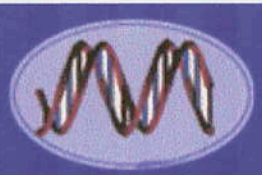
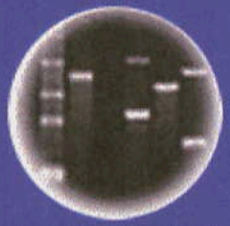


DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

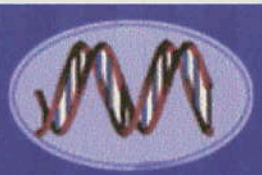
Plants of Tomorrow

HC70A Winter 2008
Genetic Engineering in Medicine, Agriculture, and Law

Professor Bob Goldberg

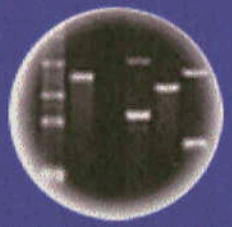
Lecture 6
Identifying Human Origins: Past and Present

70% lectures 2/12/08 to pg 9
Need 2 full lectures for Lecture 5+6 - maybe 3



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

THEMES

1. Review-Origin & Detection of DNA Variation
2. DNA Sequencing-The Ultimate Measure of DNA Variation
3. How Are Genomes Sequenced?
4. What Are the Different DNA Sequencing Methods?
5. How Many Whole Genomes Have Been Sequenced to Date?
6. How Can Comparative DNA Sequencing Provide Insights Into Human Origins?
7. How Has Massively Parallel DNA Sequencing Revolutionized Our Ability to Sequence Whole Genomes?
8. How Will the HapMap Project Provide Insights Into Human Genetic Variability and Disease?

STOP 2/12/07 1.25 hr lecture

9. How Can SNP Chips Measure Genotypes Across the Entire Genome?
10. How Can DNA Sequence Variation be Used to Trace Human Ancestry?
11. Are There "Races?"

Genetic Variability Gives Us our Individuality :

a. Yes
b. No

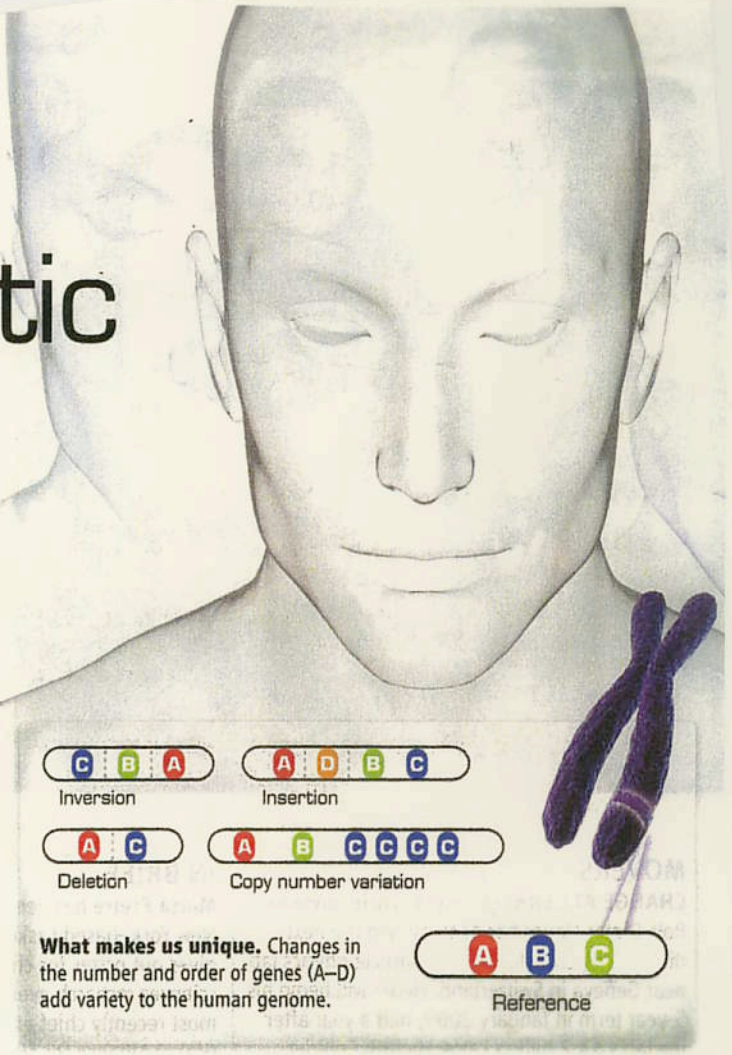
iClicker

Human Gene Variation is what makes us Individuals

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



What makes us unique. Changes in the number and order of genes (A–D) add variety to the human genome.

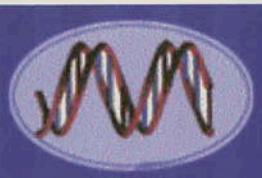
THE UNVEILING OF THE HUMAN GENOME ALMOST 7 YEARS AGO cast the first faint light on our complete genetic makeup. Since then, each new genome sequenced and each new individual studied has illuminated our genomic landscape in ever more detail. In 2007, researchers came to appreciate the extent to which 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.

It can be used to:

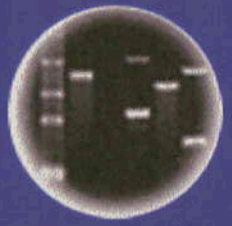
1. Trace our origins & Ancestry
2. Profile the presence of alleles that make us susceptible to Disease

i.e. Who we are & where we come from



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

Uses of SNPs

- Gene Identity-Allelic Markers
- Disease Gene Identity-Pedigrees-Testing
- Group Identity-Population History
- Individual Identity
  - Preventative Medicine
  - Pharmacogenetics
  - DNA Fingerprinting
- Group Susceptibility to Disease/Drugs

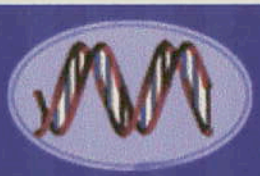
HUMAN Genetic VARIATION IS DUE PRIMARILY to SNPs & VNTRs (SSRs)

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

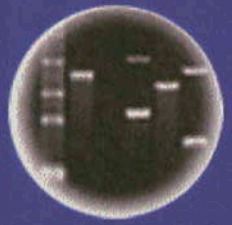
Single nucleotide polymorphism (SNP) ...GCAA T TCCCGATT...
...GCAA G TCCCGATT...

Simple sequence repeat (SSR) ...GCATTATATATATATATC...
...GCATTATAT[ ]C...



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

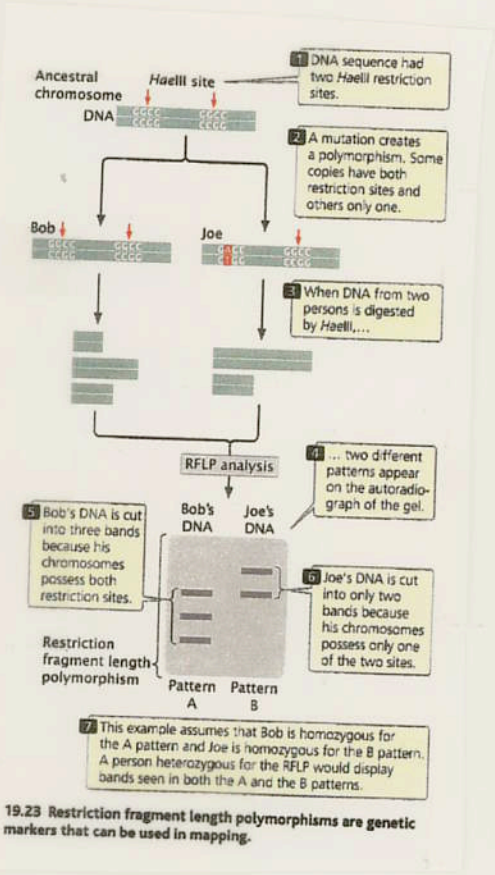
Detection of SNPs in the Human Genome

- DNA Blot-RFLP Analysis
- PCR-RFLP Analysis
- ASO Analysis
- Whole Genome Sequence Comparisons
- Whole Genome SNP Chip Comparisons

HISTORICALLY - DIFFERENCES IN SNPs & VNTRs WERE DETERMINED BY DNA BLOTS, PCR, & ASO ANALYSES

RFLP BLOT

RFLP PCR ANALYSIS



19.23 Restriction fragment length polymorphisms are genetic markers that can be used in mapping.

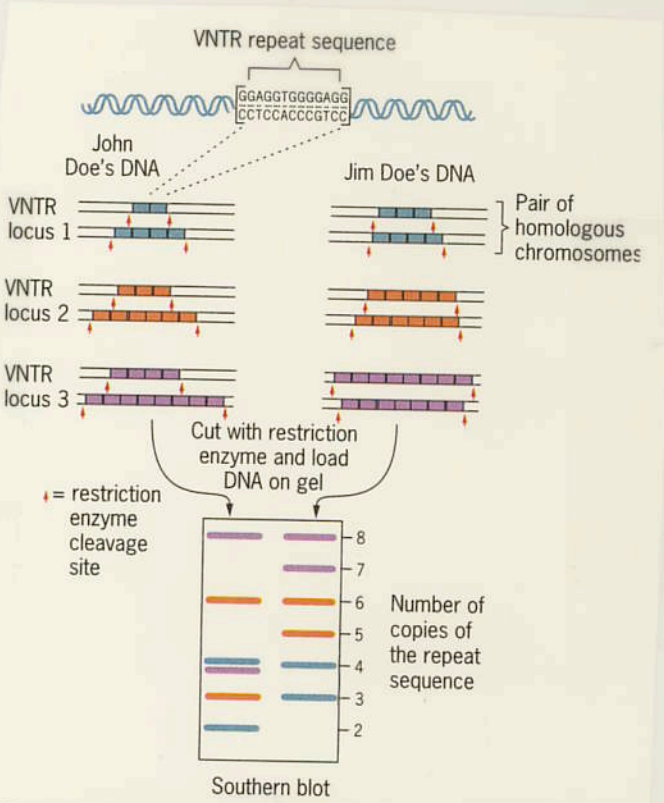


Figure 22.8 Simplified diagram of the use of variable number tandem repeats in preparing DNA fingerprints.

BOTH DEPEND UPON RFLPs
Distinguished by Size
using Gel Electrophoresis

ASOs Discriminate Between Two Alleles By Hybridization

(a)

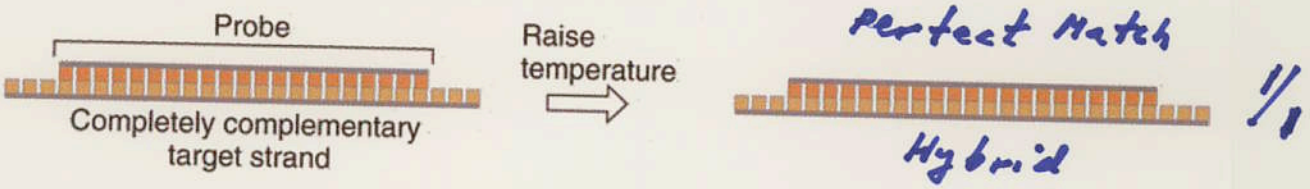
5′ GACTCCTGAGGAGAAGTGC

ASO for allele 1

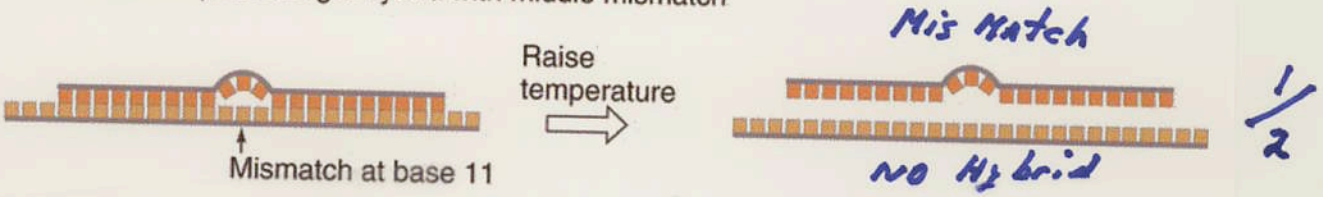
5′ GACTCCTGTGGAGAAGTGC

ASO for allele 2

(a) 1. 21-base probe/target hybrid with no mismatches



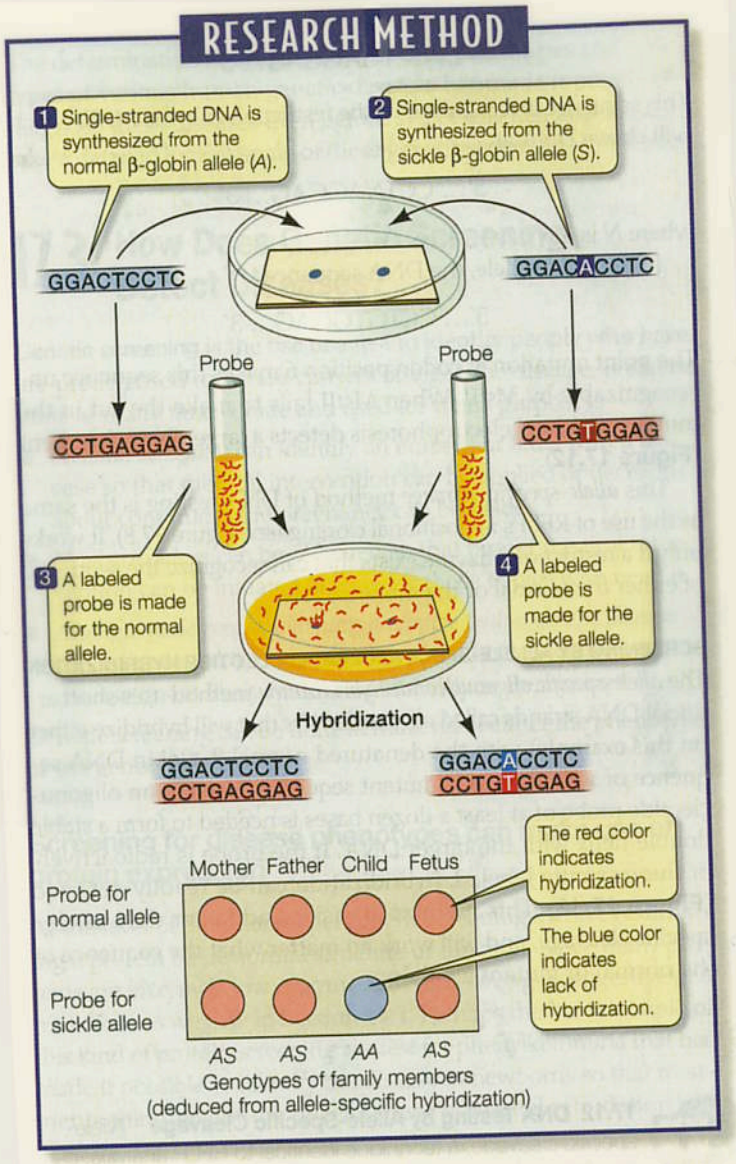
2. 21-base probe/target hybrid with middle mismatch



Principle: Only a **Perfect-Match** Hybrid CAN FORM Under Specific Conditions Using Short Oligonucleotides

BASIS OF SNP Chips!

Using ASOs to Detect Sickle-Cell Globin Allele



17.13 DNA Testing by Allele-Specific Oligonucleotide Hybridization Testing of this family reveals that three of them are heterozygous carriers of the sickle allele. The first child, however, has inherited two normal alleles and is neither affected by the disease nor a carrier.

Using ASOs to detect for Presence of the Cystic Fibrosis Disease Allele

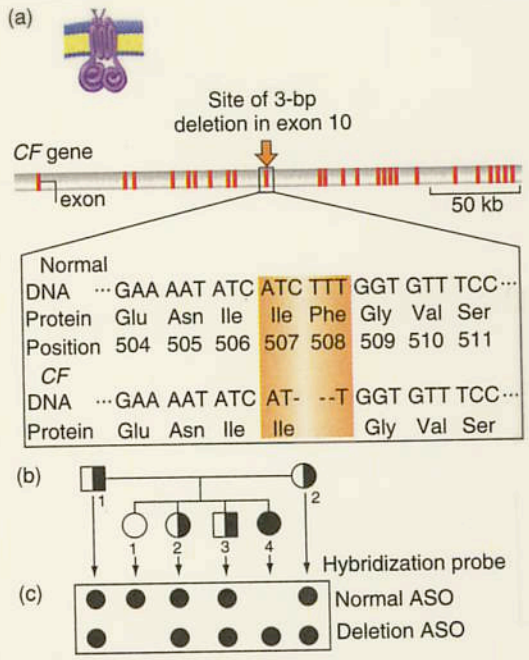
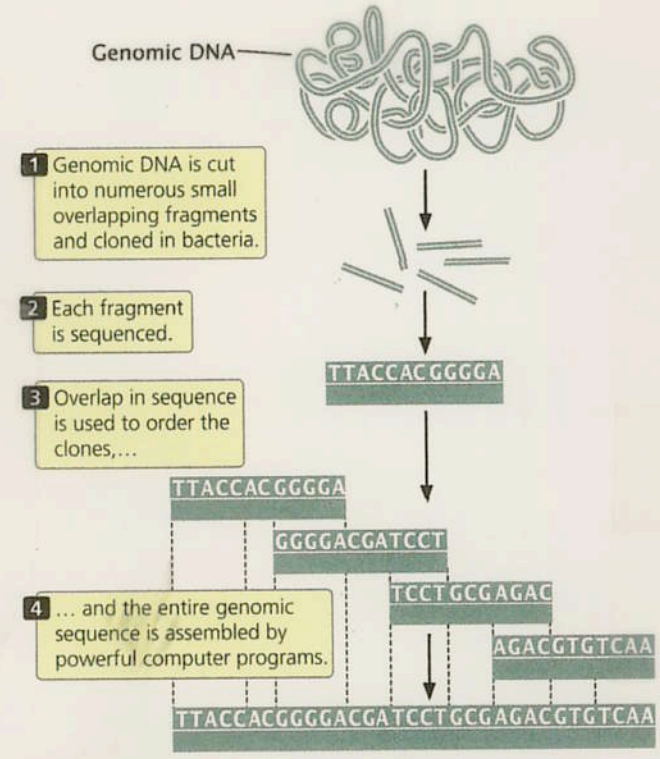


Figure 9.17 Direct detection of the most common cystic fibrosis mutation. The *CF* gene extends across 250,000 base pairs and is organized in 27 exons. It encodes a protein with 1480 amino acids. (a) The most common disease-causing mutation in the *CF* gene is a deletion of three bases in exon 10. It is possible to amplify the region containing the site of the most common mutation by PCR and divide the PCR products into two aliquots. You then blot the aliquots onto filter paper and probe with ASOs for the wild-type and mutant alleles. The ASO for the mutant allele differs by the absence of three bases from the ASO for the wild-type allele. (b) Pedigree of a family in which one daughter (child 4) has cystic fibrosis. (c) Analysis of the results of an ASO hybridization test provide direct information on the *CF* genotype of all family members.

On a whole Genome Scale "ASOs" can Scan for Thousands of Disease Alleles

The Shotgun Approach is now Used Almost Exclusively to SEQUENCE Genomes



20.10 Whole-genome shotgun sequencing utilizes sequence overlap to align sequenced fragments.

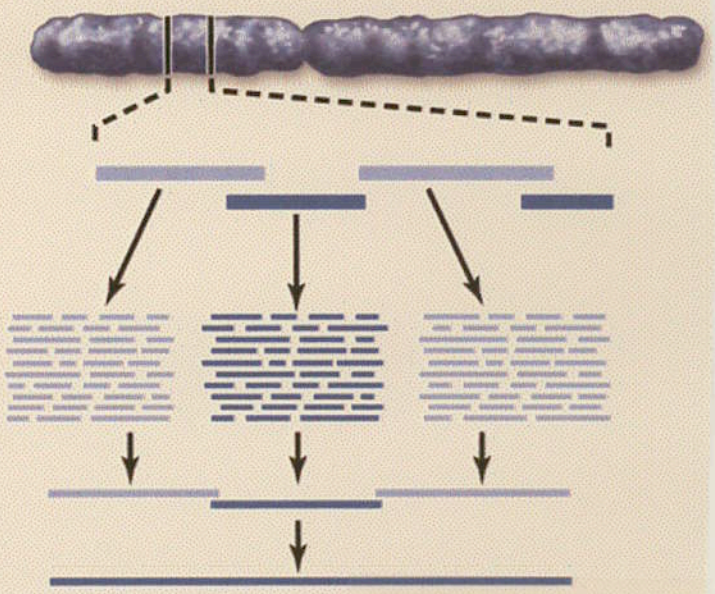
Because of: (1) Advances in Speed + Cost of Sequencing - i.e., can generate 1X Human Genome Equivalent sequences in a day on one Machine

(2) Large Data Bases to "find" + Identify DNA Sequences!

TWO SEQUENCING APPROACHES FOR WHOLE GENOMES

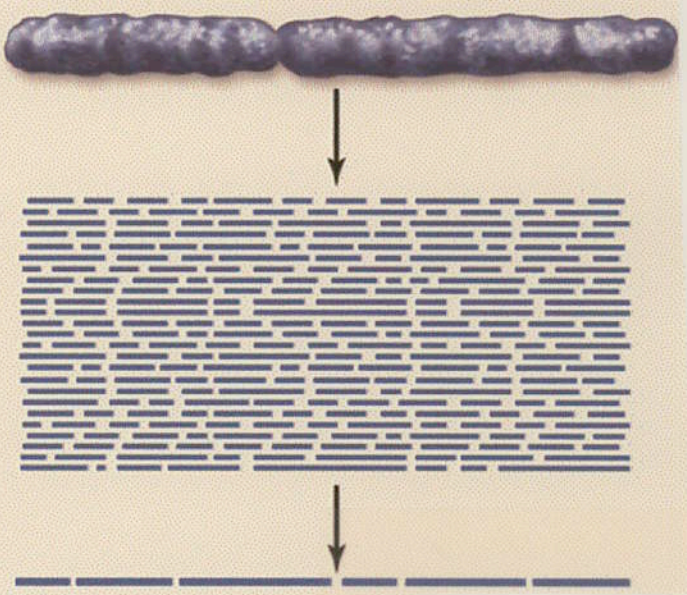
Clone-by-Clone Method TOP DOWN

1. Large DNA clones are first isolated. These are arranged into contiguous sequences based on overlapping tagged sites.
2. Large clones are fragmented into smaller clones for sequencing.
3. The entire sequence is assembled from the overlapping larger clones.

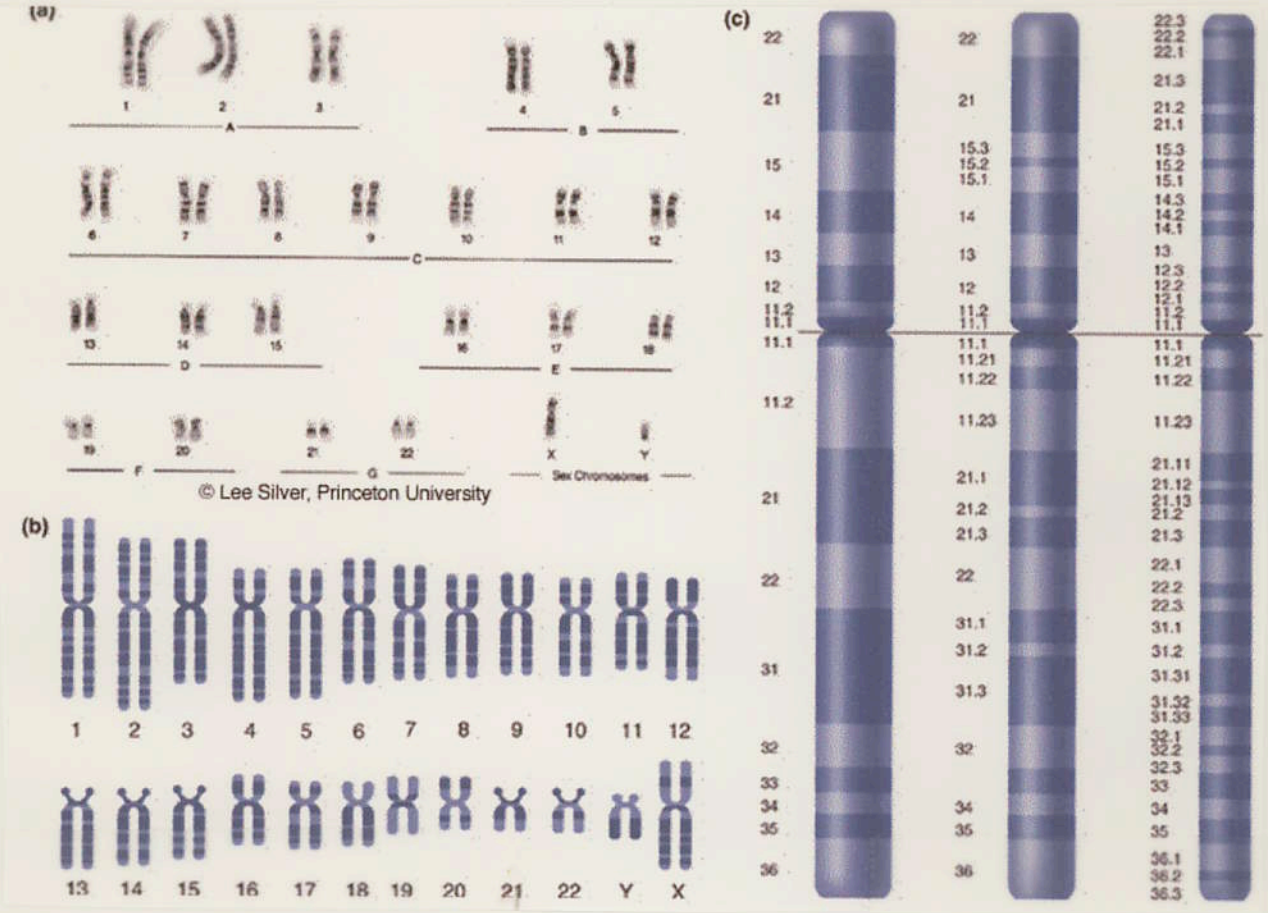


Shotgun Method BOTTOM UP

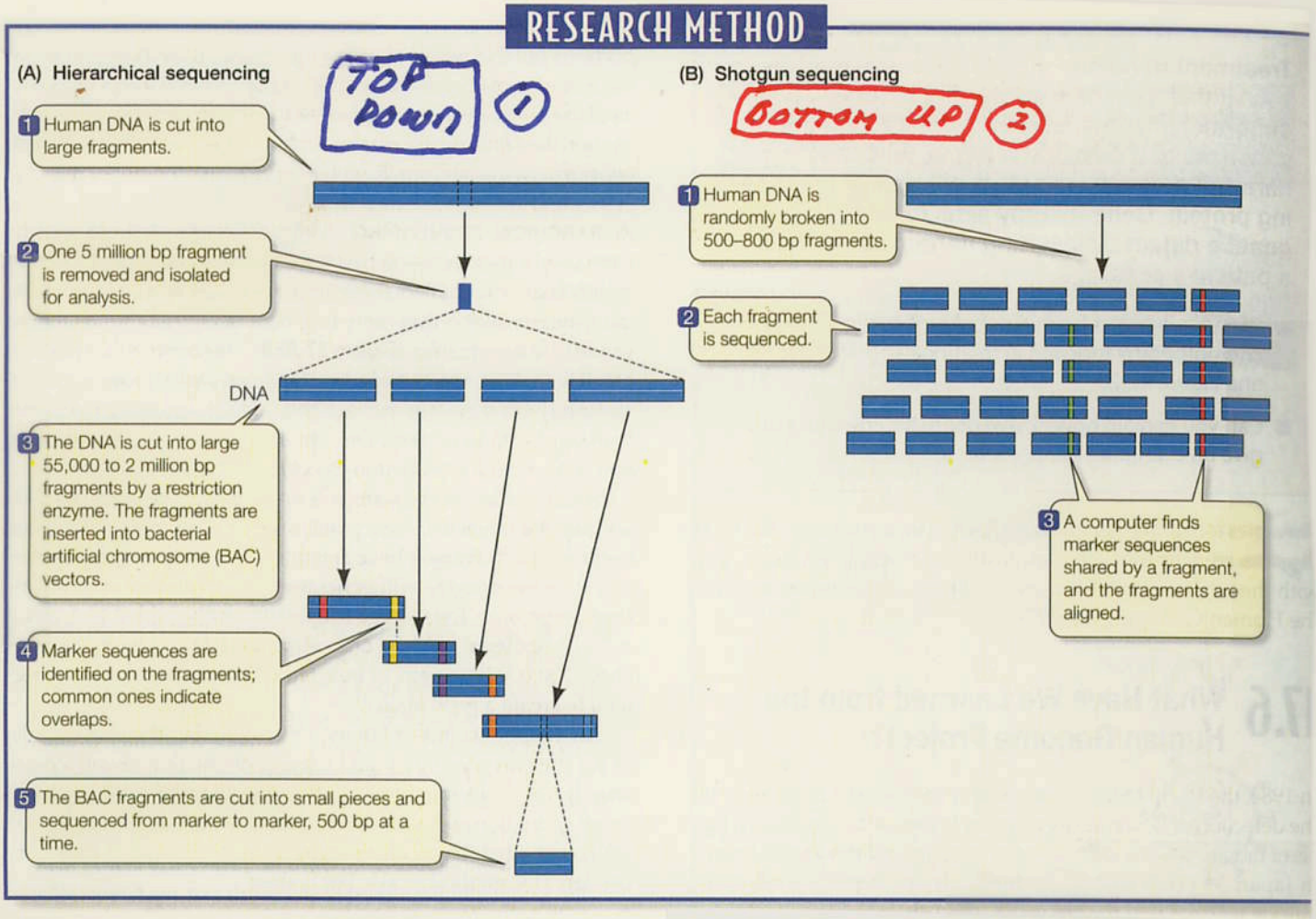
1. Cut DNA of entire chromosome into small fragments and clone.
2. Sequence each segment and arrange based on overlapping nucleotide sequences.



Genetic VARIation in a Genome is ULtimately Determined By Sequencing Several or More INdividual Genomes



Historically, Two Approaches were used to SEQUENCE The HUMAN Genome (and other Genomes)



17.22 Two Approaches to Sequencing DNA (A) In the hierarchical approach to genome sequencing, genetic markers are mapped, and DNA fragments are then aligned by matching overlapping sites with the same markers. (B) In the shotgun approach, the DNA is fragmented and a computer is used to find overlapping markers.

BOTH DEPEND ON MAKING Genome Libraries i.e., need Clone to Sequence Until 2008!

The Human Genome Sequence published in 2001 came from a mixture of five individuals DNA - 3♀ + 2♂ - 2 Caucasians, 1 African American, 1 Chinese, & 1 Hispanic

Uncovered 1 SNP / 1250 bp

Now have 1 SNP / 700 bp

HISTORICAL DEVELOPMENTS IN DNA SEQUENCING TECHNOLOGY

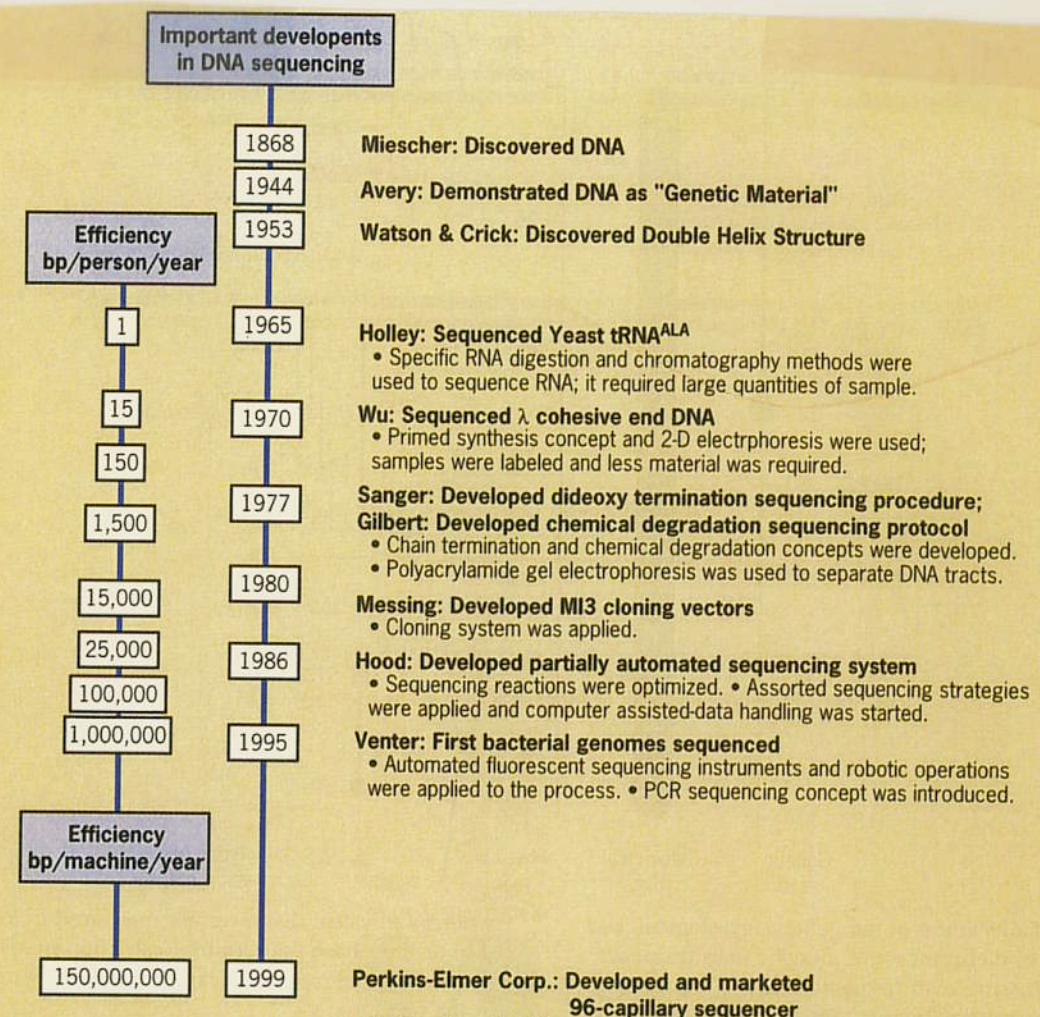


Figure 2. Advances in DNA sequencing efficiency and some of the technological developments that enhanced the productivity of sequencers. Initially, all the steps in DNA sequencing were performed manually, making it a very labor-intensive process. However, fully automated sequencing machines have now largely replaced human sequencers, greatly increasing efficiency.

1×10^{12} bp/year/Machine 2009

(300 Human Genomes!!)

Massively Parallel Pyrosequencing

Nuclear and Mitochondrial DNA Replication Are Semi-Conservative Processes?

a. Yes
b. No

iClicker

DNA Synthesis Occurs in a 5' to 3' Direction?

a. Yes
b. No

iClicker

DNA Synthesis Results in the Formation of Phosphodiester Bonds and Requires Free 3'OH Groups on the Deoxynucleotide Sugar:

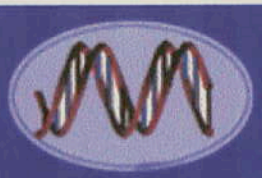
a. Yes
b. No

iClicker

To Sequence a DNA Fragment, a Clone of the DNA Must Obtained First:

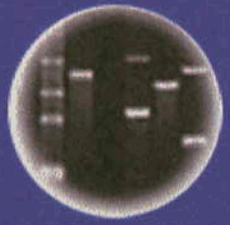
a. Yes
b. No

iClicker



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

All DNA Sequencing Methods Require:

- A DNA Clone (Until Recently)
- A Fixed-End Reference Point
- Single-Stranded DNA Template(s)
- Methods to Distinguish Each Base At Specific Positions (A,G, C, or T)
- Gel Electrophoresis to Separate DNA Fragments by Size
- Method of Detecting DNA Fragment Sizes and/or Positions of Specific Bases

MAXAM-GILBERT SEQUENCING RELIES ON Chemical Breakage of Phosphodiester Bonds at Specific BASES

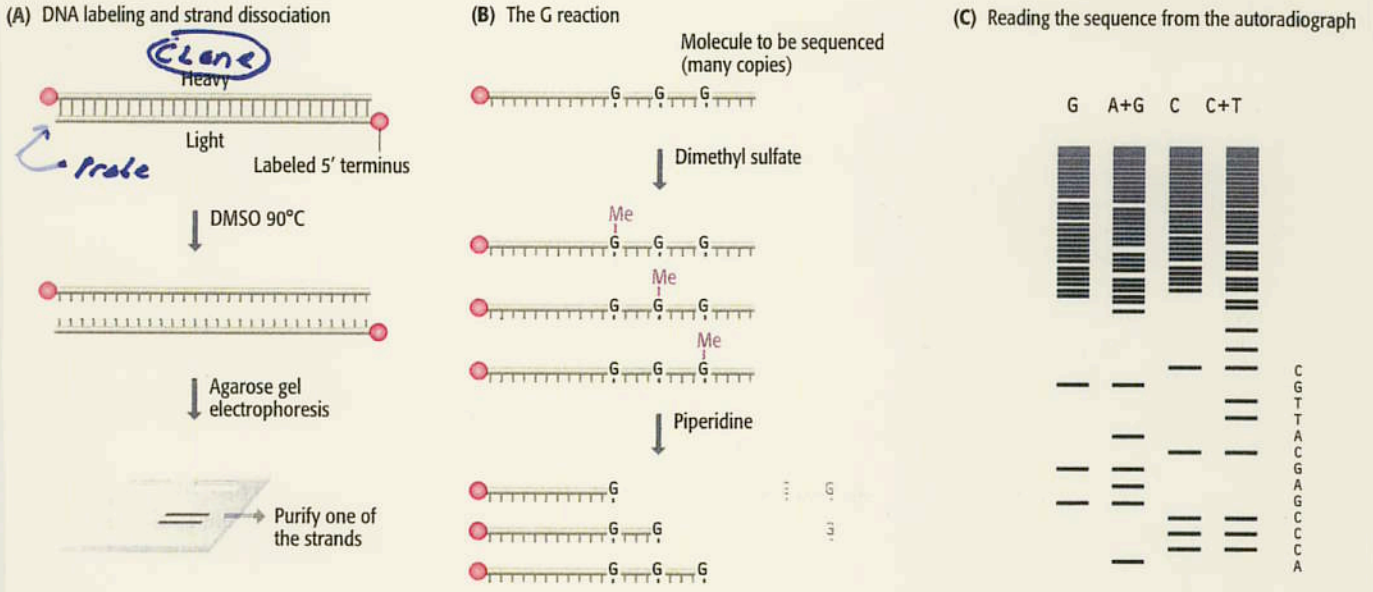


Figure 4.8 Chemical degradation sequencing.

Chemical Reactions Randomly Break DNA strands from fixed labeled end to Read Sequence

no Longer in Use!

TWO DIFFERENT APPROACHES TO SEQUENCE DNA WERE INVENTED 30 YEARS AGO!

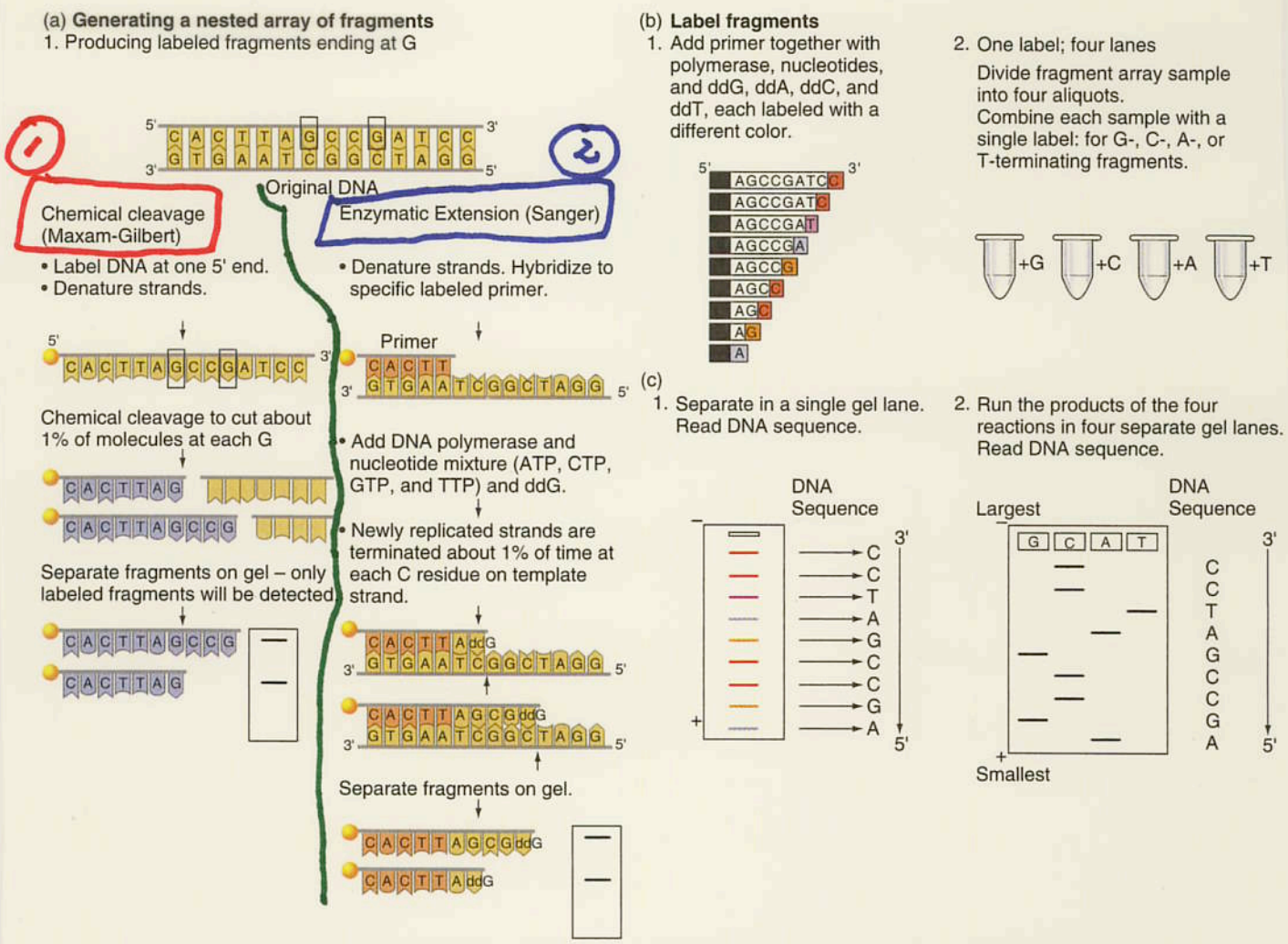
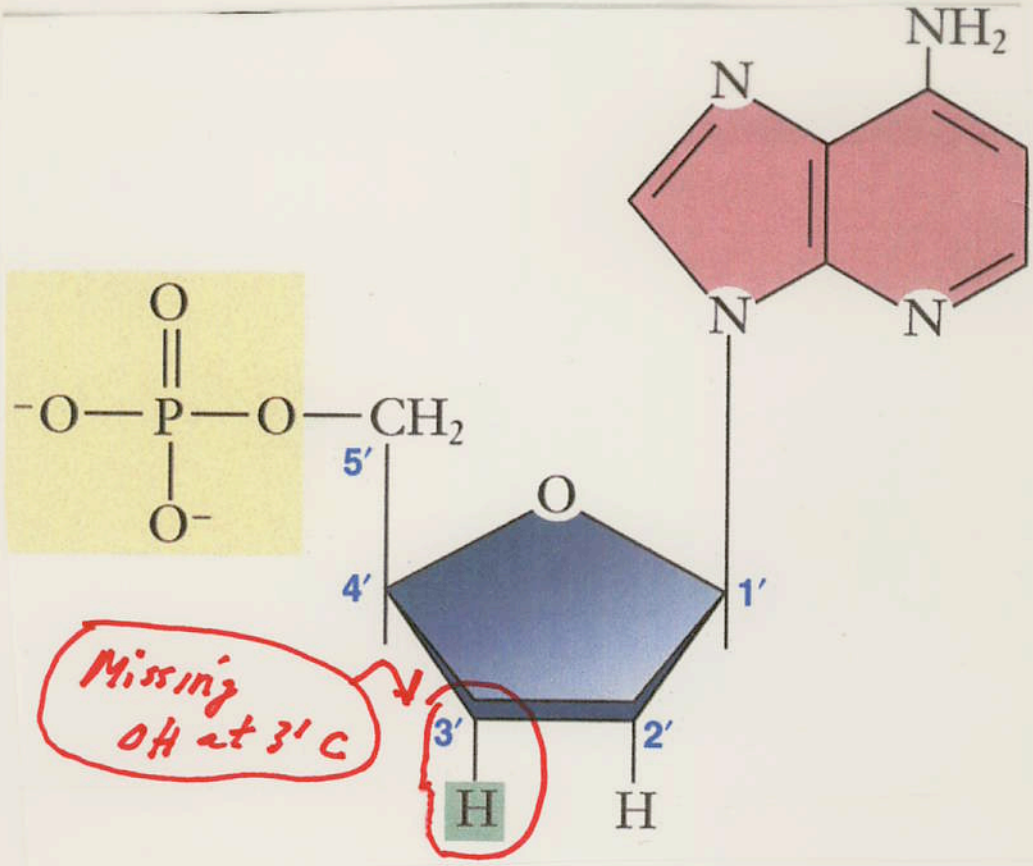


Figure 8.20 General principles of DNA sequencing. (a) All sequencing protocols rely on the production of nested arrays of fragments; the fragments all begin at the same 5′ position in a sequence but end at different 3′ positions. (1) There are two basic ways to produce the fragments: chemical cleavage, a technique developed by A. M. Maxam and W. Gilbert, and enzymatic extension, developed by F. Sanger. The chemical method uses specific chemicals to cleave the DNA after each type of base—A, G, T, or C; it requires the control of conditions such that cleavage affects only a small fraction of the bonds at a particular base. The enzymatic method depends on the ability to terminate DNA synthesis with a nucleotide base lacking a hydroxyl group that is critical for chain extension; conditions are controlled so that only a small fraction of the growing DNA chains are terminated with the addition of the chain-terminating base. (2) To generate nested arrays from the fragments produced by both of these procedures, researchers use polyacrylamide gel electrophoresis under conditions that allow the separation of DNA molecules differing in length by a single nucleotide. (b.1) It is possible to generate nested fragments ending in all four nucleotides in a single reaction, with each of the terminating nucleotides labeled with a different colored fluorescent dye. (b.2) Alternatively, four separate reactions can be performed with a different similarly labeled terminating nucleotide in each reaction. (c.1) If you generate a nested array in a single reaction, you can separate the reaction products in a single lane, with the color of each fragment indicating the associated terminating nucleotide. (c.2) Alternatively, if you use four different reactions to generate the nested array, you must use four adjacent lanes to separate the reaction products; at each position on the gel, a band will appear in only one of four lanes; its presence specifies the nucleotide at that position.

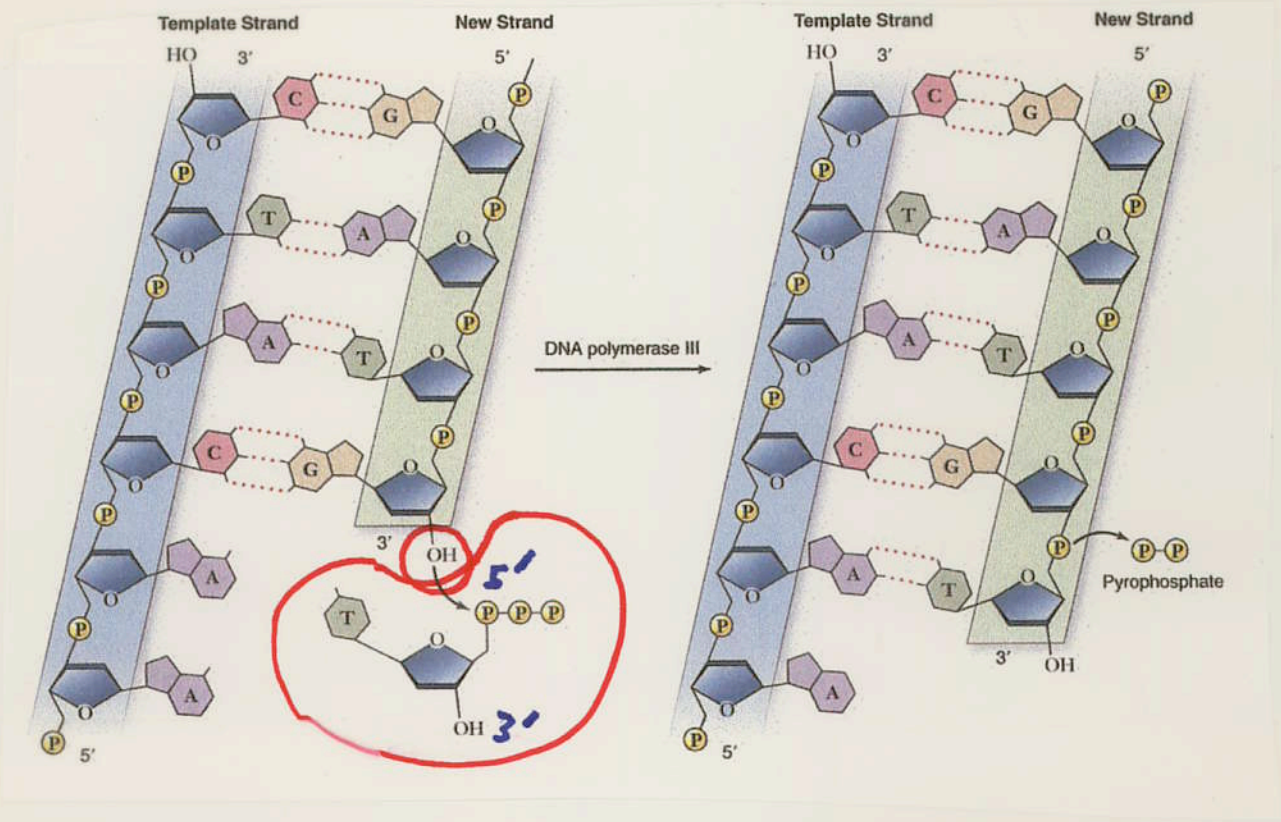
SANGER SEQUENCING WAS USED FOR THE HUMAN GENOME PROJECT!

Dideoxy nucleotides Terminate DNA Synthesis



Phosphodiester Bond CANNOT FORM

SANGER CHAIN-TERMINATING SEQUENCING RELIES ON DNA Synthesis Reactions

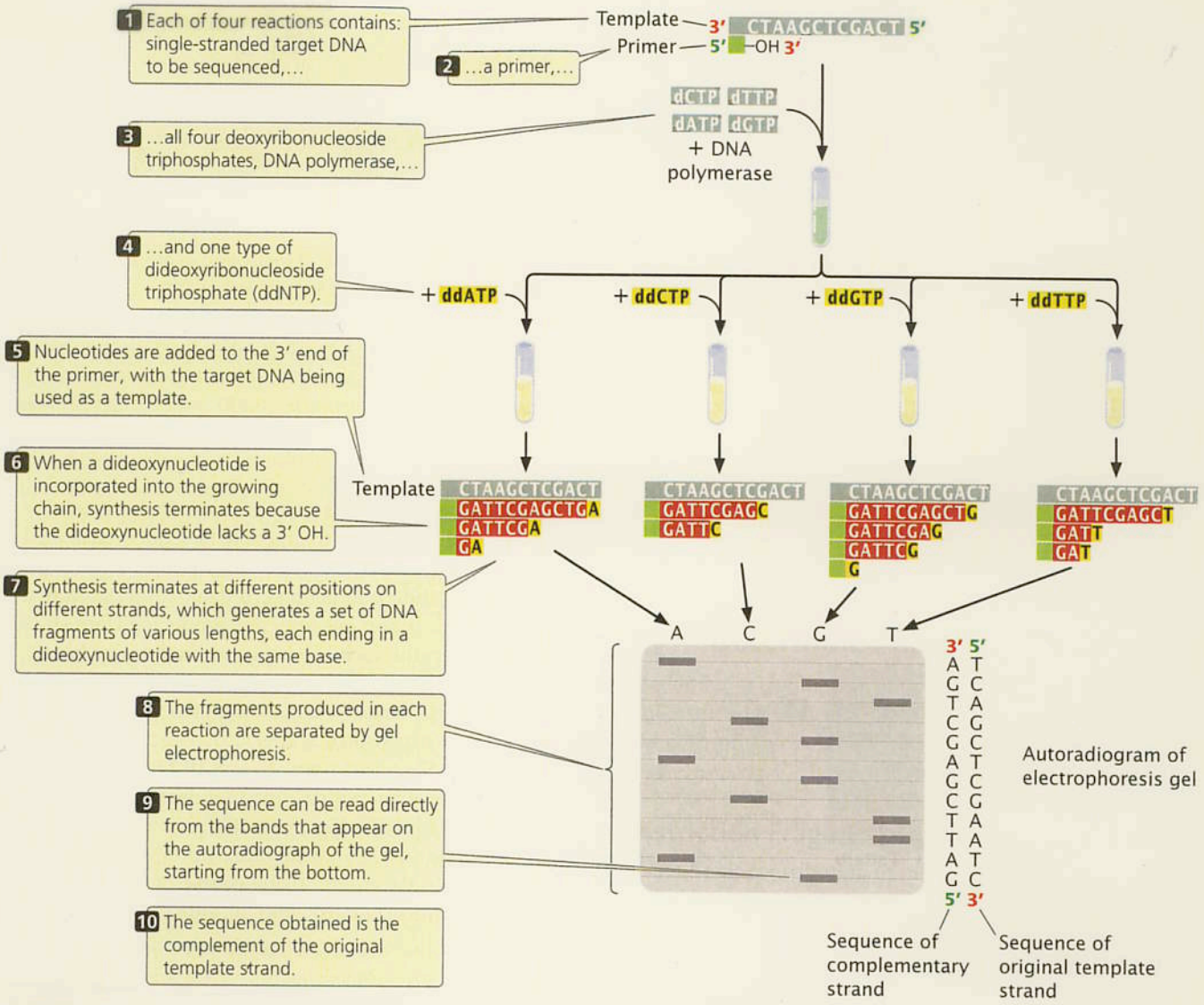


Recall! DNA Synthesis is Semi-Conservative & Proceeds 5' → 3'

REQUIRES 3'OH to FORM Phosphodiester Bond!

DNA Synthesis CANNOT PROCEED without 3'OH

SANGER DIDEOXY NUCLEOTIDE CHAIN TERMINATING SEQUENCING



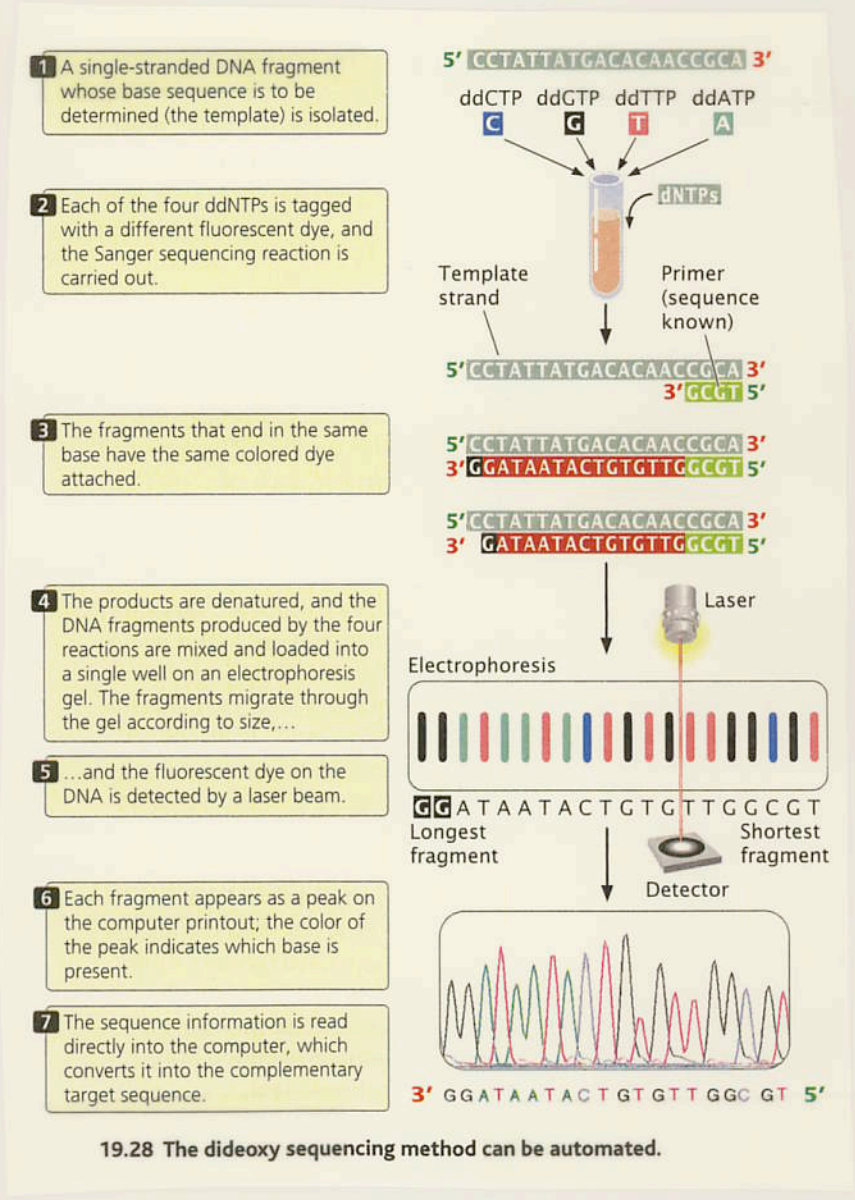
19.26 The dideoxy method of DNA sequencing is based on the termination of DNA synthesis.

RECALL: DNA Synthesis Requires

(1) Template (2) Primer (3) dNTPs

(4) DNA Polymerase

The Use of FLUORESCENT ddNTPs ALLOWED Sanger Sequencing to be AUTOMATED & MADE HIGH Throughput



19.28 The dideoxy sequencing method can be automated.

ONLY USED FOR ROUTINE DNA SEQUENCING TODAY

NOT Genome Projects!

PROCESS COULD BE AUTOMATED BY USE OF Robots & Computers & NEW Software to Analyze DNA Sequences

↳ Enabled Whole-Genome Projects!

A Complete Genome Sequence Requires Sequencing

a. 1X the Amount of DNA in the Genome
b. 10X the Amount of DNA in the Genome
c. 100X the Amount of DNA in the Genome

iClicker

GENOME PROJECTS HAVE EXPLODED

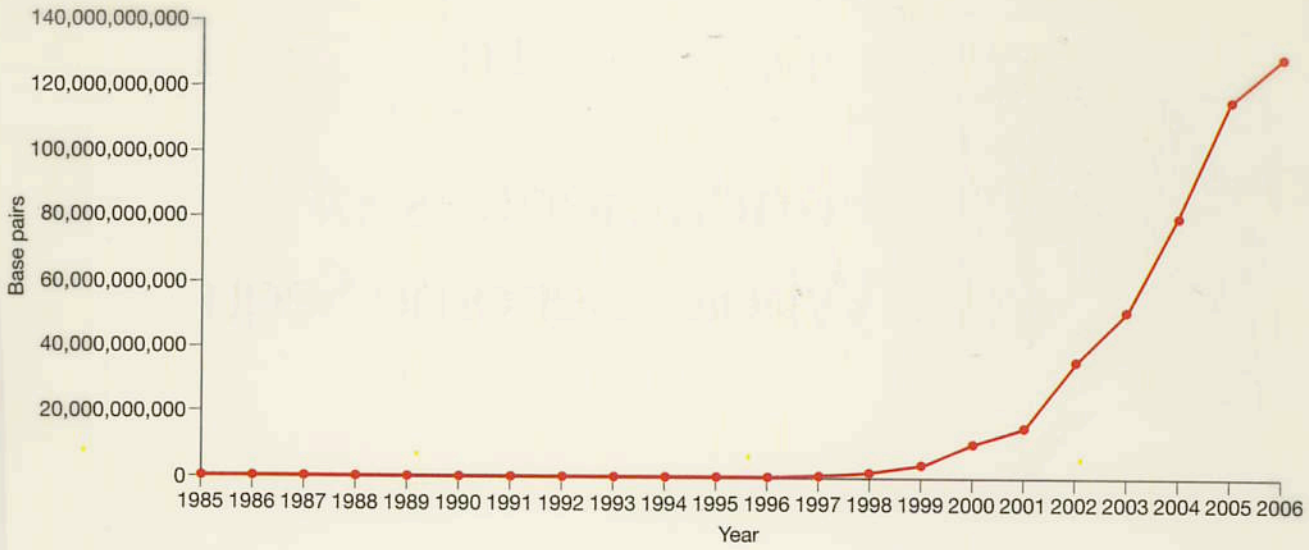
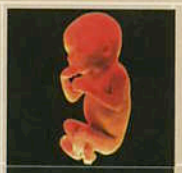


FIGURE 10-1

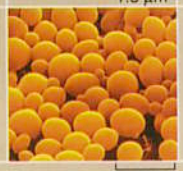
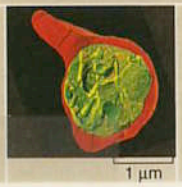
Growth of DNA sequences deposited into the public databases (GenBank, EBI, DBJL) from 1985 to 2006. GenBank and the other public databases began their operations—storing and disseminating most of the DNA sequences available—in the early 1980s. The increase in sequence data results from the massive increase in sequencing capabilities that came with the Human Genome Project, particularly in the late 1990s when sequencing of the human genome began in earnest. A milestone was reached in 2005 when the 100,000,000,000th (one hundred billionth) base pair was deposited into the databases.

TABLE 24.1 Milestones for Comparative Eukaryotic Genomics

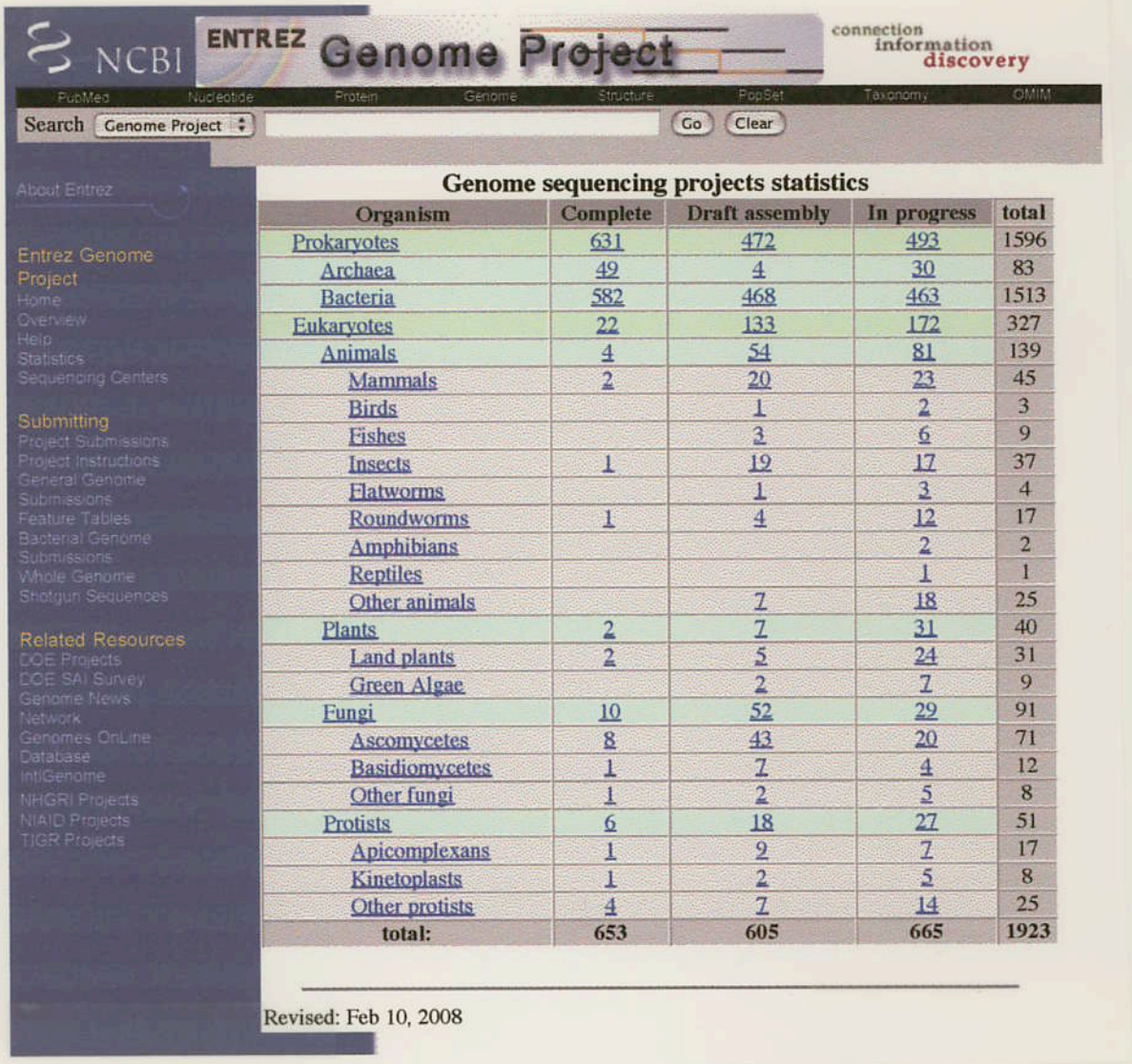
Organism		Estimated Genome Size (Mb)	Estimated Number of Genes	Year Sequenced
Vertebrates				
Homo sapiens (human) HUMANS		2,900	20,000–25,000	2001
Mus musculus (mouse) MOUSE		2,600	30,000	2002
Fugu rubripes (pufferfish)		365	33,609	2002
Rattus norvegicus (rat) RAT		2,750	20,973	2004
Pan troglodytes (chimpanzee) CHIMP		3,100	20,000–25,000	2005

GENOME PROJECTS HAVE EXPLODED DUE TO increase in Speed + Reduction in COST OF SEQUENCING

TABLE 24.1 Milestones for Comparative Eukaryotic Genomics, continued

Organism		Estimated Genome Size (Mb)	Estimated Number of Genes	Year Sequenced
Vertebrates, continued				
CHICKEN *Gallus gallus* (red jungle fowl)		1,000	20,000–23,000	2004
Invertebrates				
Drosophila melanogaster (fruit fly)		137	13,600	2000
Anopheles gambiae (mosquito)		278	46,000–56,000	2002
Fungi				
Schizosaccharomyces pombe (fission yeast)		13.8	4,824	2002
Saccharomyces cerevisiae (brewer's yeast)		12.7	5,805	1997
Plants				
Arabidopsis thaliana (wall cress)		125	25,498	2000
Oryza sativa (rice)		430	41,000	2002
Protists				
Plasmodium falciparum (malaria parasite)		23	5,300	2002

TO DATE 1,000 GENOMES HAVE BEEN SEQUENCED PROVIDING NOVEL NEW INSIGHTS INTO HOW GENOMES EVOLVED



NCBI ENTREZ Genome Project connection information discovery

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM

Search Genome Project Go Clear

About Entrez

Entrez Genome Project
Home
Overview
Help
Statistics
Sequencing Centers

Submitting
Project Submissions
Project Instructions
General Genome Submissions
Feature Tables
Bacterial Genome Submissions
Whole Genome Shotgun Sequences

Related Resources
DOE Projects
DOE SAI Survey
Genome News Network
Genomes OnLine Database
IntlGenome
NHGRI Projects
NIAID Projects
TIGR Projects

Genome sequencing projects statistics

Organism	Complete	Draft assembly	In progress	total
Prokaryotes	631	472	493	1596
Archaea	49	4	30	83
Bacteria	582	468	463	1513
Eukaryotes	22	133	172	327
Animals	4	54	81	139
Mammals	2	20	23	45
Birds		1	2	3
Fishes		3	6	9
Insects	1	19	17	37
Flatworms		1	3	4
Roundworms	1	4	12	17
Amphibians			2	2
Reptiles			1	1
Other animals		7	18	25
Plants	2	7	31	40
Land plants	2	5	24	31
Green Algae		2	7	9
Fungi	10	52	29	91
Ascomycetes	8	43	20	71
Basidiomycetes	1	7	4	12
Other fungi	1	2	5	8
Protists	6	18	27	51
Apicomplexans	1	9	7	17
Kinetoplasts	1	2	5	8
Other protists	4	7	14	25
total:	**653**	**605**	**665**	**1923**

Revised: Feb 10, 2008

AND UNCOVERED 1,000s of NOVEL Genes That May Have Useful Applications in Medicine, Agriculture, + Industrial Biotechnology

Comparative Sequencing of Mammalian Genomes Should Provide Insights Into Human Origins

a. Yes
b. No

iClicker

COMPARATIVE DNA SEQUENCING SHOULD EVENTUALLY PROVIDE CLUES TO HOW HUMANS EVOLVED

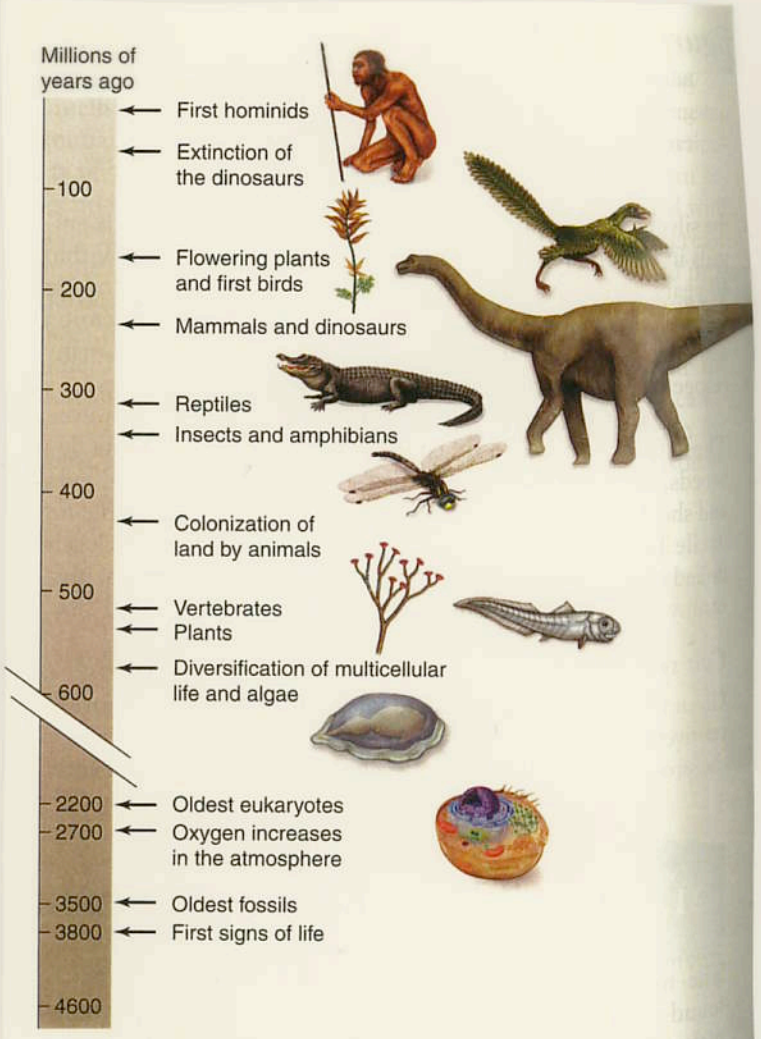


figure 21.10

HISTORY OF EVOLUTIONARY CHANGE AS REVEALED BY THE FOSSIL RECORD.

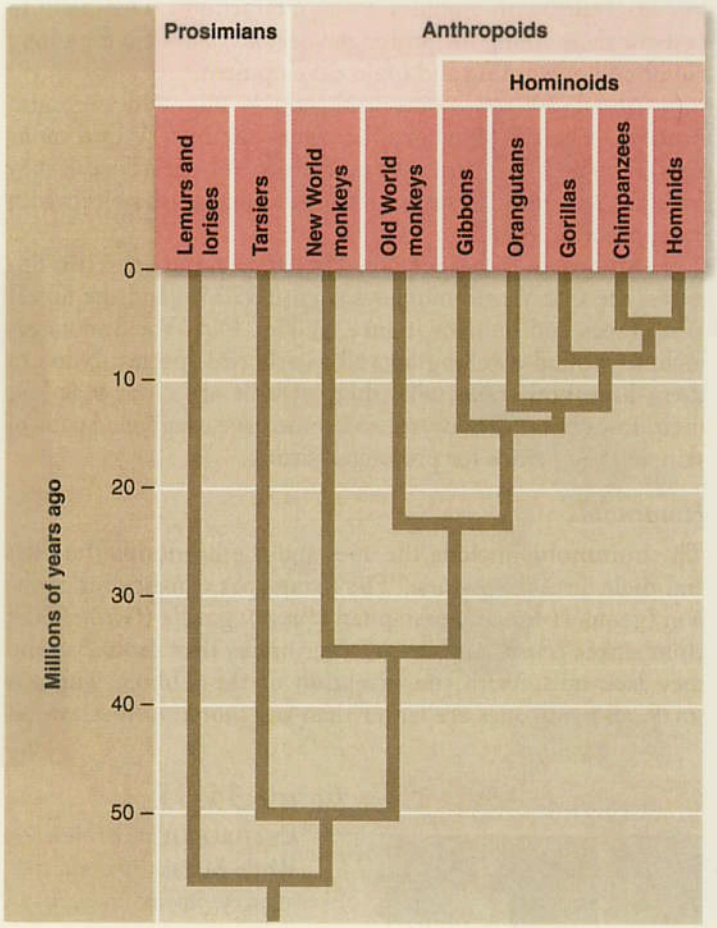


figure 35.36

A PRIMATE EVOLUTIONARY TREE. Prosimians diverged early in primate evolution, whereas hominids diverged much more recently. Apes constitute a paraphyletic group because some apes are more closely related to nonape species (hominids) than they are to other apes.

PARTICULARLY BY COMPARING Whole Genomes

HUMANS & CHIMPS DIVERGED
~ 6 MILLION years AGO FROM
a COMMON ANCESTOR

(a) (b) (c) (d)

Figure 22.5 Humans diverged from an ancestor shared with chimpanzees about 6 million years ago. Representatives of primates alive today: **(a)** orangutan, **(b)** gorilla, **(c)** chimpanzee, and **(d)** human.

There is only 1.2% Difference
in HUMAN & Chimp DNA
SEQUENCES

Human and Chimp Genomes Differ Primarily in Their

a. Genes
b. Regulatory Networks
c. Replication Processes
d. Size

iClicker

Vol 437|1 September 2005|doi:10.1038/nature04072 nature

ARTICLES

Initial sequence of the chimpanzee genome and comparison with the human genome

The Chimpanzee Sequencing and Analysis Consortium*

Here we present a draft genome sequence of the common chimpanzee (*Pan troglodytes*). Through comparison with the human genome, we have generated a largely complete catalogue of the genetic differences that have accumulated since the human and chimpanzee species diverged from our common ancestor, constituting approximately thirty-five million single-nucleotide changes, five million insertion/deletion events, and various chromosomal rearrangements. We use this catalogue to explore the magnitude and regional variation of mutational forces shaping these two genomes, and the strength of positive and negative selection acting on their genes. In particular, we find that the patterns of evolution in human and chimpanzee protein-coding genes are highly correlated and dominated by the fixation of neutral and slightly deleterious alleles. We also use the chimpanzee genome as an outgroup to investigate human population genetics and identify signatures of selective sweeps in recent human evolution.

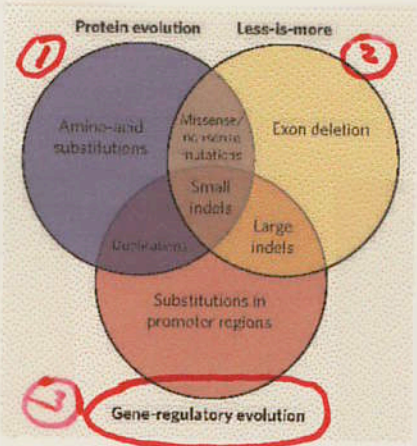


Figure 2 | Hypotheses to explain the genetic underpinnings of human-specific traits. Each of the three hypotheses — protein evolution, 'less-is-more', and gene-regulatory evolution — is depicted by a circle, with note of the mechanisms or processes that could underlie the evolutionary change. A missense mutation causes an amino-acid change; a nonsense mutation causes a sense codon to change into a stop codon, resulting in premature termination of DNA transcription. Indels are insertions/deletions of DNA segments; exons are coding sequences; promoter regions regulate gene activity in various ways.

What ALLOWED US TO BECOME HUMAN?

ANSWERS ARE "Hidden" in the Genome Differences Between Chimps & Humans

SWITCHES!

24

AN EXAMPLE OF A POTENTIALLY SIGNIFICANT Difference BETWEEN HUMAN & Chimp GENOMES

Vol 443|14 September 2006|doi:10.1038/nature05113 nature

ARTICLES

An RNA gene expressed during cortical development evolved rapidly in humans

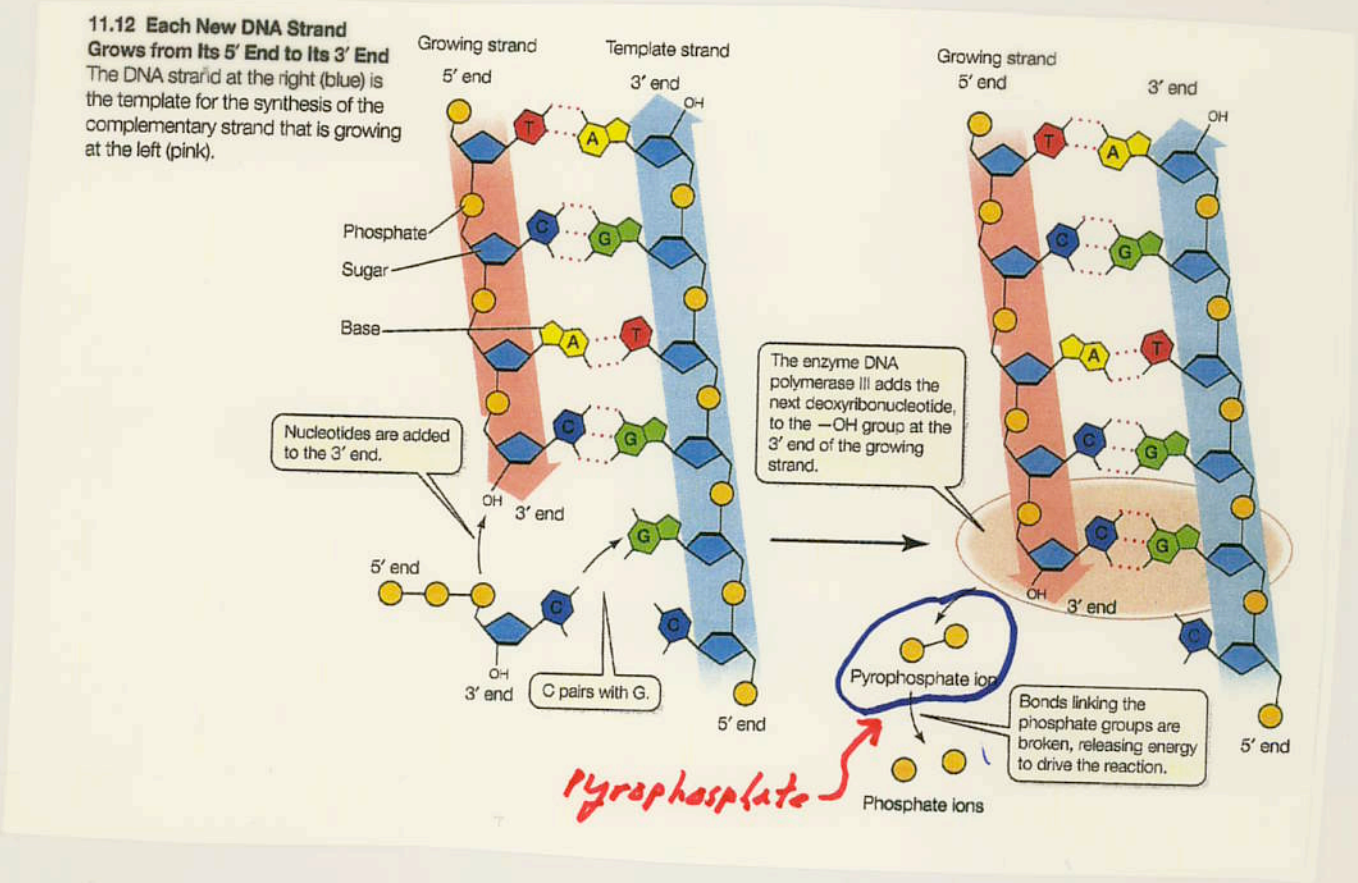
Katherine S. Pollard[1]*†, Sofie R. Salama[1,2]*, Nelle Lambert[4,5], Marie-Alexandra Lambot[4], Sandra Coppens[4], Jakob S. Pedersen[1], Sol Katzman[1], Bryan King[1,2], Courtney Onodera[1], Adam Siepel[1]†, Andrew D. Kern[1], Colette Dehay[6,7], Haller Igel[3], Manuel Ares Jr[3], Pierre Vanderhaeghen[4] & David Haussler[1,2]

The developmental and evolutionary mechanisms behind the emergence of human-specific brain features remain largely unknown. However, the recent ability to compare our genome to that of our closest relative, the chimpanzee, provides new avenues to link genetic and phenotypic changes in the evolution of the human brain. We devised a ranking of regions in the human genome that show significant evolutionary acceleration. Here we report that the most dramatic of these 'human accelerated regions', HAR1, is part of a novel RNA gene (*HAR1F*) that is expressed specifically in Cajal–Retzius neurons in the developing human neocortex from 7 to 19 gestational weeks, a crucial period for cortical neuron specification and migration. *HAR1F* is co-expressed with reelin, a product of Cajal–Retzius neurons that is of fundamental importance in specifying the six-layer structure of the human cortex. HAR1 and the other human accelerated regions provide new candidates in the search for uniquely human biology.

COMPARATIVE SEQUENCE STUDIES OF MAMMALIAN GENOMES SHOULD PROVIDE NOVEL INSIGHTS in What MAKES a HUMAN!!

The Ability to SEQUENCE Whole Genomes Has BEEN REVOLUTIONIZED in the PAST TWO years

MASSIVELY Parallel Pyrosequencing

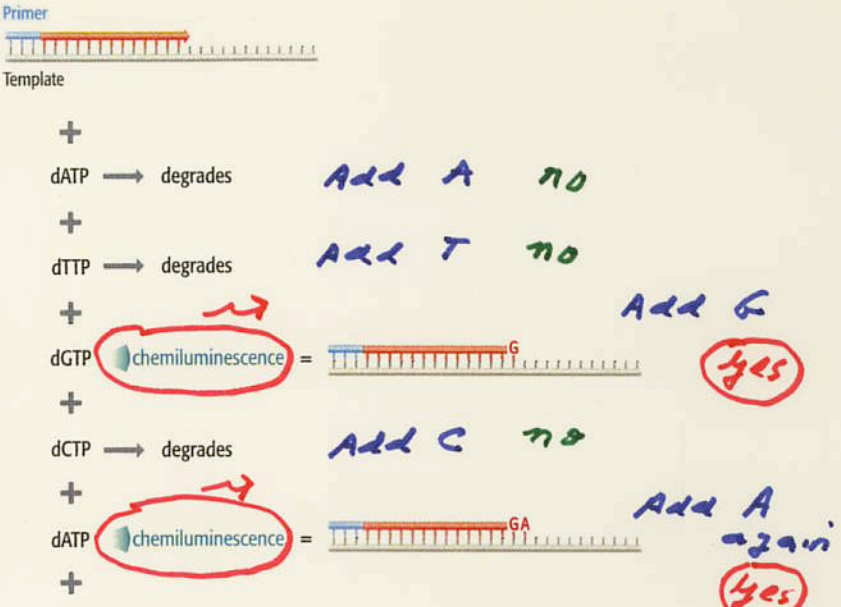


NOTE the release of PPi from dNTPs as dNMP gets incorporated into DNA Chain!

USE Pyrophosphate as "Marker" to detect nucleotide incorporation into DNA Chain!

Pyrosequencing has revolutionized ability to sequence whole genomes

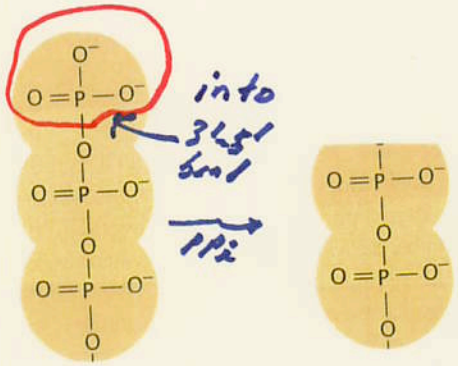
Figure 4.9 Pyrosequencing. The strand synthesis reaction is carried out in the absence of dideoxynucleotides. Each deoxynucleotide is added individually, along with a nucleotidase enzyme that degrades the deoxynucleotide if it is not incorporated into the strand being synthesized. Incorporation is detected by a flash of chemiluminescence induced by the pyrophosphate released from the deoxynucleotide. The order in which deoxynucleotides are added to the growing strand can therefore be followed.



Sequential synthesis

add T, G, C, etc. one by one!

When dNTP turns Phosphodiester Bond



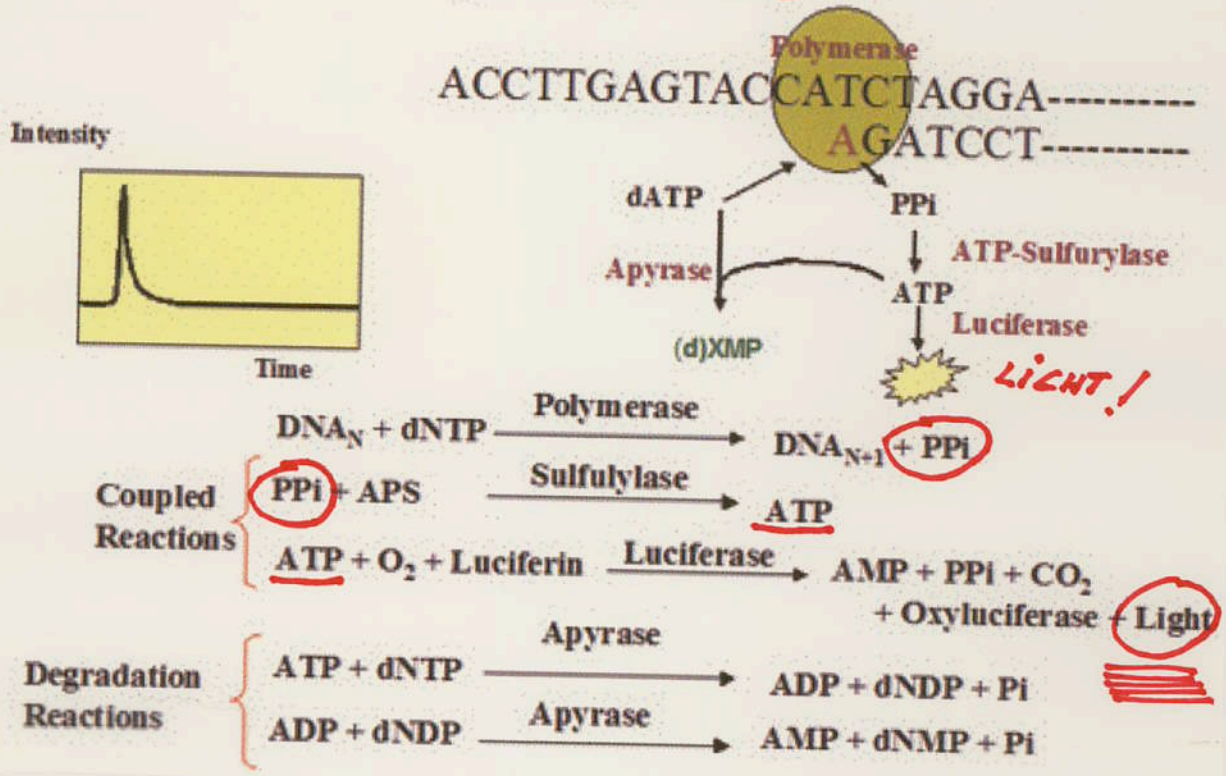
Degrade enzymatically → Light Emitted

Pyrosequencing Advantages

1. FAST
2. CHEAP
3. AUTOMATED
4. MASSIVELY PARALLEL – 100000s of DNA FRAGMENTS CAN BE SEQUENCED AT ONE TIME!

Pyrosequencing SEQUENTIALLY ADDS dNTPs to DNA Synthesis Reaction - Release of PPi Generates Light to detect SEQUENCE

Pyrosequencing



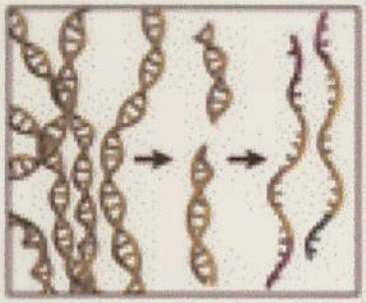
No Light if dNTP NOT incorporated

MASSIVELY Parallel Pyrosequencing OR 454 SEQUENCING

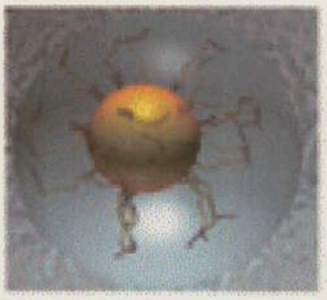
NO DNA CLONING! A FIRST!

(1) FRAGMENT DNA

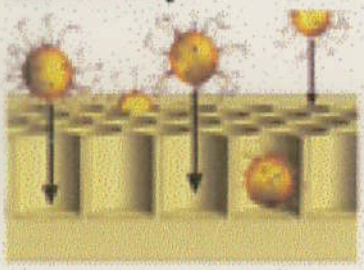
(2) AMPLIFY DNA by PCR



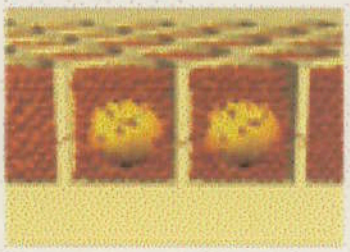
1) Prepare adapter-ligated ssDNA library



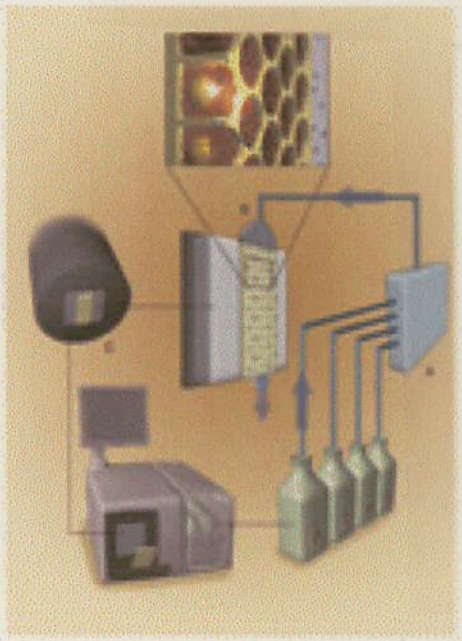
2) Clonal amplification on 28µ beads



3) Load beads in PicoTiter Plate™



4) Load enzymes beads in PicoTiter Plate™



5) Perform sequencing by synthesis on the 454 Instrument

(3) Put Each Colony in one Bead-Well

(4) CARRY OUT AMPLIFICATION and Synthesis Reactions in Each Well in Parallel

(5) Analyze DATA

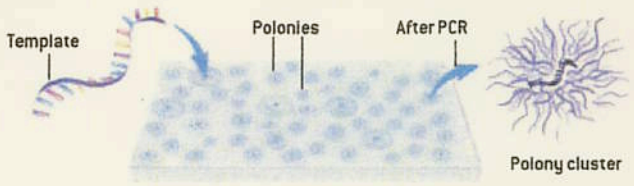
COMBINES!

1. Pyrosequencing
2. PCR Technology
3. NANO technology
4. Robotics
5. Computer Analysis

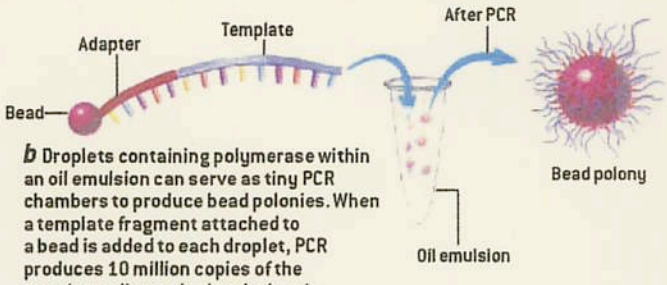
454 MASSIVELY PARALLEL Pyrosequencing
CAN SEQUENCE 1.2×10^8 bp in
TWO HOURS FOR ~ $5,000!

AMPLIFICATION

Because light signals are difficult to detect at the scale of a single DNA molecule, base-extension or ligation reactions are often performed on millions of copies of the same template strand simultaneously. Cell-free methods (*a* and *b*) for making these copies involve PCR on a miniaturized scale.



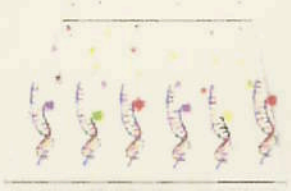
a Polonies—polymerase colonies—created directly on the surface of a slide or gel each contain a primer, which a template fragment can find and bind to. PCR within each polony produces a cluster containing millions of template copies.



b Droplets containing polymerase within an oil emulsion can serve as tiny PCR chambers to produce bead polonies. When a template fragment attached to a bead is added to each droplet, PCR produces 10 million copies of the template, all attached to the bead.

MULTIPLEXING

Sequencing thousands or millions of template fragments in parallel maximizes speed. A single-molecule base-extension system using fluorescent-signal detection, for example, places hundreds of millions of different template fragments on a single array (*below left*). Another method immobilizes millions of bead polonies on a gel surface for simultaneous sequencing by ligation with fluorescence signals, shown in the image at right below, which represents 0.01 percent of the total slide area.



Single-molecule array

Bead polonies

This HAS REVOLUTIONIZED BIOLOGY!

BECAUSE OF PCR - NEED ONLY A TINY AMOUNT OF STARTING DNA!!!

The Age of Personal GENOMICS HAS BEGUN!

Home | About | FAQ | Contact

Knome™

Know thyself.

Knome is the first personal genomics company to offer whole-genome sequencing and comprehensive analysis services for individuals.

Based in Cambridge, Massachusetts, we work alongside leading geneticists, clinicians and bioinformaticians from Harvard and MIT to enable our clients to obtain, understand, and share their genomic information in a manner that is both anonymous and secure.

We are currently offering 20 individuals the opportunity to participate in our initial launch phase. By being amongst the first individuals in history to have their whole genome sequenced, these participants will help pioneer the emerging field of personal genomics.

Recent News

January 22, 2008: Knome Commences Whole-Genome Sequencing Process For First Clients

January 10, 2008: Knome and the Beijing Genomics Institute Enter into Exclusive Strategic Alliance

November 29, 2007: Knome Launches First Commercial Whole-Genome Sequencing and Analysis Service for Individuals

$350,000

Problems?

THE 1,000 GENOMES PROJECT WILL PROVIDE NOVEL INSIGHT IN HUMAN GENOMES

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.

"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."

ONLY POSSIBLE USING 454 SEQUENCING

COMPLEMENT HAPMAP PROJECT

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.

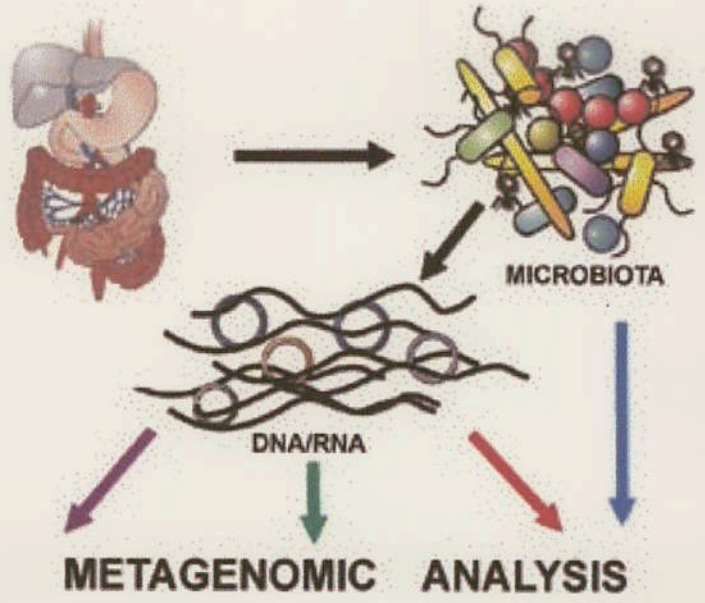
If I Could Sequence My Genome For $10 I Would Have It Done:

a. Yes
b. No

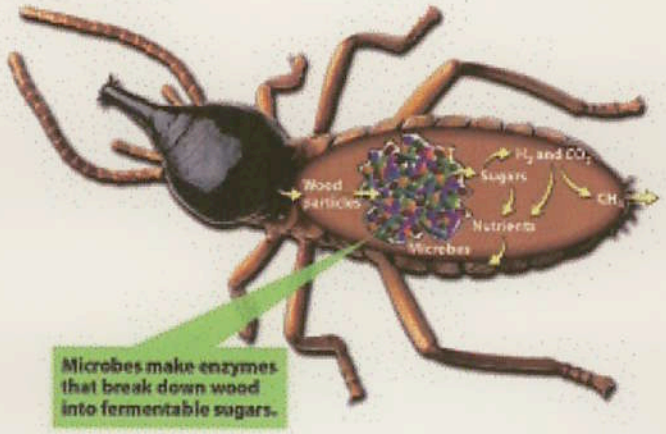
iClicker

METAGENOME ANALYSIS – SEQUENCING MANY GENOMES AT ONCE

MADE POSSIBLE USING 454 DNA SEQUENCING



COMPUTER ANALYSIS + LARGE DNA DATA BASES SORT GENOME SEQUENCE OUT



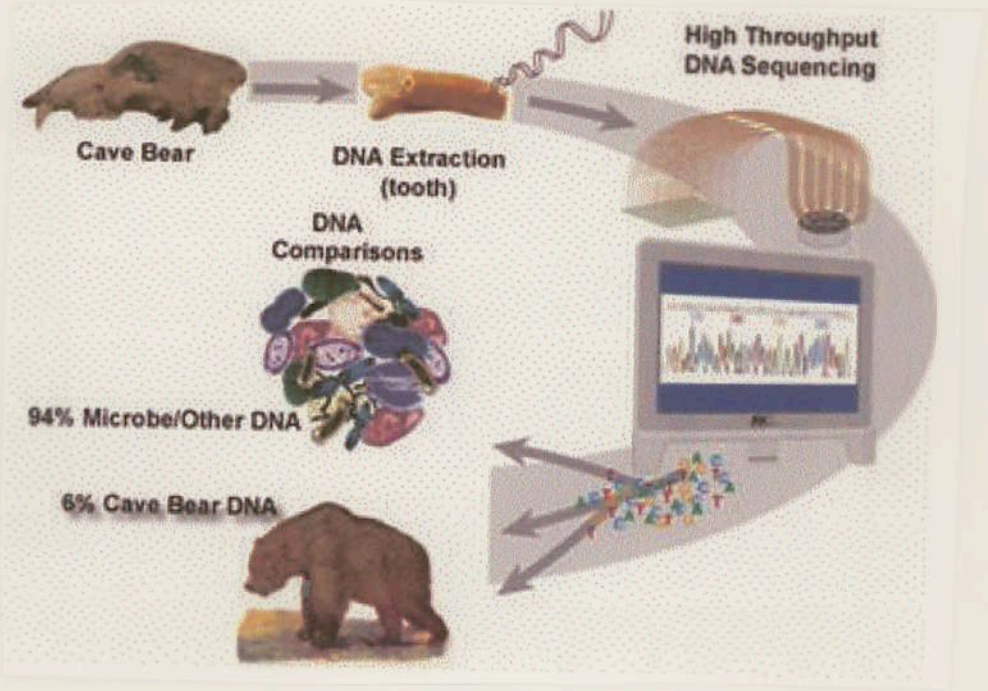
METAGENOME PROJECTS ON UNCULTIVATED MICROORGANISMS

Table 1 | **Assembled genomes of uncultivated microbes**

Species	Genome size	Host or habitat	Separation technique	Refs
Treponema pallidum	1.1 Mb	Human, rabbit	Dissection, differential lysis	20
Rickettsia prowazekii	1.1 Mb	Human, chicken	Differential centrifugation	21
Mycobacterium leprae	3.3 Mb	Human, armadillo	Gradient centrifugation	22
Tropheryma whipplei	0.9 Mb	Human	Differential centrifugation	23
Buchnera aphidicola str. APS	0.6 Mb	Aphid (*Acyrthosiphon pisum*)	Dissection, differential lysis, filtration	24
Buchnera aphidicola str. Sg	0.6 Mb	Aphid (*Schizaphis graminum*)	Gradient centrifugation	25
Wigglesworthia glossinidia brevipalpis	0.7 Mb	Tsetse fly (*Glossina brevipalpis*)	Dissection, differential lysis	26
Blochmannia floridanus	0.7 Mb	Carpenter ants	Differential lysis	27
Buchnera aphidicola str. Bp	0.6 Mb	Aphid (*Baizongia pistaciae*)	Differential lysis, filtration	28
Wolbachia pipientis wMel	1.27 Mb	Fly (*Drosophila melanogaster*)	Differential lysis, pulsed-field gel electrophoresis	29
Wolbachia pipientis wAna	1.4 Mb	Fly (*Drosophila ananassae*)	None	30
Wolbachia pipientis wBm	1.1 Mb	Parasitic nematode worm (*Brugia malayi*)	BAC library screening	31
Phytoplasma asteris, line OY-M	0.9 MB	Plants and leafhoppers	Differential lysis, pulsed-field gel electrophoresis	32
Nanoarchaeum equitans	0.5 Mb	*Ignicoccus* sp. co-culture	Differential centrifugation	59
Ferroplasma acidarmanus type II	1.8 Mb	Acid-mine biofilm	None	45
Leptospirillum sp. Group II	2.2 Mb	Acid-mine biofilm	None	45
Burkholderia sp.	~8.8 Mb	Sargasso Sea	Filtration	4
Shewanella sp.	~5 Mb	Sargasso Sea	Filtration	4

CAN DISCOVER NEW MICROBES/ORGANISMS IN ANY ENVIROMENT! SOIL, OCEAN, AIR, etc.!

USING MASSIVELY PARALLEL DNA SEQUENCING to Study ANCIENT GENOMES!



SEQUENCE & ALLOW COMPUTER ANALYSIS & DATABASES SORT OUT DNA SEQUENCES

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

ARTICLES

Analysis of one million base pairs of Neanderthal DNA

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]

Neanderthals are the extinct hominid group most closely related to contemporary humans, so their genome offers a unique opportunity to identify genetic changes specific to anatomically fully modern humans. We have identified a 38,000-year-old Neanderthal fossil that is exceptionally free of contamination from modern human DNA. Direct high-throughput sequencing of a DNA extract from this fossil has thus far yielded over one million base pairs of hominoid nuclear DNA sequences. Comparison with the human and chimpanzee genomes reveals that modern human and Neanderthal DNA sequences diverged on average about 500,000 years ago. Existing technology and fossil resources are now sufficient to initiate a Neanderthal genome-sequencing effort.

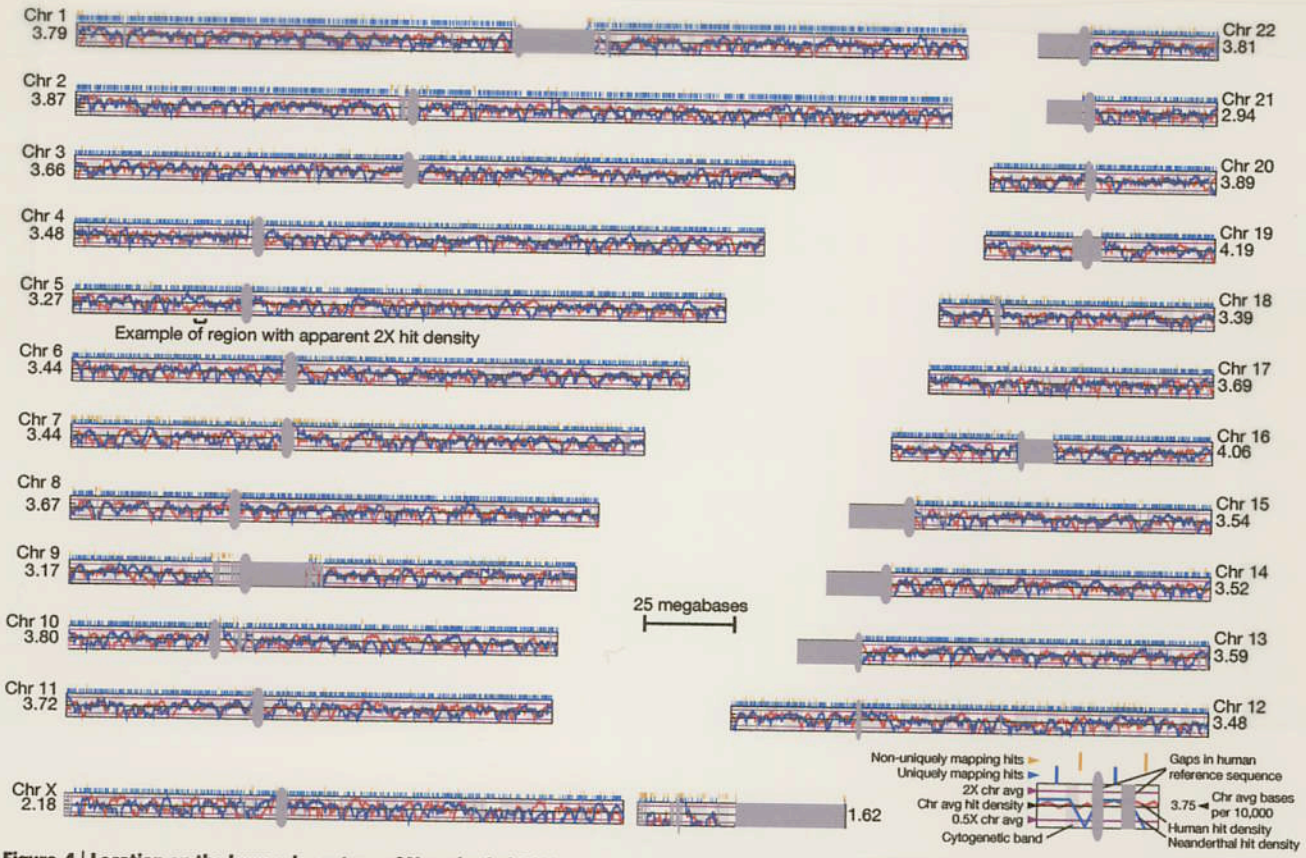
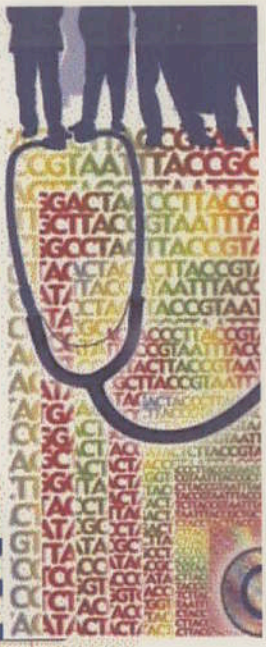
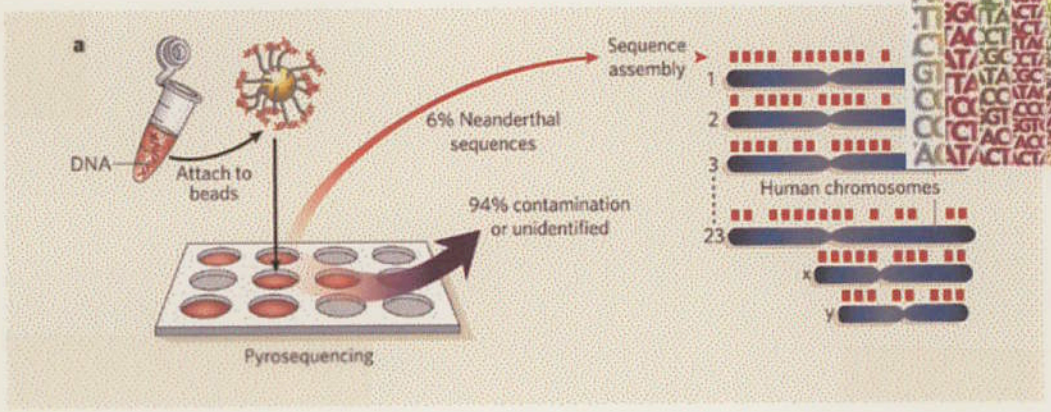


Figure 4 | Location on the human karyotype of Neanderthal DNA sequences. All sequences longer than 30 nucleotides whose best alignments were to the human genome are shown. The blue lines above each chromosome mark the position of all alignments that are unique in terms of bit-score within the human genome. Orange lines are alignments that have more than one alignment of equal bit-score. To the left of each chromosome, the average number of Neanderthal bases per 10,000 is given. Lines (Neanderthal, blue; human, red) within each chromosome show the hit density, on a log-base 2 scale, within sliding windows of 3 megabases along each chromosome. The centre black lines indicate the average hit-density for the chromosomes. The purple lines above and below indicate hit densities of 2X and 1/2X the chromosome average, respectively. On chromosome 5, an example of a region of increased sequence density is highlighted. Sequence gaps in the human reference sequence are indicated by dark grey regions. Chromosomal banding pattern is indicated by light grey regions.

SEQUENCING THE NEANDERTHAL GENOME!

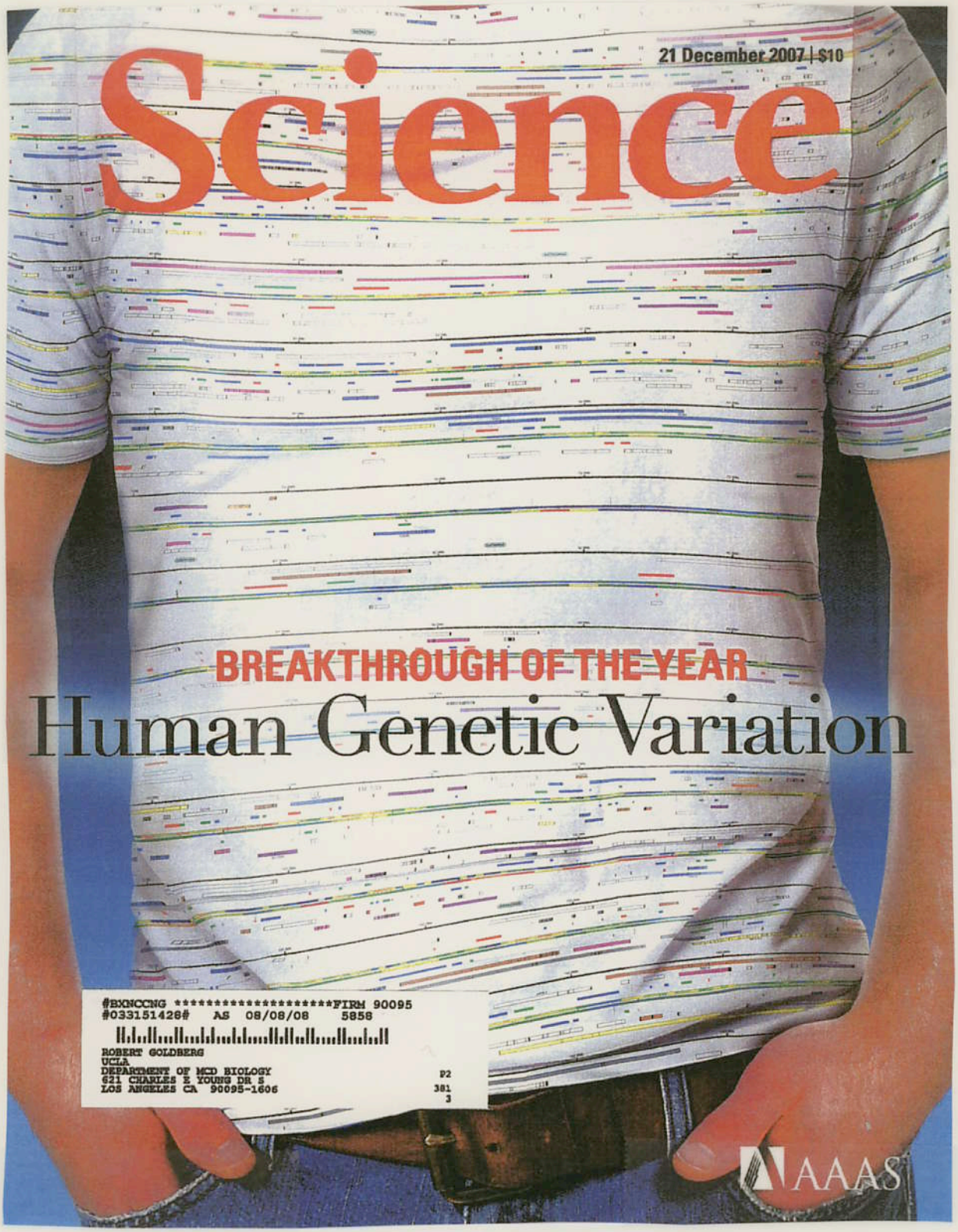




The Entire Neanderthal Genome should be complete in ~ 2 years!

How Did We Evolve & What Makes Us HUMAN?

HUMAN GENOME PROJECTS HAVE GIVEN NEW INSIGHTS INTO DNA VARIATION IN HUMAN POPULATIONS



TO UNDERSTAND WHO WE ARE, WHERE WE COME FROM, & THE NATURE OF GENETIC DISEASES

The Human HapMap Project Catalogs:

a. DNA Sequence Variants in Human Populations
b. Protein Variants in Human Populations

iClicker

International HapMap Project

The 0.1% That's Different Between Individuals!

Vol 449|18 October 2007|doi:10.1038/nature06258 nature

ARTICLES

A second generation human haplotype map of over 3.1 million SNPs

The International HapMap Consortium*

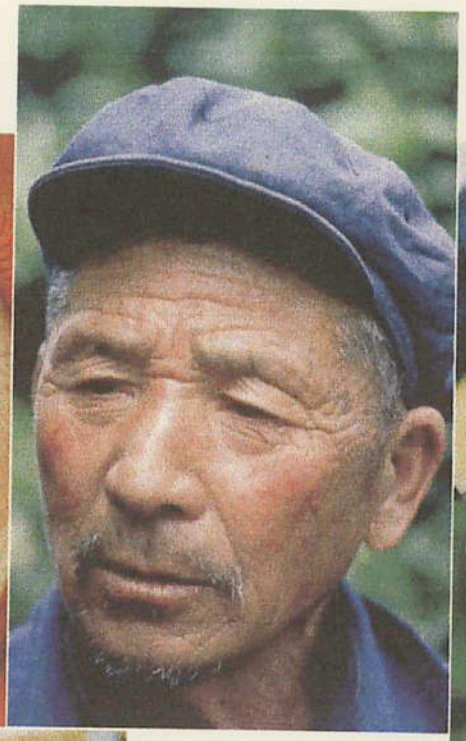
We describe the Phase II HapMap, which characterizes over 3.1 million human single nucleotide polymorphisms (SNPs) genotyped in 270 individuals from four geographically diverse populations and includes 25–35% of common SNP variation in the populations surveyed. The map is estimated to capture untyped 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.

(1) Discover Genetic VARIANTS Associated with Disease & Individual Responses to therapeatic Agents (pharmacogenetics)

(2) New Medical treatments

(3) STRUCTURE of HUMAN Populations or Ancestry!

International HapMap Groups



FAR LEFT, L. GEORGIA/CORBIS; LEFT, CORBIS

TL RIGHT, J. SLATER/CORBIS; FAR RIGHT, O. FRANKEN/CORBIS

World view: the HapMap initiative will gather genetic data from African, Asian and ancestrally European populations.

1. Northern European
2. Western European
3. Yoruba/Nigerian/African
4. Japanese + Han Chinese/Asian

"GRSup" Genetic Diversity to Disease & other Aspects of Biology

HAPLOTYPES OR SNPs on a Chromosome Inherited as a Unit

There are Millions of SNPs that Differ Among Individuals ....

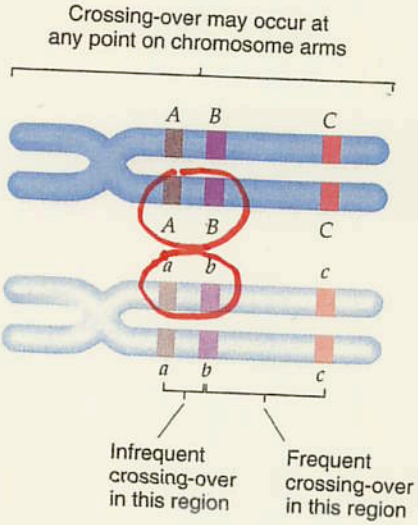
.... But a small few reflect our Ancestry & "travel" in groups on chromosomes — are linked & may show specific gene linkages!

Haplotypes!

CLOSELY-LINKED SNPS ARE INHERITED AS A UNIT

Figure 5.3

The relationship between recombination and map distance. The farther apart two genes are, the greater the number of possible sites for recombination. Thus, the probability of recombination occurring between genes *A* and *B* is much less than that between genes *B* and *C*. The percentage of recombinants can provide information about the relative genetic distance between two linked genes.

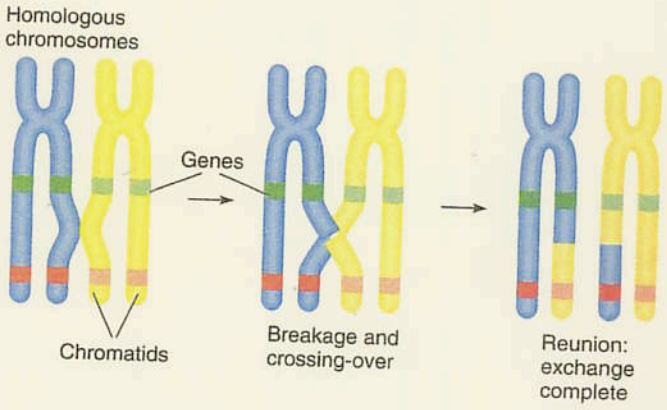


Haplotype
AB
vs.
ab

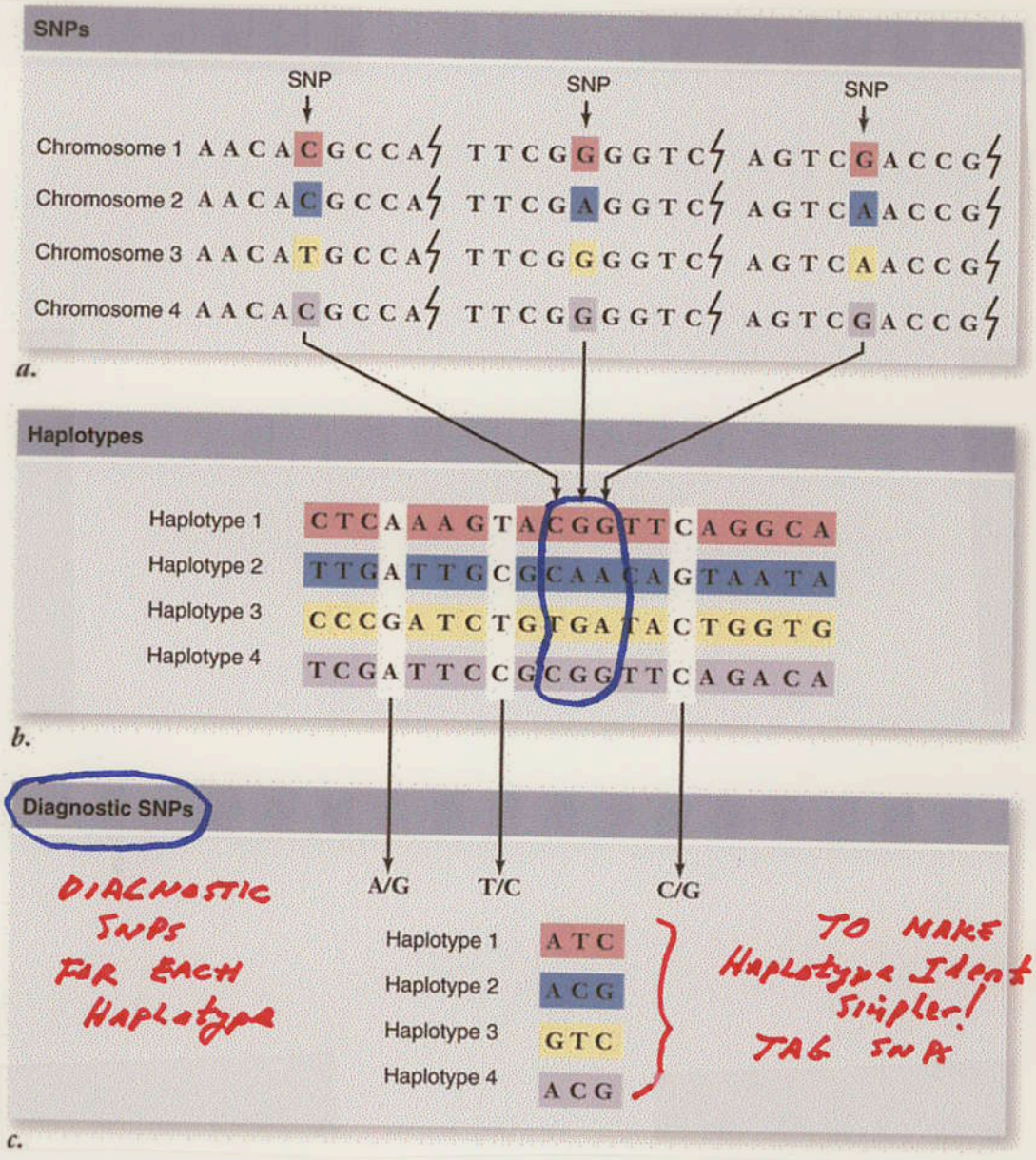
NO CROSSING OVER < 5kb
∴ HAPLOTYPE OF COMPLEX POLYMORPHIC LOCUS

Figure 5.2

Mechanism of crossing-over. A highly simplified diagram of a crossover between two nonsister chromatids during meiotic prophase, giving rise to recombinant (nonparental) combinations of linked genes.



HAPLOTYPES ARE MARKERS THAT CAN BE ASSOCIATED WITH DISEASES +/OR GROUPS OF Individuals



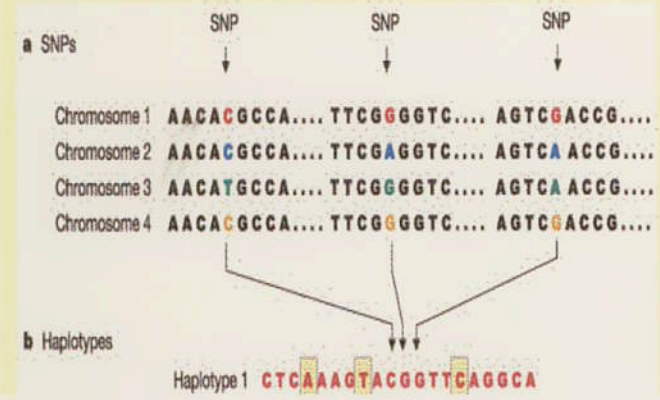
SNPs That Vary ACROSS 5,000 bp

DIAGNOSTIC SNPS FOR EACH Haplotype

TO MAKE Haplotype Identification Simpler! TAG SNPs

The 0.1% that's Different Between Individuals! The HapMap CATALOGS COMMON GENETIC VARIANTS in HUMAN BEINGS

International HapMap Project



What Is the HapMap?

The HapMap is a catalog of common genetic variants that occur in human beings. It describes what these variants are, where they occur in our DNA, and how they are distributed among people within populations and among populations in different parts of the world. The International HapMap Project is not using the information in the HapMap to establish connections between particular genetic variants and diseases. Rather, the Project is designed to provide information that other researchers can use to link genetic variants to the risk for specific illnesses, which will lead to new methods of preventing, diagnosing, and treating disease.

The DNA in our cells

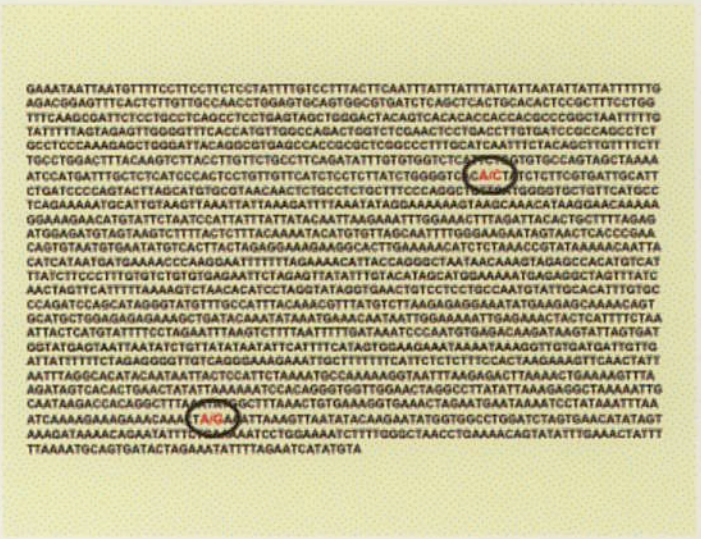


Figure 1: When DNA sequences on a part of chromosome 7 from two random individuals are compared, two single nucleotide polymorphisms (SNPs) occur in about 2,200 nucleotides.

contains long chains of four chemical building blocks -- adenine, thymine, cytosine, and guanine, abbreviated A, T, C, and G. More than 6 billion of these chemical bases, strung together in 23 pairs of chromosomes, exist in a human cell. (See **http://www.dnaftb.org/dnaftb/** for basic information about genetics.) These genetic sequences contain information

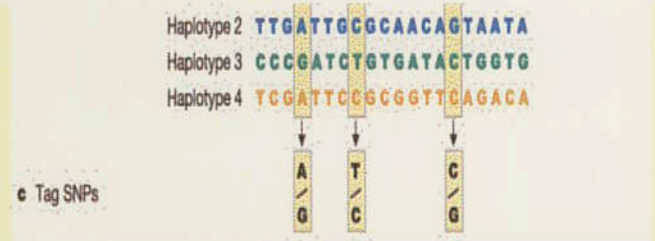


Figure 2: The construction of the HapMap occurs in three steps. (a) Single nucleotide polymorphisms(SNPs) are identified in DNA samples from multiple indivduals. (b)Adjacent SNPs that are inherited together are compiled into "haplotypes." (c)"Tag" SNPs within haplotypes are identified that uniquely identify those haplotypes. By genotyping the three tag SNPs shown in this figure, researchers can identify which of the four haplotypes shown here are present in each individual.

that influences our physical traits, our likelihood of suffering from disease, and the responses of our bodies to substances that we encounter in the environment.

The genetic sequences of different people are remarkably similar. When the chromosomes of two humans are compared, their DNA sequences can be identical for hundreds of bases. But at about one in every 1,200 bases, on average, the sequences will differ (Figure 1). One person might have an A at that location, while another person has a G, or a person might have extra bases at a given location or a missing segment of DNA. Each distinct "spelling" of a chromosomal region is called an allele, and a collection of alleles in a person's chromosomes is known as a genotype.

Differences in individual bases are by far the most common type of genetic variation. These genetic differences are known as single nucleotide polymorphisms, or SNPs (pronounced "snips"). By identifying most of the approximately 10 million SNPs estimated to occur commonly in the human genome, the International HapMap Project is identifying the basis for a large fraction of the genetic diversity in the human species.

SNPs Are Markers

For geneticists, SNPs act as markers to locate genes in DNA sequences. Say that a spelling change in a gene increases the risk of suffering from high blood pressure, but researchers do not know where in our chromosomes that gene is located. They could compare the SNPs in people who have high blood pressure with the SNPs of people who do not. If a particular SNP is more common among people with hypertension, that SNP could be used as a pointer to locate and identify the gene involved in the disease.

However, testing all of the 10 million common SNPs in a person's chromosomes would be extremely expensive. The development of the HapMap will enable geneticists to take advantage of how SNPs and other genetic variants are organized on chromosomes. Genetic variants that are near each other tend to be inherited together. For example, all of the people who have an A rather than a G at a particular location in a chromosome can have identical genetic variants at other SNPs in the chromosomal region surrounding the A. These regions of linked variants are known as haplotypes (Figure 2).

In many parts of our chromosomes, just a handful of haplotypes are found in humans. [**See The Origins of Haplotypes**.] In a given population, 55 percent of people may have one version of a haplotype, 30 percent may have another, 8 percent may have a third, and the rest may have a variety of less common haplotypes. The International HapMap Project is identifying these common haplotypes in four populations from different parts of the world. It also is identifying "tag" SNPs that uniquely identify these haplotypes. By testing an individual's tag SNPs (a process known as genotyping), researchers will be able to identify the collection of haplotypes in a person's DNA. The number of tag SNPs that

contain most of the information about the patterns of genetic variation is estimated to be about 300,000 to 600,000, which is far fewer than the 10 million common SNPs.

Once the information on tag SNPs from the HapMap is available, researchers will be able to use them to locate genes involved in medically important traits. Consider the researcher trying to find genetic variants associated with high blood pressure. Instead of determining the identity of all SNPs in a person's DNA, the researcher would genotype a much smaller number of tag SNPs to determine the collection of haplotypes present in each subject. The researcher could focus on specific candidate genes that may be associated with a disease, or even look across the entire genome to find chromosomal regions that may be associated with a disease. If people with high blood pressure tend to share a particular haplotype, variants contributing to the disease might be somewhere within or near that haplotype.

Please send questions and comments on website to
help@hapmap.org

About the HapMap
What is the HapMap?
Origins of Haplotypes
Health Benefits
Populations Sampled
Ethical Issues
Consent Forms
Community Advisory Groups(CAG)
Data Release Policy
Guidelines For Data Use
Guidelines For Referring to HapMap Populations

Project Information
About the Project
HapMap Publications
HapMap Tutorial
HapMap Mailing List
HapMap Project Participants
HapMap Mirror Site in Japan

Useful Links
HapMap Project Press Release
NHGRI HapMap Page
NCBI Variation Database (dbSNP)
Japanese SNP Database (JSNP)

Which Populations Are Being Sampled

The International HapMap Project is analyzing DNA from populations with African, Asian, and European ancestry. Together, these DNA samples should enable HapMap researchers to identify most of the common haplotypes that exist in populations worldwide. [**See What Is the HapMap?**]

Because of the history of the human species, most of the common haplotypes in human chromosomes occur in all human populations. [**See The Origin of Haplotypes.**] However, any given haplotype may be more common in one population and less common in another, and newer haplotypes may be found in just a single population. Efficiently choosing the tag SNPs needed to identify haplotypes therefore requires looking at haplotype frequencies in multiple populations. Also, genetic data from more than one population will enhance the ability of researchers to study the genetic contributions to diseases that are more or less prevalent in different groups.

The DNA samples for the HapMap have come from a total of 270 people. The Yoruba people of Ibadan, Nigeria, provided 30 sets of samples from two parents and an adult child (each such set is called a trio). In Japan, 45 unrelated individuals from the Tokyo area provided samples. In China, 45 unrelated individuals from Beijing provided samples. Thirty U.S. trios provided samples, which were collected in 1980 from U.S. residents with northern and western European ancestry by the Centre d'Etude du Polymorphisme Humain (CEPH).

The blood samples are being converted into cell lines, which are used to make DNA, by the non-profit **Coriell Institute for Medical Research**. Coriell provides DNA and cell lines from the samples for research projects that have been approved by the appropriate ethics committees. The samples and cell lines are not linked to any individual in the populations studied. However, the samples and cell lines are identified as coming from one of the four populations participating in the study, which raises ethical issues associated with conducting genetic research in named populations. [**See How Are Ethical Issues Being Addressed?** and **Guidelines for Referring to the HapMap Populations in Publications and Presentations**.]

To assess how much additional information would be gained by genotyping other populations, haplotypes in a set of chromosomal regions are being analyzed in samples from several additional populations.

47

How Will the HapMap Benefit Human Health?

The International HapMap Project will benefit human health by providing an extensive resource that researchers can use to discover the genetic variants involved in disease and individual responses to therapeutic agents. Once such variants have been discovered, researchers can learn much more about the origins of illnesses and about ways to prevent, diagnose, and treat those illnesses.

The goal of the Project is not to identify these disease-related genes directly. Rather, by identifying haplotypes, the HapMap provides a tool that can be used in what are called association studies. For these studies, researchers will compare the haplotypes in individuals with a disease to the haplotypes of a comparable group of individuals without a disease (the controls). If a particular haplotype occurs more frequently in affected individuals compared with controls, a gene influencing the disease may be located within or near that haplotype.

Common diseases such as cancer, stroke, heart disease, diabetes, depression, and asthma usually result from the combined effects of a number of genetic variants and environmental factors. According to an idea known as the common disease-common variant hypothesis, the risk of contracting common diseases is influenced by genetic variants that are relatively common in populations. Not enough data are yet available to evaluate the generality of this hypothesis, but more and more widely distributed genetic variants associated with common diseases are being discovered, including variants that contribute to autoimmune diseases, schizophrenia, diabetes, asthma, stroke, and heart attacks. One of the many benefits of the International HapMap Project will be the use of the HapMap to learn more about the links between these common disorders and our genes.

Knowledge derived from use of the HapMap also will result in advances that are difficult to predict today. Medical treatments could be customized, based on a patient's genetic make-up, to maximize effectiveness and minimize side effects. Genetic variants contributing to longevity or resistance to disease could be identified, leading to new therapies with widespread benefits. As with any new body of knowledge, the HapMap is likely to lead to both new challenges and to unexpected and unprecedented opportunities.

variation present in the ancestral human population. Also, as humans migrated out of Africa, they carried with them part but not all of the genetic variation that existed in the ancestral population. As a result, the haplotypes seen outside Africa tend to be subsets of the haplotypes inside Africa. In addition, haplotypes in non-African populations tend to be longer than in African populations, because populations in Africa have been larger through much of our history and recombination has had more time there to break up haplotypes.

As modern humans spread throughout the world, the frequency of haplotypes came to vary from region to region through random chance, natural selection, and other genetic mechanisms. As a result, a given haplotype can occur at different frequencies in different populations, especially when those populations are widely separated and unlikely to exchange much DNA through mating. Also, new changes in DNA sequences, known as mutations, have created new haplotypes, and most of the recently arising haplotypes have not had enough time to spread widely beyond the population and geographic region in which they originated.

Home | About the Project | Data | Publications | Tutorial

Please send questions and comments on website to help@hapmap.org

Association Studies Correlate Diseases with specific Haplotypes (SNP)

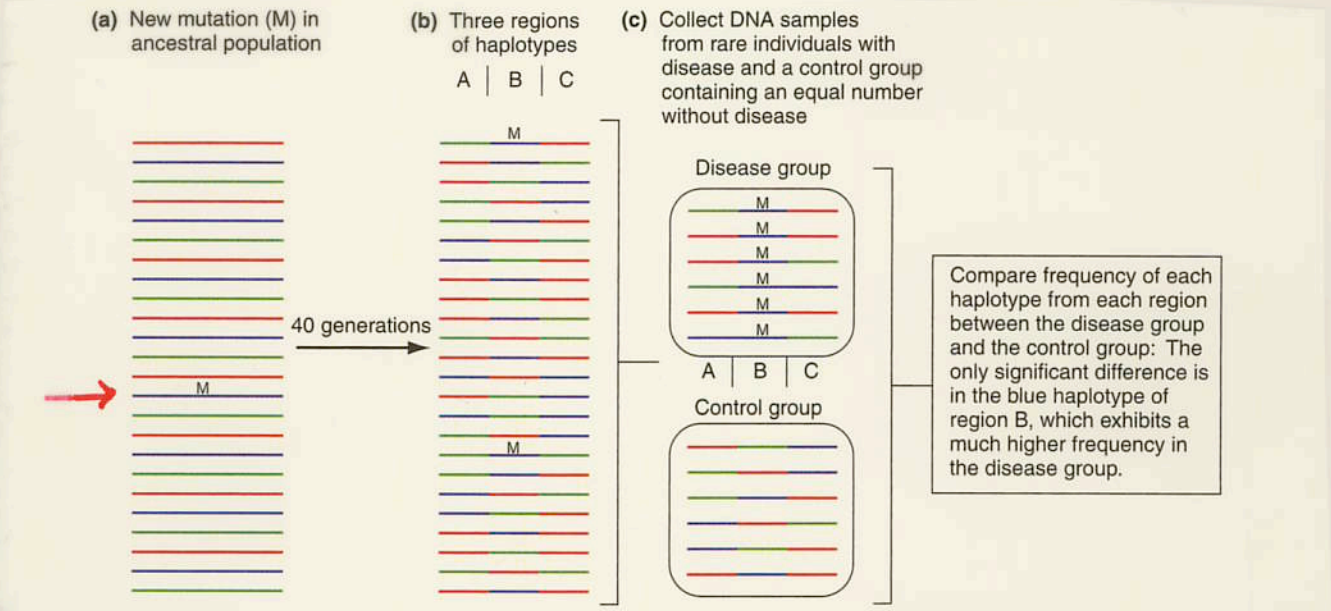
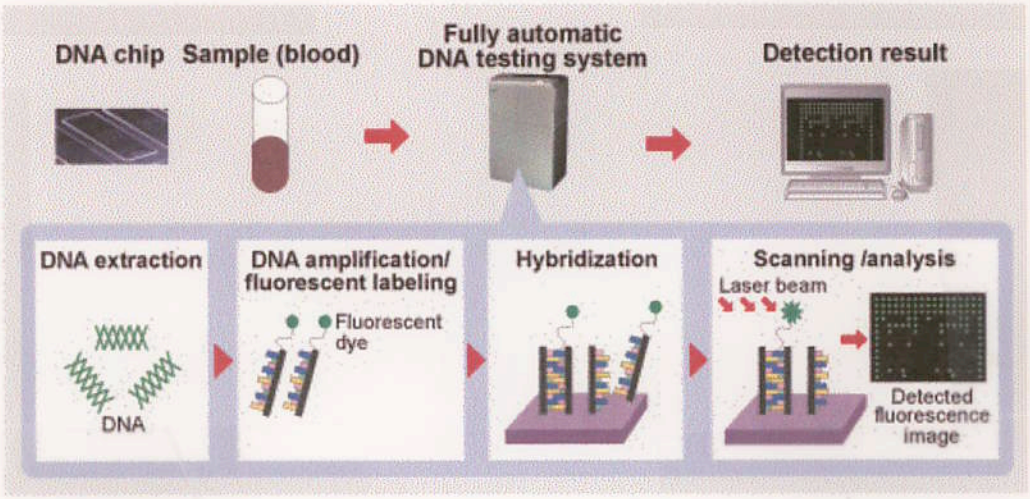


Figure 11.26 **Haplotype association allows high-resolution gene mapping.** **(a)** Representation of the same subchromosomal region in different individuals within an ancestral population that lived several thousand years ago. Lines of the same color have the same set of alleles because of common descent. The disease mutation (M) occurred on one ancestral chromosome. **(b)** After 40 generations, the original subchromosomal region has broken apart through recombination into three smaller regions (*A on the left, B in the middle, and C on the right*). Each smaller region still occurs in three different haplotypes. **(c)** You now collect DNA samples from a number of unrelated people with the disease and from an equal number of people without the disease and compare the frequency of each haplotype from regions A, B, and C in the disease and control groups. Genotyping of the samples will enable you to compare haplotype frequencies in the two groups. Haplotypes of regions not closely linked to the disease locus will occur with the same frequency in both groups. The only significant difference will be in the one haplotype that encompasses the disease locus (the *blue* haplotype of region B); it has a much higher frequency (here 100%) in the disease group than in the control group.

1. Provide Disease Genetic Markers
2. If Close Linkage - Lead to Gene Identity

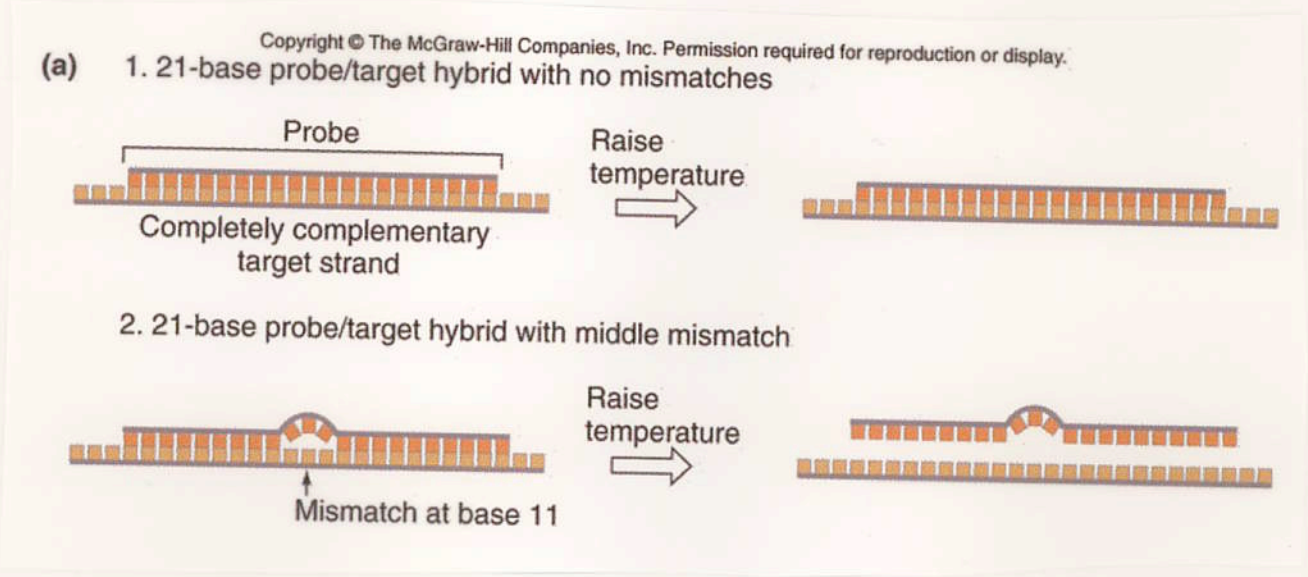
Population Studies

Gene Chips Are Used To Measure SNPs Across the Entire Genome

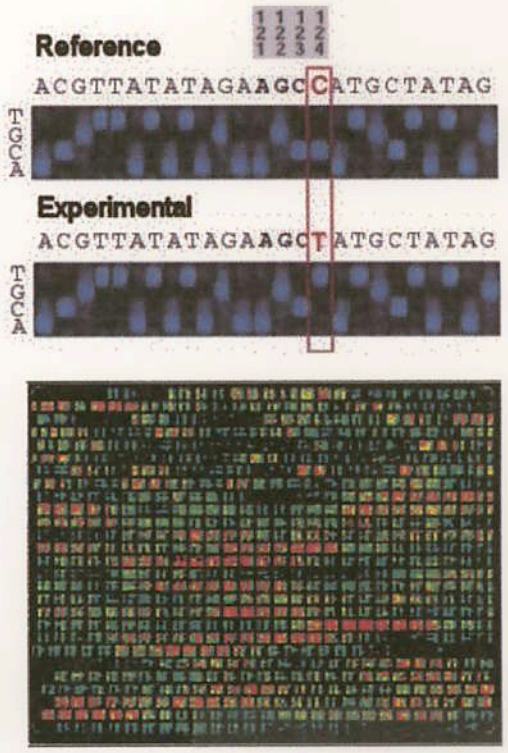


Contain 500,000 or More SNPs that can be detected By Hybridization to an Individual's DNA

DNA Chips Are BASED ON
ASO Technology
But Made to be High Throughput

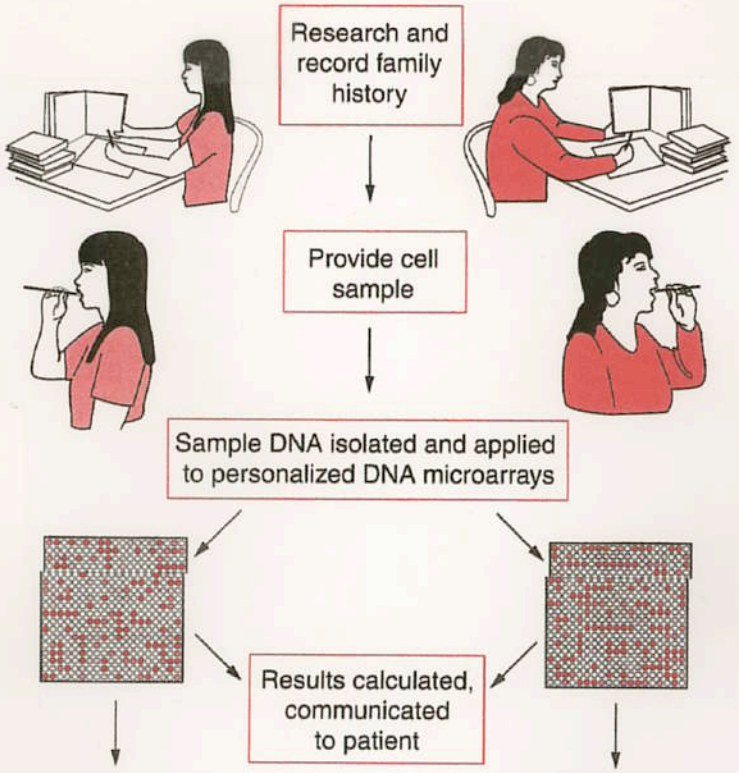


DNA Chips CAN detect SNP Genotypes (or Haplotypes) Across AN Individual's Genome



This CAN Then BE CORRELATED WITH DISEASE ASSOCIATIONS +/or Other TRaits

Whole Genome SNP Chips Have The Potential to Create a Genetic Profile



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

Figure 11.6 Using Gene Microarrays to Create a Genetic Profile

As SEQUENCING GETS CHEAPER & CHEAPER Chips will Probably Be Replaced By Whole Genome Sequences

1866: Gregor Mendel discovers the laws of inheritance.

200,000 years ago: *Homo sapiens* walks the Earth.

2003: The Human Genome Project maps a single person's genome.

2007: 23andMe introduces the first Personal Genome Service.

Unlock the secrets of your own DNA. Today.

175,000 years ago: The mother of all present-day humans is born in Africa.

1953: Watson and Crick uncover the double-helix structure of DNA.

Welcome to 23andMe, a web-based service that helps you read and understand your DNA. After providing a saliva sample using an at-home kit, you can use our interactive tools to shed new light on your distant ancestors, your close family and most of all, yourself.

news **What's new at 23andMe**

- Jan 22, 2008:23andMe now available in Canada and Europe.
- Jan 18, 2008:23andMe launches its blog, the spittoon.

Gene Journal

What do your genes say about you?

Gene Journal

Gene Journal

Getting Started With 23andMe

https://www.23andme.com/ourservice/process/

Order Form

First Name **Last Name**

You have not added any kits yet. Click the b

Order Summary
Kits in Order:

0 kits

Price per Kit:

$999.00 USD

Problems with this?

Private DNA Testing/Genotyping/Sequencing Services Should Regulated?

a. Yes
b. No

iClicker

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

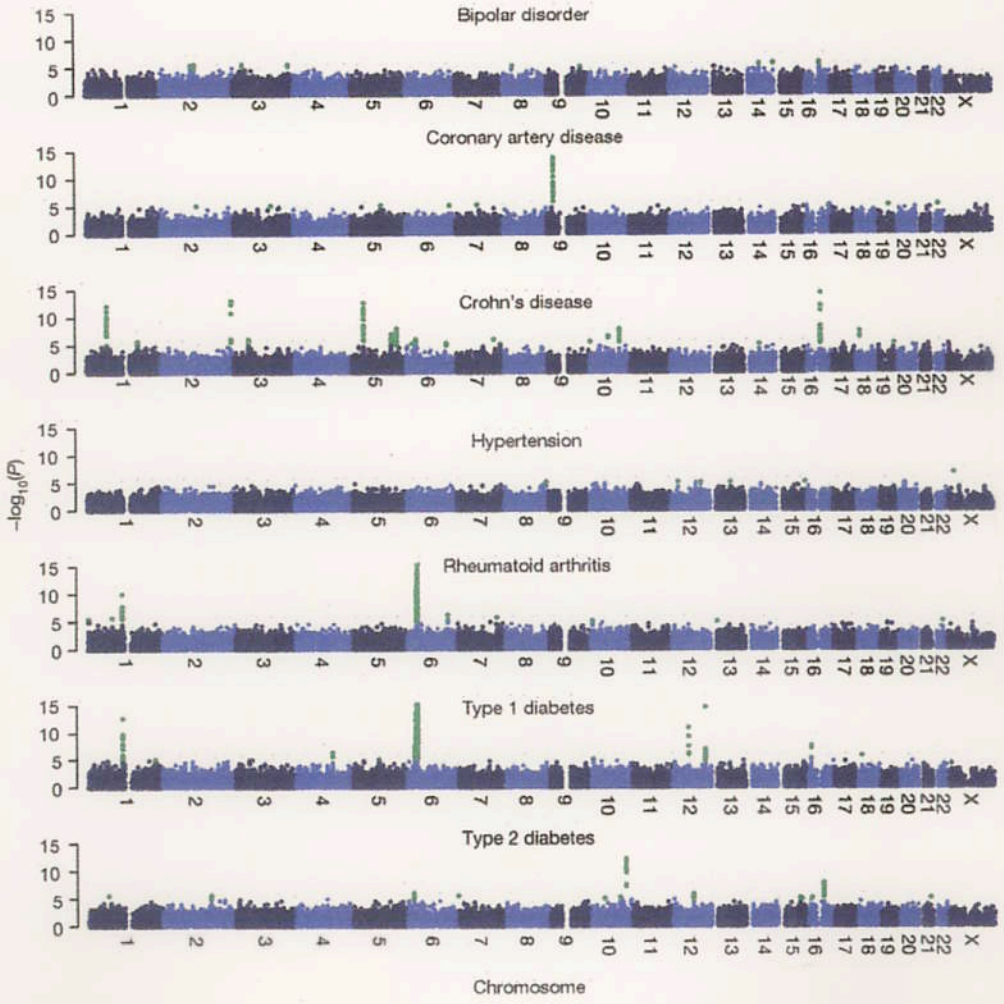
The Wellcome Trust Case Control Consortium*

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

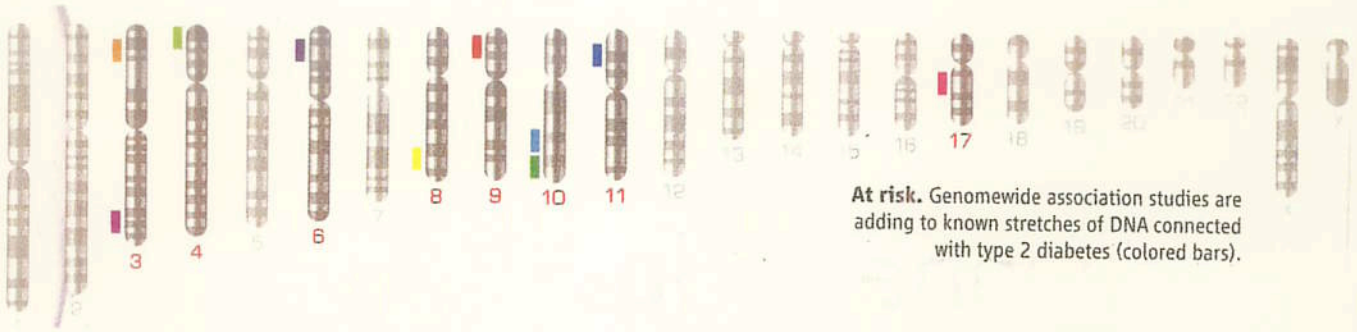
CORRELATING SNPs with Specific Diseases Using SNP Chips & Association Studies

Figure 4 : Genome-wide association study of 14,000 cases of seven... http://www.nature.com/nature/journal/v447/n7145/fig_tab/nature05...



SNPs May Be Near OR in Relevant Genes

USING LARGE POPULATIONS SNPs CAN BE USED AS MARKERS FOR SPECIFIC GENES/TRAITS



At risk. Genomewide association studies are adding to known stretches of DNA connected with type 2 diabetes (colored bars).

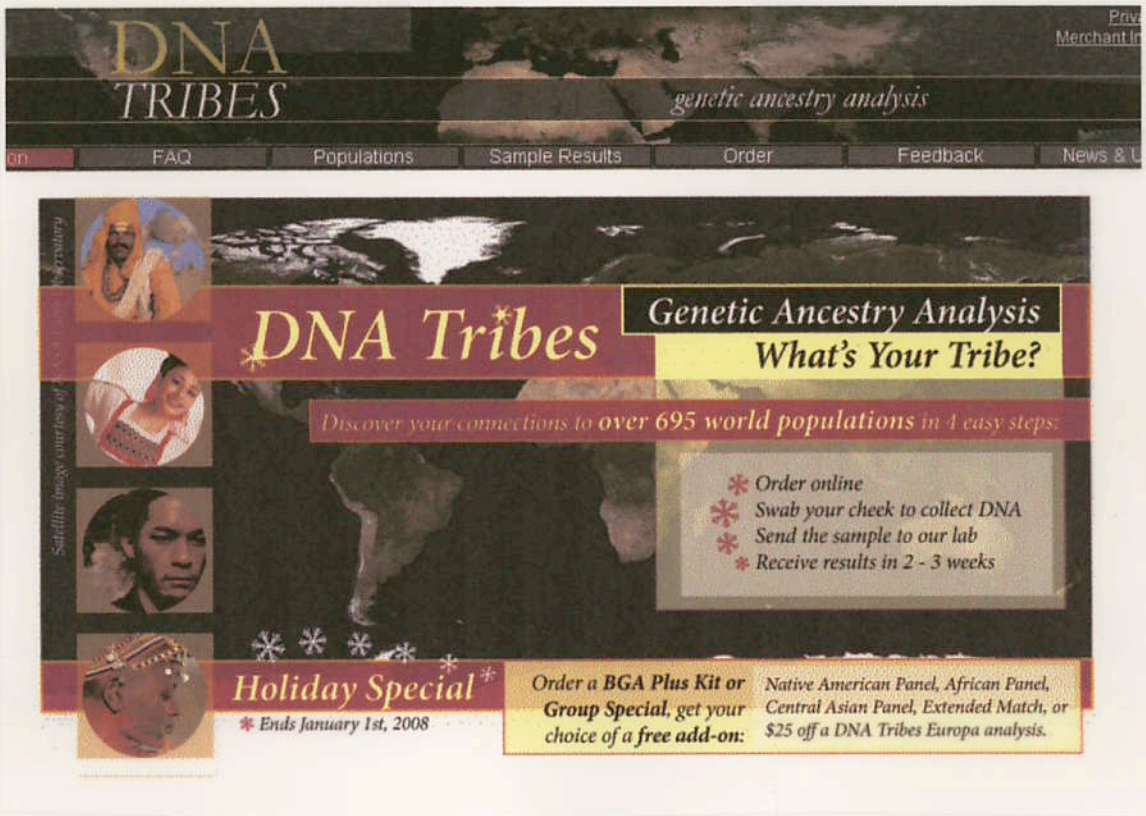
SNPedia

- 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 Good Are the Correlations?
What do with Information?
PRIVACY Issues?
GROUP Differences? Discrimination?

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

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.

Most studies of human variation begin by sampling from predefined "populations." These populations are usually defined on the basis of culture or geography and might not reflect underlying genetic relationships (*1*). Because knowledge about genetic structure of modern human populations can aid in inference of human evolutionary history, we used the HGDP-CEPH Human Genome Diversity Cell Line Panel (*2, 3*) to test the correspondence of predefined groups with those inferred from individual multilocus genotypes (supporting online text).

The average proportion of genetic differences between individuals from different human populations only slightly exceeds that between unrelated individuals from a single population (*4–9*). That is, the within-population component of genetic variation, estimated here as 93 to 95% (Table 1), accounts for most of human genetic diversity. Perhaps as a result of differences in sampling schemes (*10*), our estimate is higher than previous estimates from studies of comparable geographic coverage (*4–6, 9*), one of which also used microsatellite markers (*6*). This overall similarity of human populations is also evident in the geographically widespread nature of most alleles (fig. S1). Of 4199 alleles present more than once in the sample, 46.7% appeared in all major regions represented: Africa, Europe, the Middle East, Central/

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!

Within Population Differences Account for 95% of Human Genetic Variation

(20)

Mitochondrial DNA is MATernally Inherited

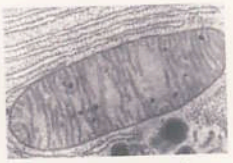
Mitochondria

Use oxygen to produce energy efficiently (aerobic metabolism).
-muscle cells are loaded with them.

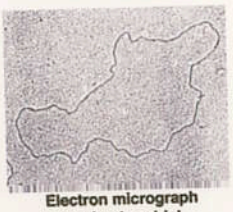
Contain own small genome ~17,000 bp circular DNA in humans

Encodes 2 rRNAs and 22 tRNAs for protein synthesis
13 proteins for energy metabolism

99.9% of the mitochondrial proteins encoded by nuclear genes.

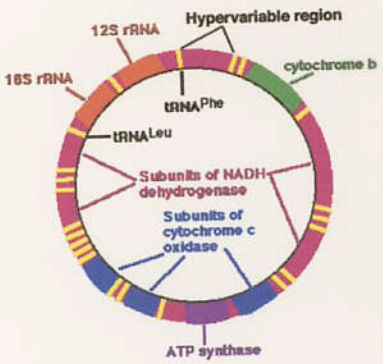


Electron micrograph of a mitochondrion



Electron micrograph of a mitochondrial DNA

Map of human mtDNA



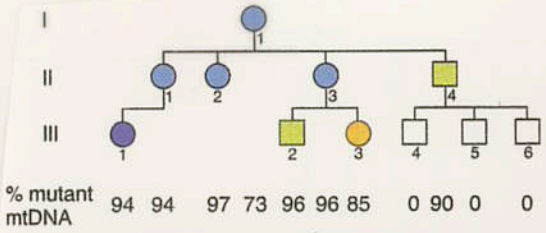


Figure 14.15 Maternal inheritance of the mitochondrial disease MERRF. (a) Transmission electron micrograph of muscle. Mitochondria from patients expressing MERRF. Mutant mitochondria are highly abnormal, showing paracrystalline arrays and crista degeneration. (b) Pedigree of family showing inheritance of MERRF. Pedigree shows typical pattern of maternal transmission observed with mitochondrial mutations. Mothers transmit the disease to all children, both boys and girls. Diseased fathers never transmit the disease to any children.

∴ CAN Be Used to TRACE Original "EVE"

By COMPARING SEQUENCE VARIANTS in Different POPULATIONS

VARIATION IN BOTH THE MITOCHONDRIAL AND NUCLEAR GENOMES CAN BE USED TO TRACE ANCESTRY

FAST FORWARD

Mitochondrial DNA Sequences Shed Light on Human Evolution

The mitochondrial DNA of all humans alive today traces back through maternal lineages to the mtDNA of a human population living in Africa some 200,000 years ago. Such is the startling conclusion of two papers published in 1987 and 1991 by Allan C. Wilson and colleagues. The carrier of this ancestral mtDNA, dubbed "mitochondrial Eve," probably lived in a population of 10,000–50,000 people. (In this context, a *population* is a group of interbreeding individuals of the same species who inhabit the same space.)

How mtDNA Variations Suggest the Region Where Modern Humans Emerged

In their studies supporting an African origin for modern humans, Wilson and coworkers first looked at restriction fragment length polymorphisms (RFLPs) in the mtDNA of 143 subjects, including Americans of African, Asian, European, and Middle Eastern origin, as well as Aboriginal women in New Guinea and Australia. Four years later, the group followed up their original 1987 study with sequence analyses of a rapidly evolving, highly polymorphic noncoding segment of mtDNA from 189 individuals, including 121 native Africans from various parts of the continent, Papua New Guineans, Europeans, Asians, African-Americans, and a native Australian. In both studies, the researchers found greater sequence differences among Africans, particularly sub-Saharan Africans, than among Asians or Europeans. Because mutations accumulate over time, they concluded that the African population has had the longest time to evolve variation and thus modern humans originated in Africa.

Statistical Calculations of the Number and Rate of Mutations Suggest When Modern Humans Appeared

Having proposed that modern humans first appeared in Africa, the researchers calculated the probable date of their origin by extrapolating the unknown from the known. They had observed the greatest human mtDNA variation in a sub-Saharan African population; about 2.8% of the base pairs in the mitochondrial genome varied among the individuals they studied from that population. They also knew that chimpanzees and humans diverged approximately 5 million years ago and that human mtDNA differs from that of chimpanzees in about 15% of the genome. Adjusting these data to account for multiple substitutions at the same base pair, they estimated that the mtDNA of humans and chimpanzees has been diverging at an average rate of about 13.8% per million years. To determine approximately how long ago the human population containing mitochondrial Eve lived, they divided the percentage of maximal human variation by the rate of chimp-human divergence:

$$2.8/13.8 = 0.20 \text{ million} = 200{,}000 \text{ years ago}$$

Although there has been some controversy over the statistical methods and assumptions that formed the basis of this analysis, most geneticists now agree with the conclusion that the women carrying our ancestral mtDNA lived roughly 200,000 years ago in sub-Saharan Africa. Recent studies of parts of the Y chromosome and other pieces of nuclear DNA support this conclusion. In one 1997 study, evolutionary geneticists examined two nuclear DNA segments on chromosome 12 in the genomes of 1600 individuals from 42 populations around the world. In populations outside of Africa, one combination of polymorphisms in the two elements showed up almost exclusively; by contrast, in sub-Saharan Africa, almost all the possible variants appeared. Recently, analyses of Y chromosome polymorphisms uncovered two ancient DNA markers shared by nonhuman primates and a small group of sub-Saharan African men. These markers sustained mutations about 100,000–200,000 years ago. Today, all men outside of Africa as well as most African men carry the mutated markers.

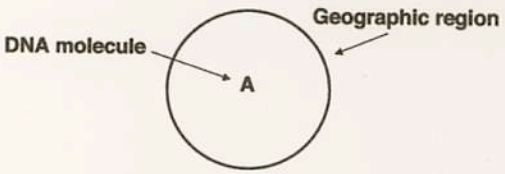
OLDEST POPULATIONS CONTAIN THE MOST DIVERSITY

Analysis of human mtDNA led to the Mitochondrial Eve Hypothesis

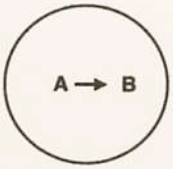
In the 1980s, Allan Wilson pioneered the use of mtDNA to study human evolution.

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.

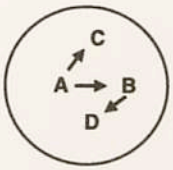
Here's the logic behind the hypothesis.



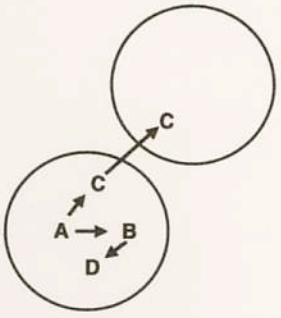
Start with A.



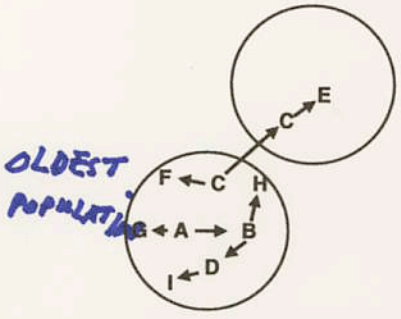
Mutation generates B from A;
now have individuals with both A and B DNAs in population.



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



C migrates to form separate population.

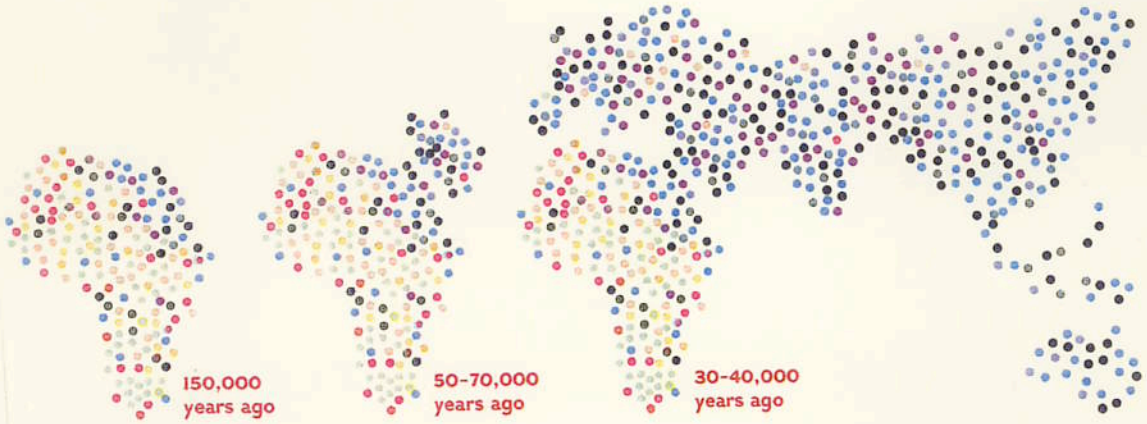


Additional mutations diversify DNAs in populations:
original population more diverse (A, B, C, D, F, G, H, I) than newer population (C, E).

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.



ORIGINS OF HUMAN POPULATIONS FROM DNA SEQUENCE COMPARISONS

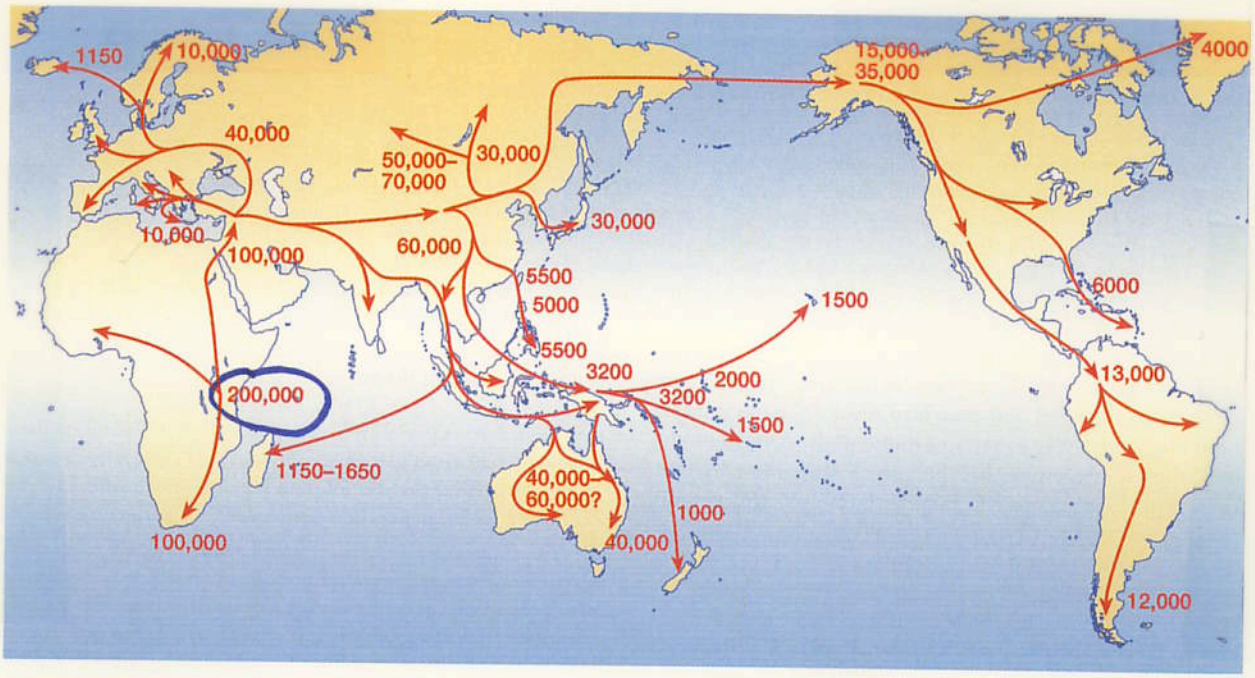


Figure 12.30: The spreading of *Homo sapiens* out of East Africa as postulated by the Recent African Origin model (uniregional hypothesis).

Dates (in years before the present) denote estimated dates of arrival of modern humans at indicated sites as supported by paleontological and archeological records. Dates of migration out of Africa suggested by genetic studies supporting the Recent African Origin model vary in the 50 000–200 000 year range. Figure adapted from Klein and Takahata (2002) *Where do we Come From? The molecular evidence for human descent,* with permission from Springer-Verlag, Berlin.

Are There Human Races?

a. Yes
b. No

iClicker

Human Races Have a Genetic Basis:

a. Yes
b. No

iClicker

ARE THERE HUMAN "RACES?"

What is the Biology + the History?

Human pigmentation genetics: the difference is only skin deep

Richard A. Sturm,[1]* Neil F. Box,[1] and Michele Ramsay[2]

Summary

There is no doubt that visual impressions of body form and color are important in the interactions within and between human communities. Remarkably, it is the levels of just one chemically inert and stable visual pigment known as melanin that is responsible for producing all shades of humankind. Major human genes involved in its formation have been identified largely using a comparative genomics approach and through the molecular analysis of the pigmentary process that occurs within the melanocyte. Three classes of genes have been examined for their contribution to normal human color variation through the production of hypopigmented phenotypes or by genetic association with skin type and hair color. The MSH cell surface receptor and the melanosomal P-protein are the two most obvious candidate genes influencing variation in pigmentation phenotype, and may do so by regulating the levels and activities of the melanogenic enzymes tyrosinase, TRP-1 and TRP-2. *BioEssays* **20**:712–721, 1998.

TABLE 1. Human Pigmentation Genes

Gene symbol	Mouse homologue	Chromosome	Phenotype	Protein	Function/ activity
TYR	Albino (c)	11q14–21	OCA1	Tyrosinase	Tyrosine hydroxylation; DOPA oxidase
TYRP1	Brown (b)	9p23	OCA3/ROCA	TRP-1	DHICA oxidase
TYRP2	Slaty (slt)	13q31–32	Unknown	TRP-2	Dopachrome tautomerase
P	Pink-eyed dilute (p)	15q11.2–12	OCA2, BOCA	P-protein	Melanosomal transmembrane protein
MC1R	Extension (e)	16q24.3	Red hair	MSHR	G-protein-coupled receptor

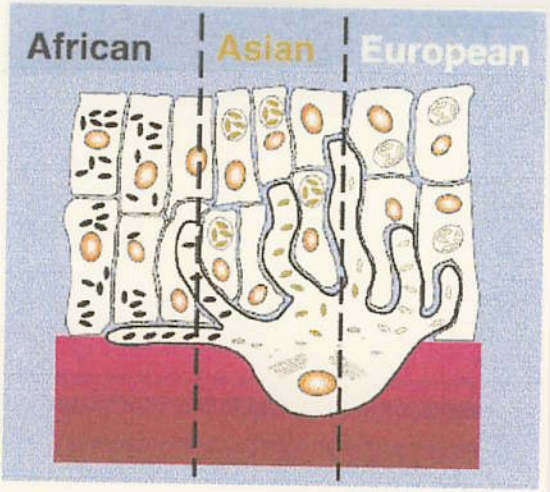


Figure 1. Variation in melanosome structure and distribution in different groups. A single skin melanocyte cell interdigitating with keratinocyte cells is partitioned into three sections. Shown within the melanocyte are the four stages of melanosome formation from budding from the Golgi apparatus, to the fully pigmented stage IV melanosomes migrating up the dendritic processes of the cell and secreted into the keratinocytes. In African populations, the melanosomes remain as singular heavily pigmented particles while in Asians and Europeans the melanosomes cluster in membrane bound organelles giving different skin complexions.

SLC24A5, a Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans

Rebecca L. Lamason,[1*] Manzoor-Ali P.K. Mohideen,[1†] Jason R. Mest,[1] Andrew C. Wong,[1‡] Heather L. Norton,[6] Michele C. Aros,[1] Michael J. Jurynec,[8] Xianyun Mao,[6] Vanessa R. Humphreville,[1§] Jasper E. Humbert,[2,9] Soniya Sinha,[2] Jessica L. Moore,[1‖] Pudur Jagadeeswaran,[10] Wei Zhao,[3] Gang Ning,[7] Izabela Makalowska,[7] Paul M. McKeigue,[11] David O'Donnell,[11] Rick Kittles,[12] Esteban J. Parra,[13] Nancy J. Mangini,[14] David J. Grunwald,[8] Mark D. Shriver,[6] Victor A. Canfield,[4] Keith C. Cheng[1,4,5¶]

Lighter variations of pigmentation in humans are associated with diminished number, size, and density of melanosomes, the pigmented organelles of melanocytes. Here we show that zebrafish *golden* mutants share these melanosomal changes and that *golden* encodes a putative cation exchanger slc24a5 (nckx5) that localizes to an intracellular membrane, likely the melanosome or its precursor. The human ortholog is highly similar in sequence and functional in zebrafish. The evolutionarily conserved ancestral allele of a human coding polymorphism predominates in African and East Asian populations. In contrast, the variant allele is nearly fixed in European populations, is associated with a substantial reduction in regional heterozygosity, and correlates with lighter skin pigmentation in admixed populations, suggesting a key role for the *SLC24A5* gene in human pigmentation.

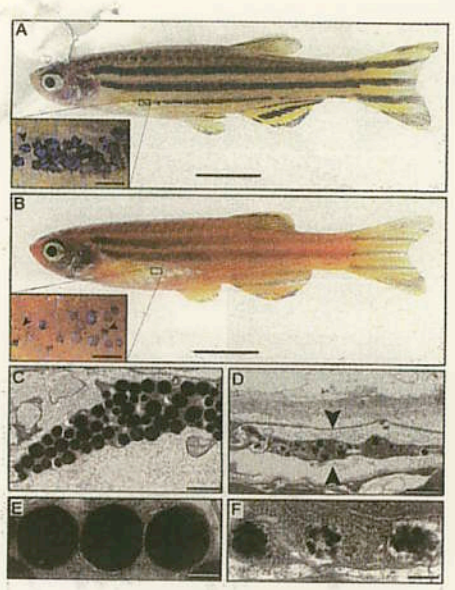


Fig. 1. Phenotype of *golden* zebrafish. Lateral views of adult wild-type (A) and *golden* (B) zebrafish. Insets show melanophores (arrowheads). Scale bars, 5 mm (inset, 0.5 mm). gol^{b1} mutants have melanophores that are, on average, smaller, more pale, and transparent. Transmission electron micrographs of skin melanophore from 55-hpf wild-type (C and E) and gol^{b1} (D and F) larvae. gol^{b1} skin melanophores (arrowheads show edges) are thinner and contain fewer melanosomes than do those of wild type. Melanosomes of gol^{b1} larvae are fewer in number, smaller, less-pigmented, and irregular compared with wild type. Scale bars in (C) and (D), 1000 nm; in (E) and (F), 200 nm.

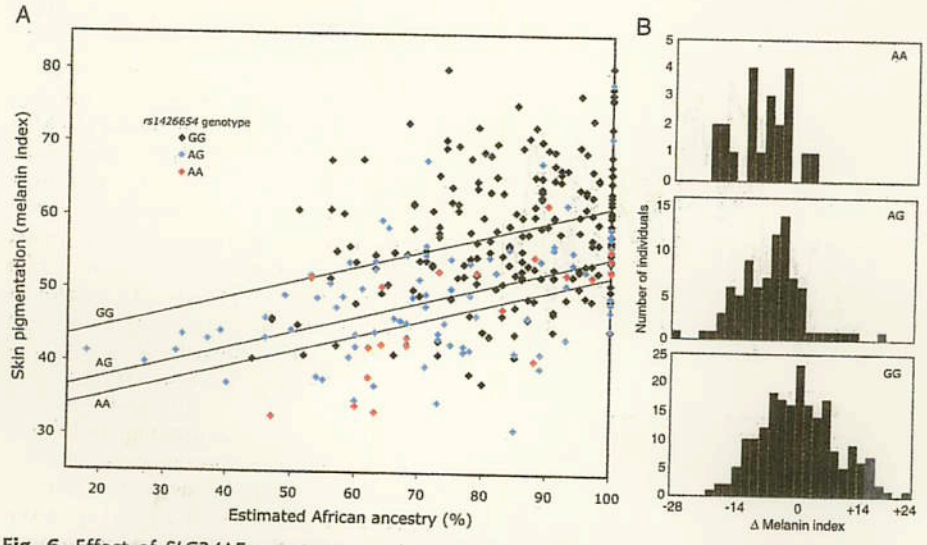


Fig. 6. Effect of *SLC24A5* genotype on pigmentation in admixed populations. (A) Variation of measured pigmentation with estimated ancestry and *SLC24A5* genotype. Each point represents a single individual; *SLC24A5* genotypes are indicated by color. Lines show regressions, constrained to have equal slopes, for each of the three genotypes. (B) Histograms showing the distribution of pigmentation after adjustment for ancestry for each genotype. Values shown are the difference between the measured melanin index and the calculated GG regression line ($y = 0.2113x + 30.91$). The corresponding uncorrected histograms are shown in fig. S7. Mean and SD (in parentheses) are given as follows: for GG, 0 (8.5), $n = 202$ individuals; for AG, –7.0 (7.4), $n = 85$; for AA, –9.6 (6.4), $n = 21$.

Race is Primarily a Sociological Concept that has caused much Human Suffering

Race is a largely non-biological concept confounded by misunderstanding and a long history of prejudice. The relationship of genomics to the concepts of race and ethnicity has to be considered within complex historical and social contexts.

Most variation in the genome is shared between all populations, but certain alleles are more frequent in some populations than in others, largely as a result of history and geography. Use of genetic data to define racial groups, or of racial categories to classify biological traits, is prone to misinterpretation. To minimize such misinterpretation, the biological and sociocultural factors that interrelate genetics with constructs of race and ethnicity need to be better understood and communicated within the next few years.

This will require research on how different individuals and cultures conceive of race, ethnicity, group identity and self-identity, and what role they believe genes or other biological factors have. It will also require a critical examination of how the scientific community understands and uses these concepts in designing research and presenting findings, and of how the media report these. Also necessary is widespread education about the biological meaning and limitations of research findings in this area (Box 6) and the formulation and adoption o public-policy options that protect agains genomics-based discrimination or maltreatment (see Grand Challenge III-1).

Genome Project Goal

(1) MOST VARIATION SHARED

(2) Differences due to Migrations & Geographic Isolation

(3) Most VARIation SHARED by groups!!

Based on a very few genes that vary between groups FAR more than Majority - VAST Majority of other Genes

HUMAN DIVERSITY

RICHARD LEWONTIN

SIMILARITY AND DIFFERENCES IN BLOOD GROUP ALLELE FREQUENCIES by "RACE" or Population

HOW DOES Genetic VARIATION (different alleles) VARY Between & within Populations?

Examples of extreme differentiation and close similarity in blood group allele frequencies in three racial groups

(1) Different variation but most Alleles present →

(2) SAME VARIATION →

Gene	Alleles	Caucasoid	Negroid	Mongoloid
Duffy	*Fy*	.0300	.9393	.0985
	Fy^a	.4208	.0607	.9015
	Fy^b	.5492	—	—
Rhesus	R_0	.0186	.7395	.0409
	R_1	.4036	.0256	.7591
	R_2	.1670	.0427	.1951
	r	.3820	.1184	.0049
	r'	.0049	.0707	0
	others	.0239	.0021	0
P	P_1	.5161	.8911	.1677
	P_2	.4839	.1089	.8323
Auberger	Au^a	.6213	.6419	
	Au	.3787	.3581	
Xg	Xg^a	.67	.55	.54
	Xg	.33	.45	.46
Secretor	*Se*	.5233	.5727	
	se	.4767	.4273	

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

(1) Most alleles in ALL "races" / populations

(2) NO Homozygousity at any locus

(3) Some Allelic Frequencies differ Between "Races" & Some are the Same! Duffy differs & X_g is the SAME. Why? Adaptive value.

(4) Auberger, X_g, & Secretor Loci show have alleles vary within populations similarly & show NO between population differences.

(5) Wide range & Different Alleles within/between "Races"

There is A Large variation in D1S80 VNTR alleles within populations but Little between

Table 15.2 Allele frequencies for *D1S80* among U.S. population groups

Repeat number	Caucasian	Hispanic	African American	Asian
14	0	0	0	0
15	0	0.001	0	0
16	0.001	0.010	0.002	0.034
17	0.002	0.009	0.028	0.025
18	0.237	0.224	0.073	0.152
19	0.003	0.005	0.003	0.022
20	0.018	0.013	0.032	0.007
21	0.021	0.028	0.115	0.034
22	0.038	0.024	0.081	0.017
23	0.012	0.009	0.014	0.017
24	0.378	0.315	0.234	0.230
25	0.046	0.072	0.045	0.027
26	0.020	0.007	0.006	0
27	0.007	0.016	0.008	0.047
28	0.063	0.078	0.130	0.076
29	0.052	0.055	0.053	0.042
30	0.008	0.039	0.009	0.123
31	0.072	0.053	0.054	0.093
32	0.006	0.005	0.007	0.012
33	0.003	0.004	0.004	0.005
34	0.001	0.006	0.086	0.005
35	0.003	0	0.002	0.005
36	0.004	0.011	0.001	0.005
37	0.001	0.004	0	0.007
38	℘	℘	℘	℘
39	0.003	0.004	0.003	0.005
40	0	0	0	0
41	0	0.002	0.002	0.007
>41	0.001	0.006	0.007	0.002
Sample size	718	409	606	204

Source: Data from B. Budowle, et al. 1995. *Journal of Forensic Science* 40:38

JUST AS Predicted FROM Class Genotype

VNTR used for HGTOA DNA Fingerprint

NO ADAPTIVE VALUE!!
Good Locus for Forensics - Genes Neutral!

Reasons for Allelic Variation Between Populations

1. **Founder Effect / Geographical Isolation**
   - ↳ Selective mating due to geography/culture or both

2. **Adaptive Value**
   - ↳ Have positive effect in specific environments
     - e.g.; HB^S Sickle-Cell Globin allele (India, Africa, Mediterranean)
     - Duffy
     - Skin Color Genes

GEOGRAPHY &/or ADAPTIVE VALUE

But do not vary across whole genome as most loci are "neutral"

ADAPTIVE VALUE OF THE Hb^S Allele

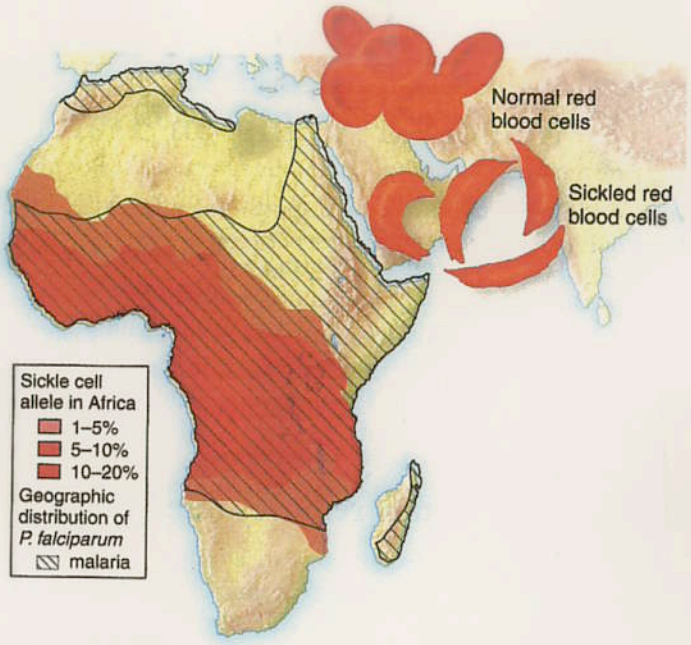


figure 20.12

FREQUENCY OF SICKLE CELL ALLELE AND DISTRIBUTION OF *PLASMODIUM FALCIPARUM* MALARIA. The red blood cells of people homozygous for the sickle cell allele collapse into sickled shapes when the oxygen level in the blood is low. The distribution of the sickle cell allele in Africa coincides closely with that of *P. falciparum* malaria.

There is More Genetic Diversity within Populations than Between Populations!! So much for the concept of racial "purity"!!!

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).

More genetic diversity within any population than between populations!

(1) 85% of Human Genetic Variation occurs within populations & between Individuals in that population!

(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 due to Differences between Races!! Geographic

CONCLUSION

(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 a "race") → e.g., FyO^{Es}

(3) Then Losing all "races" except one retains 94% of all Human Genetic Variation!

[85% + (15% − 7%)] = 94%

VARIATION That Occurred in ANCESTORAL Population

85% within population genetic variability
8% between populations of same "race"
7% between "race" genetic variability

(4) ∴ Humans 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 frequency in other "Races."
ALL Humans have SAME Genes — differ in form
Mostly within populations!

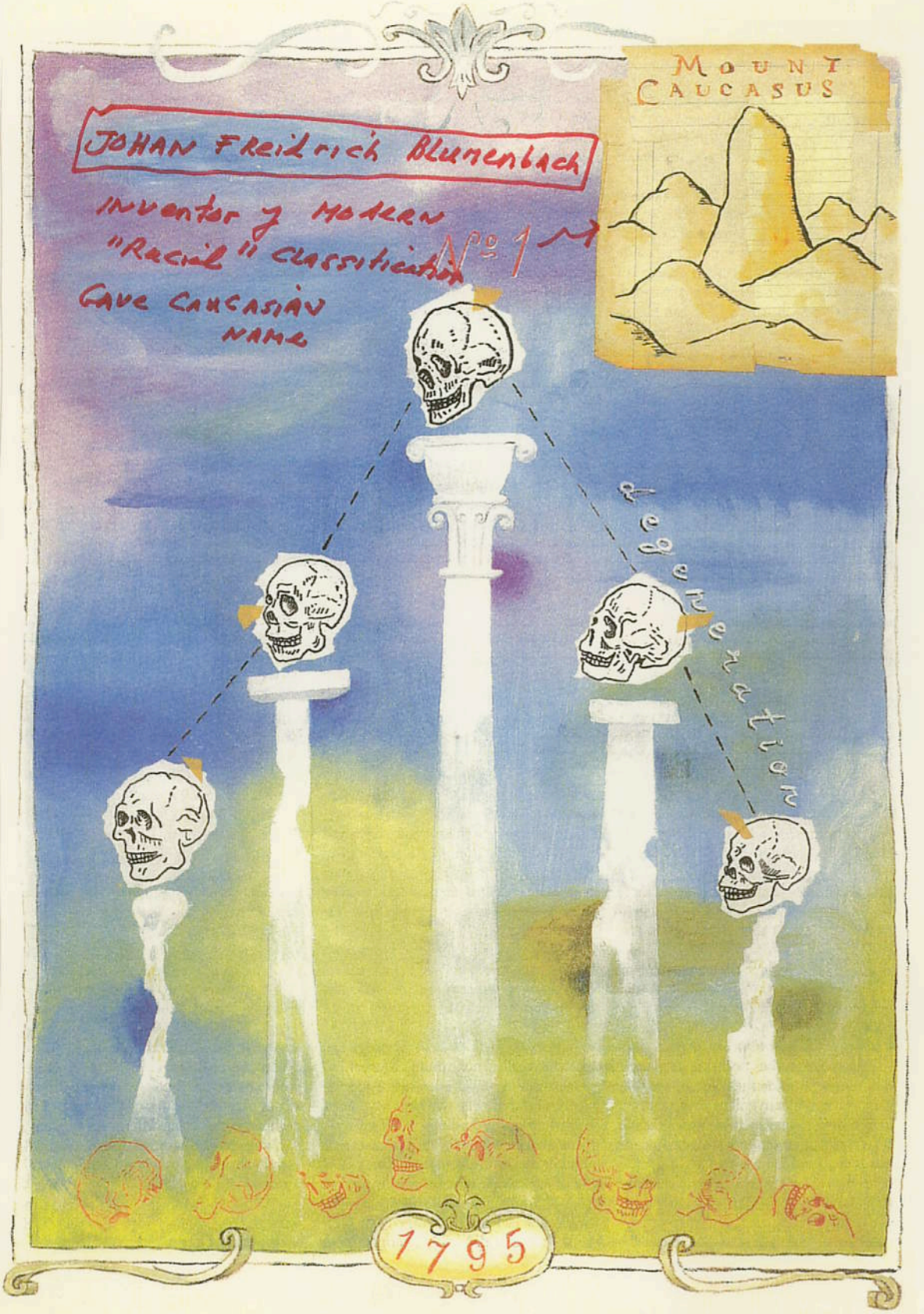
(3) Heterozygosity (Variation) high in human populations — ALL populations. None homozygous at all Loci!

(4) NO such thing as a "pure" Race — would have little variation —

(5) Genes Affecting Physical Features NOT representative of genes across genome —

Geographical Ancestry is Relevant — Many "racial" groups now have Multiple Ancestries because of Admixture & Migration

HOW DID WE GET TO WHERE WE ARE?



JOHAN FREIDRICH BLUMENBACH

(1) Changed Linnaean Neutral-geography-based Human Classifications to a value-based Classification—

(2) Said all "races" originated from one place/origin – in Europe around Mt. CAUCASUS – because they are the most "beautiful" race – even though he felt all races were equal in all respects & argued with people that didn't!

(3) Classified five "races"

- CAUCASIAN
- Mongolian
- Ethiopian
- American
- Malay

Value-Based Classification & putting Caucasians on top – Most beautiful – had disasterous consequences!

The beautiful skull and Blumenbach's errors

Raj Bhopal reappraises Blumenbach, an important contributor to the scientific concept of race

The biological concept of human races, as subspecies characterised primarily by physique, has a stormy history.[1] The consensus after the Second World War—that race is a social construct with minor biological components—is now under academic scrutiny, as illustrated by three advances in biomedical science. Firstly, the mapping of the human genome is enabling the importance of genetics in creating and perpetuating differences between populations to be analysed thoroughly.[2] Secondly, personalised medicine has been rejuvenated by pharmacogenomics, which is finding racial classification, for all its weaknesses, a convenient though crude route to understanding differences in drug response. Thirdly, in 2005, the Food and Drug Administration of the United States licensed the drug BiDil (a combination of hydralazine and isosorbide dinitrate) exclusively for the (self defined) black population.[3]

Meanwhile, in support of race as a social construct, and to counter racism, race equality has been enshrined in international and national laws and in governmental and institutional policies.

Blumenbach's human varieties

First edition of his MD thesis: four varieties according to geography (1775)
- People from Europe
- People from Asia to the Ganges and some parts of North America
- People from Africa
- People from North America

Second edition: five varieties according to geography (1781)
- People from Europe (primeval) including north India, North Africa, North America (for example, Esquimaux)
- People from the rest of Asia, beyond the Ganges river
- People from Africa (except the north)
- People from the rest of America
- People from the southern world (such as the Philippines)

Third edition: five generic varieties (1795)
- Caucasians
- Mongolians
- Ethiopians
- Americans
- Malays

Race, and the related and newer concept of ethnicity (subgrouping human populations using cultural and physical features, thereby subsuming race), are prominent in modern multi-ethnic societies.[1] Race and racism are topical subjects in the United Kingdom because of the 200th anniversary of the UK's 1807 Abolition of the Slave Trade Act. It is therefore a good time to re-examine the role of perhaps the most important contributor to the scientific concept of race, Blumenbach, whose insights and errors provide important lessons for us today.[4]

A scientist and humanitarian

For this article, I draw largely on Blumenbach's collected treatises,[4] edited by Thomas Bendyshe. This book includes two memoirs on Blumenbach, one by Professor K F H Marx and the other by M Flourens; the first and third editions of Blumenbach's MD thesis; some other works by Blumenbach; and an essay by Dr John Hunter also on the varieties of humans published in 1775.

Marx's introduction states that in Blumenbach's time "negroes and savages" were considered half animal, and the idea of emancipating slaves was alien. While this was an exaggeration, as the emancipation movement was already gathering momentum, it does reflect the ethos of those times. Blumenbach proclaimed, unequivocally, that such people were only separated from other humans by opportunity. This contribution alone is notable. Blumenbach was revered for his humanity and his science, as indicated not only by the two memoirs in Bendyshe's volume, but also other sources.

Blumenbach's thesis was published in three editions. In 1775 the first edition discussed four geographically defined varieties of humans, while the second edition in 1881 outlined five geographically defined varieties. He developed this classification further in the third edition, which is the definitive volume, where he provided generic rather than geographical labels.

On the natural varieties of mankind (1775)

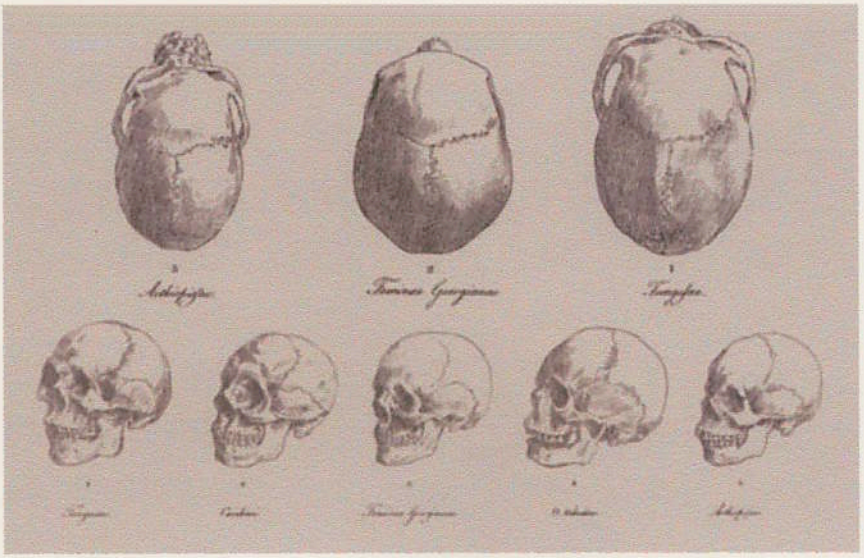
The first edition starts with the potential of crossing between species and mentions humans mating with animals. Blumenbach found no evidence for this. He concluded that humans are a unique species, with no intermediate forms that are partly non-human. He identified the major unique features of humans as the large brain, speech, erect posture, two free hands, naked skin, and the hymen in women (and possibly menstruation).

Blumenbach's central question, one of great interest at the time and still rarely discussed in science, was whether contemporary humans comprised one or more species.[4] Plurality of human species (polygeny) was the popular view in the 18th century. Blumenbach emphasised the unity of humanity, however. He saw gradations among humans, but no distinct species or subspecies. None the less, in the first edition he ventured to describe—cautiously and somewhat reluctantly—four varieties of humans relating to four geographical regions (box). In the second edition in 1781 (also appearing in a footnote in a reprinting of the first edition he identified five varieties of humans relating to five geographical regions (box).

Blumenbach attributed differences between these human types—such as variations in stature and colour—largely to climate. He dismissed leucoplakia, a condition characterised by loss of skin pigmentation, as merely a disease and not even a variety of humanity, never mind a subspecies, as Blumenbach interpreted others' work. He noted that many plants and animals in northern latitudes are white, especially in winter, and also that humans are all born red. Colour, he said, cannot constitute a species or a variety. He attributed the shape of the skull to environmental factors, an observation that threatened the foundations of craniology but was not properly heeded by craniologists (and possibly by himself). He identified the important role that culture plays in changing the body.

Five varieties of humans (1795)

The third edition starts with a letter from Blumenbach to Sir Joseph Banks, which clarified the notion that humans have their own order of mammalia, gave credit to Linnaeus for being the first to arrange mankind in certain varieties, and making the argument for fresh thinking on this issue.[4 5] He also lists his scientific methods, which include examining skulls, fetuses, hair, anatomical preparations, and pictures and drawings.



Blumenbach thought the skull of a Georgian female (middle, top and bottom) was the fairest of them all

He is more systematic than in the first edition; for example, the differences between humans and animals are listed as erect position, broad flat pelvis, two hands, and regular close set rows of teeth. He emphasises that no clear cut subdivisions of human species exist, but that the "varieties . . . run into one another by insensible degrees." None the less, he now discerns five varieties—and these labels have stuck to this day (box). The conceptual underpinning of Blumenbach's classification is largely forgotten and misrepresented, while his classification is mentioned by most people reviewing the topic of race.

He gave examples of people fitting his five varieties. He stated that Turkish and Hindostan women were Caucasians but that people from Bengal and Esquimaux people were Mongolians. He identified New Zealanders (Maoris) as Malays. He thought that Egyptians could be Ethiopian, Indian, or a type with "short chin and prominent eyes." He was surprised that other people attributed Egyptians to one type. Blumenbach recognised the heterogeneity within populations in one land or nation, something that was overlooked in his time, as it often is now.

Skulls and blunders

Blumenbach put special emphasis on the study of skulls and he reduced a diversity of skulls to five main varieties. The two key plates are reproduced in the figure (plates III and IV in the treatise). Surprisingly, although he had noted that environment influenced skull shape, he drew major and firm conclusions from his skulls. He wrote, "The meaning and use of this will easily be seen by an examination of plate III, which represents, by way of specimen, three skulls disposed in the order mentioned. The middle one (2) is a very symmetrical and beautiful one of a Georgian female; on either side are two skulls differing from it in the most opposite way. The one elongated in front, and as it were keeled, is that of an Ethiopian female of Guinea (3); the other dilated outwardly toward the sides, and as it were flattened, is that of a Reindeer Tungus (2). In the first, the margin of the orbits, the beautifully narrowed malar bones, and the mandibles themselves under the bones, are concealed by the periphery of the moderately expanded forehead; in the second, the maxillary bones are compressed laterally, and project; and in the third, the malar bones, placed in nearly the same horizontal plane with the little bones of the nose and the glabella, project enormously, and rise on each side."

Mostly, Blumenbach's writing retained a scientific stance, but he exposed his bias on beauty when he wrote that the Causasian skull of a Georgian female was the "most handsome and becoming." He stated that the most beautiful people live in the Southern slope of Mount Caucasus—that is, the Georgian people. He then speculated on the origins of humans and made his second error, by going beyond the available evidence. White, to quote Blumenbach, "we may fairly assume to have been the primitive colour of mankind." His reasoning was that it is easy to change from white to brown but not vice versa. Time has shown that this view was wrong.

These errors were not the result of colour prejudice. Blumenbach refuted the notion that Ethiopians were inferior to other races. Blumenbach wrote favourably about "negroes," extolling their beauty, mental abilities, and achievements in literature and other fields. He pointed to variations in opportunity as the cause of differences. His viewpoint on Africans was out of tune with that of the times[6] and more in line with that seen during the movements for civil rights and equality in the 1960s.

Blumenbach's legacy

Blumenbach wrote, in a pleased tone, that he had made no striking new discovery but had reached a satisfactory conclusion that all humans are one species. His view on the unity of humanity (monogeny) was a timely correction of the erroneous movement claiming that humans comprised several species.[6]

Blumenbach's work was a turning point in the history of race and science, although it was nearly 200 years before the lessons were properly absorbed. Blumenbach's legacy is tarnished by biases and errors, and it teaches us that even great scientists can be led astray by personal views (such as notions about beauty) shaped by the ethos of their times. His original words also show how the simple, clear cut classification of five distinct human races displaced—against Blumenbach's repeated warnings—the complex reality of gradations and the unity of humanity (including equal potential). Blumenbach's name has been associated with scientific racism, but his arguments actually undermined racism. Blumenbach could not have foreseen the coming abuse of his ideas and classification in the 19th and (first half of the) 20th centuries.

We continue to struggle with the complexity of the concepts of race and ethnicity, and the resultant imperfect classifications.[1] Now Blumenbach's varieties of humanity can be seen in virtually every major city, and through the visual media, globally. Blumenbach's thinking, despite its faults, continues to be relevant, inspiring, and illuminating.

Raj Bhopal is Bruce and John Usher professor of public health at the university of Edinburgh (Raj.bhopal@ed.ac.uk).

Thanks to Anne Houghton for secretarial assistance, the library of the University of Edinburgh for reproducing Blumenbach's plates electronically, and Aziz Sheikh and Iain Milne for helpful comments on an earlier draft.

A biography of Blumembach is in the version on bmj.com

Competing interests: None declared.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 Bhopal RS. *Ethnicity, race, and health in multicultural societies; foundations for better epidemiology, public health, and health care.* Oxford: Oxford University Press, 2007:357.
2 Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic structure of human populations. *Science* 2002;298:2381-5.
3 Rahemtulla T, Bhopal R. Pharmacogenetics and ethnically targeted therapies. *BMJ* 2005;330:1036-7.
4 Blumenbach JF. *The anthropological treatises of Johann Friedrich Blumenbach.* London: Anthropological Society, 1865.
5 Linnaeus C. *A general system of nature through the three grand kingdoms of animals, vegetables, and minerals. Systema naturae.* London: Lackington Allen, 1806.
6 Kitson PJ. Bales of leading anguish: representations of race and the slave in romantic writing. *ELH* 2000;67:515-37.

81

CONCLUDE

(1) Races are Arbitrary Entities - Social Constructs that are Culturally generated.

(2) Yes - there is genetic differences between "races" or relatively inbred geographical populations that can lead to physical & other differences - due to very small # genes & not reflective of whole genome

(3) Within population genetic variation Much greater than between population genetic variation - Many loci have same allele frequencies - some differ

(4) Only minor differences between genomes of different people or groups of people - Unity >>> differences!

(5) We are all the same - but different. you now know why!

Race classifications Arbitrary, unscientific & divisive!