et al. Nature 463, 943-947 (2010)

Each Genome Contains One Million SNPs Not Found in Any Other Genome

Conclusions

1. If 85% of Human Genetic Variation Occurs Between Different People Within Any Given Population (localized)
2. If only 7% of Human Genetic Variation Occurs Between "Races" (novel alleles specific to "races") e.g. F_{yB} ^{ES}
3. Then Losing all "Races" Except One Retains 94% of all Human Genetic Variation!

[85% + (15%-7%)]= 94%

Variation That Occurs in Ancestral Population

85% Within Population genetic variability

8% Between Populations of Same "Race"

7% Between "Race" Genetic Variability

4. ∴ Human Highly Heterozygous or Hybrids- & If Above Not True- Most of Us Would Not Be Here- Need Genetic Variation to Survive!

So What is a "Race"?

1. Primarily a sociological concept- but could be a localized or inbred population that has a higher frequency of alleles at a very small number of loci. Affects few physical features.
2. High frequency alleles in one "race" are present at lower frequencies in other "races". All humans have same genes- differ in form mostly within populations!
3. Heterozygosity (variation) high in human populations- all populations. None homozygous at all loci!
4. No such thing as a "pure" race - would have little variation
5. Genes affecting physical features not representation of genes across genome-

Geographical Ancestry is relevant-many "racial" groups now have multiple ancestries because of admixture and migration

A Better Term is POPULATIONS!


Knowledge or Certainty: The Ascent of Man Series
Jacob Bronowski, 1973