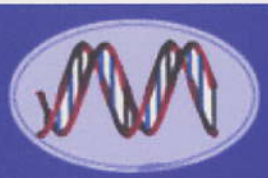
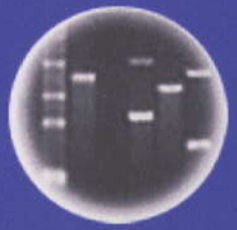


DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

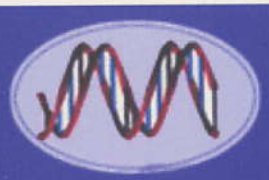
Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

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

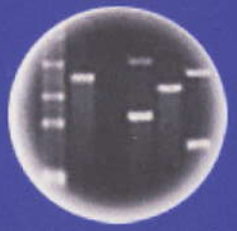
Professor Bob Goldberg

Lecture 5
The Age of Genomics: Your Personal Genome



DNA
Genetic Code of Life

Entire Genetic Code
of a Bacteria



DNA Fingerprinting

Cloning: Ethical Issues
and Future Consequences

Plants of Tomorrow

THEMES

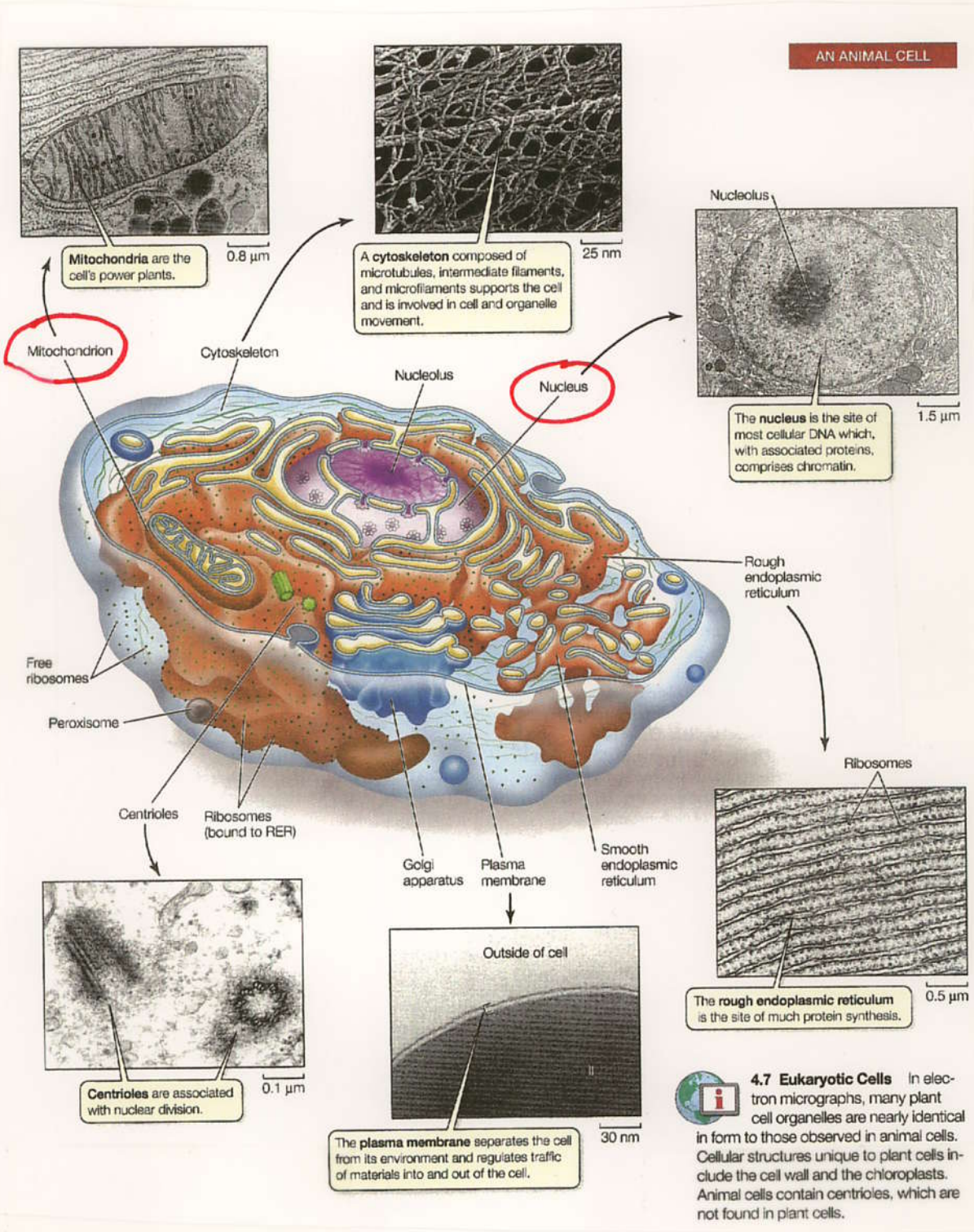
1. Two Genomes in a Cell!
2. What is the Mitochondrial Genome and How is it inherited?
3. What Are the Characteristics of the Human Genome?
4. How Does Genetic Variation Arise in the Human Genome?
5. How to Detect DNA Sequence Variation: SNPs and VNTRs?
6. How to Use DNA Markers to Find Human Disease Gene Alleles?
7. What is a SNP Map of the Human Genome and What is it Used For?
8. How Has the Age of Personal Genome Sequences Arrived?
9. How Are DNA Fragments Sequenced?
10. How is a Genome Sequenced?
11. How Has Massively Parallel DNA Sequencing "Changed the Game?"

You are Reading Your Text Before Class and Discussion?

a. Yes
b. No

iClicker

HUMAN CELLS HAVE TWO GENOMES



4.7 Eukaryotic Cells In electron micrographs, many plant cell organelles are nearly identical in form to those observed in animal cells. Cellular structures unique to plant cells include the cell wall and the chloroplasts. Animal cells contain centrioles, which are not found in plant cells.

NUCLEAR & MITOCHONDRION

There Are Two Genomes in Different Parts of a Cell?:

a. Yes
b. No

iClicker

The Nuclear and Mitochondrial Genomes Differ in Size & Shape

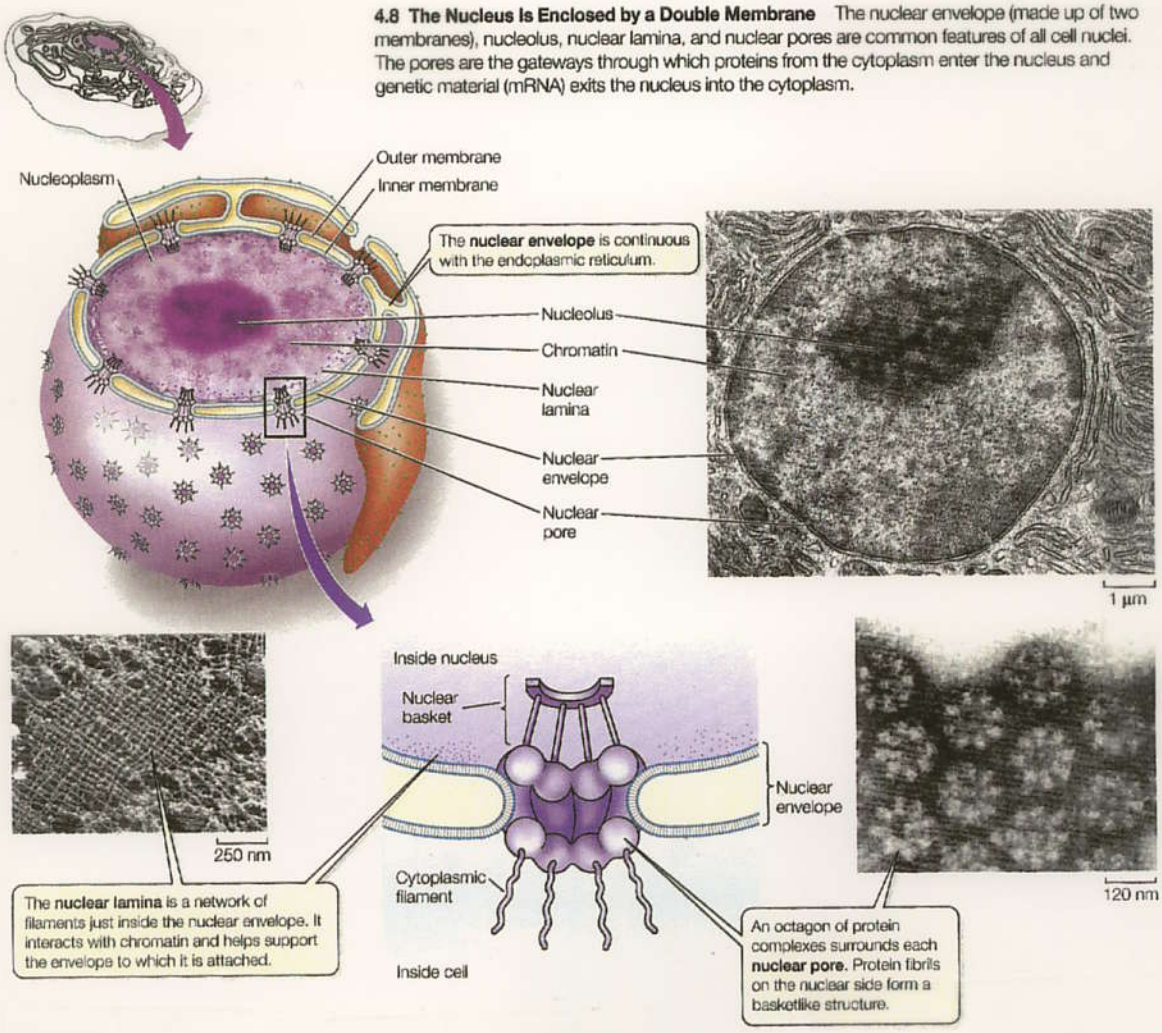
NUCLEAR
3.2 Mb
25,000 Genes
24 Linear Pieces

Mitochondrion
17 kb
37 Genes
Circle

Table 9.1: The human nuclear and mitochondrial genomes

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

The Nucleus is a complex organelle with 23 pairs of chromosomes (Humans)



4.8 The Nucleus Is Enclosed by a Double Membrane The nuclear envelope (made up of two membranes), nucleolus, nuclear lamina, and nuclear pores are common features of all cell nuclei. The pores are the gateways through which proteins from the cytoplasm enter the nucleus and genetic material (mRNA) exits the nucleus into the cytoplasm.

Packing Problem

figure 10.6

A HUMAN KARYOTYPE. The individual chromosomes that make up the 23 pairs differ widely in size and in centromere position. In this preparation, the chromosomes have been specifically stained to indicate differences in their composition and to distinguish them clearly from one another. Notice that members of a chromosome pair are very similar but not identical.

The Human Genome End to End is

NOTE: Chromosome sizes & bands = markers

Mitochondria Power Human Cells and Contain a Circular Genome

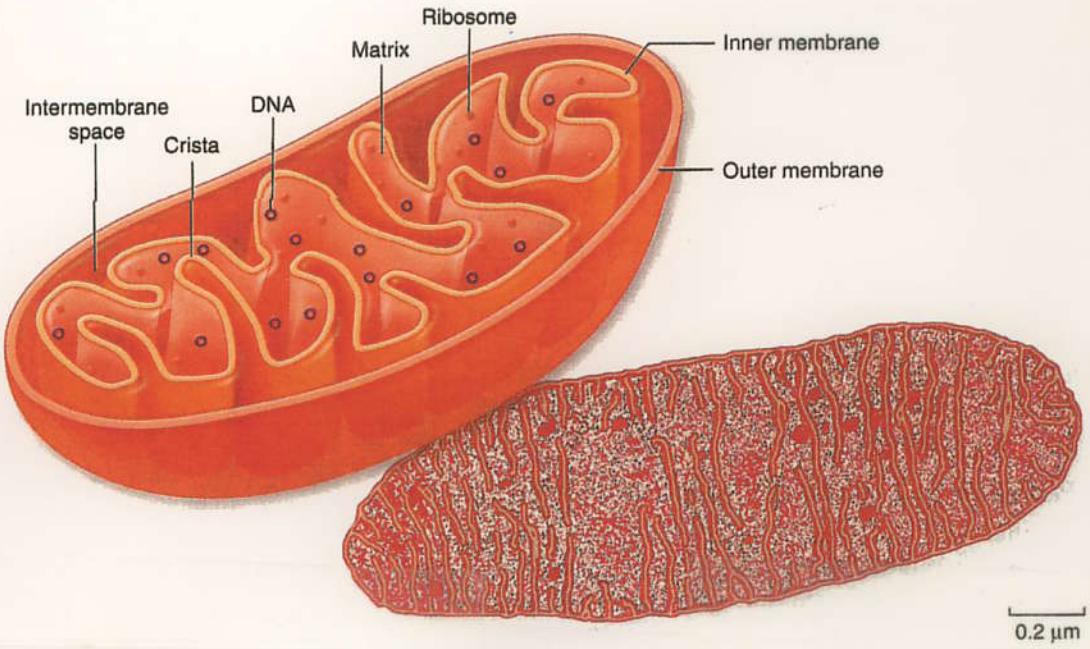


figure 4.17

MITOCHONDRIA. The inner membrane of a mitochondrion is shaped into folds called cristae that greatly increase the surface area for oxidative metabolism. A mitochondrion in cross section and cut lengthwise is shown colored red in the micrograph.

SEMI-AUTONOMOUS GENOME

DNA DIVIDES TRANSCRIPTION TRANSLATION

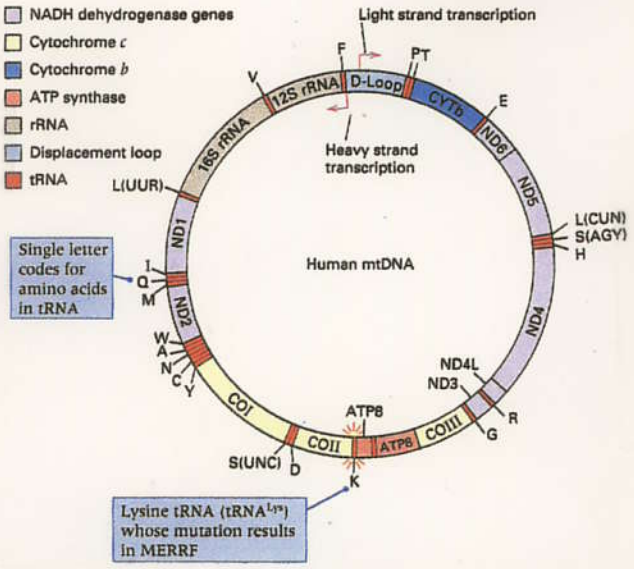


Figure 16.3 Genes in human mitochondrial DNA. The tRNA genes are indicated by the one-letter amino acid symbols; hence tRNA^{Lys} is denoted K. The positions of these and other genes in the mitochondrial DNA are indicated by color according to the key at the upper left. The arrows indicate the promoters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

MITOCHONDRIAL PROTEINS

Mitochondrial Genes Are Inherited:

a. Paternally
b. Maternally

iClicker

DEFECTS IN MITOCHONDRIAL GENES CAN LEAD TO SERIOUS HUMAN GENETIC DISEASES

Table 16.1 Phenotypes associated with some mitochondrial mutations

Nucleotide changed	Mitochondrial component affected	Phenotype[a]
3460	ND1 of Complex I[b]	LHON
11778	ND4 of Complex I	LHON
14484	ND6 of Complex I	LHON
8993	ATP6 of Complex V[b]	NARP
3243	$tRNA^{Leu(UUR)}$[c]	MELAS, PEO
3271	$tRNA^{Leu(UUR)}$	MELAS
3291	$tRNA^{Leu(UUR)}$	MELAS
3251	$tRNA^{Leu(UUR)}$	PEO
3256	$tRNA^{Leu(UUR)}$	PEO
5692	$tRNA^{Asn}$	PEO
5703	$tRNA^{Asn}$	PEO, myopathy
5814	$tRNA^{Cys}$	Encephalopathy
8344	$tRNA^{Lys}$	MERRF
8356	$tRNA^{Lys}$	MERRF
9997	$tRNA^{Gly}$	Cardiomyopathy
10006	$tRNA^{Gly}$	PEO
12246	$tRNA^{Ser(AGY)}$[c]	PEO
14709	$tRNA^{Glu}$	Myopathy
15923	$tRNA^{Thr}$	Fatal infantile multisystem disorder
15990	$tRNA^{Pro}$	Myopathy

[a]LHON Leber's hereditary optic neuropathy; NARP Neurogenic *muscle weakness, ataxia, retinitis pigmentosa;* MERRF Myoclonic epilepsy and ragged-red fiber syndrome; MELAS Mitochondrial myopathy, encephalopathy, lactic acidosis, stroke-like episodes; PEO Progressive external ophthalmoplegia

[b]Complex I is NADH dehydrogenase. Complex V is ATP synthase.

[c]In $tRNA^{Leu(UUR)}$, the R stands for either A or G; in $tRNA^{Ser(AGY)}$, the Y stands for either T or C.

HOW ARE MITOCHONDRIAL GENE DEFECTS INHERITED?

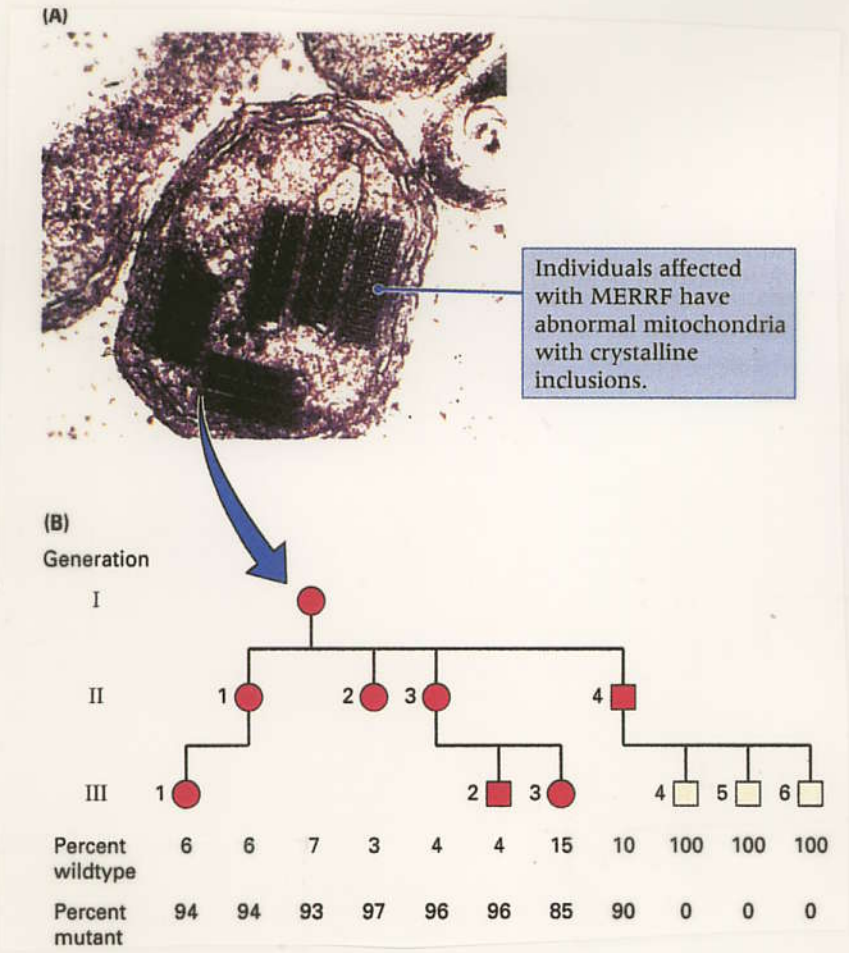


Figure 16.2 Inheritance of myoclonic epilepsy with ragged-red fiber disease (MERRF) in humans. (A) Electron micrograph of an abnormal MERRF mitochondrion containing paracrystalline inclusions. (B)The pedigree shows inheritance of MERRF in one family and the percentage of the mitochondria in each person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

RFLPs CAN BE USED TO IDENTIFY INDIVIDUALS USING MITOCHONDRIAL DNAs

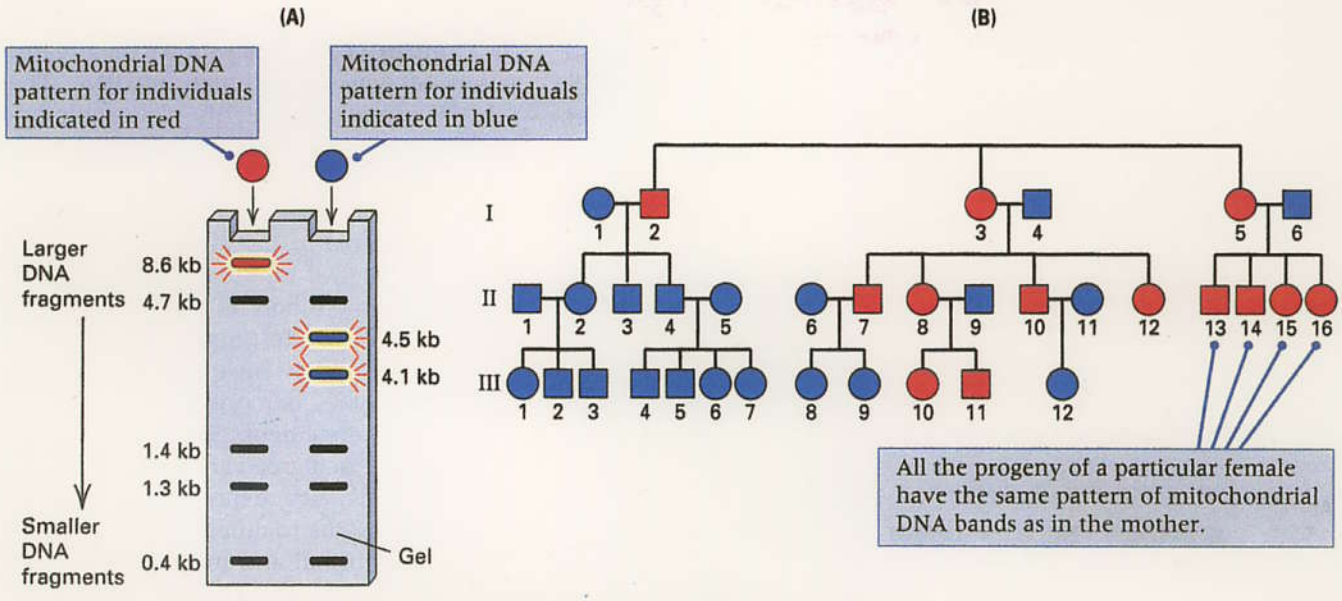


Figure 16.1 Maternal inheritance of human mitochondrial DNA. (A) Pattern of DNA fragments obtained when mitochondrial DNA is digested with the restriction enzyme *Hae*II. The DNA type at the left includes a fragment of 8.6 kb (red). The DNA type at the right contains a cleavage site for *Hae*II within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

NOTICE HOW MITOCHONDRIAL RFLP MARKERS ARE INHERITED!!

THE HUMAN GENOME HAS BEEN SEQUENCED

The Human Genome Project

ws
Print"

The New York Times

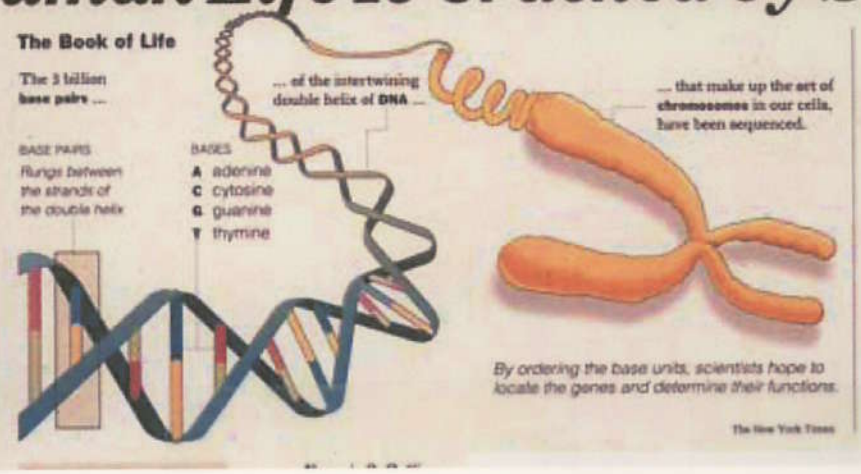
National Edition

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

. No. 51,432 Copyright © 2000 The New York Times TUESDAY, JUNE 27, 2000 Printed in Arizona ONE DOLL

tic Code of Human Life Is Cracked by Scientist

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



A SHARED SUCCE

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

By NICHOLAS WADE

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

PUBLIC & PRIVATE EFFORT USING DIFFERENT STRATEGIES - A RACE!

The public Human Genome Project cost 3 Billion dollars & took 15 years

What was the the Human Genome Project's budget?

U.S. Human Genome Project Funding ($Millions)			
FY	DOE	NIH*	U.S. Total
1988	10.7	17.2	27.9
1989	18.5	28.2	46.7
1990	27.2	59.5	86.7
1991	47.4	87.4	134.8
1992	59.4	104.8	164.2
1993	63.0	106.1	169.1
1994	63.3	127.0	190.3
1995	68.7	153.8	222.5
1996	73.9	169.3	243.2
1997	77.9	188.9	266.8
1998	85.5	218.3	303.8
1999	89.9	225.7	315.6
2000	88.9	271.7	360.6
2001	86.4	308.4	394.8
2002	90.1	346.7	434.3
2003	64.2	372.8	437
Note: These numbers do not include construction funds, which are a very small part of the budget.			

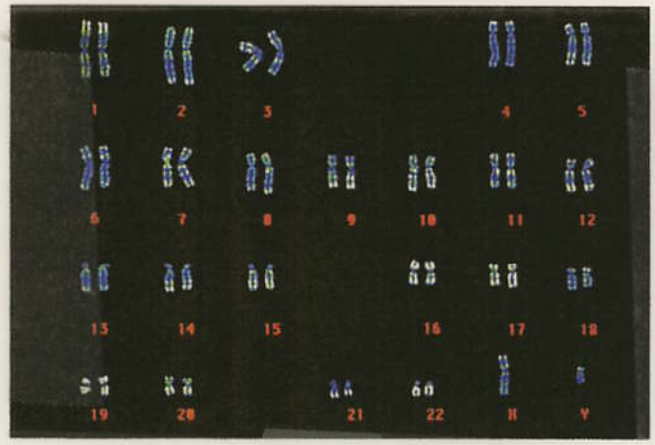
Technology development & sequencing

ONLY A SMALL FRACTION OF THE HUMAN GENOME ENCODES PROTEINS

TABLE 18.1 Classes of DNA Sequences Found in the Human Genome

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

THE HUMAN GENOME HAS DNA SEQUENCES PRESENT ONCE AS WELL AS REPEATED MANY TIMES



9.1 A human karyotype consists of 46 chromosomes. A karyotype for a male is shown here; a karyotype for a female would have two X chromosomes. *[ISM/Phototake.]*

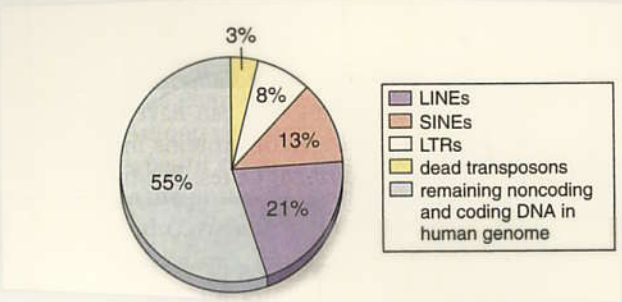


Table 20.6 Average characteristics of genes in the human genome

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

HUMAN GENES ARE LARGE BUT CONTAIN MOSTLY INTRONS

All DNA Sequences in the Human Genome Are Present Once:

a. Yes
b. No

iClicker

The HUMAN GENOME CONTAINS SINGLE-COPY AND REPETITIVE SEQUENCES Interspersed AMONG EACH other

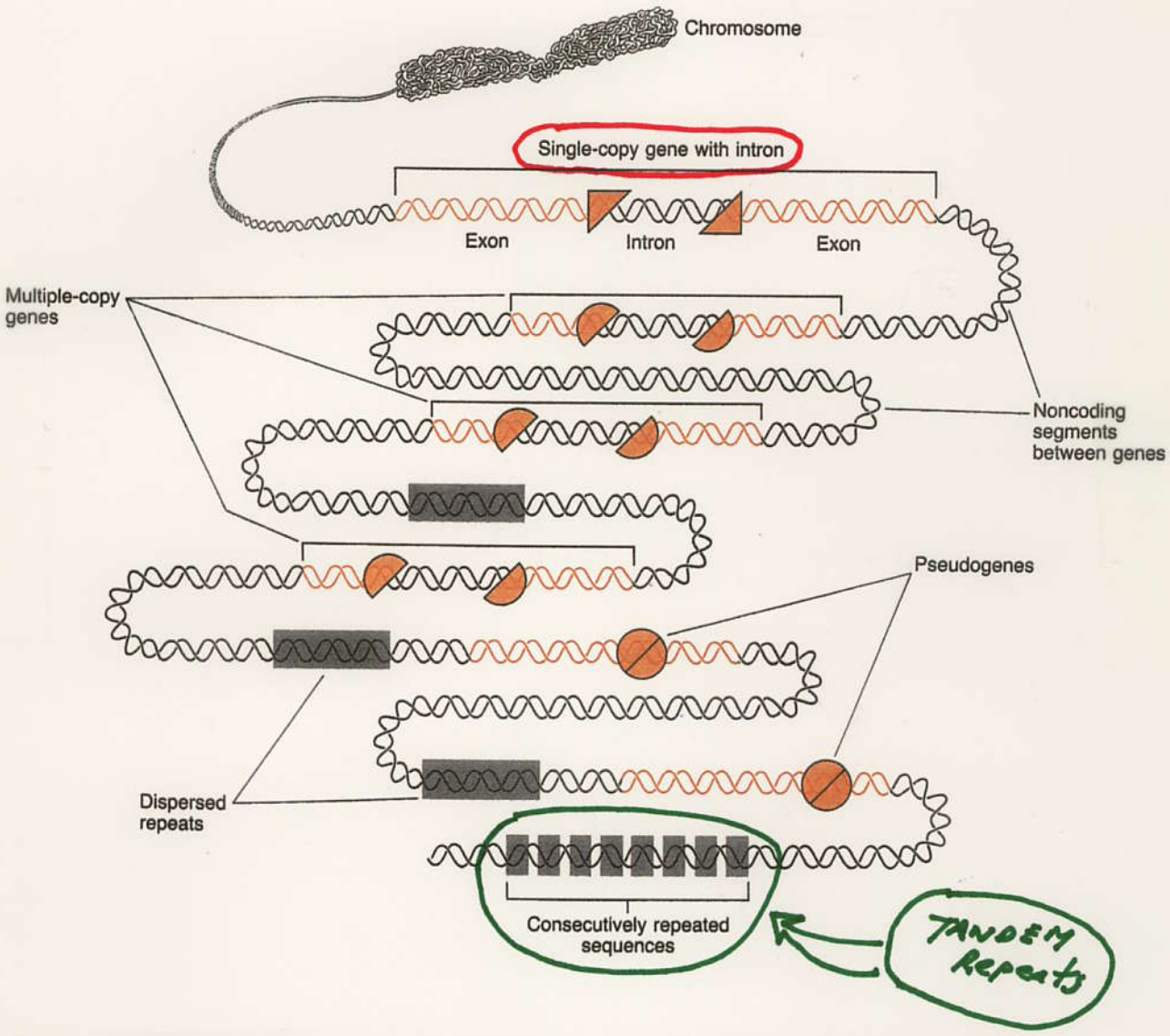


Figure 7.1 Occurrence of different kinds of unique and repeated DNA segments on chromosomal DNA.

TANDEM REPEATS ARE USEFUL FOR DNA Fingerprinting Studies!

e.g. D1S80 on Chromosome 1
core = 16 bp

The HUMAN GENOME Contains ~ 25,000 Different Genes

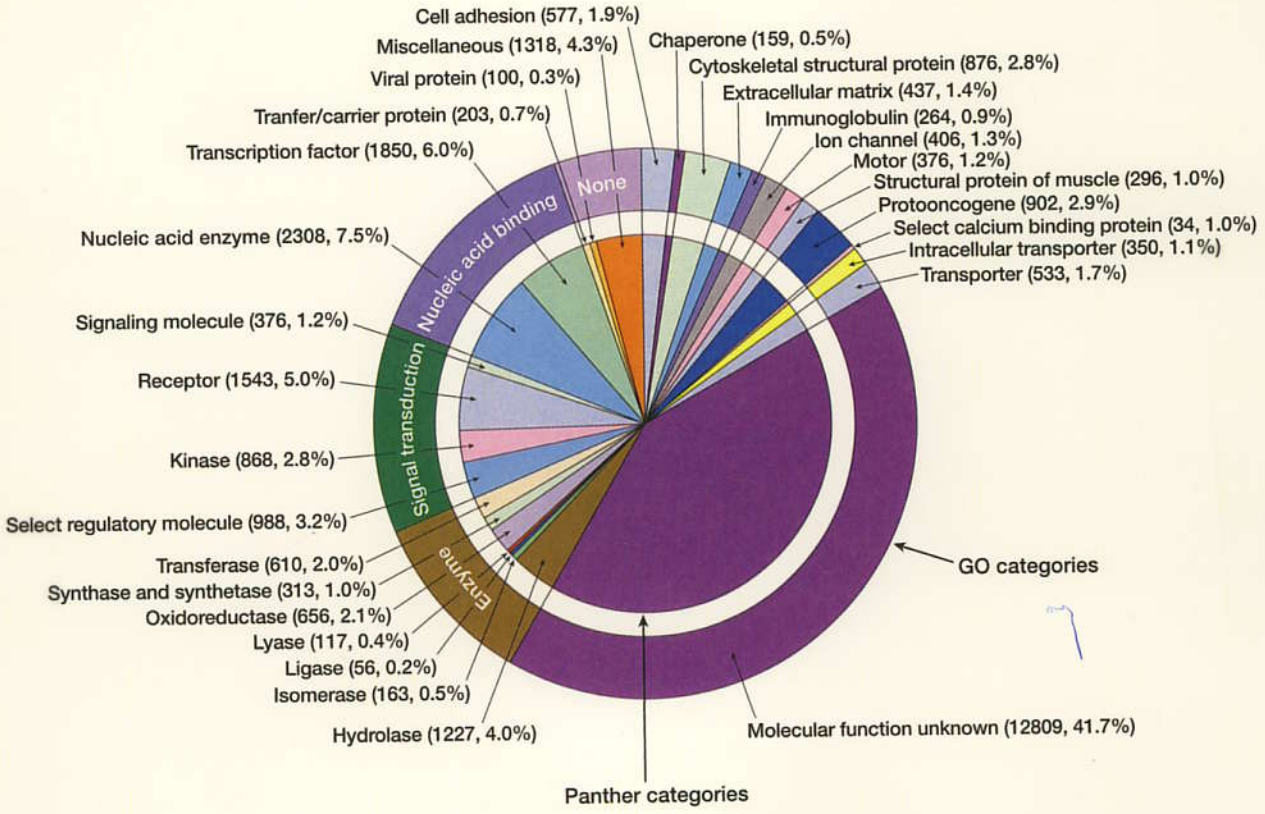


Figure 12.20: A preliminary functional classification of human polypeptide-encoding genes.

Known or predicted functions for 26 383 human polypeptide-encoding genes. Classification is according to the GO molecular function categories as shown in the outer circle (Gene Ontology classification – see Section 8.3.6) or to Celera's Panther molecular function categories (inner circle). Reproduced from Venter *et al.* (2001) *Science* **291**, 1304–1351, with permission from the American Association for the Advancement of Science.

HOW MANY Encoded Proteins?

There ARE LARGE VARIATIONS IN DNA SEQUENCES in HUMAN Populations

DNA Sequence VARIATION MAKES US INDIVIDUALS!

Genetic VARIABILITY - Allelic Differences

Different Forms of a Gene are Called Alleles:

a. Yes
b. No

iClicker

Alleles and Homologous Chromosomes

MARKER

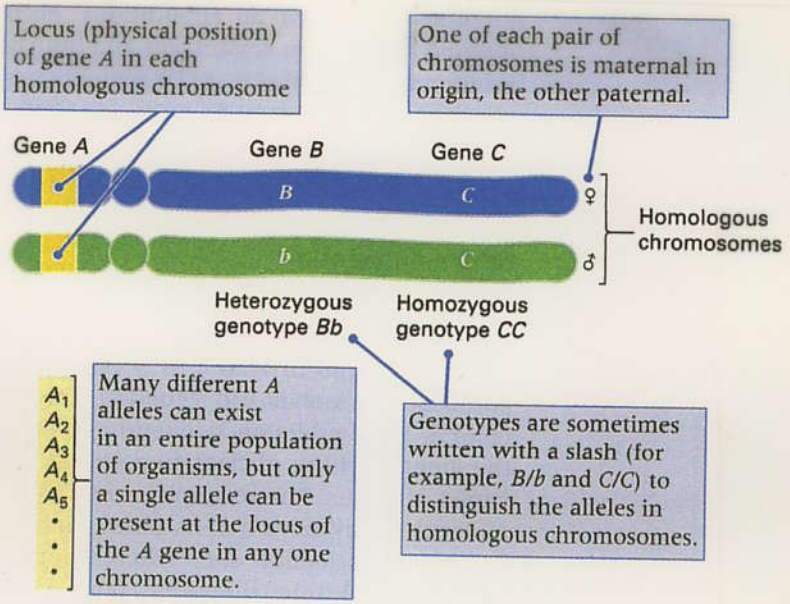


Figure 2.22 Key concepts and terms used in modern genetics. Note that a single gene can have any number of alleles in the population as a whole, but no more than two alleles can be present in any one individual.

Individuals may contain two different alleles at any DNA location

There can be an infinite # of alleles for any gene (or DNA sequence) in a population

Genetic Variability Gives Us our Individuality :

a. Yes
b. No

iClicker

MOST DNA VARIATION BETWEEN INDIVIDUALS OCCURS BECAUSE of BASE-PAIR CHANGES in NON-CODING REGIONS of the GENOME

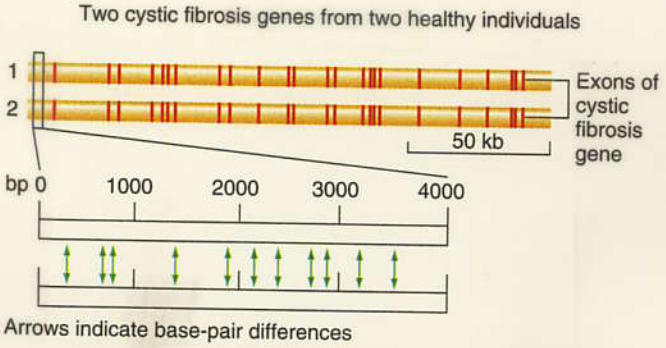


Figure 9.2 Base-pair differences between DNA cloned from the cystic fibrosis locus of two healthy individuals. These base-pair differences have no phenotypic effect; apparently they neither encode nor regulate expressed regions of the gene.

SNPs or Single Nucleotide Polymorphisms

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

$(3 \times 10^9) \times 0.8 = 2.4 \times 10^9$, $(2.4 \times 10^9) \times \frac{1}{700} = 3.4$ million.

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

This is what makes us unique individuals

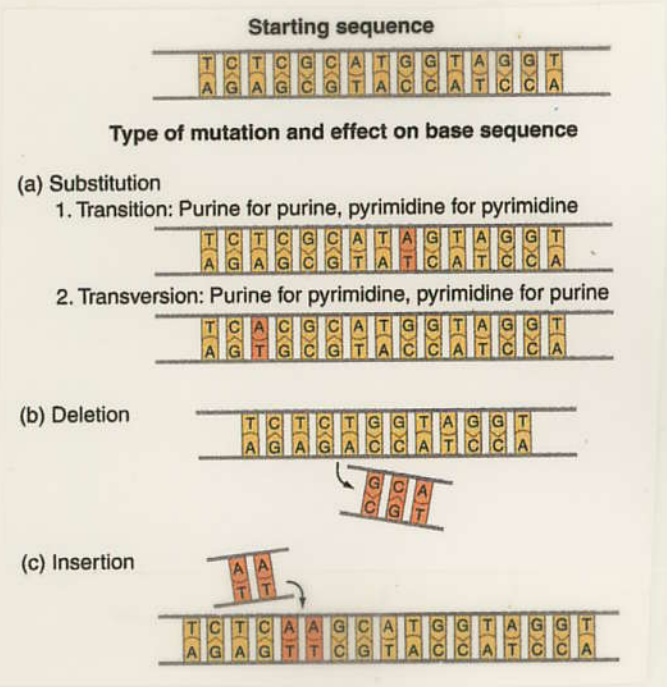
There is ~ 1bp change / 700 bp in Human Genomes or ~ 3.4 Million bp DIFFERENCES Between Individuals ~ 0.1% of Genome

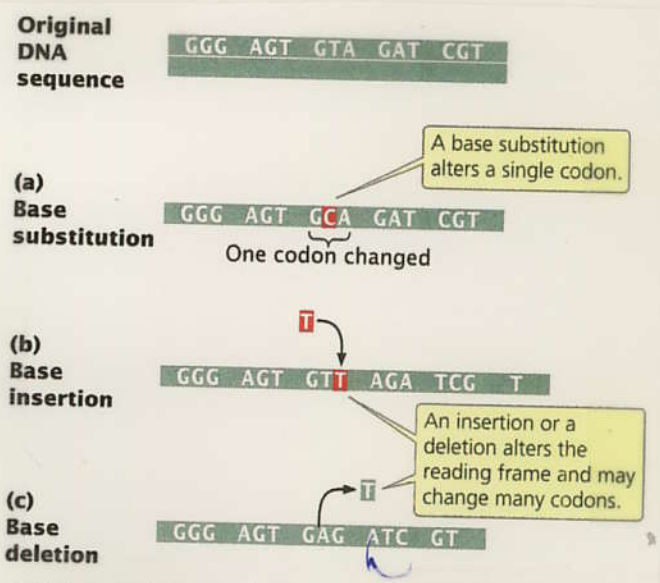
Mutation in DNA Occur During: ?

a. Transcription
b. Translation
c. Replication
d. Alternative Splicing

iClicker

How do SNPs arise in the Human Genome during DNA Replication?





18.2 Three basic types of gene mutations are base substitutions, insertions, and deletions.

Most SNPs are single Nucleotide changes that Have <u>No</u> Effect on the Phenotype or Gene Function!

Different "Forms" of the Same SNP = Allele!

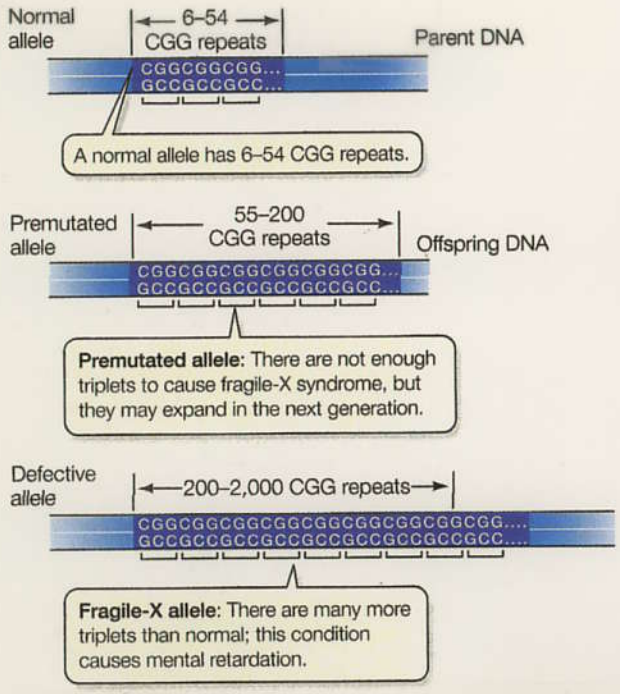
DNA Sequence Change in the Genome Are RARE

TABLE 9.1 Five Classes of DNA Polymorphism

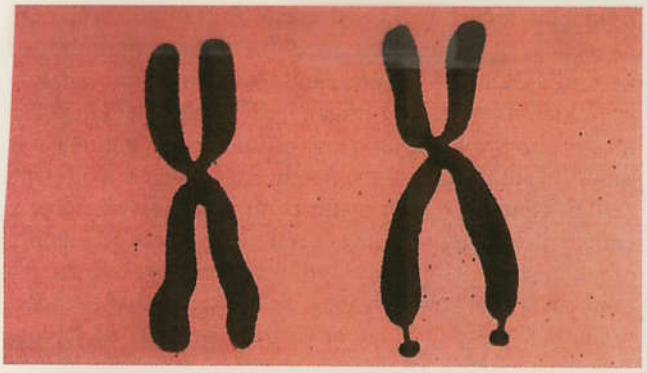
Class	Cause	Rate of Mutation per Locus per Gamete	Frequency in Genome	Number per Human Genome (on average)
Single base (SNP)	Mutagens or replication errors	10^{-8}–10^{-9}	1/700 bp	3 million
Microsatellite (VNTR or SSR)	Slippage during replication	10^{-3}	1/30,000 bp	100,000
Minisatellite	Unequal crossovers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

ONLY A few Affect Gene function & lead to a Visible Mutation!

Changes in Trinucleotide Repeats within Some Genes Can Lead to Serious Human Diseases



17.10 The CGG Repeats in the *FMR1* Gene Expand with Each Generation The genetic defect in fragile-X syndrome is caused by 200 or more repeats of the CGG triplet.



18.4 The fragile-X chromosome is associated with a characteristic constriction (fragile site) on the long arm. *[Visuals Unlimited.]*

Repeat in coding sequence!

Table 18.1 Examples of genetic diseases caused by expanding trinucleotide repeats

Disease	Repeated Sequence	Number of Copies of Repeat	
		Normal Range	Disease Range
Spinal and bulbar muscular atrophy	CAG	11–33	40–62
Fragile-X syndrome	CGG	6–54	50–1500
Jacobsen syndrome	CGG	11	100–1000
Spinocerebellar ataxia (several types)	CAG	4–44	21–130
Autosomal dominant cerebellar ataxia	CAG	7–19	37–220
Myotonic dystrophy	CTG	5–37	44–3000
Huntington disease	CAG	9–37	37–121
Friedreich ataxia	GAA	6–29	200–900
Dentatorubral-pallidoluysian atrophy	CAG	7–25	49–75
Myoclonus epilepsy of the Unverricht-Lundborg type*	CCCCGCCCCGCG	2–3	12–13

TRIPLET REPEATS in Genes Increases FROM Generation to Generation Leading to More Severe Forms OF the Disease

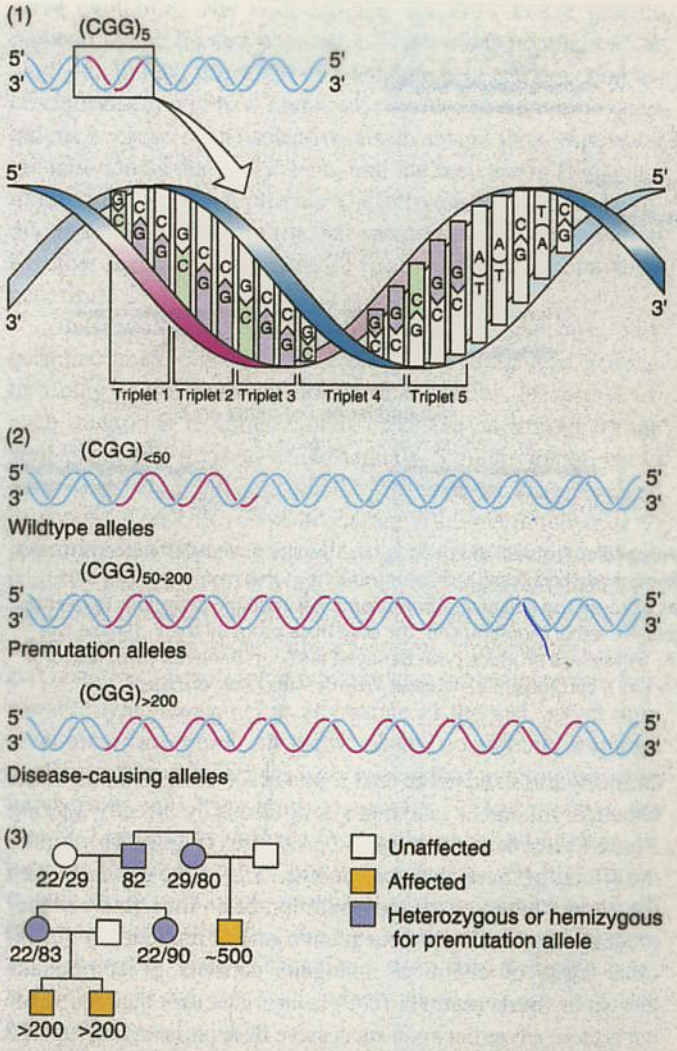


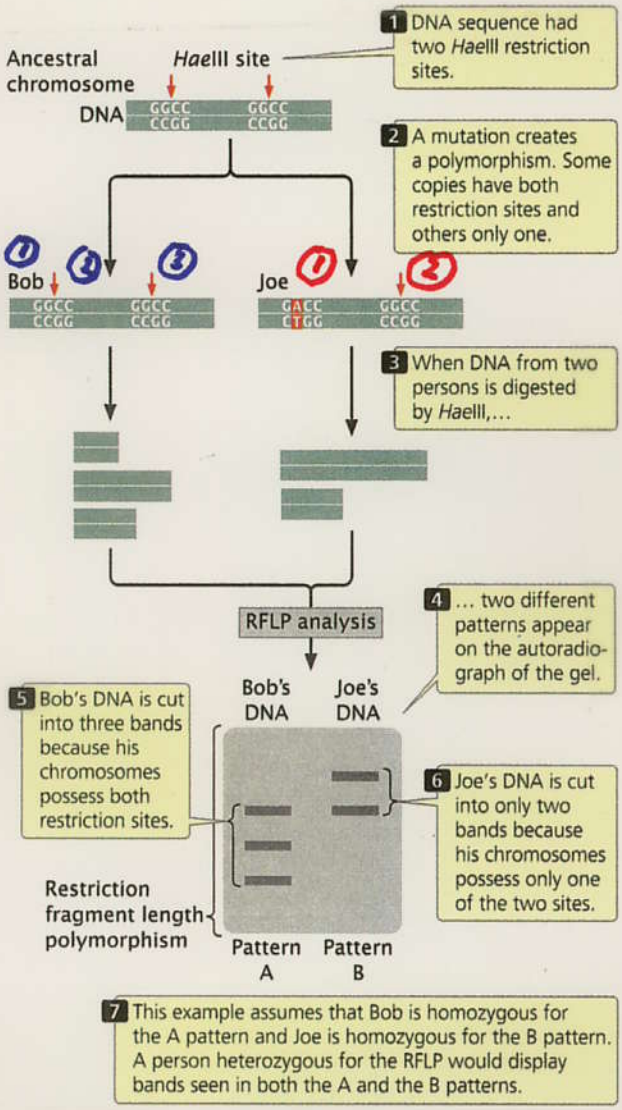
Figure B Amplification of the triplet repeat CGG correlates with the fragile X syndrome. (1) Region of an X chromosome containing a normal *FMR-1* gene with 5 repeats of the CGG sequence on one strand. (2) *FMR-1* genes in unaffected people generally have fewer than 50 repeats; unstable premutation alleles of the gene have between 50 and 200 copies of the repeat. Full-blown disease-causing alleles have more than 200 CGG repeats; some mutant alleles have more than 4000. (3) A fragile X pedigree showing the number of CGC repeats in different individuals. Note that individuals affected by fragile X syndrome are almost always the progeny of mothers who carried premutation alleles.

"Anticiptin" of Disease

How detect on A VARIATION in Individuals?

RFLPs Identify Individuals in A Population

RFLP at SAME Locus in Each Individual



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

Mitochondrial RFLPs can be used to identify Individuals Too!

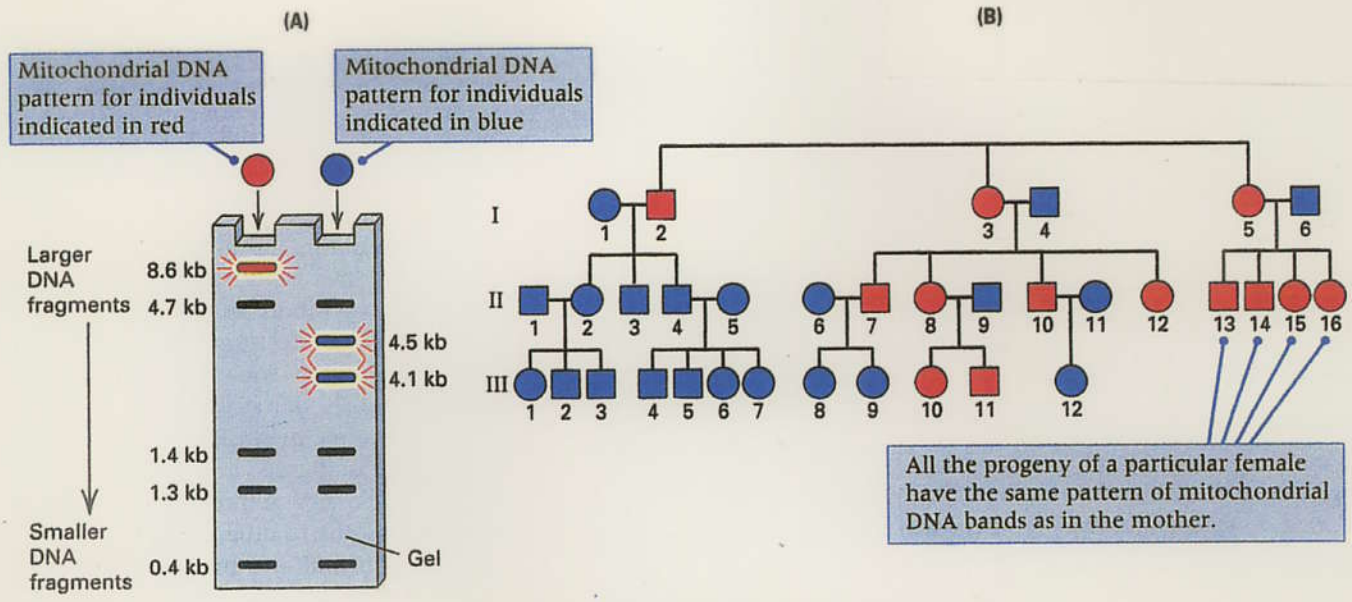


Figure 16.1 Maternal inheritance of human mitochondrial DNA. (A) Pattern of DNA fragments obtained when mitochondrial DNA is digested with the restriction enzyme *Hae*II. The DNA type at the left includes a fragment of 8.6 kb (red). The DNA type at the right contains a cleavage site for *Hae*II within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

Mitochondrial MARKERS FOLLOW The Maternal "line" BACK to Eve!

PCR CAN BE USED TO Identify RFLPs AND DNA VARIation

Two fragments

One Large Fragment

1. PCR
2. Restriction Digest
3. Gel

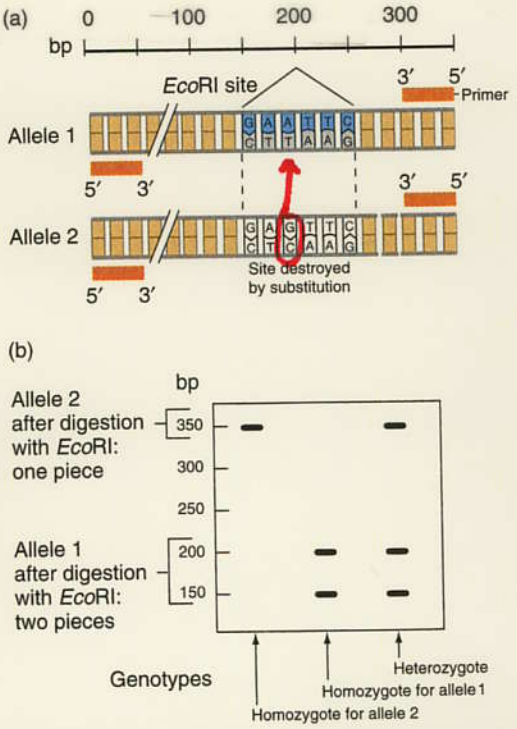


Figure 9.7 Restriction site polymorphisms can be detected most efficiently with PCR-based protocols. (a) PCR amplification of two alleles of a DNA locus with a restriction site polymorphism. Allele 1 has an *Eco*RI site that is eliminated in allele 2. The PCR products amplified from both alleles are identical in size. (b) Exposure of these PCR products to *Eco*RI causes cleavage of the allele 1 product but not the allele 2 product. Gel electrophoresis and ethidium bromide staining distinguish the three genotypes possible with the two alleles at this locus.

VNTRs, STRs, SSRs can be assayed using PCR

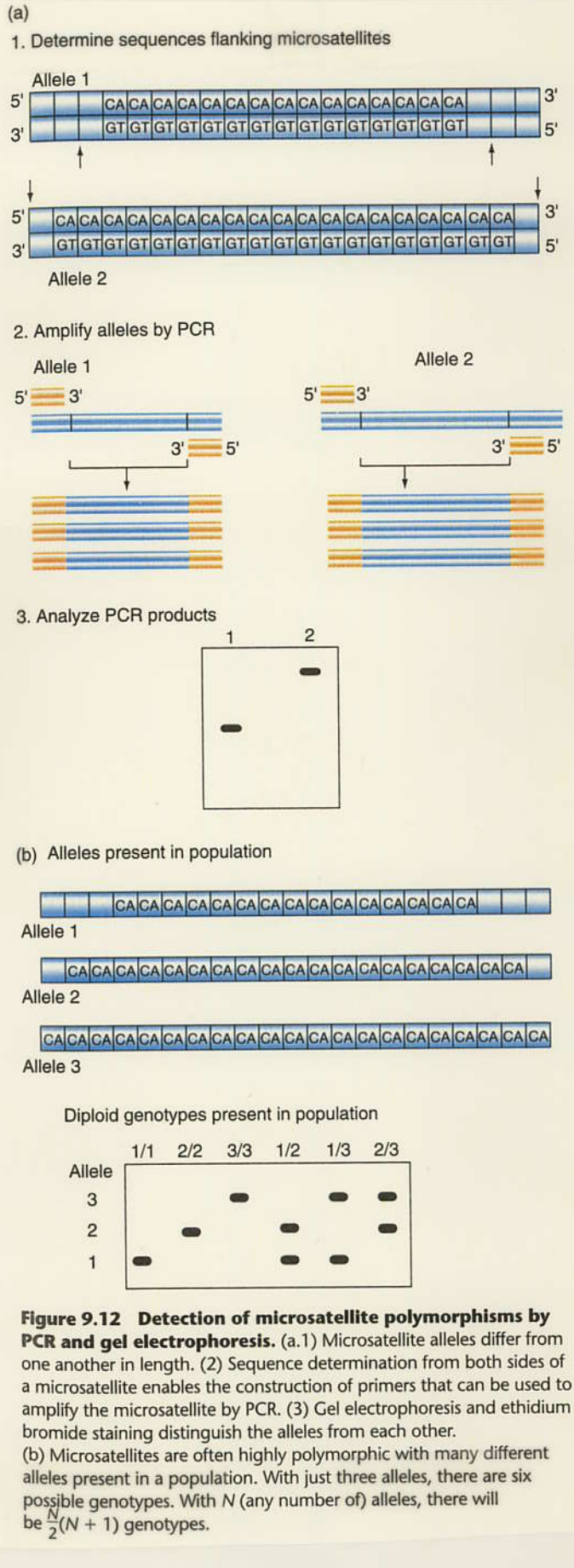


Figure 9.12 Detection of microsatellite polymorphisms by PCR and gel electrophoresis. (a.1) Microsatellite alleles differ from one another in length. (2) Sequence determination from both sides of a microsatellite enables the construction of primers that can be used to amplify the microsatellite by PCR. (3) Gel electrophoresis and ethidium bromide staining distinguish the alleles from each other.
(b) Microsatellites are often highly polymorphic with many different alleles present in a population. With just three alleles, there are six possible genotypes. With N (any number of) alleles, there will be $\frac{N}{2}(N + 1)$ genotypes.

15 Copies

19 Copies

neighbor sequences the same

Size Differences on gel

Because they vary by length in individuals

USED TO IDENTIFY individuals

Using VNTRs at one locus to Identify Individuals

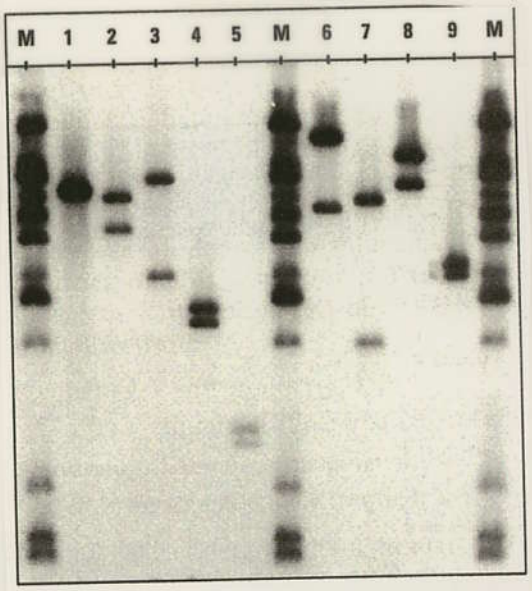
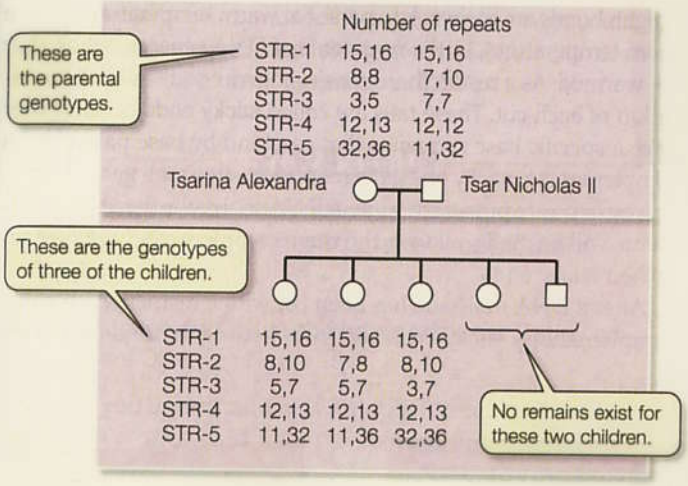


Figure 17.12 Genetic variation in a VNTR used in DNA typing. Each numbered lane contains DNA from a single person; the DNA has been cleaved with a restriction enzyme, separated by electrophoresis, and hybridized with radioactive probe DNA. The lanes labeled M contain molecular-weight markers. [Courtesy of R. W. Allen.]

VNTRs Used to Verify Remains of Russian Royal Family



Identify Remains

SEVERAL UNTR (STR) LOCI CAN BE USED AT ONCE OR SEPARATELY TO MAKE IDENTIFICATION STRONGER!

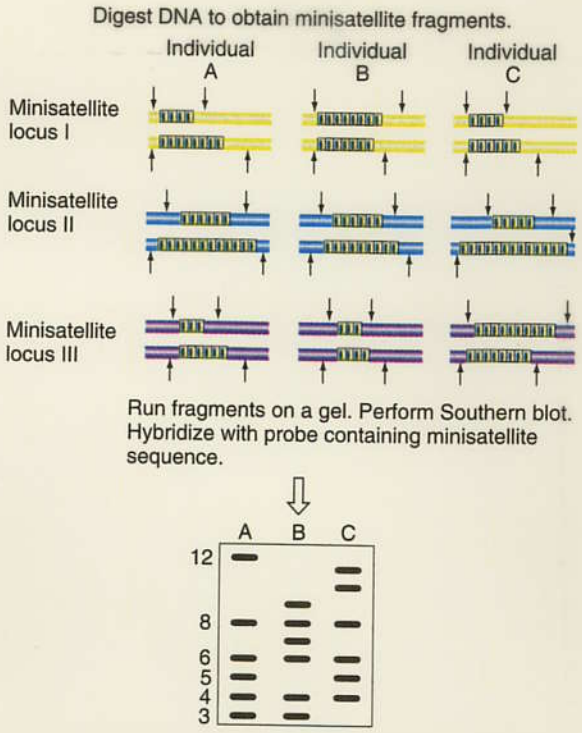
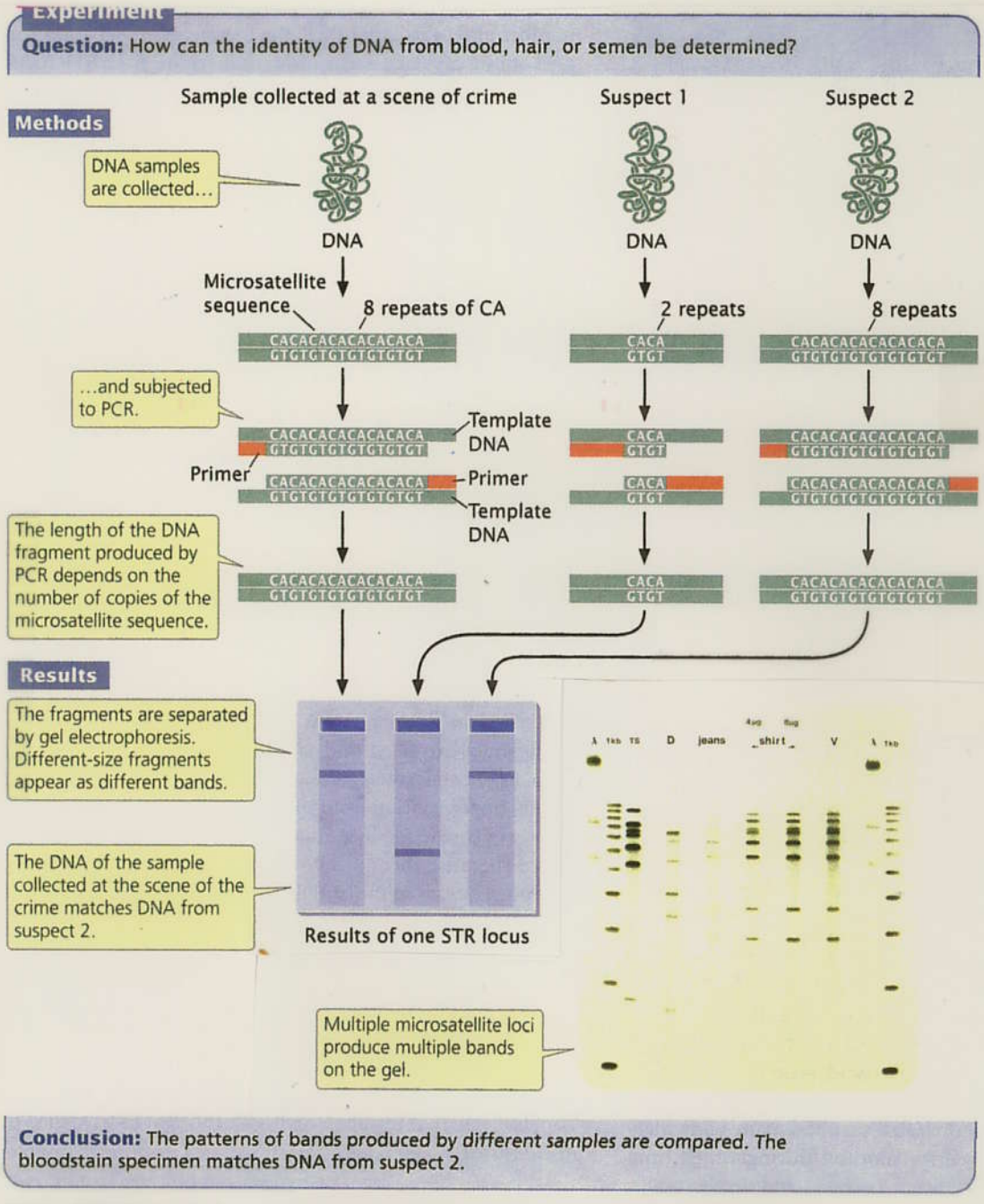


Figure 9.13 Minisatellite analysis provides a means of simultaneously detecting polymorphisms at multiple loci. A hypothetical minisatellite sequence is present at three genomic loci in three individuals. Each of the loci can be polymorphic, with different alleles defined by different numbers of repeating units. The three individuals are all heterozygous at each locus. To detect the minisatellite polymorphisms, you digest the genomic DNA from each individual with a restriction enzyme that cuts outside the minisatellite region. You then separate the digested DNA by gel electrophoresis, make a blot, and probe the blot with the minisatellite sequence. In this example, each individual produces a different pattern of six gel bands. The purpose of this protocol is not to assign a specific band to a specific locus, but rather to obtain a whole genome fingerprint based on simultaneously detected polymorphisms at multiple loci.

i.e. - greater probability that individual is unique!

HOW MANY FRAGMENTS per PERSON?
EACH FRAGMENT length POLYMORPHISM = ALLELE

Using Short TANDem Repeats to Identify Individuals Who Committed a Crime

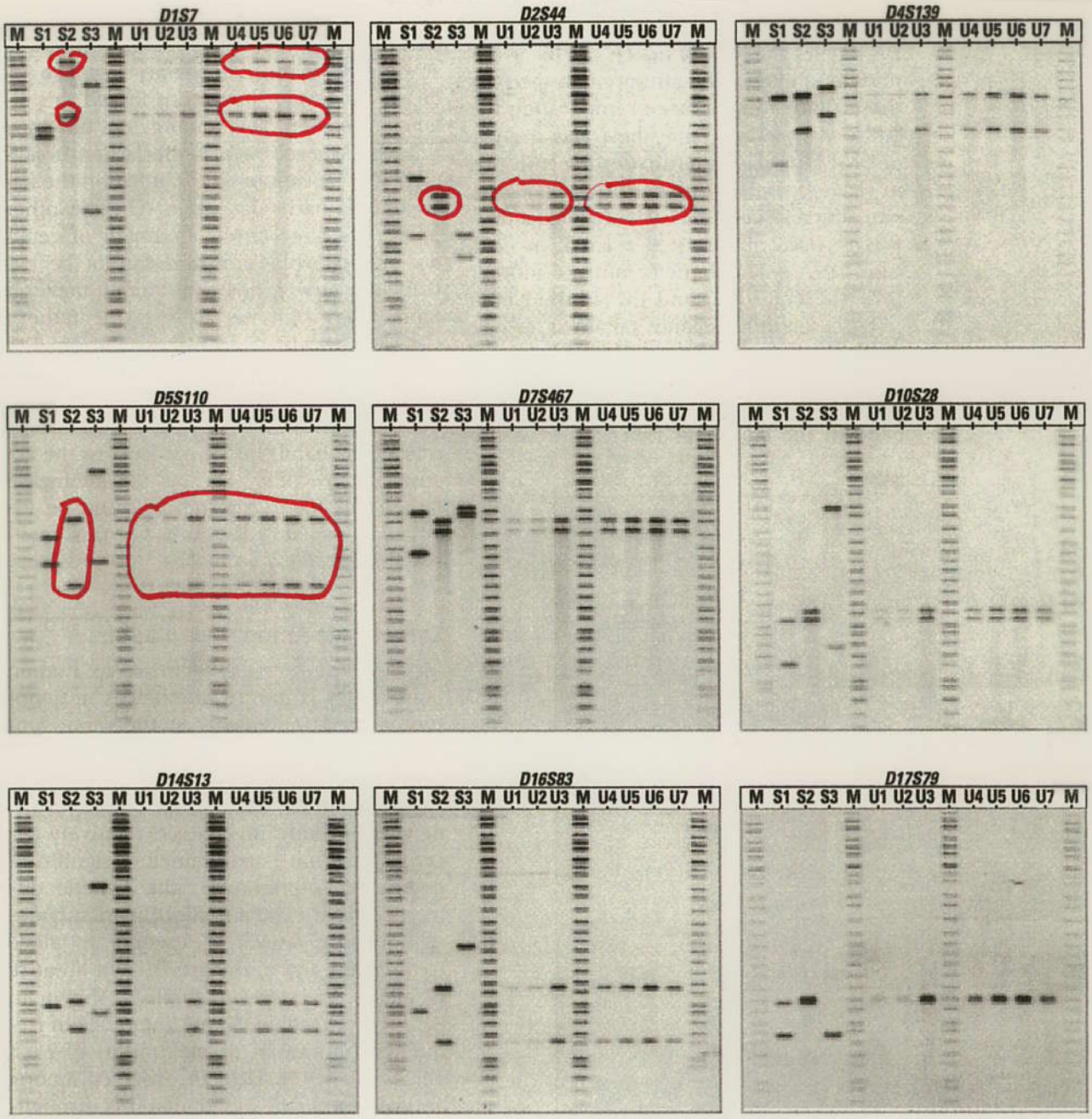


19.30 DNA fingerprinting can be used to identify a person. *[Gel courtesy of Orchard Cellmark, Germantown, Maryland.]*

Very Sensitive Using PCR! One Cell!

Recall - Must have DNA Sequence from a Recombinant Clone to do this!

Using VNTRs to Identify the Suspect who committed the Crime



Any Doubt?

Figure 17.13 An example of DNA typing in a criminal case. Each panel is the result of DNA typing for a different VNTR. The lanes marked S1, S2, and S3 contain DNA from blood samples of three male suspects; those in columns U1 through U7 contain DNA from semen samples collected from seven female victims of rape. The lanes marked M contain molecular-weight markers. In each case, the DNA from suspect S2 matches the samples obtained from the victims. [Courtesy of Steven J. Redding, Office of the Hennepin County District Attorney, Minneapolis, and Lowell C. Van Berkom and Carla J. Finis, Minnesota Bureau of Criminal Apprehension.]

Nine VNTR Loci used – from Different Genome Regions!

Identifying DNA VARIATION BETWEEN Individuals Has MANY USES

Epidemiology and food safety science. DNA typing also has important applications in tracking the spread of viral and bacterial epidemic diseases, as well as in identifying the source of contamination in contaminated foods.

Human population history. DNA polymorphisms are widely used in anthropology to reconstruct the evolutionary origin, global expansion, and diversification of the human population.

Improvement of domesticated plants and animals. Plant and animal breeders have turned to DNA polymorphisms as genetic markers in pedigree studies to identify, by genetic mapping, genes that are associated with favorable traits in order to incorporate these genes into currently used varieties of plants and breeds of animals.

History of domestication. Plant and animal breeders also study genetic polymorphisms to identify the wild ancestors of cultivated plants and domesticated animals, as well as to infer the practices of artificial selection that led to genetic changes in these species during domestication.

DNA polymorphisms as ecological indicators. DNA polymorphisms are being evaluated as biological indicators of genetic diversity in key indicator species present in biological communities exposed to chemical, biological, or physical stress. They are also used to monitor genetic diversity in endangered species and species bred in captivity.

Evolutionary genetics. DNA polymorphisms are studied in an effort to describe the patterns in which different types of genetic variation occur throughout the genome, to infer the evolutionary mechanisms by which genetic variation is maintained, and to illuminate the processes by which genetic polymorphisms within species become transformed into genetic differences between species.

William H. Mullins/Photo Researchers, Inc.

A paternity test–for wines. Genetic researchers have shown that the notion of "purity of breed" thought to be so critical to the commercial value of certain wines is ill-founded. DNA tests indicate that wine grapes such as the cabernet sauvignon and chardonnay (shown here) are in fact hybrids.

(+) Forensics
Wildlife Identification (Poachers)
Breeding
paternity Clone Identification etc.

HOW USE RFLPs TO IDENTIFY specific Disease Genes?

RFLPs or DNA Markers can be used to follow/identify Gene Alleles it linked

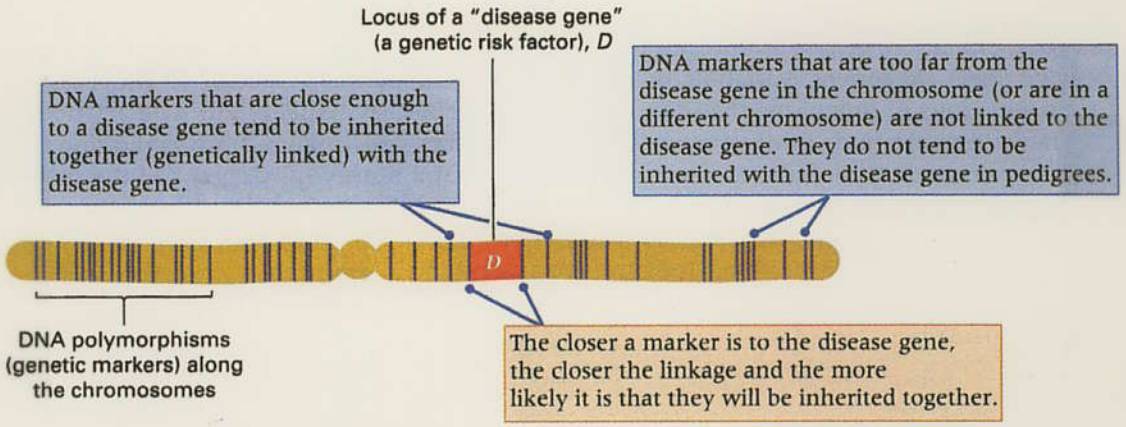


Figure 2.29 Concepts in genetic localization of genetic risk factors for disease. Polymorphic DNA markers (indicated by the vertical lines) that are close to a genetic risk factor (*D*) in the chromosome tend to be inherited together with the disease itself. The genomic location of the risk factor is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

MARKERS FOR WRINKLED PEAS

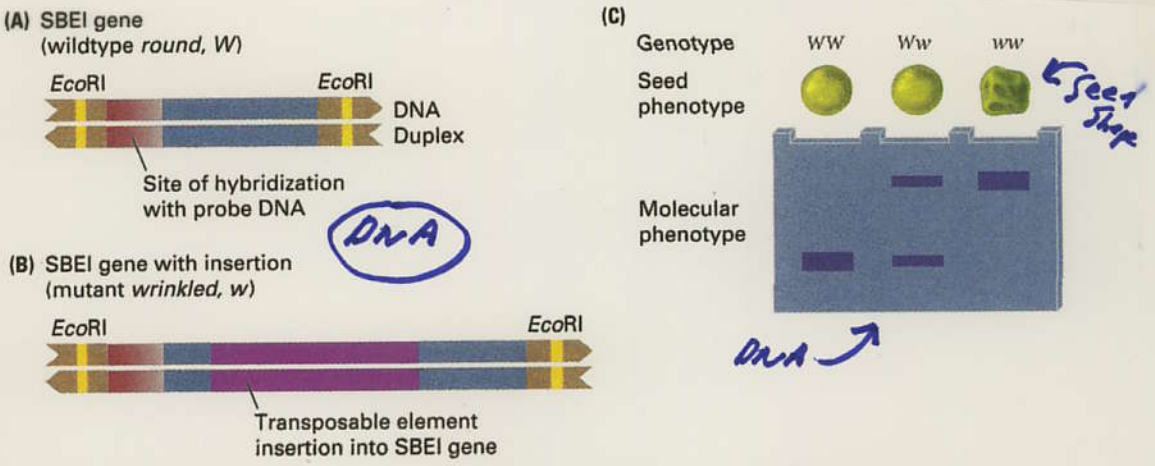


Figure 3.2 (A) *W* (round) is an allele of a gene that specifies the amino acid sequence of starch branching enzyme I (SBEI). (B) *w* (wrinkled) is an allele that encodes an inactive form of the enzyme because its DNA sequence is interrupted by the insertion of a transposable element. (C) At the level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

Useful for DNA Testing + Genetic Diagnosis!

USING RFLPs & MARKERS TO IDENTIFY THE Sickle Cell. ALLele

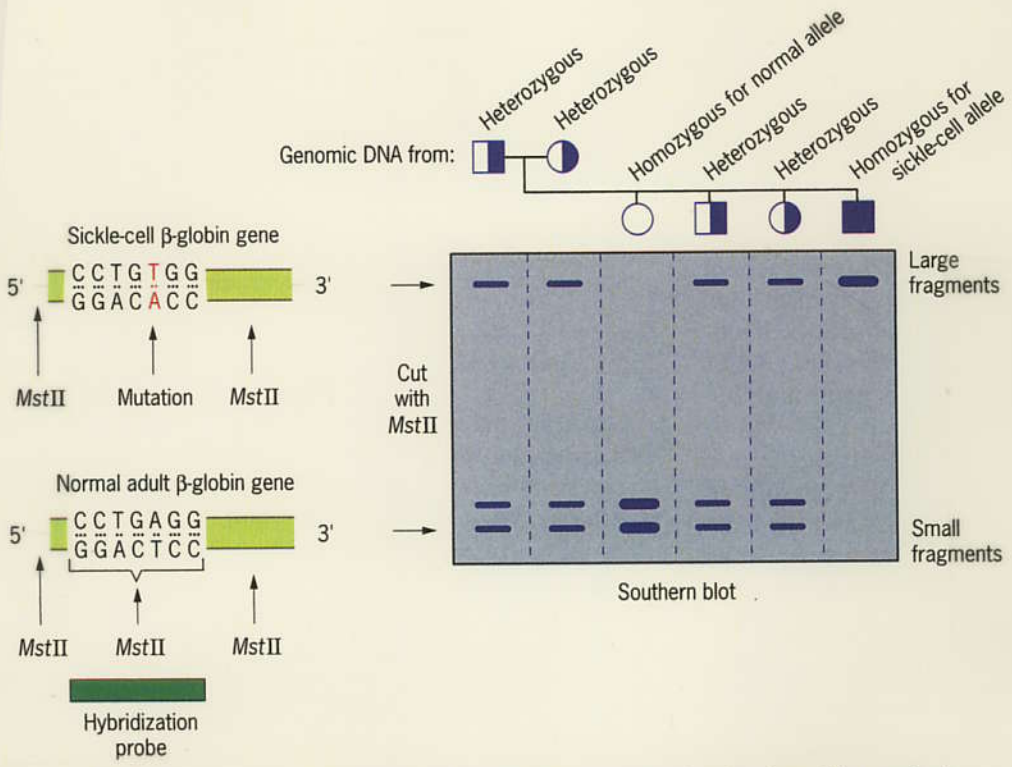


Figure 22.7 Detection of the sickle-cell hemoglobin mutation by Southern blot analysis of genomic DNAs cut with restriction enzyme *Mst*II.

Using PCR & RFLPs to Identify Hemophilia A Alleles

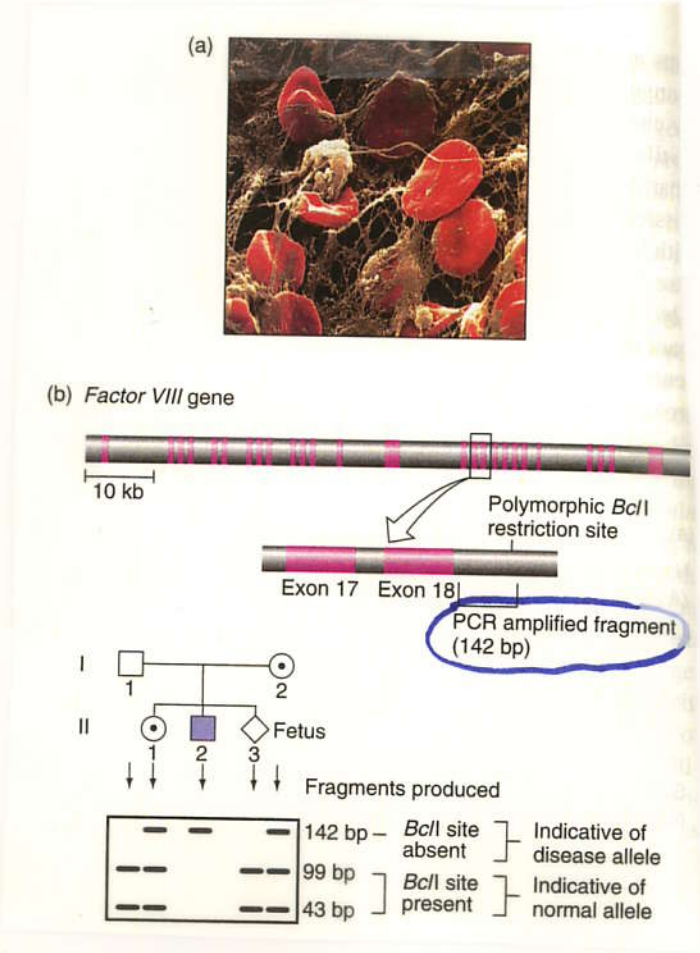


Figure 9.18 Diagnosis of hemophilia through the indirect detection of genotype at the factor VIII locus. The factor VIII protein participates in a cascade of reactions that result in formation of a blood clot. (a) A polymorphic *Bcl*I restriction site within intron 18 of the *factor VIII* gene has no effect on gene function but can provide a marker to follow the segregation of the gene from parents to children. (b) The family described by the pedigree has two healthy parents, but the mother is an obligate carrier of the disease mutation because she has passed this X-linked disease on to her son; her carrier status is signified by a circle with a dot in the middle. By comparing the RFLP pattern obtained from the mother's DNA with the pattern from her son's DNA, you can see that the disease allele is associated with the 142-bp *Bcl*I restriction fragment, and the wild-type allele in the mother's genome contains a *Bcl*I restriction site that causes this fragment to be cut into two pieces, one 43 bp and the other 99 bp in length. Using this information, you can determine that the firstborn sister is a carrier like her mother, while the male fetus will be disease free.

Using Allele-Specific Oligonucleotides (ASOs) to Identify Specific Alleles of Disease Genes

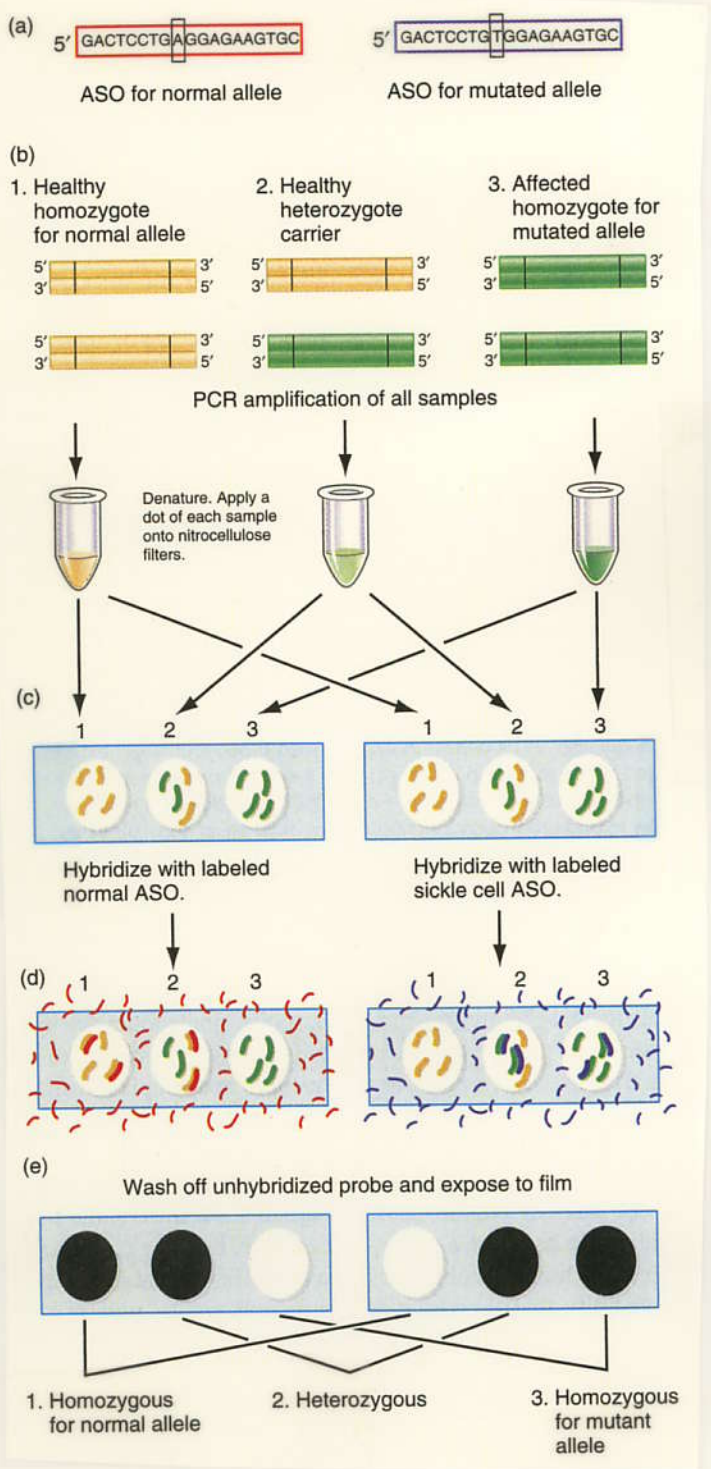


Figure 9.9 Using PCR with ASOs to determine genotype at the β-globin locus. (a) Before performing the genotyping protocol, it is necessary to synthesize two oligonucleotides that differ at only a single base; one of these oligonucleotides is complementary to the wild-type β-globin allele, the other is complementary to the sickle-cell allele. These two synthetic DNA molecules serve as the ASOs for the sickle-cell genotype assay. (b) Genomic DNA samples obtained from individual people are subjected to PCR amplification with primers complementary to nonpolymorphic sequences that flank the base that mutates to cause sickle-cell anemia. (c) The amplified sample from each individual is divided into two aliquots that are blotted directly to filter paper. (d) One aliquot from each sample is hybridized to the wild-type ASO; the other aliquot is hybridized to the sickle-cell ASO. (e) Autoradiography indicates the β-globin genotype of each individual.

Using ASOs to Identify Cystic Fibrosis Genes With PCR

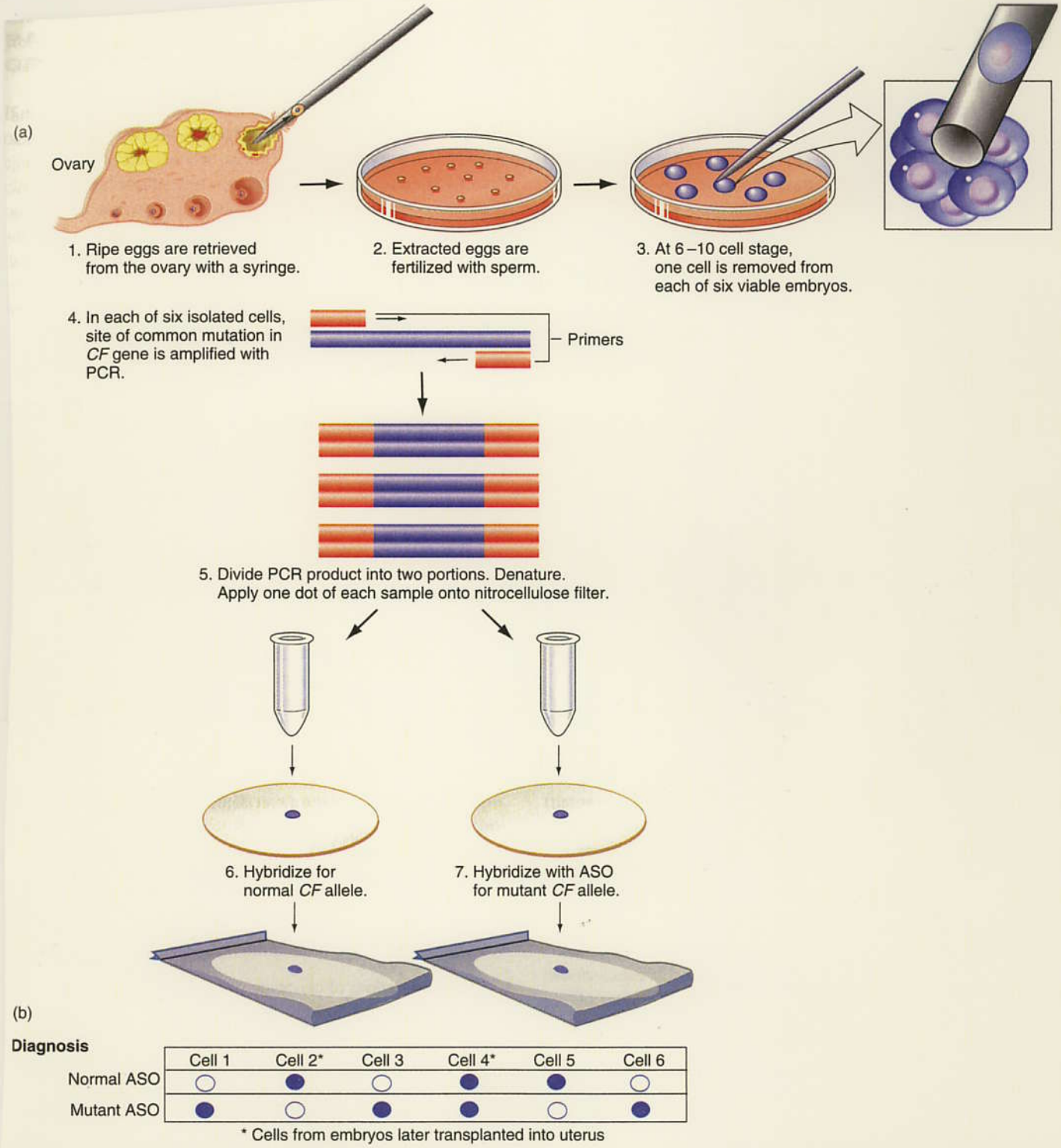


Figure 9.1 Detecting the cystic fibrosis genotype of embryonic cells. (a) *In vitro* fertilization and preimplantation diagnosis. (b) Cell 2 is homozygous for the normal allele; cell 4 is heterozygous for the *CF* mutation.

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

a. Yes
b. No

iClicker

DNA Testing Results Should Be Made Waidely Available?

a. Yes
b. No

iClicker

Identify SNPs in Human Genome.

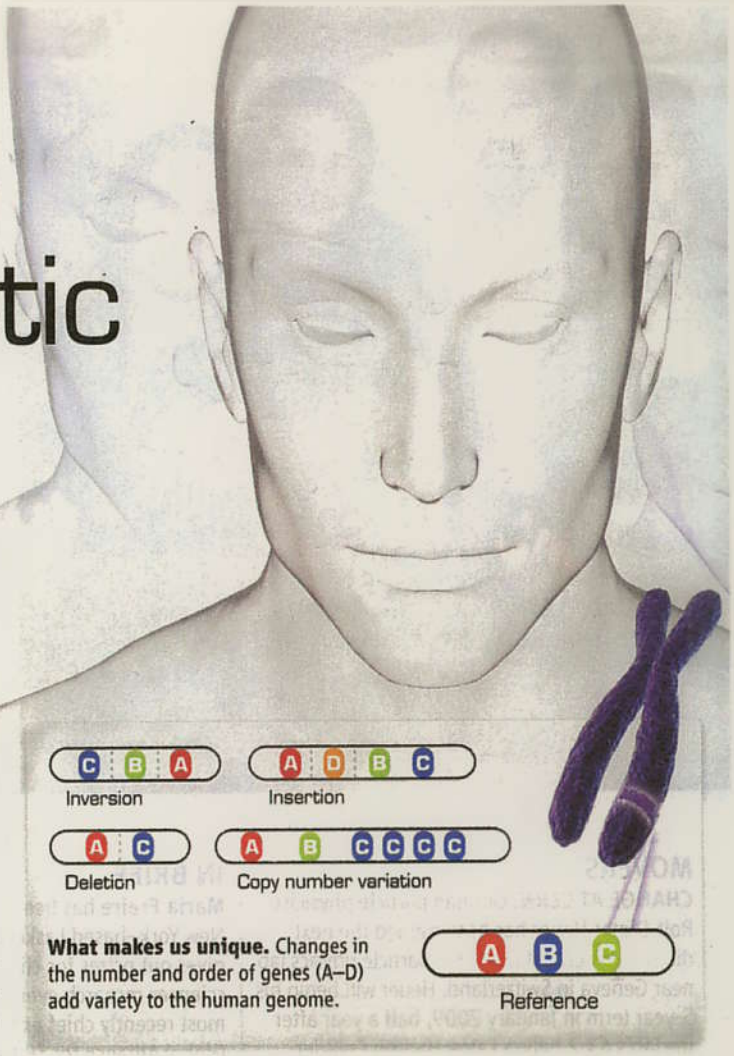
BREAKTHROUGH OF THE YEAR

Human Genetic Variation

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

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

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



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

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.

SNPs BECOME PERSONAL!

It's All About Me

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

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

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

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

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

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

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

–JOCELYN KAISER

OCA2

From SNPedia

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

is a	gene
is	mentioned by
wikipedia	OCA2 (http://en.wikipedia.org/wiki/OCA2)
google	OCA2 (http://www.google.com/search?hl=en&q=OCA2)
GeneRIF	4948 (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=4948&ordinalpo
dbSNP	4948 (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=4948&chooseRs=all)
PubMed	4948 (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=Link&LinkName=gene_pubmed&from_uid=4948)

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

- rs7495174
- rs6497268
- rs11855019

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

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

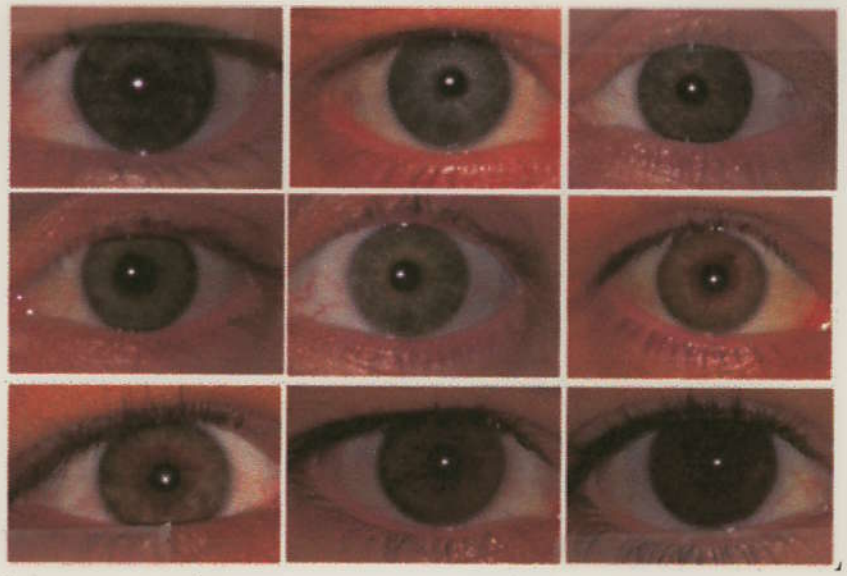
The haplotypes shown in ***bold italics*** represent the ones reported by the authors of this study to be most associated with brown eye color. Furthermore, the haplotypes shown above are as published, and the associated SNPs - which have since changed # as well - are not in the orientation shown in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

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

- rs12913832
- rs1129038

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

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



human eye color.

[replay animation]

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

1866: Gregor Mendel discovers the laws of inheritance.

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?

42

SNP Chips can scan SNPs across the Genome

Spit Kit

Joining 23andMe is easy - once you've placed your order and signed our online consent form, all you do is spit in a plastic tube included with the kit we ship to you. Each kit is labeled with the name of the person it is designated for along with a claim code. Just use the kit's pre-paid, pre-addressed shipping envelope to send your sample to our contracted laboratory.

Genotyping

After receiving your sample, lab professionals extract DNA from cells in your saliva. Your DNA is then chopped up into shorter strands and copied many times via a process called amplification. Next, your DNA is washed over a small microchip-like device that contains short strands of synthetic DNA. The synthetic DNA fragments latch onto the pieces of your DNA that are a complementary match. Then a laser-scanning step reveals which strands of synthetic DNA are stuck to your DNA to determine your genotype. The chip used in our process is the Illumina HumanHap550+ BeadChip, which reads more than 550,000 SNPs (single nucleotide polymorphisms) plus a 23andMe custom-designed set that analyzes more than 30,000 additional SNPs. What this means is that the laboratory process reads nearly 600,000 data points on your genome. Find out more about our genotyping process.

Whole Genome SNP Chips

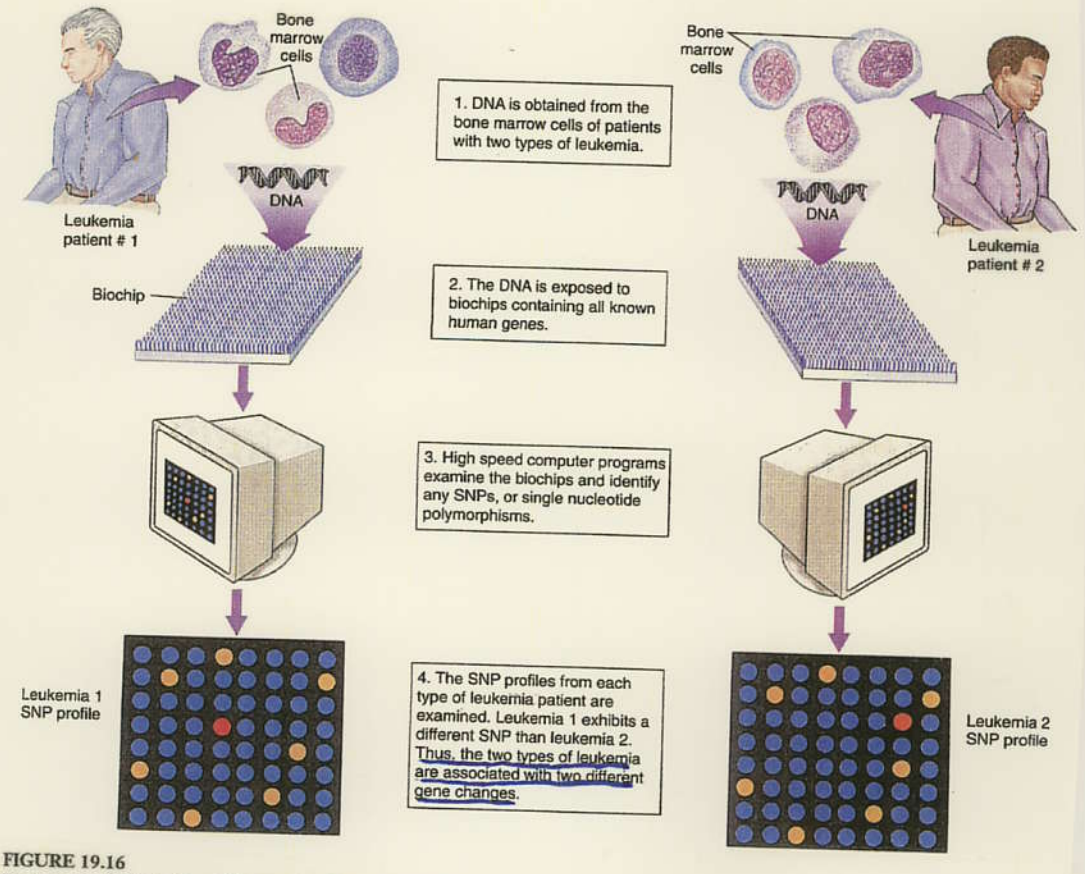


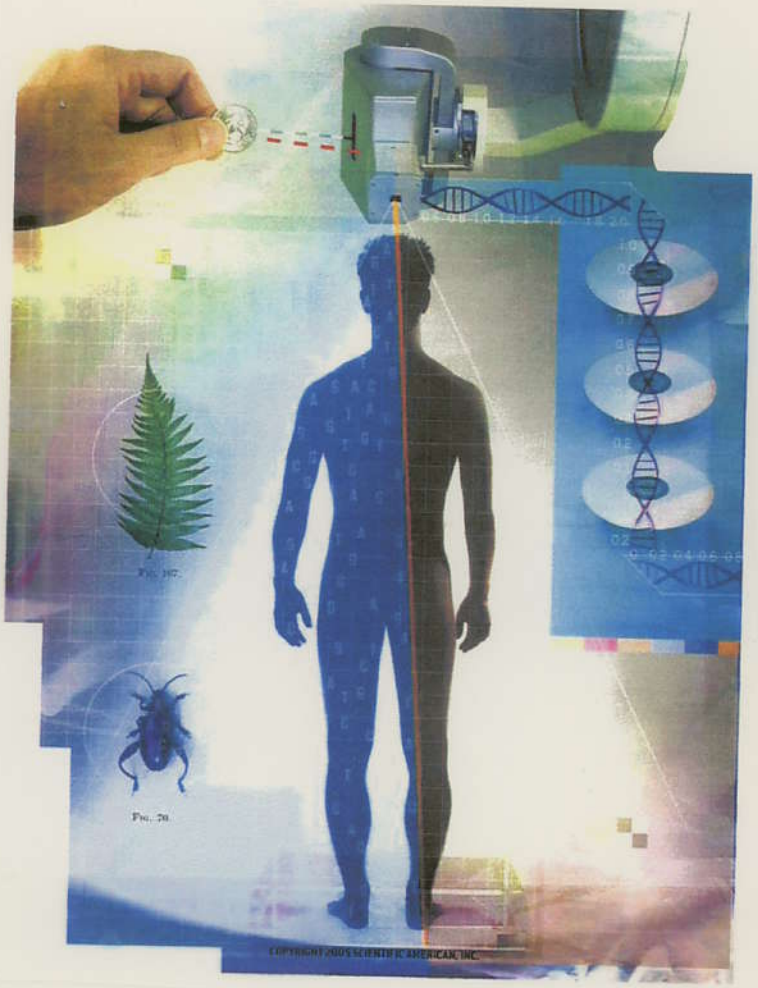
FIGURE 19.16
Biochips can help in identifying precise forms of cancer.

The ultimate Measure of Individuality
Personal Genome Sequence

Genomes *for* ALL

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

Find DNA VARIABILITY IN ALL Genes & Associate w. th Specific Traits!



DNA Sequencing Has become Inexpensive + High Throughput

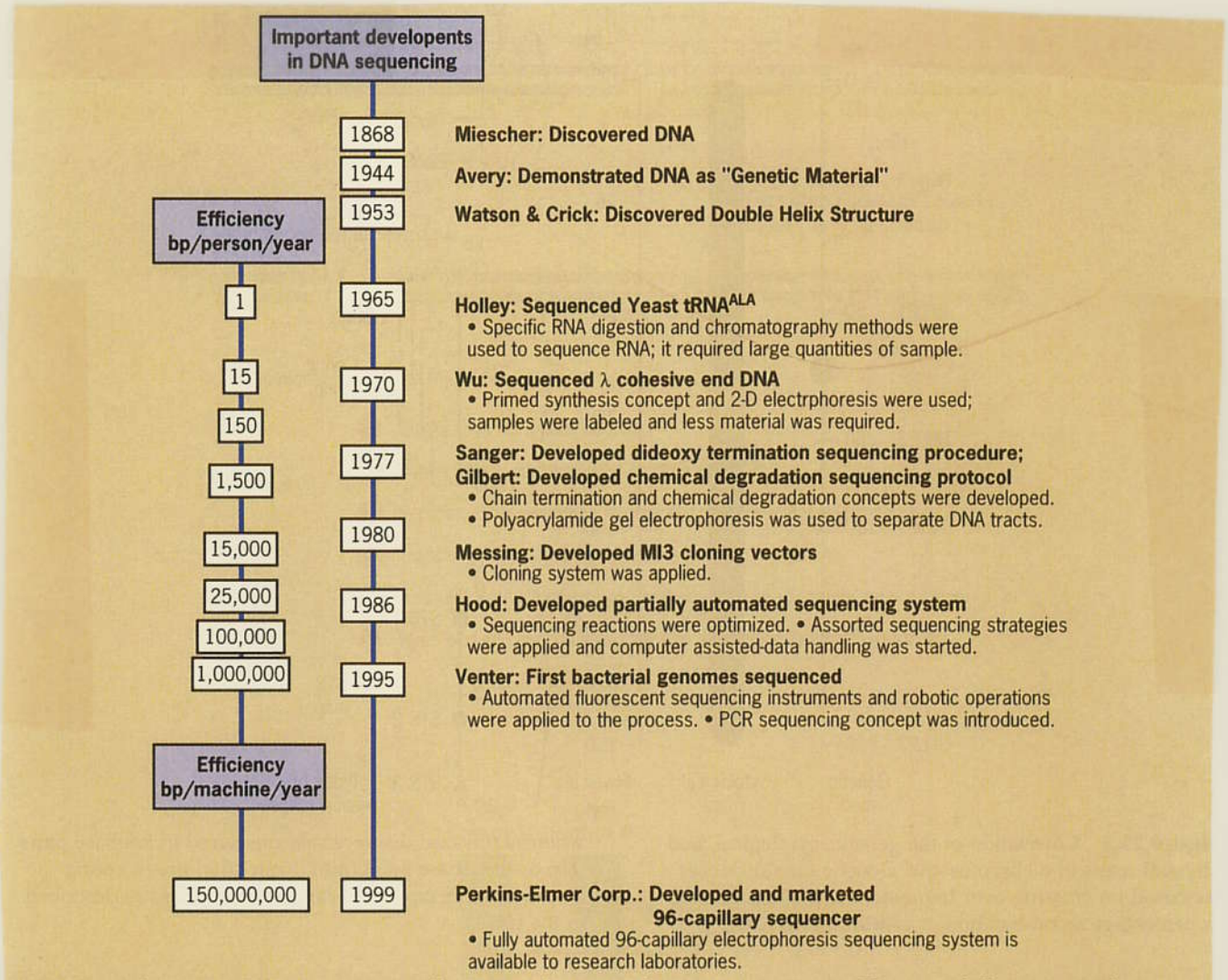


Figure 2. Advances in DNA sequencing efficiency and some of the technological developments that enhanced the productivity of sequencers. Initially, all the steps in DNA sequencing were performed manually, making it a very labor-intensive process. However, fully automated sequencing machines have now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

TODAY - 1×10^8 bp/hour OR 1×10^{12} bp/year per machine

≈ 1 human genome/day for ≈ $10,000 !!

GENE DATABASES ARE EXPLODING!

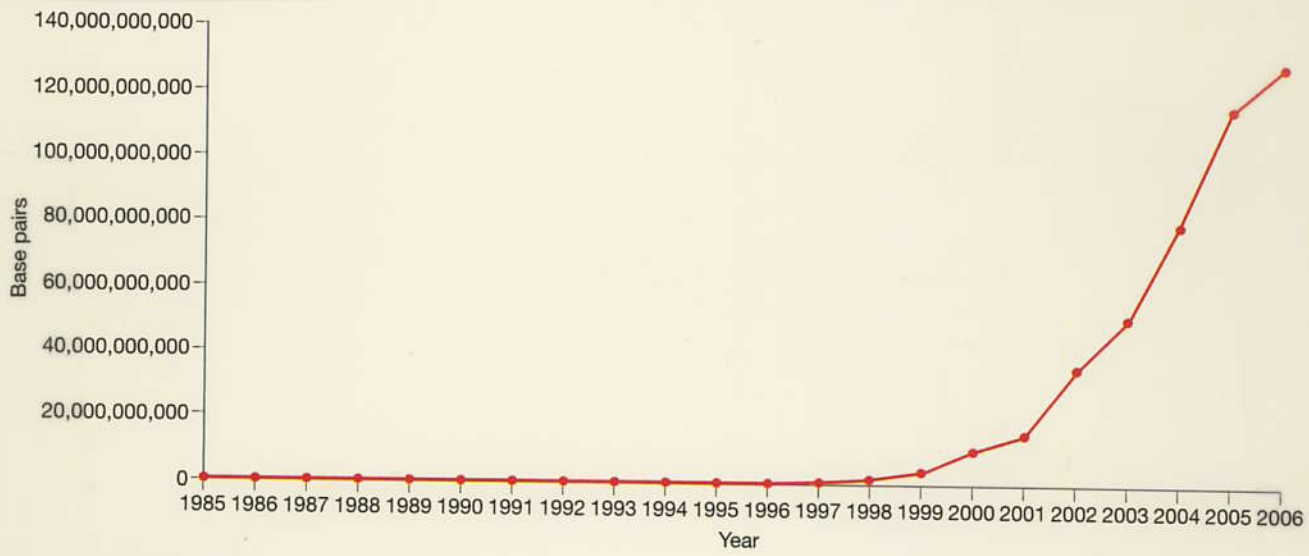


FIGURE 10-1

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.

figure 18.4

AUTOMATED SEQUENCING. This sequence facility simultaneously runs multiple automated sequencers, each processing 96 samples at a time.

HOW IS DNA SEQUENCED?

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

a. Yes
b. No

iClicker

MAXAM-Gilbert Chain Breakage

Reactions Break Phosphodiester Bonds

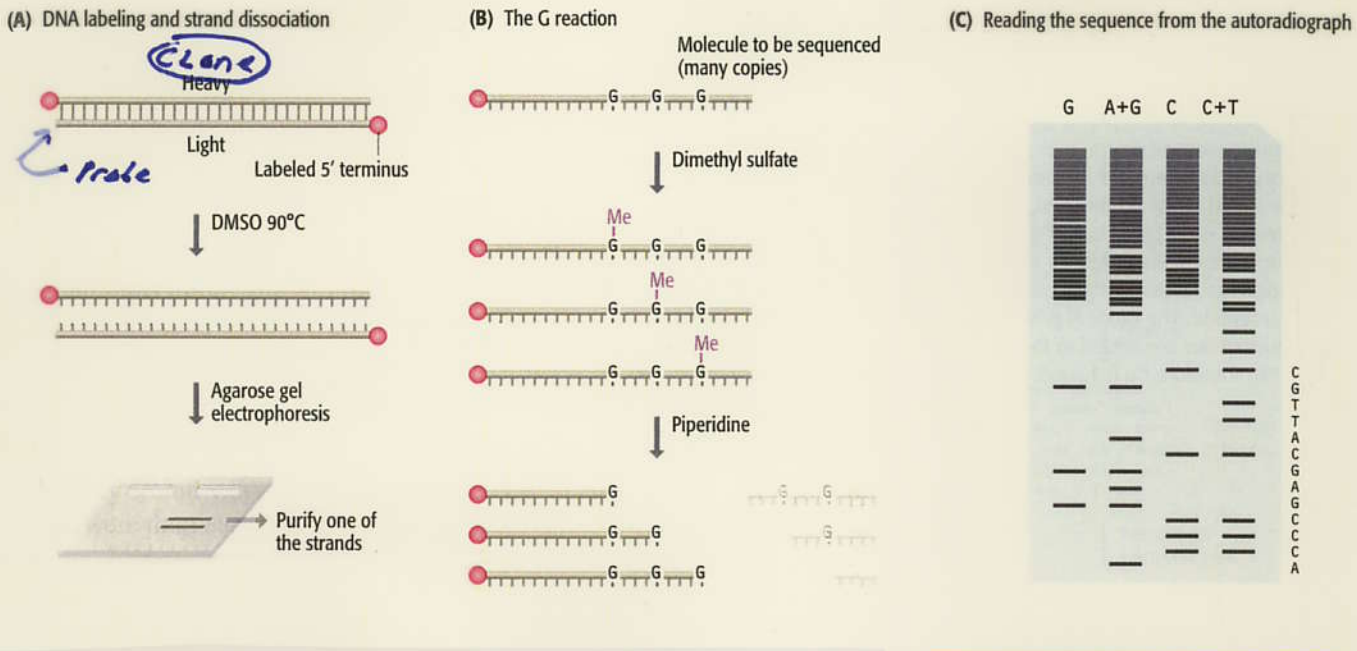
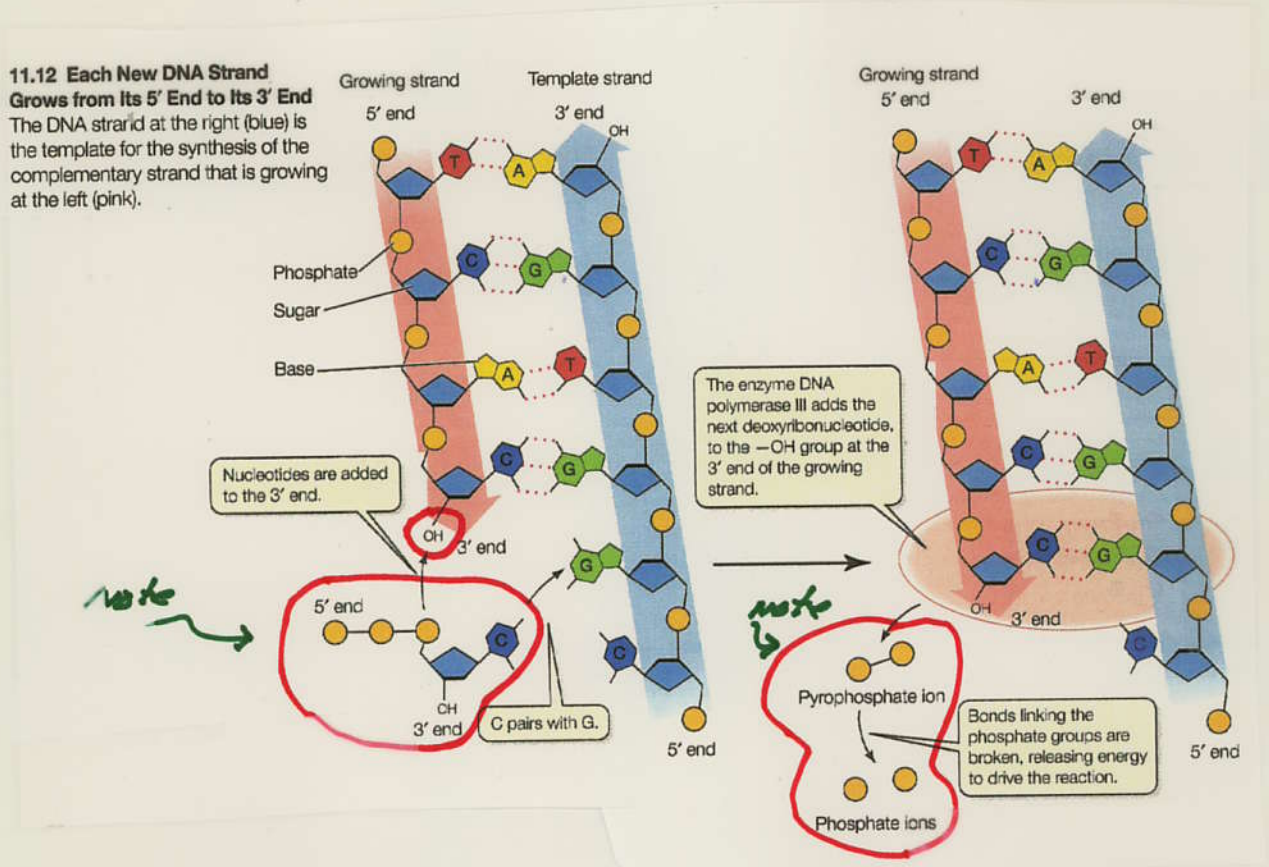


Figure 4.8 Chemical degradation sequencing.

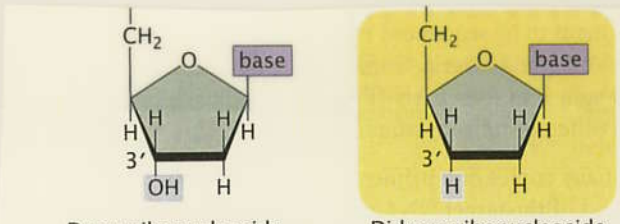
AT SPECIFIC BASES RANDOMLY IN POPULATION OF DNA Molecules

DNA Synthesis Sequencing



1. Need Primer
2. Need Template
3. Need dNTPs

Sanger Chain-Termination Dideoxy Nucleotide Sequencing

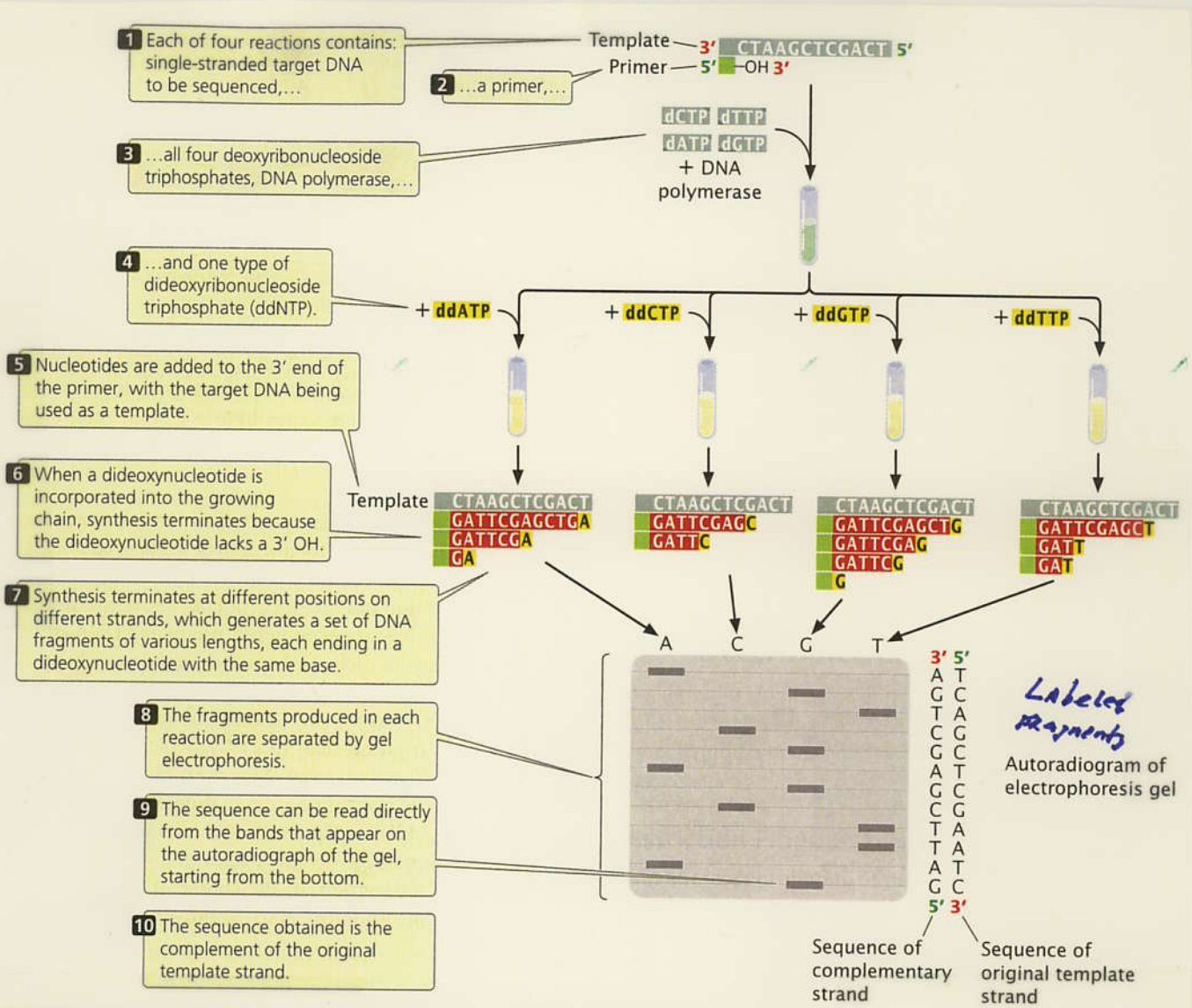


Deoxyribonucleoside triphosphate (dNTP)

Dideoxyribonucleoside triphosphate (ddNTP)

19.25 The dideoxy sequencing reaction requires a special substrate for DNA synthesis. (a) Structure of deoxyribonucleoside triphosphate, the normal substrate for DNA synthesis. (b) Structure of dideoxyribonucleoside triphosphate, which lacks an OH group on the 3′-carbon atom.

Clone first or PCR



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

Changing the Game - High Throughput DNA Sequencing → Automation!

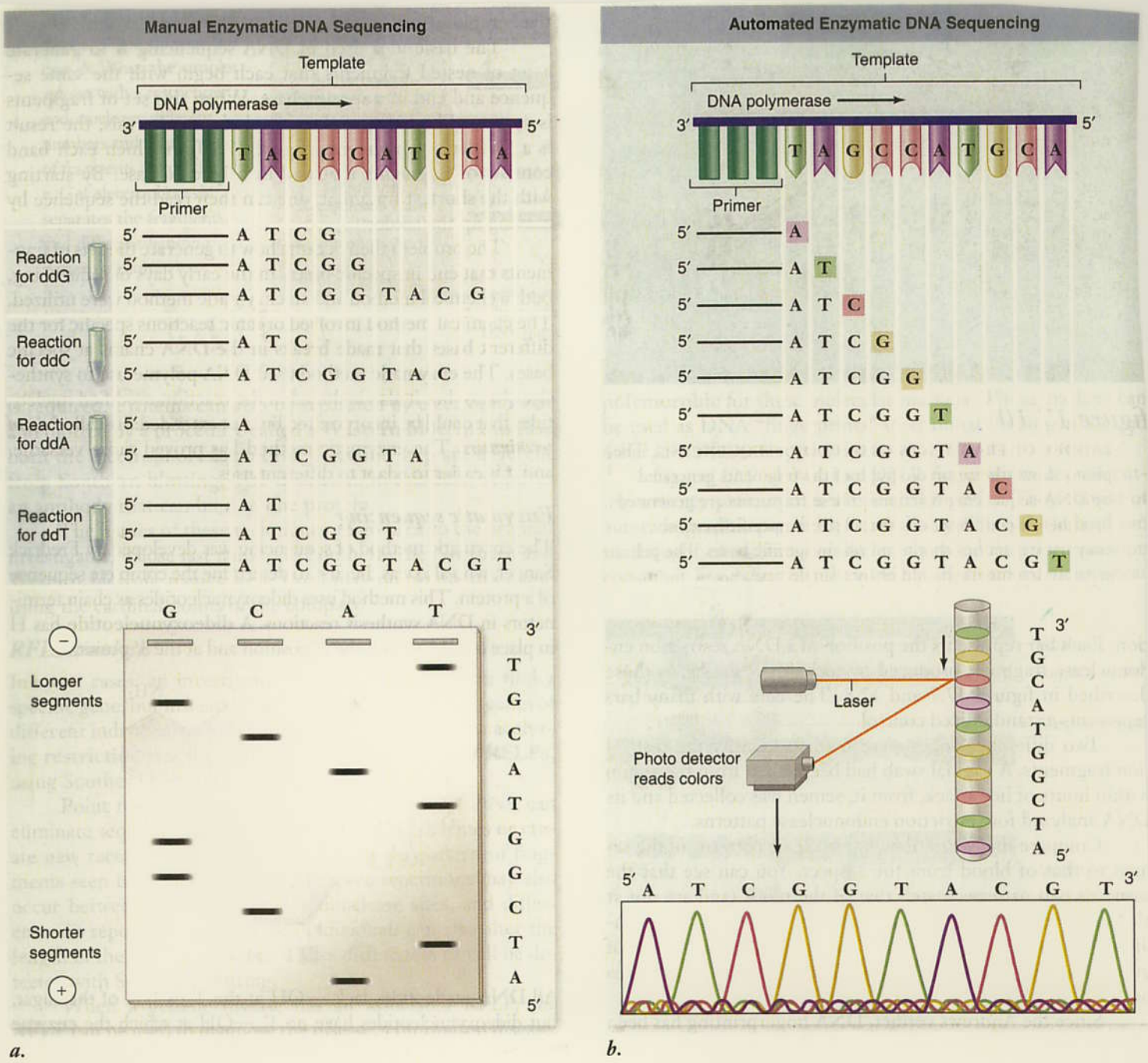
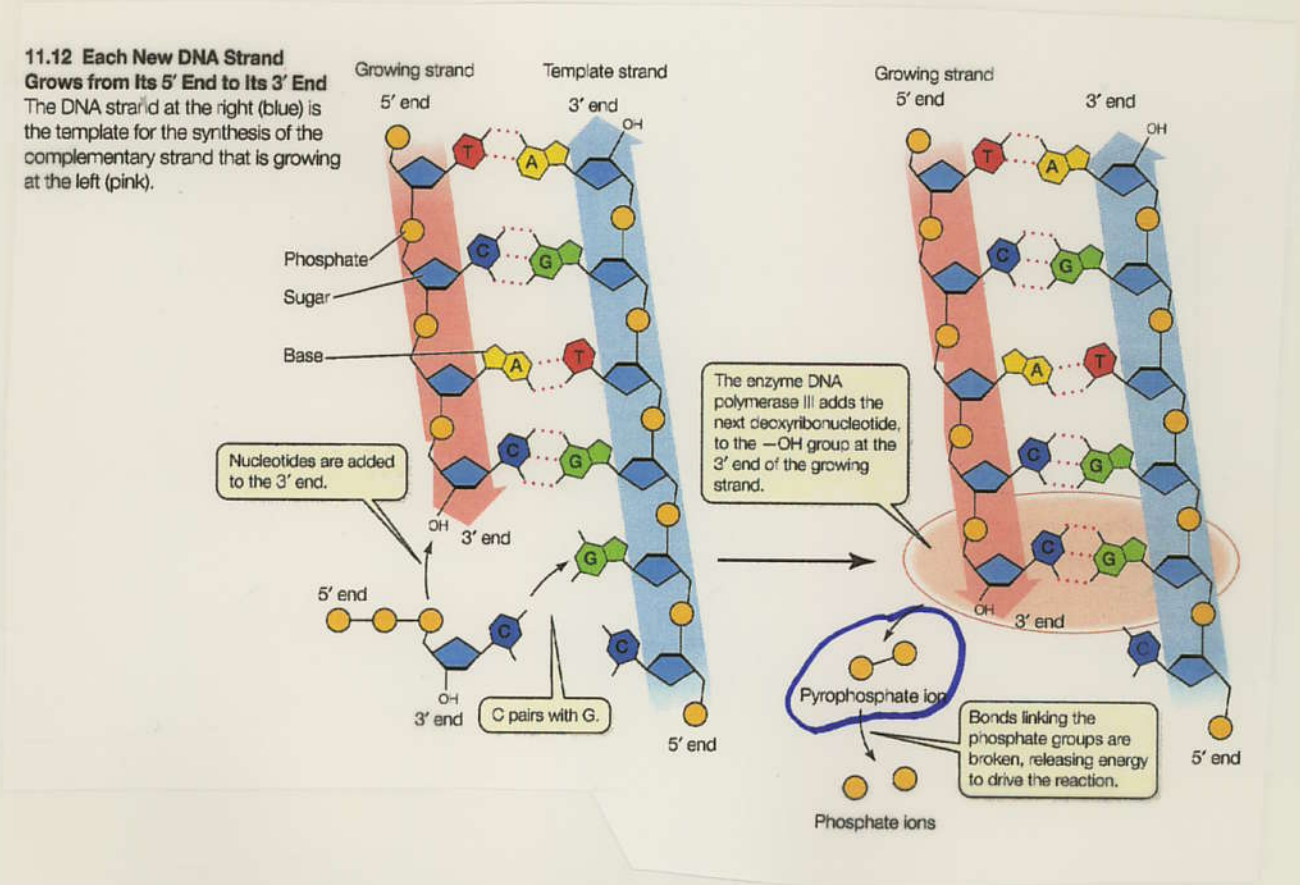


figure 17.11

MANUAL AND AUTOMATED ENZYMATIC DNA SEQUENCING. The sequence to be determined is shown at the top as a template strand for DNA polymerase with a primer attached. ***a.*** In the manual method, four reactions were done, one for each nucleotide. For example, the A tube would contain dATP, dGTP, dCTP, dTTP, and ddATP. This leads to fragments that end in A due to the dideoxy terminator. The fragments generated in each reaction are shown along with the results of gel electrophoresis. ***b.*** In automated sequencing, each ddNTP is labeled with a different color fluorescent dye, which allows the reaction to be done in a single tube. The fragments generated by the reactions are shown. When these are electrophoresed in a capillary tube, a laser at the bottom of the tube excites the dyes, and each will emit a different color that is detected by a photodetector.

Changing the GAME Again - MASSIVELY Parallel Pyrosequencing!

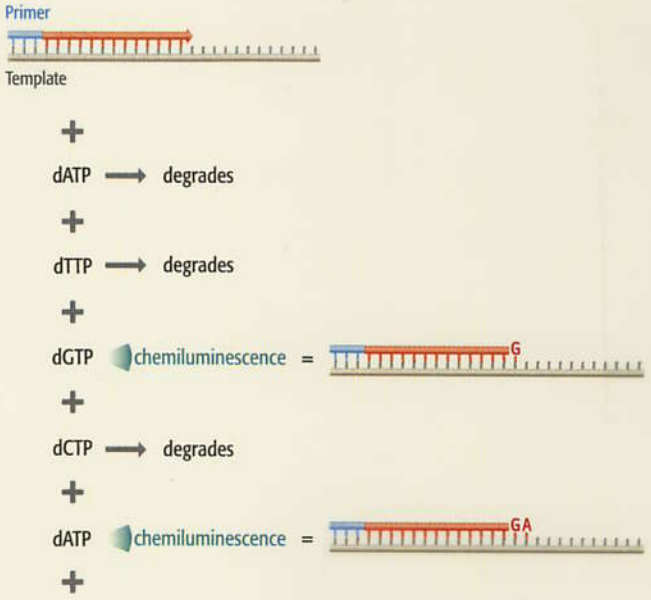


11.12 Each New DNA Strand Grows from Its 5′ End to Its 3′ End
The DNA strand at the right (blue) is the template for the synthesis of the complementary strand that is growing at the left (pink).

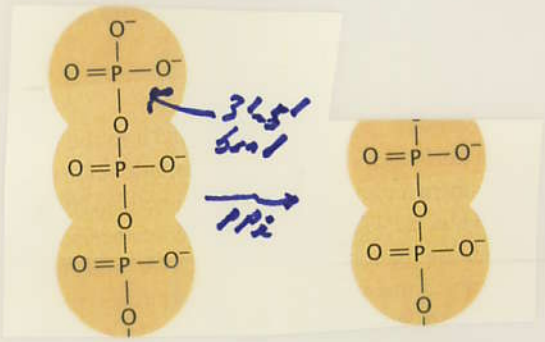
PP_i released when dNTP Forms Phosphodiester bond

SEQUENTIAL PYROSEQUENCING

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.



When dNTP forms Phosphodiester Bond

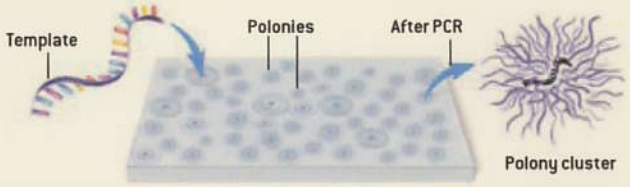


→ Degrade Enzymatically → Light Emitted

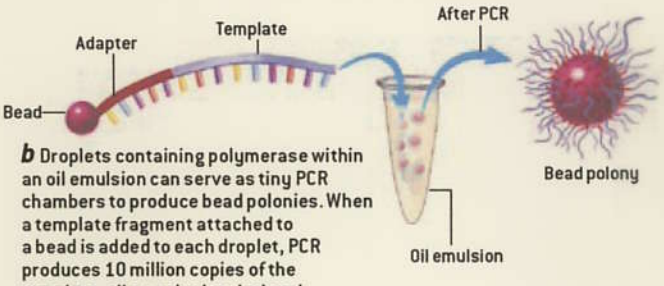
MASSIVELY PARALLEL SEQUENCING USES PCR AND DNA SYNTHESIS TO SEQUENCE HUNDREDS & THOUSANDS OF DNA FRAGMENTS AT ONE TIME!

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.



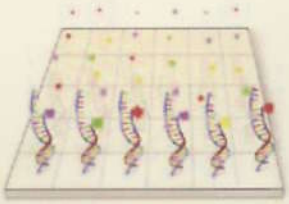
a Polonies—polymerase colonies—created directly on the surface of a slide or gel each contain a primer, which a template fragment can find and bind to. PCR within each polony produces a cluster containing millions of template copies.



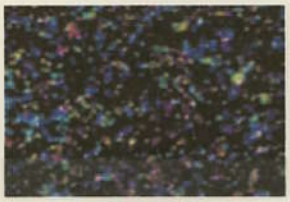
b Droplets containing polymerase within an oil emulsion can serve as tiny PCR chambers to produce bead polonies. When a template fragment attached to a bead is added to each droplet, PCR produces 10 million copies of the template, all attached to the bead.

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.



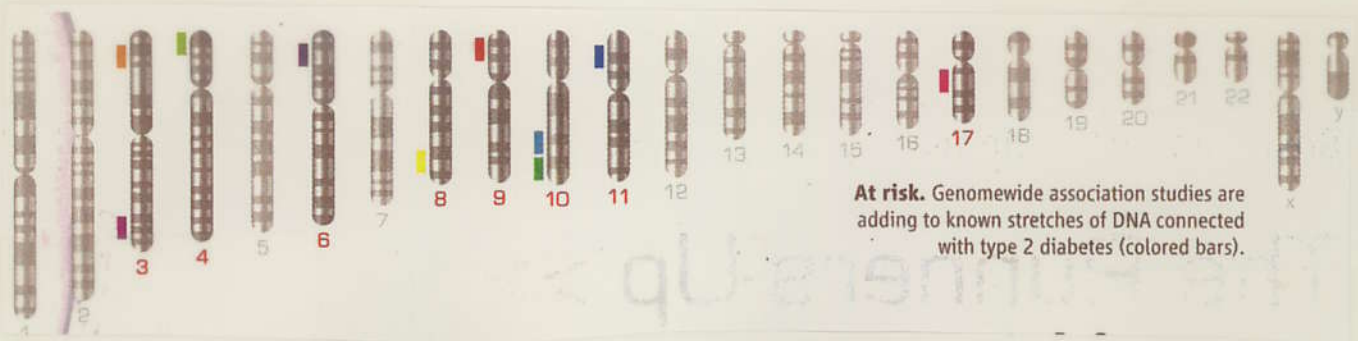
Single-molecule array



Bead polonies

CLONING ELIMINATED!
1×10^8 bp/hr for $5,000

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

The Personal Genome !

INTERNATIONAL
Herald Tribune | AMERICAS

DNA pioneer Watson gets own genome map

By Nicholas Wade
Published: June 1, 2007

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

A copy of his genome, recorded on two DVDs, was presented to Watson on Thursday in a ceremony in Houston by Richard Gibbs, director of the Human Genome Sequencing Center at the Baylor College of Medicine, and by Jonathan Rothberg, founder of the company 454 Life Sciences.

"I am thrilled to see my genome," Watson said.

Richard Carson/Reuters

James Watson, co-discoverer of the DNA helix, received his genome data Thursday.

Rothberg's company makes an innovative DNA sequencing machine, the latest version of which proved capable of decoding Watson's genome in two months for less than $1 million, said Michael Egholm, vice president for research at 454. The sequence was verified and analyzed by Gibbs's center in Houston. It was Gibbs who proposed the idea of sequencing Watson's genome.

WANT YOUR Genome Sequence!?

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

What will You Do with the INFORMation?

57

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

a. Yes
b. No

iClicker

The Race is on!

$10 Million
to the First Team to Sequence
100 Human Genomes
in
10 Days

Technology Innovation!

What can we do with the sequences of 1,000 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."

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.

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.

HOW ARE GENOMES SEQUENCED?

TODAY →

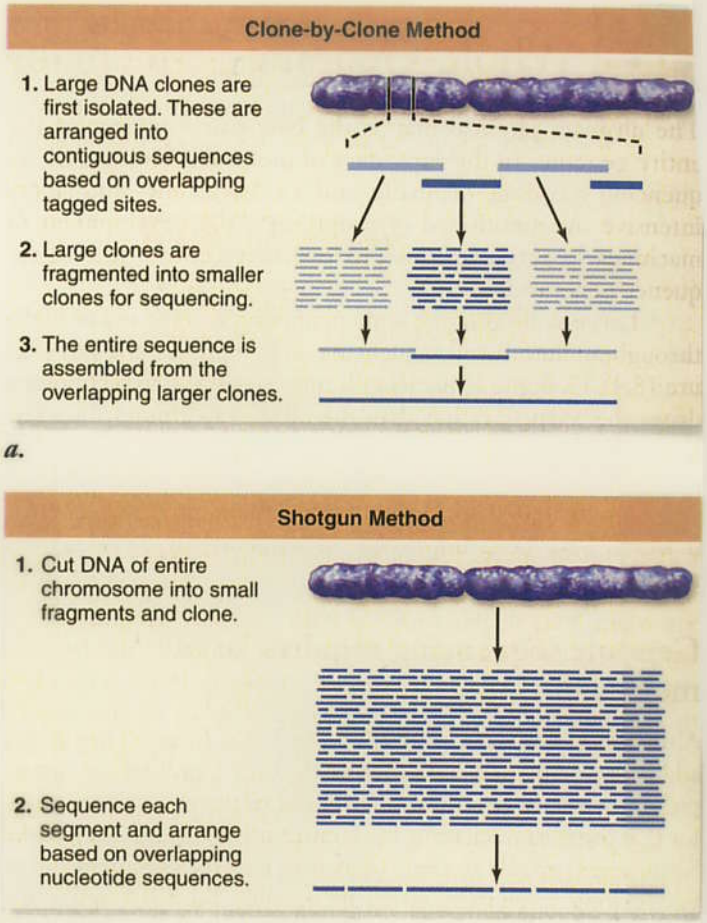
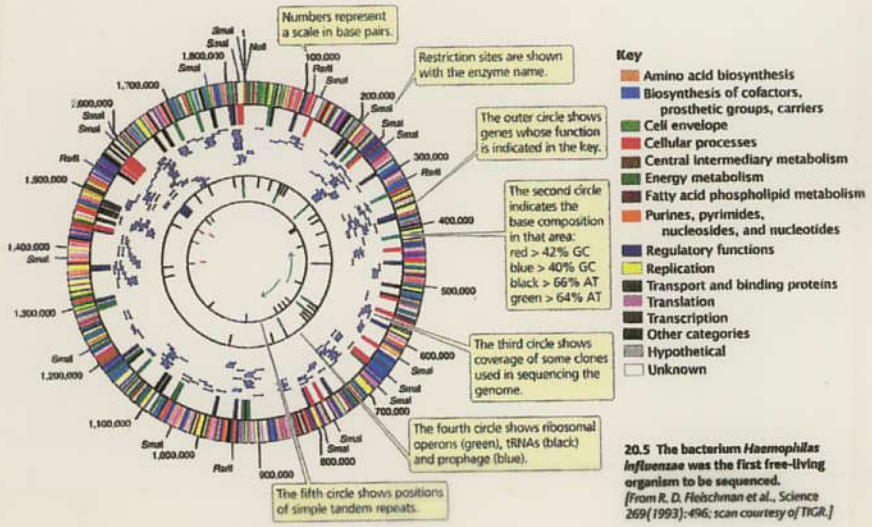


figure 18.5

COMPARISON OF SEQUENCING METHODS. ***a.*** The clone-by-clone method uses large clones assembled into overlapping regions by STSs. Once assembled, these can be fragmented into smaller clones for sequencing. ***b.*** In the shotgun method the entire genome is fragmented into small clones and sequenced. Computer algorithms assemble the final DNA sequence based on overlapping nucleotide sequences.

(2)

The First Genome Sequence



20.5 The bacterium *Haemophilus influenzae* was the first free-living organism to be sequenced. *[From R. D. Fleischman et al., Science 269(1993):496; scan courtesy of TIGR.]*

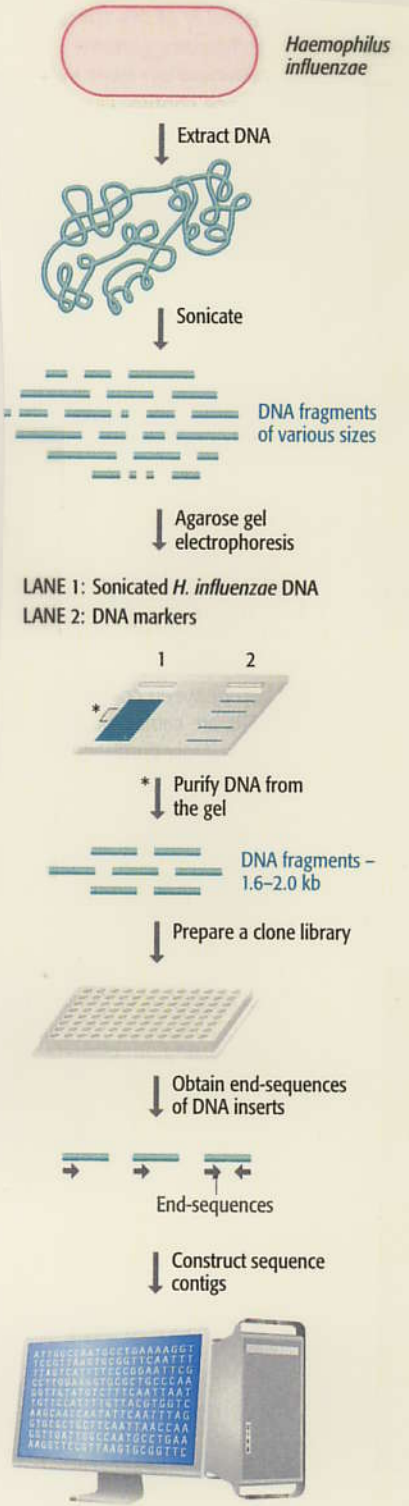


Figure 4.10 The way in which the shotgun method was used to obtain the DNA sequence of the *Haemophilus influenzae* genome. *H. influenzae* DNA was sonicated, and fragments with sizes between 1.6 kb and 2.0 kb were purified from an agarose gel and ligated into a plasmid vector to produce a clone library. End-sequences were obtained from clones taken from this library, and a computer was used to identify overlaps between sequences. This resulted in 140 sequence contigs, which were assembled into the complete genome sequence, as shown in Figure 4.11.

Some Genomes Sequenced to Date

TABLE 24.1 Milestones for Comparative Eukaryotic Genomics

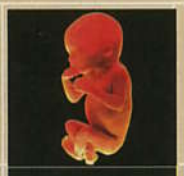
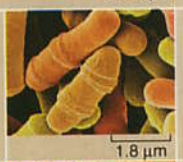
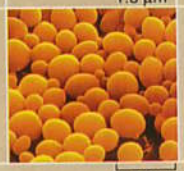
Organism		Estimated Genome Size (Mb)	Estimated Number of Genes	Year Sequenced
Vertebrates				
Homo sapiens (human)		2,900	20,000–25,000	2001
Mus musculus (mouse)		2,600	30,000	2002
Fugu rubripes (pufferfish)		365	33,609	2002
Rattus norvegicus (rat)		2,750	20,973	2004
Pan troglodytes (chimpanzee)		3,100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

TABLE 24.1 Milestones for Comparative Eukaryotic Genomics, continued

Organism		Estimated Genome Size (Mb)	Estimated Number of Genes	Year Sequenced
Vertebrates, continued				
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

By the end of 2006, the genomes of more than 300 bacterial species and 100 eukaryotic organisms were either completely sequenced or had substantial amounts of genomic sequence data available in public databases. The table demonstrates the wide range of types of organisms and the rationale for choosing some of them. Genome sizes are shown in millions of base pairs (Mb). The numbers of protein-coding genes in the genomes are estimated by annotation efforts based on cDNA sequences, comparison with other genome and gene sequences, and computation.

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

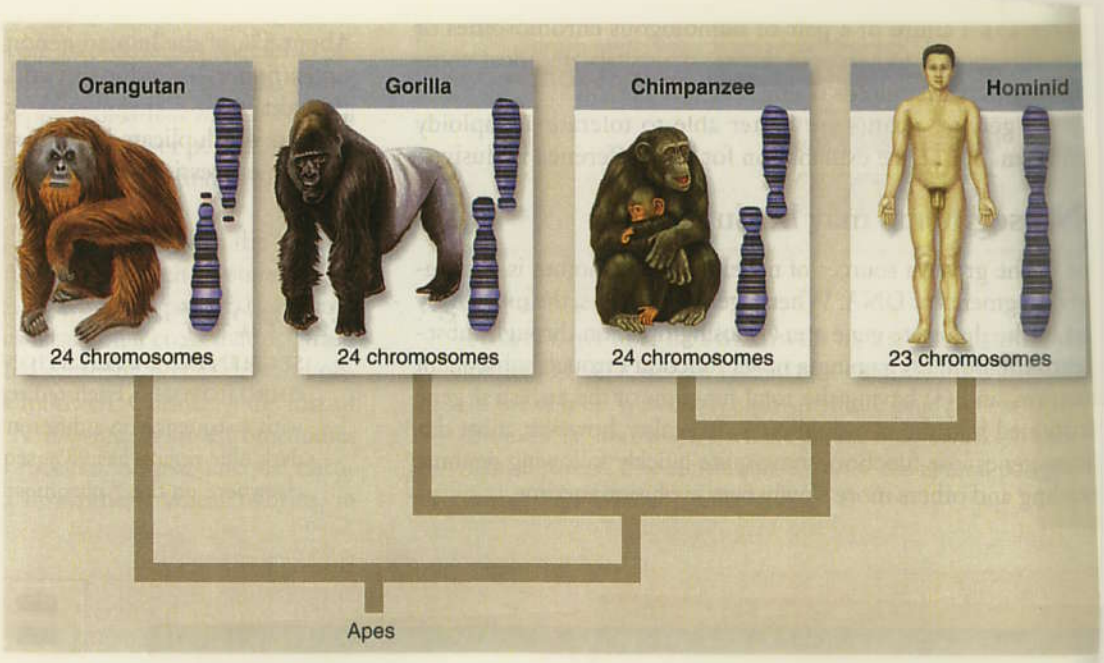
Organism	Complete	Draft assembly	In progress	total
Prokaryotes	627	460	476	1563
Archaea	49	4	30	83
Bacteria	578	456	446	1480
Eukaryotes	22	134	172	328
Animals	4	55	81	140
Mammals	2	21	23	46
Birds		1	2	3
Fishes		3	6	9
Insects	1	19	17	37
Flatworms		1	3	4
Roundworms	1	4	12	17
Amphibians			2	2
Reptiles			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:	**649**	**594**	**648**	**1891**

Revised: Jan 28, 2008

What CAN We DO with ALL OF These DNA SEQUENCES?

figure 24.8

LIVING GREAT APES. All living great apes, with the exception of humans, have a haploid chromosome number of 24. Humans have not lost a chromosome; rather, two smaller chromosomes fused to make a single chromosome.



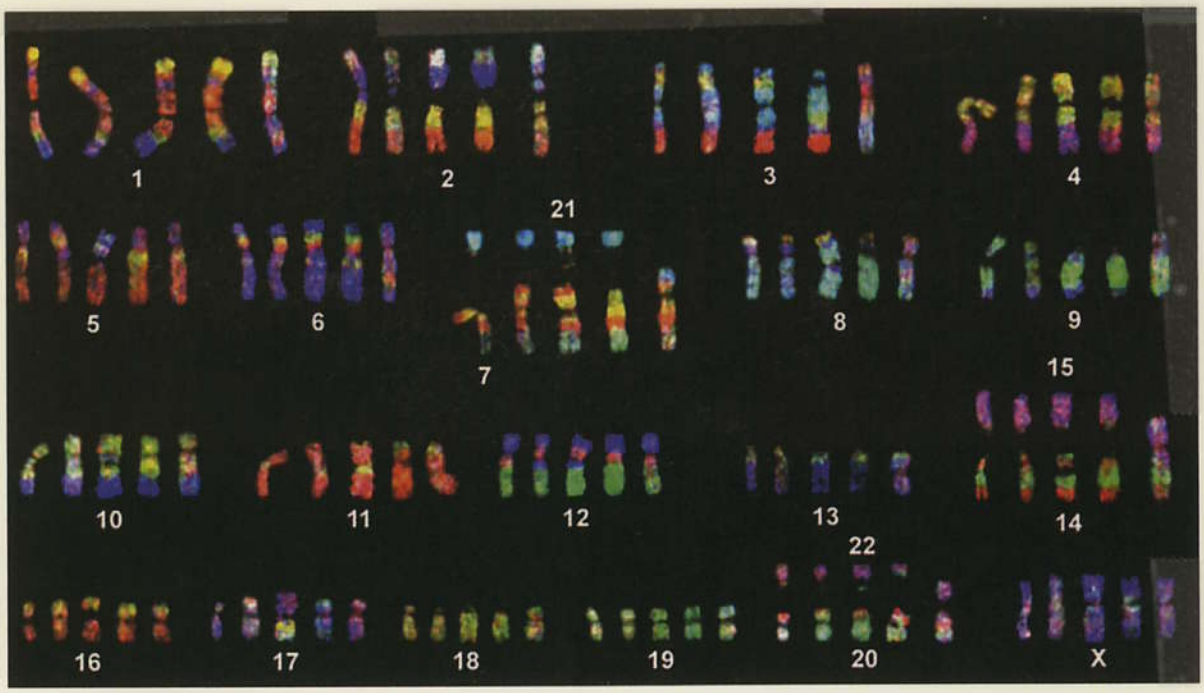


Figure 12.28: Bar-coding of primate chromosomes as a way of revealing structural differences.

Cross-species color banding (*Rx-FISH*) profiles show alignment of orthologous primate chromosomes with human chromosomes 1-22 and X. Chromosome sets (with numbering according to the human homologs) show from left to right: human, chimpanzee, gorilla, orang-utan and macaque. To improve comparisons, chromosomes 2p/2q, 7/21, 14/15 and 20/22, which are single chromosomes in human or in the macaque are shown together with the great ape homologs. This type of analysis has suggested a preliminary ancestral karyotype for human and great apes. Reproduced from Muller and Wienberg (2002) *Hum. Genet.* **109**, 85–94 with permission from Springer-Verlag.

What Makes "A Man a Man" and "A Mouse A Mouse?"

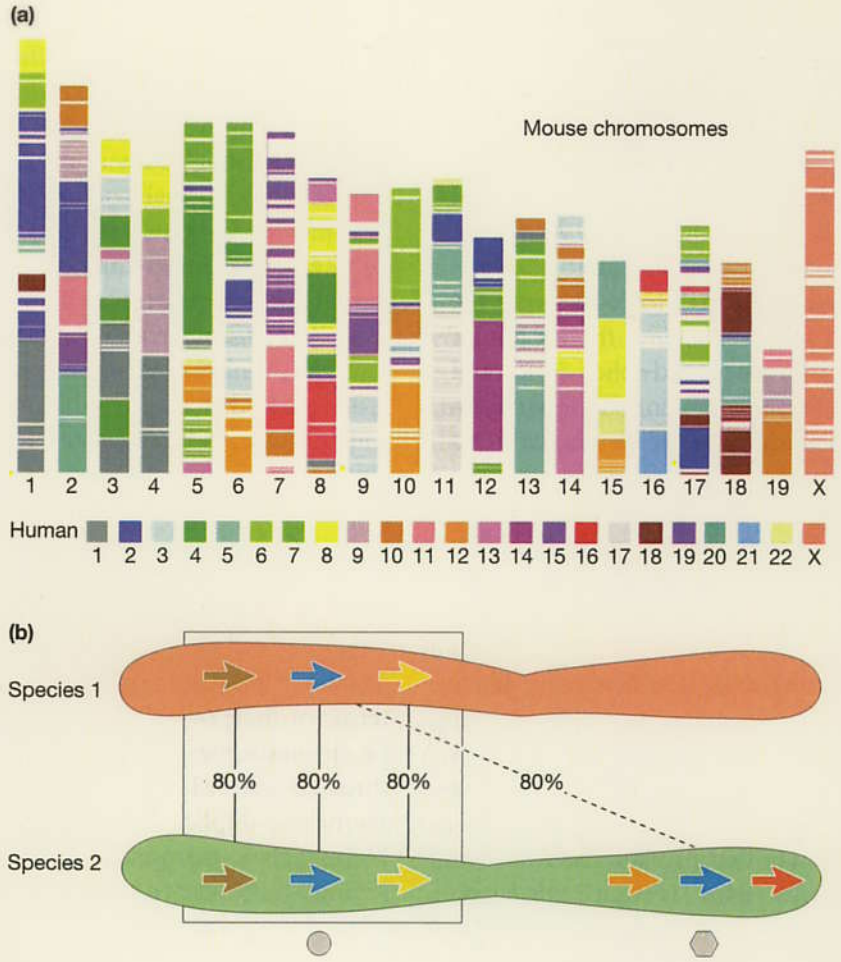


FIGURE 12-10

Syntenic relationships between genes in different species. Mammalian genomes are characterized by large chromosomal segments in which a few up to several thousand genes, spanning hundreds of thousands to millions of base pairs, are in the same order and orientation in distinct species. (*a*) Each of the 20 mouse chromosomes (autosomes 1–19 and X) is colored according to the chromosome(s) in the human genome with which it shares syntenic blocks. The key to the colors of the human chromosomes (1–22 and the X) is shown at the *bottom*. For example, there is a large block at the tip of mouse chromosome 1 that is orthologous to a segment from human chromosome 8 (*yellow*). These regions were part of the same chromosome in the last common ancestor of humans and mice. The complete synteny between the X chromosomes of mice and human beings is striking. This is characteristic of the X chromosomes of placental mammals—all are derived from the same ancestral chromosome and there is great evolutionary pressure to keep X-chromosome genes together and not mixed in with autosomal genes. (*b*) Synteny can be used to identify the likely ortholog of a protein. The *arrows* represent genes. When the sequence of *blue* gene in Species 1 is compared to the sequence of the entire genome of Species 2, two genes in Species 2 are found to match it at the level of 80% sequence similarity (*gray circle* and *hexagon*). A similar comparison is done for the neighboring genes. One of the *blue* genes in Species 2 is found to be flanked by the same neighboring genes (*brown* and *yellow*) as is the *blue* gene in Species 1. In contrast, the second *blue* gene in Species 2 (*hexagon*) is flanked by two genes (*orange* and *red*) that have no similarity to the flanking genes in Species 1. The location of the *blue* gene between two similar genes in Species 2 makes it a very strong candidate for the ortholog of the *blue* gene in Species 1.

Comparative Sequencing Pathogen Genomes

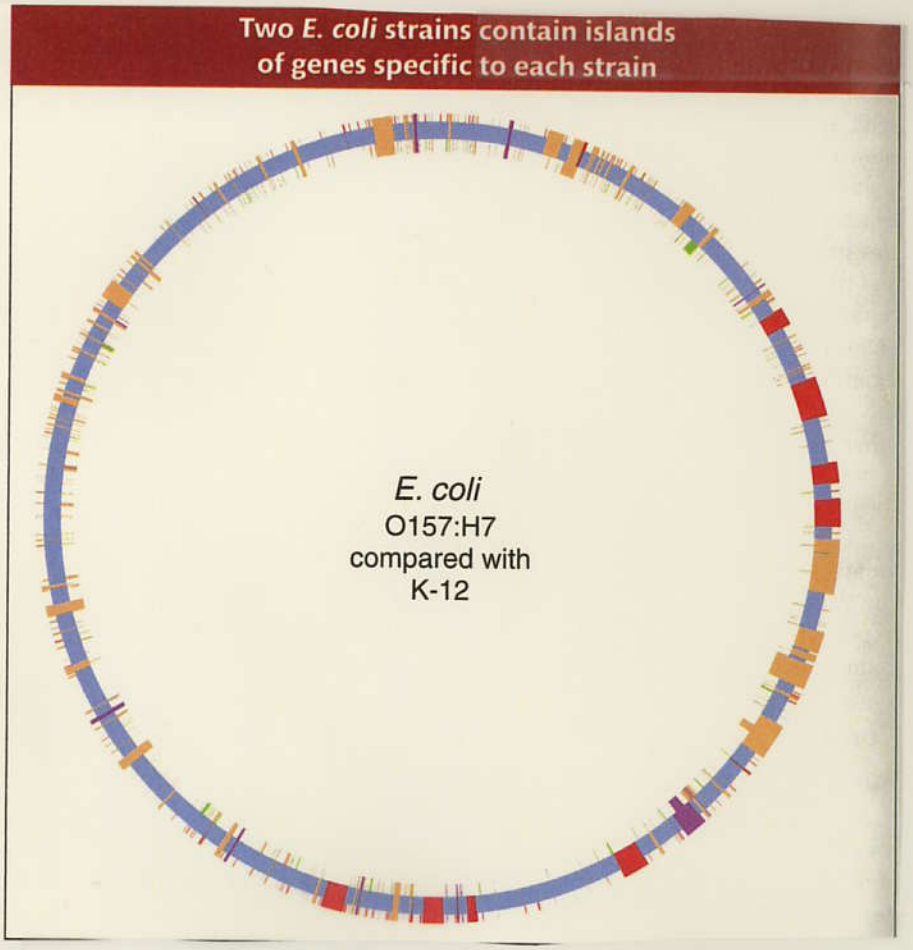
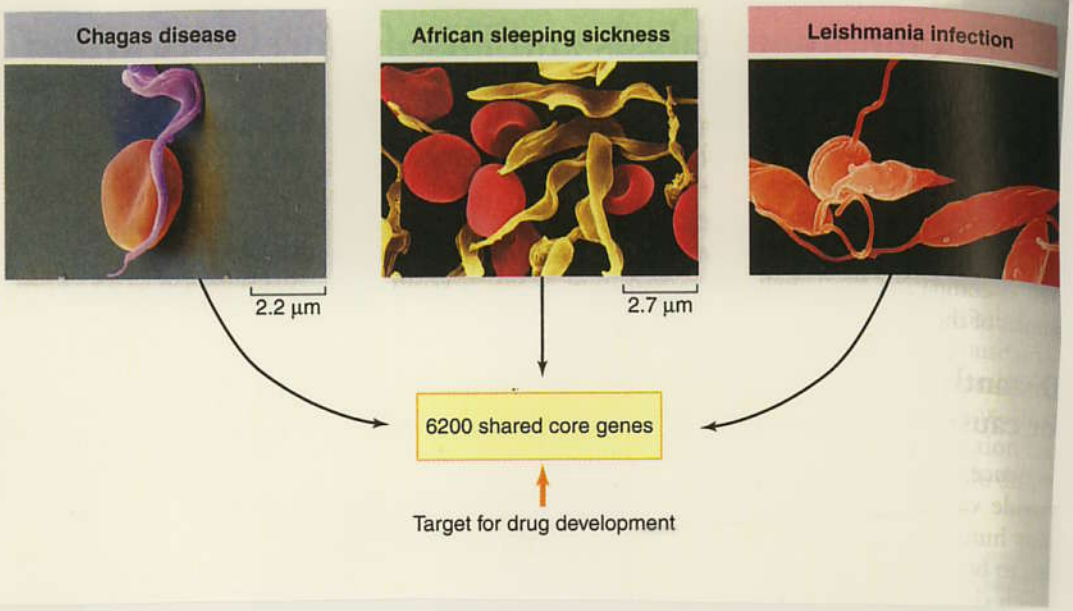


FIGURE 13-17 The circular genome maps of *E. coli* strains K-12 and O157:H7. The circle depicts the distribution of sequences specific to each strain. The colinear backbone common to both strains is shown in blue. The positions of O157:H7-specific sequences are shown in red. The positions of K-12-specific sequences are shown in green. The positions of O157:H7- and K-12-specific sequences at the same location are shown in tan. Hypervariable sequences are shown in purple. [After N. T. Perna et al., "Genome Sequence of Enterohaemorrhagic *Escherichia coli* O157:H7," *Nature* 409, 2001, 7529–7533. Courtesy of Guy Plunkett III and Frederick Blattner.]

figure 24.14

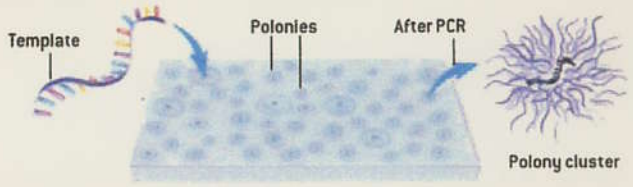
COMPARATIVE GENOMICS MAY AID IN DRUG DEVELOPMENT. Chagas disease, African sleeping sickness, and Leishmania, which claim millions of lives in developing nations each year, share 6200 core genes. Drug development targeted at proteins encoded by the shared core genes could yield a single treatment for all three diseases.



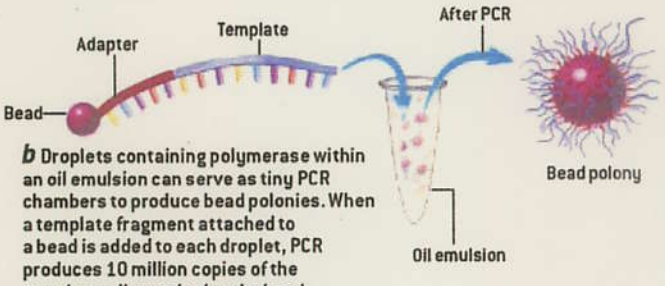
The future only Arrived A couple 7 years Ago!

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.



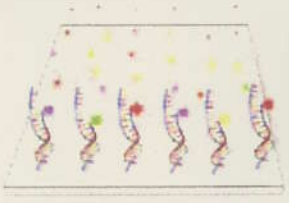
a Polonies—polymerase colonies—created directly on the surface of a slide or gel each contain a primer, which a template fragment can find and bind to. PCR within each polony produces a cluster containing millions of template copies.



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

SEQUENCING The Genomes of Extinct Organisms is Now Possible!

NATIONAL GEOGRAPHIC NEWS

REPORTING YOUR WORLD DAILY

NEWS HOME ANIMAL NEWS ANCIENT WORLD ENVIRONMENT NEWS CULTURES N

Mammoths to Return? DNA Advances Spur Resurrection Debate

Mason Inman
for National Geographic News
June 25, 2007

Today the only place to see woolly mammoths and people side-by-side is on *The Flintstones* or in the movies.

But researchers are on the verge of piecing together complete genomes of long-dead species such as Neandertals and mammoths. (See a brief overview of human genetics.)

Enlarge Photo

So now the big question is, Will we soon be able to bring such extinct species back to life?

Researchers are divided over how they might try to do this and whether it's even feasible. (Related: "Woolly Mammoth Resurrection, 'Jurassic Park' Planned [April 8, 2005].)

The Age of the Synthetic Genome Has Arrived!

Friday, January 25, 2008

Synthesizing a Genome from Scratch

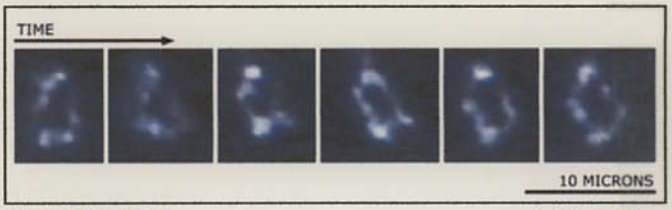
Scientists say the results represent a new stage in synthetic biology.

By Emily Singer

NATIONAL GEOGRAPHIC NEWS
REPORTING YOUR WORLD DAILY

NEWS HOME ANIMAL NEWS ANCIENT WORLD ENVIRONMENT NEWS CULTURES NEWS

Entire Synthetic Genome Created



Email to a Friend << Back to Full Story

A micrograph shows a synthetic *Mycoplasma genitalium* genome during a roughly 0.6-second period.

The feat is the first time an entire genome has been stitched together in the lab—an important step toward creating synthetic life, according to the creators.

FROM Pure A, G, C, & T.!

Synthetic Genomes Can Lead to New Forms of Life:

a. Yes
b. No

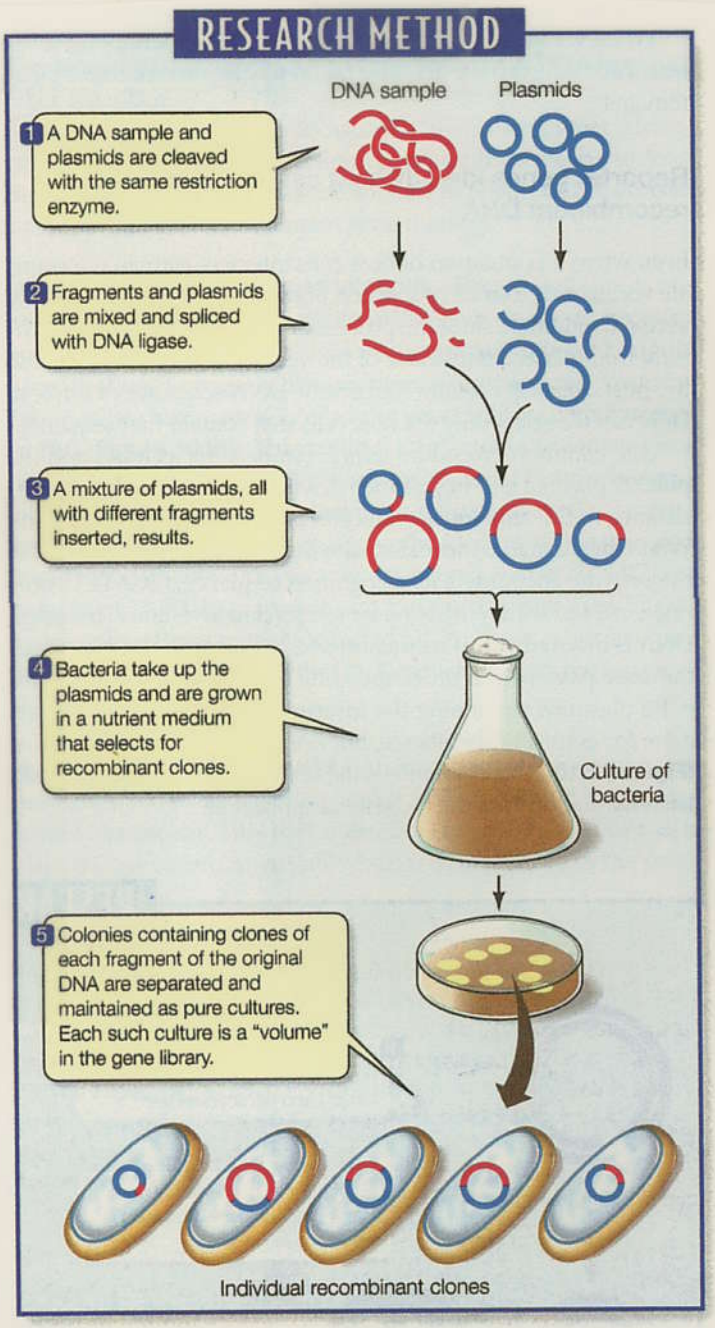
iClicker

Synthetic Biology Should Be Regulated:

a. Yes
b. No

iClicker

AND IT ALL STARTED WITH RECOMBINANT DNA in 1973!



16.11 Constructing a Gene Library Human chromosomal DNA is isolated and broken up into fragments using restriction enzymes. The fragments are inserted into vectors (plasmids are shown here) and taken up by host bacterial cells, each of which then harbors a single fragment of the human DNA. The information in the resulting bacterial cultures and sets of colonies constitutes a gene library.

Who could have dreamed what it lead to & what's next?!!