HC70A Winter 2004
Professor Bob Goldberg

Lectures 8 & 9
The Human Genome Project & detecting Changes in the Human Genome

THEMES/CONCEPTS

1. Genomes in Human cells
2. Mitochondrial Genome in Medicine, Forensics & Evolution
3. Human Genome Project - How Done?
4. Characteristics of the Human Genome
5. Human Genome vs. Mouse & Chimp Genomes

12.5hr
Stop 3/2/04

6. DNA Sequence organization in Human Genome
7. VNTRs & Their Utility in DNA Testing & Populations
8. Human Chromosomes
9. Detecting Large Changes in Human Chromosomes
10. Disease & Changes in Human Chromosome Number
11. Mutations in the Human Genome - Detection & Frequency
12. Usefulness of SNPs as Genetic Markers / Diversity within Populations

1 2hr lecture

Stop 3/4/04

HUMAN GENES ARE PRESENT
IN TWO COMPARTMENTS --
THE NUCLEUS & THE MITOCHONDRIA

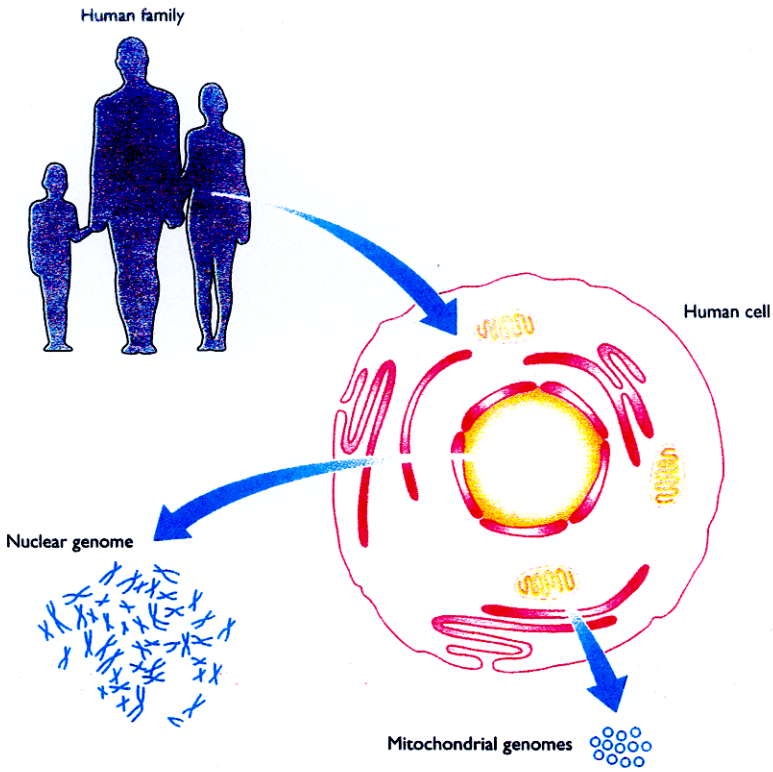
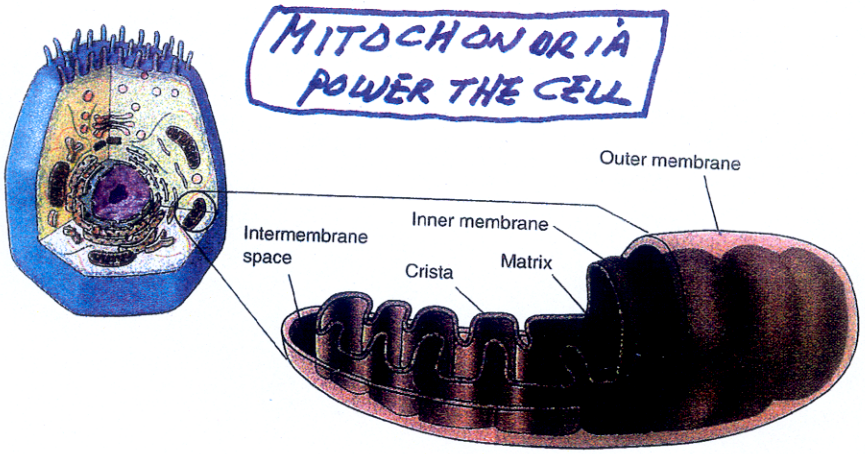


Figure 1.3 The nuclear and mitochondrial components of the human genome.
For more details on the anatomy of the human genome, see Section 6.1.

Genes in BOTH compartments are critical for human development.

MITOCHONDRIA POWER THE CELL



(b)

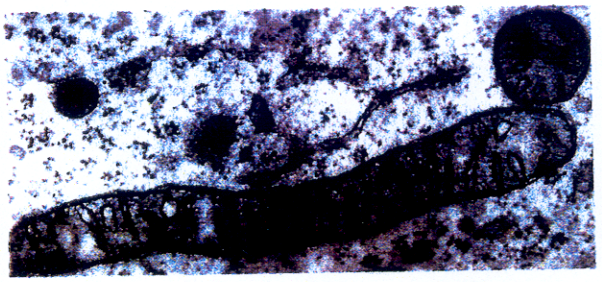


FIGURE 5.21
Mitochondria. (*a*) The inner membrane of a mitochondrion is shaped into folds called cristae, which greatly increase the surface area for oxidative metabolism. (*b*) Mitochondria in cross-section and cut lengthwise (70,000×).

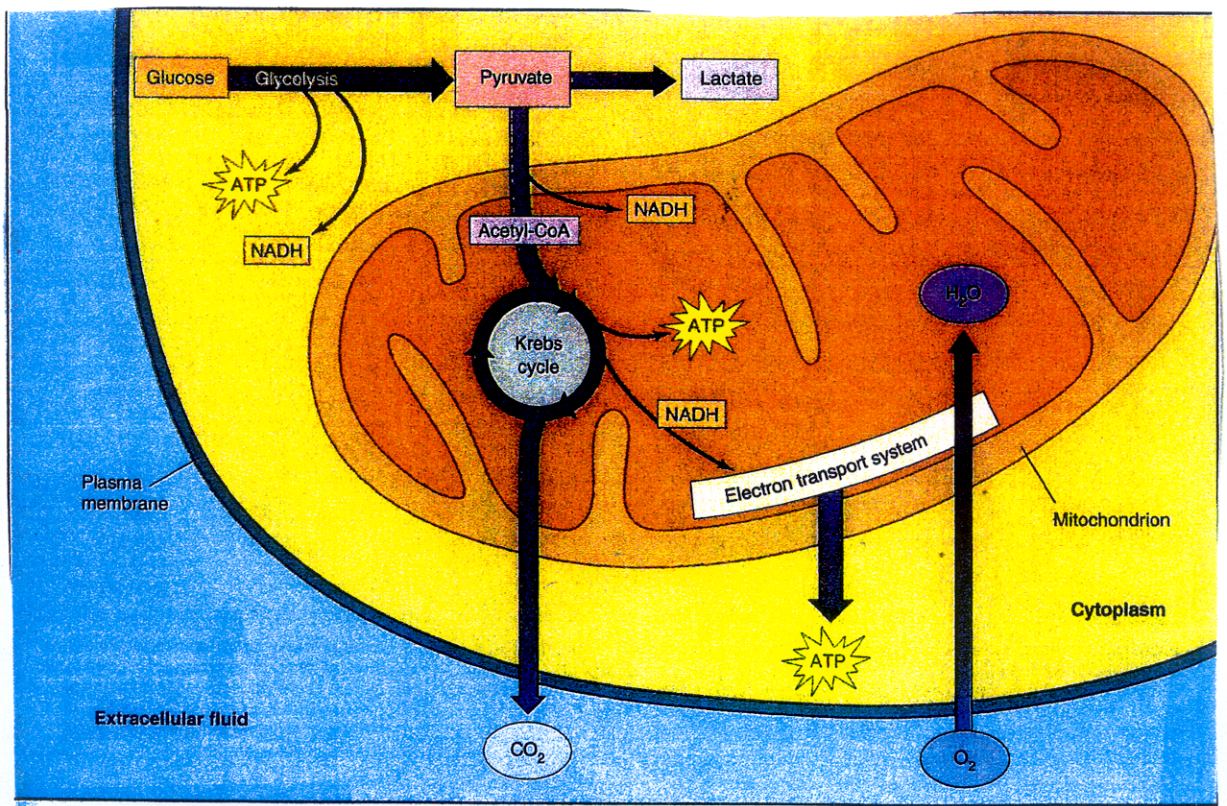


FIGURE 9.6
An overview of aerobic respiration.

2

MITOCHONDRIA ARE SEMI-AUTONOMOUS ORGANELLES

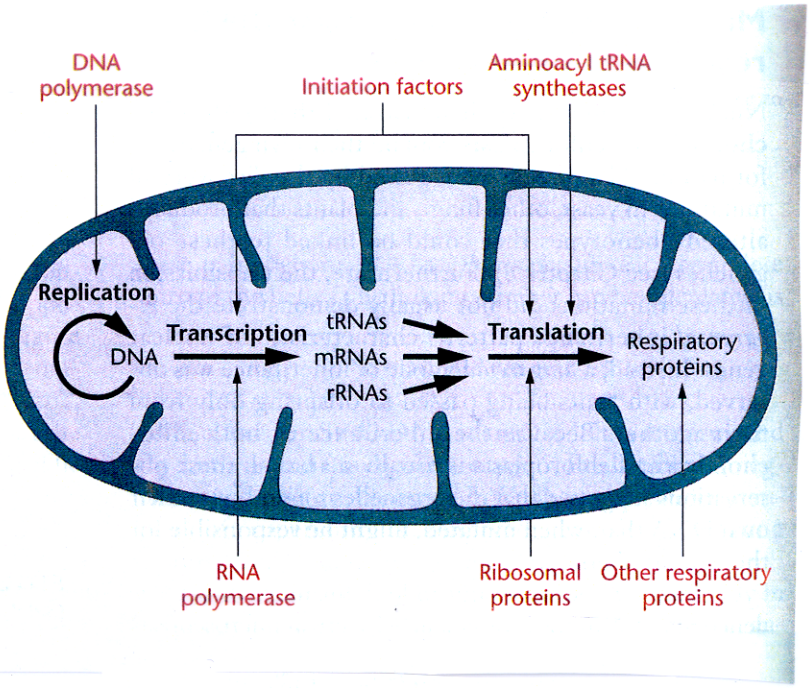


FIGURE 19.6 A comparison of the origin of gene products that are essential to mitochondrial function. Those shown entering the organelle are derived from the cytoplasm and encoded by the nucleus.

They Divide - have Genes - Encode Proteins - undergo Protein Synthesis

What ARE THE CHARACTERISTICS of The HUMAN Nuclear & Mitochondrial Genomes?

Table 7.1: The human nuclear and mitochondrial genomes

	Nuclear genome	Mitochondrial genome
Size	3300 Mb	16.6 kb
No. of different DNA molecules	23 (in XX) or 24 (in XY) cells, all linear	One circular DNA molecule
Total no. of DNA molecules per cell	23 in haploid cells; 46 in diploid cells	Several thousand
Associated protein	Several classes of histone and nonhistone protein	Largely free of protein
Number of genes	~65 000–80 000	37
Gene density	~1/40 kb	1/0.45 kb
Repetitive DNA	Large fraction, see *Figure 7.1.*	Very little
Transcription	The great bulk of genes are transcribed individually	Continuous transcription of multiple genes
Introns	Found in most genes	Absent
% of coding DNA	~3%	~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

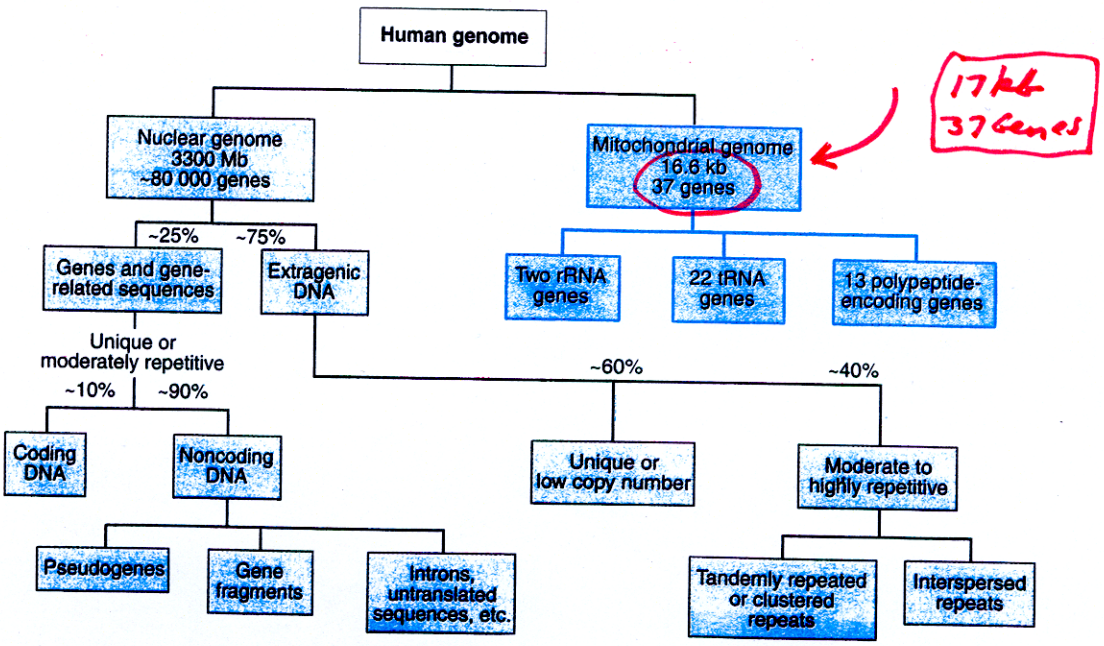


Figure 7.1: Organization of the human genome.

The Mitochondrial Genome is A SMALL CIRCLE containing only 37 Genes

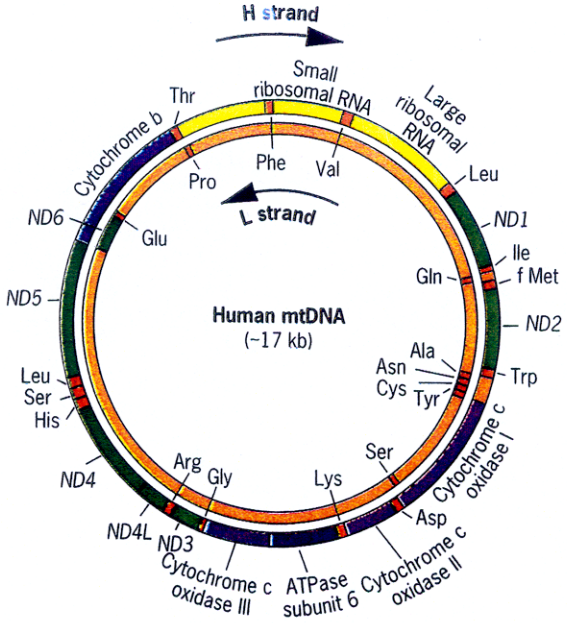
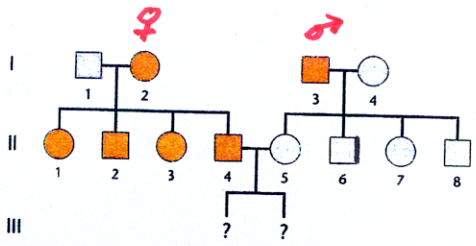


Figure 19.14 Map of human mtDNA showing the pattern of transcription. Genes on the inner circle are transcribed from the L strand of the DNA, whereas genes on the outer circle are transcribed from the H strand of the DNA. Arrows show the direction of transcription. *ND1–6* are genes encoding subunits of the enzyme NADH reductase; the tRNA genes in the mtDNA are indicated by abbreviations for the amino acids.

MtDNA is a CIRCLE

Mitochondrial Genes Are Inherited MATERNALLY



PASSED DIRECTLY FROM Mother to Children

Hypothesis to Explain?

IN A CLONING EXPERIMENT MOST OF MITOCHONDRIA COMES FROM EGG DONOR!

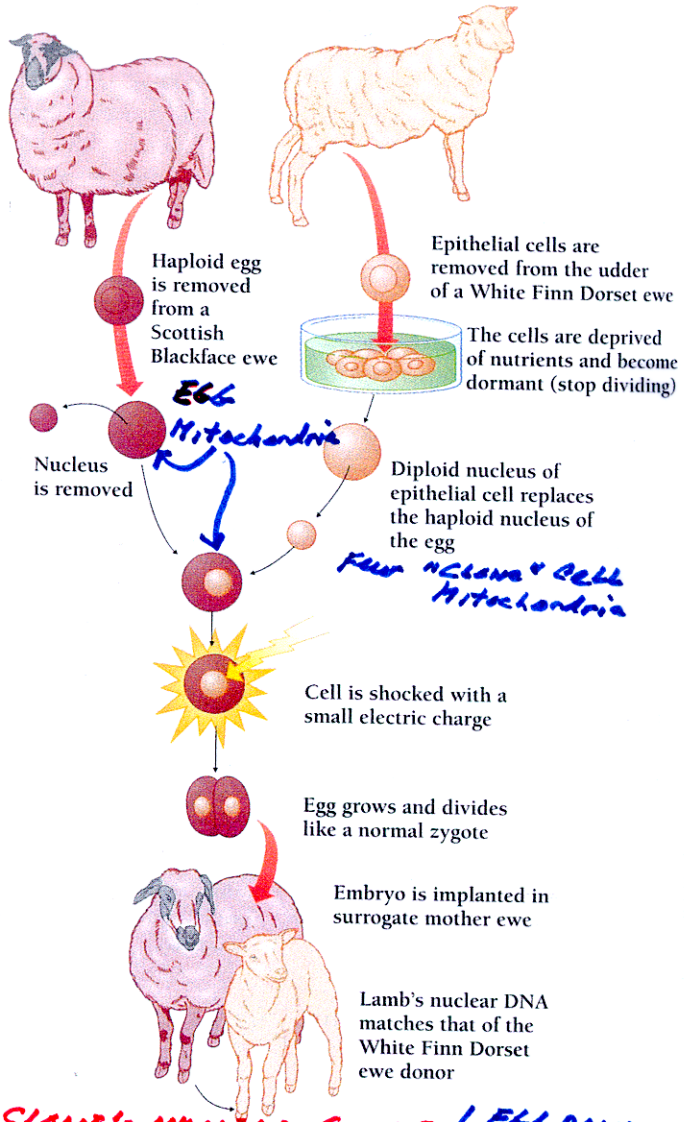


Figure 44-16 Cloning Dolly. The trick to cloning Dolly was to make differentiated cells less differentiated. By depriving the cultured udder cells of nutrients, the researchers induced the nuclei to enter a dormant state.

∴ CLONE NOT TECHNICALLY A CLONE WITH RESPECT TO MITOCHONDRIAL GENES

WILL BE MOSAIC - MOST MT FROM EGG DONOR - FEW MT FROM CELL USED FOR CLONE'S GENOME!!

Another Potential Problem Source for Clone development!

Several Mitochondrial Diseases Occur in Humans

In order for a human disorder to be attributable to genetically altered mitochondria, several criteria must be met.

1. Inheritance must exhibit a maternal rather than a Mendelian pattern.
2. The disorder must reflect a deficiency in the bioenergetic function of the organelle.
3. There must be a specific genetic mutation in one of the mitochondrial genes.

Thus far, several cases are known to demonstrate these characteristics. For example, **myoclonic epilepsy and ragged red fiber disease (MERRF)** demonstrates a pattern of inheritance consistent with maternal inheritance. Only offspring of affected mothers inherit the disorder; the offspring of affected fathers are all normal. Individuals with this rare disorder express deafness, dementia, and seizures. Both muscle fibers and mitochondria are affected; the aberrant mitochondria characterize what are described as ragged red fibers (**RRFs**) of skeletal muscle (Figure 8.5). Analysis of mtDNA has revealed a mutation in one of the mitochondrial genes encoding a transfer RNA. This genetic alteration apparently interferes with translation within the organelle, which in turn leads to the various manifestations of the disorder.

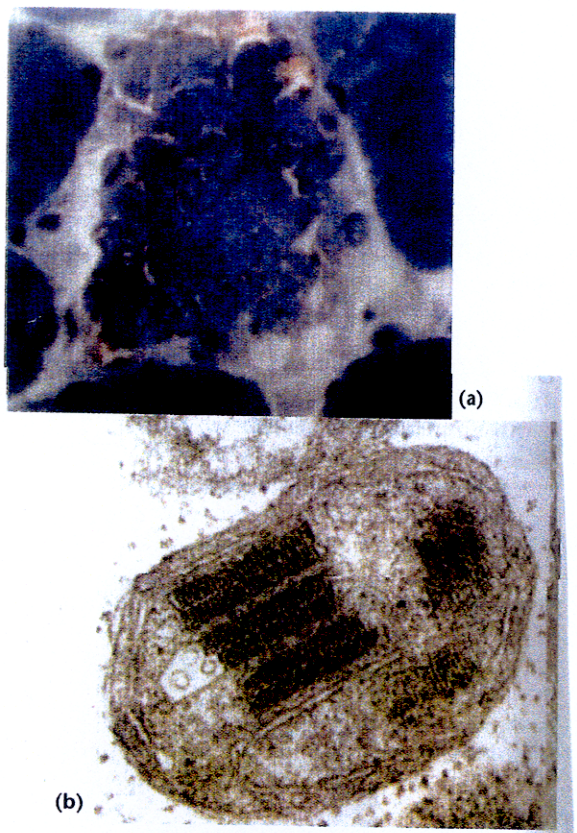


FIGURE 8.5 Mitochondrial myopathy in skeletal muscle cells of a patient with MERFF. Part (a) shows a ragged red fiber with abnormal mitochondria. Part (b) shows an abnormal mitochondrion revealing paracrystalline arrays within it.

A second disorder, **Leber's hereditary optic neuropathy (LHON)**, also exhibits maternal inheritance as well as mtDNA lesions. The disorder is characterized by sudden bilateral blindness. The average age of vision loss is 27, but onset is quite variable. Four mutations have been identified, all of which disrupt normal oxidative phosphorylation. Over 50 percent of cases are due to a mutation at a specific position in the mitochondrial gene encoding a subunit of NADH dehydrogenase so that the amino acid arginine is converted to histidine. This mutation is transmitted to all maternal offspring. It is interesting to note that in many instances of LHON, there is no family history; a significant number of cases appear to result from "new" mutations.

Individuals severely affected by a third disorder, **Kearns-Sayre syndrome (KSS)**, lose their vision, undergo hearing loss, and display heart conditions. The genetic basis of KSS involves deletions at various positions within mtDNA. Many KSS patients are symptom-free as children but display progressive symptoms as adults. The proportion of mtDNAs that reveal deletions increases as the severity of symptoms increases.

The study of hereditary mitochondrial-based disorders provides insights into the importance and genetic basis of this organelle during normal development, as well as the relationship between mitochondrial function and neuromuscular disorders. Such study has also suggested a hypothesis for aging based on the progressive accumulation of mtDNA mutations and the accompanying loss of mitochondrial function.

Mitochondrial Mutations LEADING TO 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.

MAP OF MITOCHONDRIAL GENES AND CORRESPONDING DISEASES

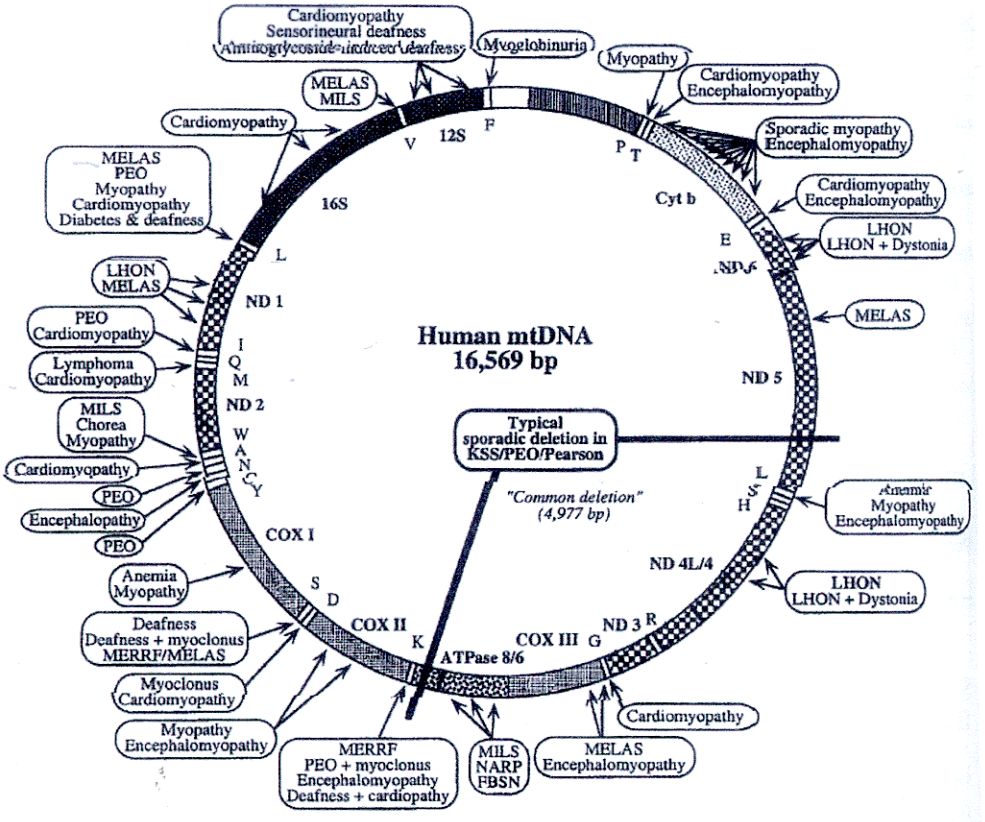


FIGURE 18.16. Morbidity map of the human mitochondrial genome. Abbreviations are for the genes encoding seven subunits of complex I (ND), three subunits of cytochrome c oxidase (COX), cytochrome b (Cyt b), and the two subunits of ATP synthase (ATPase 6 and 8). 12S and 16S refer to ribosomal RNAs; 22 transfer RNAs are identified by the one-letter codes for the corresponding amino acids. FBSN, familial bilateral striatal necrosis; KSS, Kearns–Sayre syndrome; LHON, Leber hereditary optic neuropathy; MELAS, mitochondrial encephalomyopathy, lactic acidosis, and strokelike episodes; MERRF, myoclonic epilepsy with ragged-red fibers; MILS, maternally inherited Leigh syndrome; NARP, neuropathy, ataxia, retinitis pigmentosa; PEO, progressive external ophthalmoplegia. From DiMauro and Schon (2001). Used with permission.

Mitochondrial Diseases ARE Inherited maternally

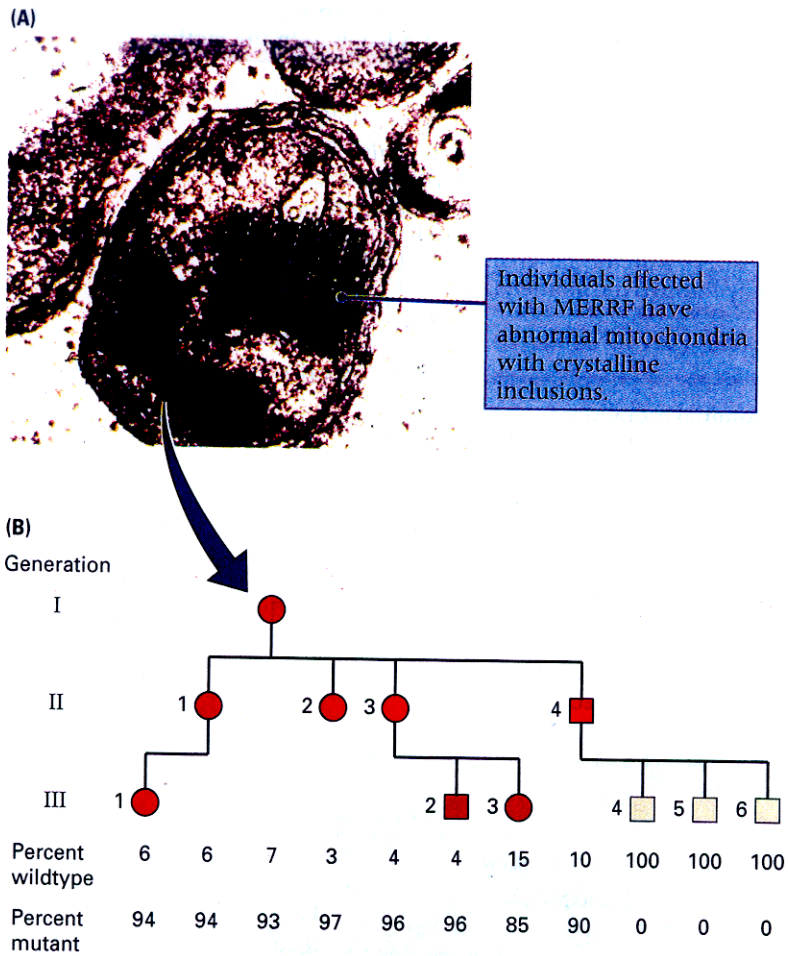


Figure 16.2 Inheritance of myoclonic epilepsy with ragged-red fiber disease (MERRF) in humans. (A) Electron micrograph of an abnormal MERRF mitochondrion containing paracrystalline inclusions. (B)The pedigree shows inheritance of MERRF in one family and the percentage of the mitochondria in each person 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.]

Never passed on from diseased ♂ to children

Mitochondria Absent from Sperm "Head"

Mitochondrial RFLP Markers can be used to Follow Diseased Mt Genes

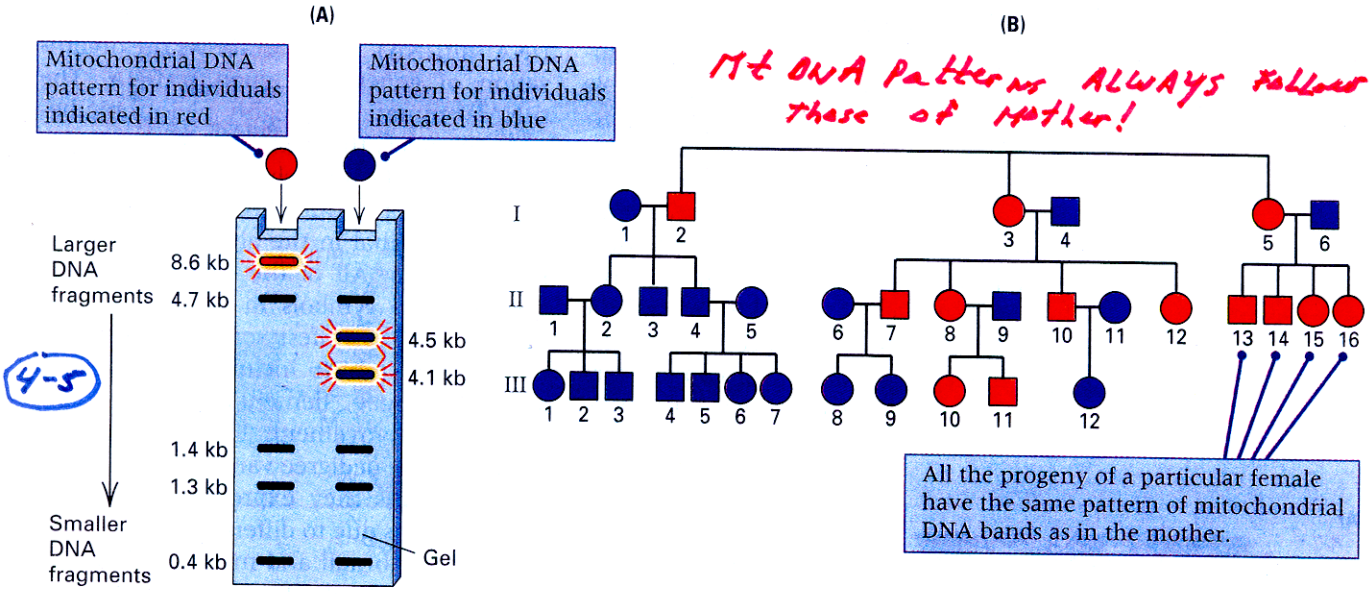


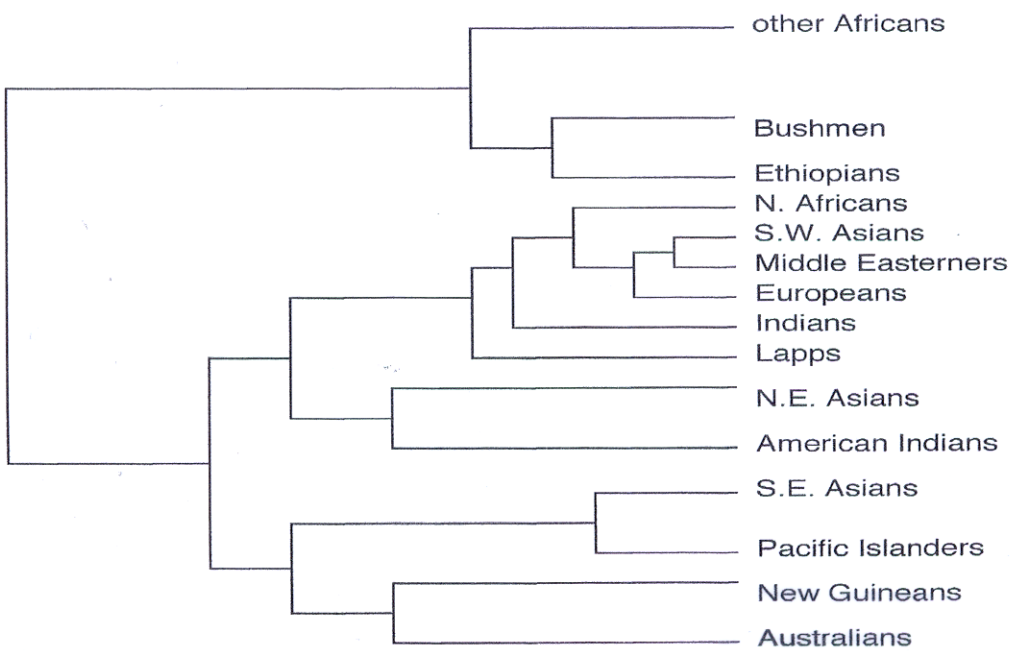
Figure 16.1 Maternal inheritance of human mitochondrial DNA. (A) Pattern of DNA fragments obtained when mitochondrial DNA is digested with the restriction enzyme *Hae*II. The DNA type at the left includes a fragment of 8.6 kb (red). The DNA type at the right contains a cleavage site for *Hae*II within 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.]

Because Mt Genome is SMALL (16kb) — Restriction fragments can be seen directly in Gel — no BLOT needed

How Many EcoRI fragments in Mt Genome with 50% G+C ?

Using Mt DNA Polymorphisms to construct trees of Human origins

Evolutionary Tree



MITOCHONDRIAL FINGERPRINTS / POLYMORPHISMS CAN BE USED TO STUDY HUMAN ORIGINS OR FIND THE FIRST "EVE"

HUMAN GENETICS SIDELIGHT

Using Mitochondrial DNA to Study Human Evolution

In biology few subjects are more fascinating than that of human evolution. Who are we? Where did we come from? Where are we going? Before the advent of molecular biology, the study of human evolution depended on the analysis of rare fossils—fragments of bone, a few teeth, an occasional weapon or tool. Today, human evolution can be studied by comparing DNA sequences. Each DNA sequence is descended from a sequence that was present in an ancestral organism. Thus, the DNA sequences that we find today are, in effect, living fossils—records of ancient DNA that has been transmitted through many generations to organisms currently alive. Because mutations may have occurred during this time, a modern DNA sequence is not likely to be an exact replica of its ancestor. However, by comparing modern DNA sequences, we can sometimes reconstruct features of the evolutionary process that produced them.

Some of the most insightful studies of human evolution have involved the analysis of mitochondrial DNA. There are two reasons why mtDNA is so useful: (1) it evolves faster than nuclear DNA, and (2) it is transmitted exclusively through the female. The rapidity of mtDNA evolution allows a scientist to detect significant genetic changes over a relatively short period of time (in evolutionary terms), and the strict maternal transmission of mtDNA allows a researcher to trace modern DNA sequences back to a common female ancestor.

Pioneering studies of human mtDNA were carried out in the 1980s by Allan Wilson, Rebecca Cann, Mark Stoneking, and their colleagues. These studies established that there is relatively little variation in the mtDNA from different human populations and that the greatest variation is found in the mtDNA from populations in Africa. Given the rate at which mtDNA is known to evolve, these discoveries suggested that modern human beings originated rather recently, probably within the last 200,000 years, and probably in Africa. Although these conclusions were initially controversial, later work has reinforced them.[1] Wilson's laboratory collected mtDNA samples from more than 200 individuals representing many different racial and ethnic groups. The mtDNA sequences in this collection were determined biochemically and then analyzed by a computer program that arranges the sequences in a phylogenetic, or evolutionary, tree. Wilson's conclusion was startling. The mtDNA in all modern groups of humans is descended from an mtDNA molecule that existed in a single woman who lived in Africa about 200,000 years ago. Applying a biblical metaphor, the popular press nicknamed this woman "Mitochondrial Eve."

By focusing on the evolution of mtDNA, Wilson's laboratory traced human ancestry back to a point where the maternal lineages of all modern mtDNA sequences coalesce in a single common ancestor—the mitochondrial mother of us all. However, these researchers never meant to imply that a single woman alone gave rise to all modern human beings. The mass of human nuclear DNA, which is inherited equally from males and females, and which varies among the members of a breeding population, cannot be traced to a single individual.

The work of Wilson and his colleagues strongly argues that all modern humans evolved from individuals who lived in Africa less than 200,000 years ago, and possibly as recently as 120,000 years ago. Migrants from this original African population presumably founded the archaic human populations of Europe and Asia, which, in turn, founded the early human populations of Australia, Oceana, and the Americas. This evolutionary scenario has been called the "Out of Africa" hypothesis. Another hypothesis proposes that humans evolved simultaneously in many regions of the world, from groups that were long established in those regions, perhaps for many hundreds of thousands of years, and that these groups probably interbred with other archaic populations such as the Neanderthals of Europe and western Asia.

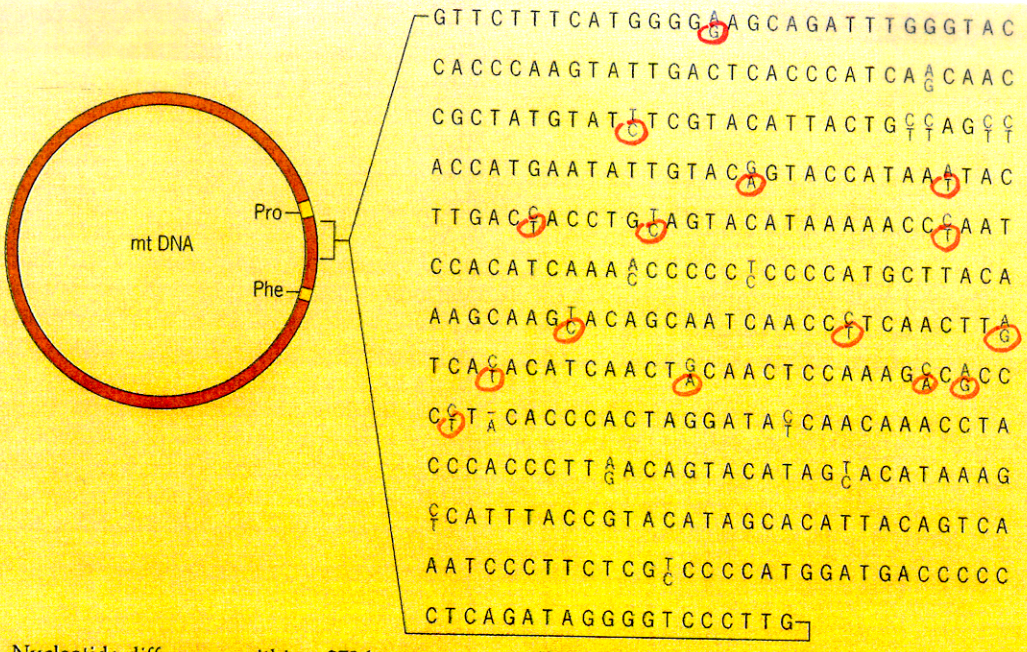
The Neanderthals have always been an enigmatic group for students of human evolution. Fossil remains indicate that they were quite different from modern humans; thicker bones, greater musculature, and different body proportions clearly set them apart. Were the Neanderthals ancestral to modern humans? Did they interbreed with the populations that ultimately produced modern humans, or were they a separate and distinct species altogether?

In 1997 Matthias Krings, Anne Stone, Ralf Schmitz, Heike Krainitzki, Mark Stoneking, and Svante Pääbo published the sequence of 379 base pairs of mtDNA extracted from a fossilized Neanderthal arm bone.[2] This particular fossil, discovered in 1856 near Dusseldorf, Germany, has been the subject of many intensive studies. After lengthy negotiations, the fossil's custodians granted Krings and co-workers permission to remove a 3.5-g piece of bone from the right humerus. Small fragments from this piece were pulverized, and the DNA remnants within them were carefully extracted. Because of the fossil's age (between 30,000 and 100,000 years), most of the DNA was expected to be degraded. However, because mtDNA is much more abundant than any particular sequence of nuclear DNA, Krings and co-workers hoped that some of it had survived. Their first step was to use a technique called the polymerase chain reaction (PCR, see Chapter 20) to amplify small segments of surviving mtDNA molecules. PCR allows a researcher to generate millions of identical DNA molecules from just a few molecules by *in vitro* replication with a bacterial DNA polymerase. The sequence of the amplified DNA can then be determined biochemically.

In carefully controlled experiments, Krings and coworkers succeeded in amplifying mtDNA remnants extracted from the fossil. Biochemical analysis of this ampli-

DNAs with shared polymorphisms are most closely related

NEANDERTHAL DNA SEQUENCES OBTAINED FROM BONES/FOSSILS USING PCR (ANCIENT DNA) indicate A SEPARATE LINE OF EVOLUTION



TOP nt alg HUMAN Mt DNA

Bottom Neanderthal Fossils

Figure 1. Nucleotide differences within a 379-bp noncoding region of the mtDNA of a Neanderthal fossil and that of a modern human being. The sequenced region lies between the genes for the phenylalanine (Phe) and proline (Pro) tRNAs. For each nucleotide difference (highlighted), the upper nucleotide is found in modern human mtDNA and the lower one is found in the Neanderthal mtDNA.

fied material showed that Neanderthal mtDNA differs from modern human mtDNA in 28 of the 379 nucleotides that were analyzed (Figure 1). The mtDNA isolated from different modern humans typically shows only 8 nucleotide substitutions in this region. Thus, Neanderthal mtDNA is significantly unlike that of modern humans. Computer analysis of the DNA sequences suggested that the human and Neanderthal mtDNA lineages began to evolve separately between 550,000 and 690,000 years ago, and that modern human mtDNAs originated between 120,000 and 150,000 years ago, apparently in Africa. Thus, Neanderthals were almost certainly not ancestral to modern humans. Rather, they evolved separately and, in the end, became extinct.

In the discussion section of their paper, Krings and coauthors concluded that "The Neanderthal mtDNA sequence thus supports a scenario in which modern humans arose recently in Africa as a distinct species and replaced Neanderthals with little or no interbreeding." They also added a caveat: "It must be emphasized that the above conclusions are based on a single individual sequence; the retrieval and analysis of mtDNA sequences from additional Neanderthal specimens is obviously desirable."[3] Of course, obtaining mtDNA sequences from other Neanderthals will entail the destruction of rare fossil material. Thus, the decision to collect such data should not be taken lightly. The benefit of collecting data from several individuals may not outweigh the cost of sacrificing so many valuable fossils. However, obtaining the sequence from at least one more Neanderthal does seem worthwhile, since this sequence could reinforce or invalidate the inferences that have to be made from the single sequence now available. We will have to wait and see if another Neanderthal fossil suitable for DNA analysis can be found. If it can, then the issue will be whether or not to allow part of that fossil to be destroyed to obtain a few molecules of mtDNA.

[1]Wilson, A. C., and R. L. Cann. 1992. The recent African genesis of humans. *Sci. Amer.*, 266(4):68–73.

[2]Krings, M., A. Stone, R. W. Schmitz, H. Krainitzki, M. Stoneking, and S. Pääbo. 1997. Neandertal DNA sequences and the origin of modern humans. *Cell* 90:19–30.

[3]ibid., p. 27.

USING MITOCHONDRIAL DNA IN FORENSICS – CZAR'S CHILDREN

TECHNICAL SIDELIGHT

DNA Tests and the Mystery of the Duchess Anastasia

According to historical records, the Russian royal family—Tsar Nicholas II, Tsarina Alexandra, and their five children: Alexis, Olga, Tatiana, Marie, and Anastasia (Figure 1)—were executed on July 16, 1918, by a revolutionary Bolshevik firing squad and then were buried in a single grave in the Ural Mountains. However, in 1920, an unknown woman, "Fraulein Unbekannt," who was pulled from a canal in Berlin in a state of hypothermia, claimed that she was the Duchess Anastasia. Although she did not speak Russian, Fraulein Unbekannt, or Anna Anderson Manahan, as she was subsequently known, was amazingly well informed about details of life in the imperial Russian court. Her claim to be Anastasia was vigorously rejected by the surviving relatives of the Russian royal family. The Grand Duke of Hesse even hired a private detective to investigate Anna's heritage. The detective concluded that Anna was really Franzisca Schanzkowska, but the dispute continued. Although little is known about Franzisca, she was born in the northern part of Germany, lived in Berlin during World War I, and was severely injured by an explosion while working in a munitions factory. She was subsequently admitted to two mental hospitals for treatment. She disappeared in 1920, about the same time that Anna Anderson Manahan was rescued from the Berlin canal and claimed to be Anastasia.

When Princess Irene of Prussia, Anastasia's aunt, was persuaded to meet with the woman who claimed to be her niece, Anna ran and hid in her room. Anna's bizarre behavior made her claim to be Anastasia difficult to evaluate, and the controversy over the identity of Anna Anderson Manahan continued for over 70 years. Was Anna really Anastasia? Her supporters were steadfast in their belief that she was indeed the Duchess. Disbelievers were equally adamant that she was not Anastasia.

In 1979, a Russian geologist discovered a shallow grave believed to contain the remains of the royal family. Because of the political climate in the Soviet Union at the time, the geologist reburied the bodies. Twelve years later, when the political climate was more favorable, the bodies were exhumed, and their authenticity was established by comparing DNA from the skeletons with DNA from surviving relatives. However, the controversy about the identity of Anna was rekindled by the absence of two bodies, those of Anastasia and her brother Alexis. Had they somehow escaped execution? Although there is still no definitive answer to this question, the results of recent DNA tests indicate that Anna Anderson Manahan was not the Duchess Anastasia.

Anna Anderson Manahan died in 1984 at the age of 83. However, during surgery performed in 1979 at the Martha Jefferson Hospital in Charlottesville, Virginia, intestinal tissues were removed, fixed in formaldehyde, and preserved

Figure 1 The children of Tsar Nicholas II: (left to right) Marie, Tatiana, Anastasia, Olga, and Alexis.

in paraffin blocks. In addition, a few of Anna's hair follicles were preserved. DNA tests—VNTR (variable number tandem repeat) prints and nucleotide sequences of specific noncoding regions of mitochondrial DNA—were performed on these preserved tissues and on relatives of Franzisca Schanzkowska and of the royal family. These tests were performed independently in three different laboratories: (1) the Armed Forces DNA Identification Laboratory in the United States, (2) the Forensic Science Service in the United Kingdom, and (3) the Department of Anthropology at Pennsylvania State University. The results obtained by the three laboratories all indicate that Anna Anderson Manahan was not Anastasia. Indeed, the results strongly suggest that Anna was Franzisca Schanzkowska.

Of five different VNTRs examined, four were inconsistent with the possibility that Anna was the daughter of Tsar Nicholas II and Tsarina Alexandra. DNA sequence comparisons also argued that Anna was not related to the royal family. Instead, the nucleotide sequence data indicated that Anna was Ms. Schanzkowska. At the six variable positions shown below, Anna's mitochondrial DNA contained the same nucleotides as the DNA from Carl Maucher, Franzisca Schanzkowska's grand nephew, and differed from those in the DNA of the Duke of Edinburgh, the grand nephew of Tsarina Alexandra.

	Variable Nucleotides in Mitochondrial DNA					
Position:	**1**	**2**	**3**	**4**	**5**	**6**
Anna Anderson Manahan	C	C	T	T	C	T
Carl Maucher (Grand nephew of Franzisca)	C	C	T	T	C	T
Duke of Edinburgh (Grand nephew of Alexandra)	T	T	C	C	T	C

NEED MT DNA + SEQUENCE OF REGION!

THE HUMAN GENOME SEQUENCE

articles

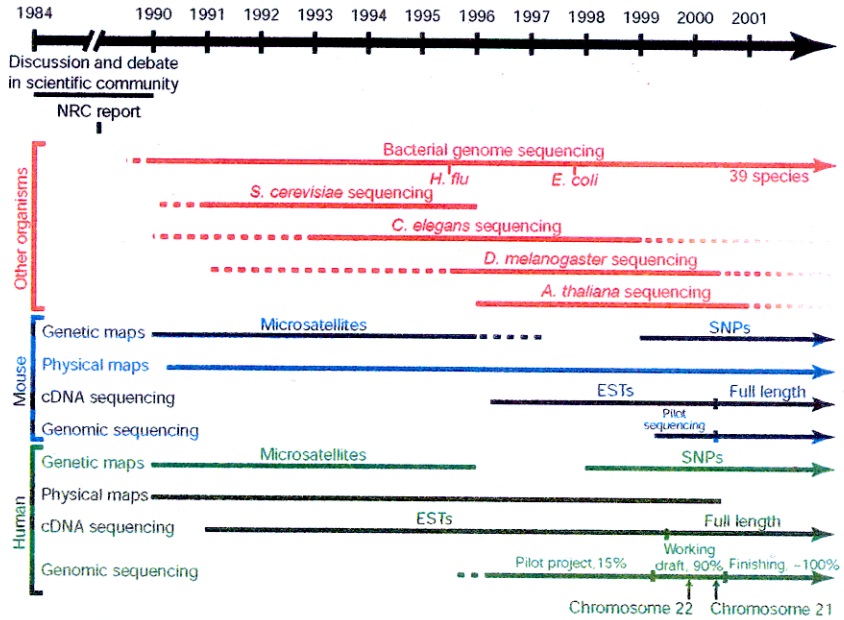
Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium*

* *A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.*

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

NATURE | VOL 409 | 15 FEBRUARY 2001 | www.nature.com



DONE!

Figure 1 Timeline of large-scale genomic analyses. Shown are selected components of work on several non-vertebrate model organisms (red), the mouse (blue) and the human (green) from 1990; earlier projects are described in the text. SNPs, single nucleotide polymorphisms; ESTs, expressed sequence tags.

16

WITHOUT AUTOMATION THE HUMAN GENOME COULD NOT HAVE BEEN SEQUENCED

PRODUCTION LINE

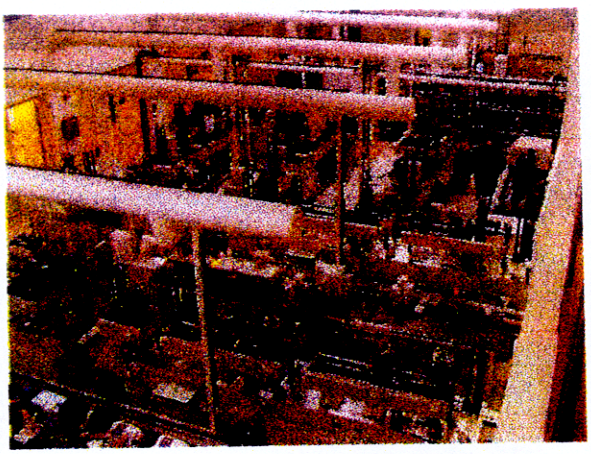


Figure 3 The automated production line for sample preparation at the Whitehead Institute, Center for Genome Research. The system consists of custom-designed factory-style conveyor belt robots that perform all functions from purifying DNA from bacterial cultures through setting up and purifying sequencing reactions.

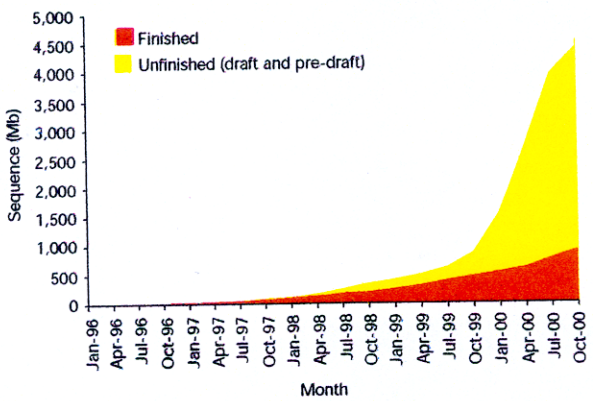


Figure 4 Total amount of human sequence in the High Throughput Genome Sequence (HTGS) division of GenBank. The total is the sum of finished sequence (red) and unfinished (draft plus predraft) sequence (yellow).

THE HUMAN GENOME SEQUENCE IS THE RESULT OF AN INTERNATIONAL COLLABORATION

Genome Sequencing Centres (Listed in order of total genomic sequence contributed, with a partial list of personnel. A full list of contributors at each centre is available as Supplementary Information.)

Whitehead Institute for Biomedical Research, Center for Genome Research: Eric S. Lander[1]*, Lauren M. Linton[1], Bruce Birren[1]*, Chad Nusbaum[1]*, Michael C. Zody[1]*, Jennifer Baldwin[1], Keri Devon[1], Ken Dewar[1], Michael Doyle[1], William FitzHugh[1]*, Roel Funke[1], Diane Gage[1], Katrina Harris[1], Andrew Heaford[1], John Howland[1], Lisa Kann[1], Jessica Lehoczky[1], Rosie LeVine[1], Paul McEwan[1], Kevin McKernan[1], James Meldrim[1], Jill P. Mesirov[1]*, Cher Miranda[1], William Morris[1], Jerome Naylor[1], Christina Raymond[1], Mark Rosetti[1], Ralph Santos[1], Andrew Sheridan[1], Carrie Sougnez[1], Nicole Stange-Thomann[1], Nikola Stojanovic[1], Aravind Subramanian[1] & Dudley Wyman[1]

The Sanger Centre: Jane Rogers[2], John Sulston[2]*, Rachael Ainscough[2], Stephan Beck[2], David Bentley[2], John Burton[2], Christopher Clee[2], Nigel Carter[2], Alan Coulson[2], Rebecca Deadman[2], Panos Deloukas[2], Andrew Dunham[2], Ian Dunham[2], Richard Durbin[2]*, Lisa French[2], Darren Grafham[2], Simon Gregory[2], Tim Hubbard[2]*, Sean Humphray[2], Adrienne Hunt[2], Matthew Jones[2], Christine Lloyd[2], Amanda McMurray[2], Lucy Matthews[2], Simon Mercer[2], Sarah Milne[2], James C. Mullikin[2]*, Andrew Mungall[2], Robert Plumb[2], Mark Ross[2], Ratna Shownkeen[2] & Sarah Sims[2]

Washington University Genome Sequencing Center: Robert H. Waterston[3]*, Richard K. Wilson[3], LaDeana W. Hillier[3]*, John D. McPherson[3], Marco A. Marra[3], Elaine R. Mardis[3], Lucinda A. Fulton[3], Asif T. Chinwalla[3]*, Kymberlie H. Pepin[3], Warren R. Gish[3], Stephanie L. Chissoe[3], Michael C. Wendl[3], Kim D. Delehaunty[3], Tracie L. Miner[3], Andrew Delehaunty[3], Jason B. Kramer[3], Lisa L. Cook[3], Robert S. Fulton[3], Douglas L. Johnson[3], Patrick J. Minx[3] & Sandra W. Clifton[3]

US DOE Joint Genome Institute: Trevor Hawkins[4], Elbert Branscomb[4], Paul Predki[4], Paul Richardson[4], Sarah Wenning[4], Tom Slezak[4], Norman Doggett[4], Jan-Fang Cheng[4], Anne Olsen[4], Susan Lucas[4], Christopher Elkin[4], Edward Uberbacher[4] & Marvin Frazier[4]

Baylor College of Medicine Human Genome Sequencing Center: Richard A. Gibbs[5]*, Donna M. Muzny[5], Steven E. Scherer[5], John B. Bouck[5]*, Erica J. Sodergren[5], Kim C. Worley[5]*, Catherine M. Rives[5], James H. Gorrell[5], Michael L. Metzker[5], Susan L. Naylor[6], Raju S. Kucherlapati[7], David L. Nelson, & George M. Weinstock[8]

RIKEN Genomic Sciences Center: Yoshiyuki Sakaki[9], Asao Fujiyama[9], Masahira Hattori[9], Tetsushi Yada[9], Atsushi Toyoda[9], Takehiko Itoh[9], Chiharu Kawagoe[9], Hidemi Watanabe[9], Yasushi Totoki[9] & Todd Taylor[9]

Genoscope and CNRS UMR-8030: Jean Weissenbach[10], Roland Heilig[10], William Saurin[10], Francois Artiguenave[10], Philippe Brottier[10], Thomas Bruls[10], Eric Pelletier[10], Catherine Robert[10] & Patrick Wincker[10]

GTC Sequencing Center: Douglas R. Smith[11], Lynn Doucette-Stamm[11], Marc Rubenfield[11], Keith Weinstock[11], Hong Mei Lee[11] & JoAnn Dubois[11]

Department of Genome Analysis, Institute of Molecular Biotechnology: André Rosenthal[12], Matthias Platzer[12], Gerald Nyakatura[12], Stefan Taudien[12] & Andreas Rump[12]

Beijing Genomics Institute/Human Genome Center: Huanming Yang[13], Jun Yu[13], Jian Wang[13], Guyang Huang[14] & Jun Gu[15]

Multimegabase Sequencing Center, The Institute for Systems Biology: Leroy Hood[16], Lee Rowen[16], Anup Madan[16] & Shizen Qin[16]

Stanford Genome Technology Center: Ronald W. Davis[17], Nancy A. Federspiel[17], A. Pia Abola[17] & Michael J. Proctor[17]

Stanford Human Genome Center: Richard M. Myers[18], Jeremy Schmutz[18], Mark Dickson[18], Jane Grimwood[18] & David R. Cox[18]

University of Washington Genome Center: Maynard V. Olson[19], Rajinder Kaul[19] & Christopher Raymond[19]

Department of Molecular Biology, Keio University School of Medicine: Nobuyoshi Shimizu[20], Kazuhiko Kawasaki[20] & Shinsei Minoshima[20]

University of Texas Southwestern Medical Center at Dallas: Glen A. Evans[21]†, Maria Athanasiou[21] & Roger Schultz[21]

University of Oklahoma's Advanced Center for Genome Technology: Bruce A. Roe[22], Feng Chen[22] & Huaqin Pan[22]

Max Planck Institute for Molecular Genetics: Juliane Ramser[23], Hans Lehrach[23] & Richard Reinhardt[23]

Cold Spring Harbor Laboratory, Lita Annenberg Hazen Genome Center: W. Richard McCombie[24], Melissa de la Bastide[24] & Neilay Dedhia[24]

GBF—German Research Centre for Biotechnology: Helmut Blöcker[25], Klaus Hornischer[25] & Gabriele Nordsiek[25]

* **Genome Analysis Group (listed in alphabetical order, also includes individuals listed under other headings):** Richa Agarwala[26], L. Aravind[26], Jeffrey A. Bailey[27], Alex Bateman[2], Serafim Batzoglou[1], Ewan Birney[28], Peer Bork[29,30], Daniel G. Brown[1], Christopher B. Burge[31], Lorenzo Cerutti[28], Hsiu-Chuan Chen[26], Deanna Church[26], Michele Clamp[2], Richard R. Copley[30], Tobias Doerks[29,30], Sean R. Eddy[32], Evan E. Eichler[27], Terrence S. Furey[33], James Galagan[1], James G. R. Gilbert[2], Cyrus Harmon[34], Yoshihide Hayashizaki[35], David Haussler[36], Henning Hermjakob[28], Karsten Hokamp[37], Wonhee Jang[26], L. Steven Johnson[32], Thomas A. Jones[32], Simon Kasif[38], Arek Kaspryzk[28], Scot Kennedy[39], W. James Kent[40], Paul Kitts[26], Eugene V. Koonin[26], Ian Korf[3], David Kulp[34], Doron Lancet[41], Todd M. Lowe[42], Aoife McLysaght[37], Tarjei Mikkelsen[38], John V. Moran[43], Nicola Mulder[28], Victor J. Pollara[1], Chris P. Ponting[44], Greg Schuler[26], Jörg Schultz[30], Guy Slater[28], Arian F. A. Smit[45], Ella Stupka[28], Joseph Szustakowki[38], Danielle Thierry-Mieg[26], Jean Thierry-Mieg[26], Lukas Wagner[26], John Wallis[3], Raymond Wheeler[34], Alan Williams[34], Yuri I. Wolf[26], Kenneth H. Wolfe[37], Shiaw-Pyng Yang[3] & Ru-Fang Yeh[31]

Scientific management: National Human Genome Research Institute, US National Institutes of Health: Francis Collins[46]*, Mark S. Guyer[46], Jane Peterson[46], Adam Felsenfeld[46]* & Kris A. Wetterstrand[46]; **Office of Science, US Department of Energy:** Aristides Patrinos[47]; **The Wellcome Trust:** Michael J. Morgan[48]

BUT IT WAS ALSO DONE INDEPENDENTLY BY A COMPANY – CELERA®

The Sequence of the Human Genome

J. Craig Venter,[1]* Mark D. Adams,[1] Eugene W. Myers,[1] Peter W. Li,[1] Richard J. Mural,[1] Granger G. Sutton,[1] Hamilton O. Smith,[1] Mark Yandell,[1] Cheryl A. Evans,[1] Robert A. Holt,[1] Jeannine D. Gocayne,[1] Peter Amanatides,[1] Richard M. Ballew,[1] Daniel H. Huson,[1] Jennifer Russo Wortman,[1] Qing Zhang,[1] Chinnappa D. Kodira,[1] Xiangqun H. Zheng,[1] Lin Chen,[1] Marian Skupski,[1] Gangadharan Subramanian,[1] Paul D. Thomas,[1] Jinghui Zhang,[1] George L. Gabor Miklos,[2] Catherine Nelson,[3] Samuel Broder,[1] Andrew G. Clark,[4] Joe Nadeau,[5] Victor A. McKusick,[6] Norton Zinder,[7] Arnold J. Levine,[7] Richard J. Roberts,[8] Mel Simon,[9] Carolyn Slayman,[10] Michael Hunkapiller,[11] Randall Bolanos,[1] Arthur Delcher,[1] Ian Dew,[1] Daniel Fasulo,[1] Michael Flanigan,[1] Liliana Florea,[1] Aaron Halpern,[1] Sridhar Hannenhalli,[1] Saul Kravitz,[1] Samuel Levy,[1] Clark Mobarry,[1] Knut Reinert,[1] Karin Remington,[1] Jane Abu-Threideh,[1] Ellen Beasley,[1] Kendra Biddick,[1] Vivien Bonazzi,[1] Rhonda Brandon,[1] Michele Cargill,[1] Ishwar Chandramouliswaran,[1] Rosane Charlab,[1] Kabir Chaturvedi,[1] Zuoming Deng,[1] Valentina Di Francesco,[1] Patrick Dunn,[1] Karen Eilbeck,[1] Carlos Evangelista,[1] Andrei E. Gabrielian,[1] Weiniu Gan,[1] Wangmao Ge,[1] Fangcheng Gong,[1] Zhiping Gu,[1] Ping Guan,[1] Thomas J. Heiman,[1] Maureen E. Higgins,[1] Rui-Ru Ji,[1] Zhaoxi Ke,[1] Karen A. Ketchum,[1] Zhongwu Lai,[1] Yiding Lei,[1] Zhenya Li,[1] Jiayin Li,[1] Yong Liang,[1] Xiaoying Lin,[1] Fu Lu,[1] Gennady V. Merkulov,[1] Natalia Milshina,[1] Helen M. Moore,[1] Ashwinikumar K Naik,[1] Vaibhav A. Narayan,[1] Beena Neelam,[1] Deborah Nusskern,[1] Douglas B. Rusch,[1] Steven Salzberg,[12] Wei Shao,[1] Bixiong Shue,[1] Jingtao Sun,[1] Zhen Yuan Wang,[1] Aihui Wang,[1] Xin Wang,[1] Jian Wang,[1] Ming-Hui Wei,[1] Ron Wides,[13] Chunlin Xiao,[1] Chunhua Yan,[1] Alison Yao,[1] Jane Ye,[1] Ming Zhan,[1] Weiqing Zhang,[1] Hongyu Zhang,[1] Qi Zhao,[1] Liansheng Zheng,[1] Fei Zhong,[1] Wenyan Zhong,[1] Shiaoping C. Zhu,[1] Shaying Zhao,[12] Dennis Gilbert,[1] Suzanna Baumhueter,[1] Gene Spier,[1] Christine Carter,[1] Anibal Cravchik,[1] Trevor Woodage,[1] Feroze Ali,[1] Huijin An,[1] Aderonke Awe,[1] Danita Baldwin,[1] Holly Baden,[1] Mary Barnstead,[1] Ian Barrow,[1] Karen Beeson,[1] Dana Busam,[1] Amy Carver,[1] Angela Center,[1] Ming Lai Cheng,[1] Liz Curry,[1] Steve Danaher,[1] Lionel Davenport,[1] Raymond Desilets,[1] Susanne Dietz,[1] Kristina Dodson,[1] Lisa Doup,[1] Steven Ferriera,[1] Neha Garg,[1] Andres Gluecksmann,[1] Brit Hart,[1] Jason Haynes,[1] Charles Haynes,[1] Cheryl Heiner,[1] Suzanne Hladun,[1] Damon Hostin,[1] Jarrett Houck,[1] Timothy Howland,[1] Chinyere Ibegwam,[1] Jeffery Johnson,[1] Francis Kalush,[1] Lesley Kline,[1] Shashi Koduru,[1] Amy Love,[1] Felecia Mann,[1] David May,[1] Steven McCawley,[1] Tina McIntosh,[1] Ivy McMullen,[1] Mee Moy,[1] Linda Moy,[1] Brian Murphy,[1] Keith Nelson,[1] Cynthia Pfannkoch,[1] Eric Pratts,[1] Vinita Puri,[1] Hina Qureshi,[1] Matthew Reardon,[1] Robert Rodriguez,[1] Yu-Hui Rogers,[1] Deanna Romblad,[1] Bob Ruhfel,[1] Richard Scott,[1] Cynthia Sitter,[1] Michelle Smallwood,[1] Erin Stewart,[1] Renee Strong,[1] Ellen Suh,[1] Reginald Thomas,[1] Ni Ni Tint,[1] Sukyee Tse,[1] Claire Vech,[1] Gary Wang,[1] Jeremy Wetter,[1] Sherita Williams,[1] Monica Williams,[1] Sandra Windsor,[1] Emily Winn-Deen,[1] Keriellen Wolfe,[1] Jayshree Zaveri,[1] Karena Zaveri,[1] Josep F. Abril,[14] Roderic Guigó,[14] Michael J. Campbell,[1] Kimmen V. Sjolander,[1] Brian Karlak,[1] Anish Kejariwal,[1] Huaiyu Mi,[1] Betty Lazareva,[1] Thomas Hatton,[1] Apurva Narechania,[1] Karen Diemer,[1] Anushya Muruganujan,[1] Nan Guo,[1] Shinji Sato,[1] Vineet Bafna,[1] Sorin Istrail,[1] Ross Lippert,[1] Russell Schwartz,[1] Brian Walenz,[1] Shibu Yooseph,[1] David Allen,[1] Anand Basu,[1] James Baxendale,[1] Louis Blick,[1] Marcelo Caminha,[1] John Carnes-Stine,[1] Parris Caulk,[1] Yen-Hui Chiang,[1] My Coyne,[1] Carl Dahlke,[1] Anne Deslattes Mays,[1] Maria Dombroski,[1] Michael Donnelly,[1] Dale Ely,[1] Shiva Esparham,[1] Carl Fosler,[1] Harold Gire,[1] Stephen Glanowski,[1] Kenneth Glasser,[1] Anna Glodek,[1] Mark Gorokhov,[1] Ken Graham,[1] Barry Gropman,[1] Michael Harris,[1] Jeremy Heil,[1] Scott Henderson,[1] Jeffrey Hoover,[1] Donald Jennings,[1] Catherine Jordan,[1] James Jordan,[1] John Kasha,[1] Leonid Kagan,[1] Cheryl Kraft,[1] Alexander Levitsky,[1] Mark Lewis,[1] Xiangjun Liu,[1] John Lopez,[1] Daniel Ma,[1] William Majoros,[1] Joe McDaniel,[1] Sean Murphy,[1] Matthew Newman,[1] Trung Nguyen,[1] Ngoc Nguyen,[1] Marc Nodell,[1] Sue Pan,[1] Jim Peck,[1] Marshall Peterson,[1] William Rowe,[1] Robert Sanders,[1] John Scott,[1] Michael Simpson,[1] Thomas Smith,[1] Arlan Sprague,[1] Timothy Stockwell,[1] Russell Turner,[1] Eli Venter,[1] Mei Wang,[1] Meiyuan Wen,[1] David Wu,[1] Mitchell Wu,[1] Ashley Xia,[1] Ali Zandieh,[1] Xiaohong Zhu[1]

AND COMPLETED IN ONLY Nine Months!

IT WAS A RACE!

Celera Effort

5 people's DNA combined & Sequenced

3♀ 2♂

2 Caucasian
1 African American
1 Chinese
1 Hispanic

A 2.91-billion base pair (bp) consensus sequence of the euchromatic portion of the human genome was generated by the whole-genome shotgun sequencing method. The 14.8-billion bp DNA sequence was generated over 9 months from 27,271,853 high-quality sequence reads (5.11-fold coverage of the genome) from both ends of plasmid clones made from the DNA of five individuals. Two assembly strategies—a whole-genome assembly and a regional chromosome assembly—were used, each combining sequence data from Celera and the publicly funded genome effort. The public data were shredded into 550-bp segments to create a 2.9-fold coverage of those genome regions that had been sequenced, without including biases inherent in the cloning and assembly procedure used by the publicly funded group. This brought the effective coverage in the assemblies to eightfold, reducing the number and size of gaps in the final assembly over what would be obtained with 5.11-fold coverage. The two assembly strategies yielded very similar results that largely agree with independent mapping data. The assemblies effectively cover the euchromatic regions of the human chromosomes. More than 90% of the genome is in scaffold assemblies of 100,000 bp or more, and 25% of the genome is in scaffolds of 10 million bp or larger. Analysis of the genome sequence revealed 26,588 protein-encoding transcripts for which there was strong corroborating evidence and an additional ~12,000 computationally derived genes with mouse matches or other weak supporting evidence. Although gene-dense clusters are obvious, almost half the genes are dispersed in low G+C sequence separated by large tracts of apparently noncoding sequence. Only 1.1% of the genome is spanned by exons, whereas 24% is in introns, with 75% of the genome being intergenic DNA. Duplications of segmental blocks, ranging in size up to chromosomal lengths, are abundant throughout the genome and reveal a complex evolutionary history. Comparative genomic analysis indicates vertebrate expansions of genes associated with neuronal function, with tissue-specific developmental regulation, and with the hemostasis and immune systems. DNA sequence comparisons between the consensus sequence and publicly funded genome data provided locations of 2.1 million single-nucleotide polymorphisms (SNPs). A random pair of human haploid genomes differed at a rate of 1 bp per 1250 on average, but there was marked heterogeneity in the level of polymorphism across the genome. Less than 1% of all SNPs resulted in variation in proteins, but the task of determining which SNPs have functional consequences remains an open challenge.

1% Exons
24% Introns
75% Intergenic
100% DNA

1 SNP / 1250 bp on average between two people

BUT The International Public Sequence is More Complete than the Private one That contains Gaps

READ –

"The GENOME WAR"

by James Shreeve

ISBN 0375406298

JANUARY, 2004

Private vs. Public Genome Projects!