

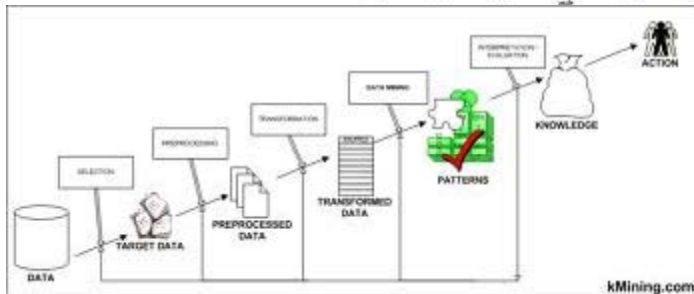
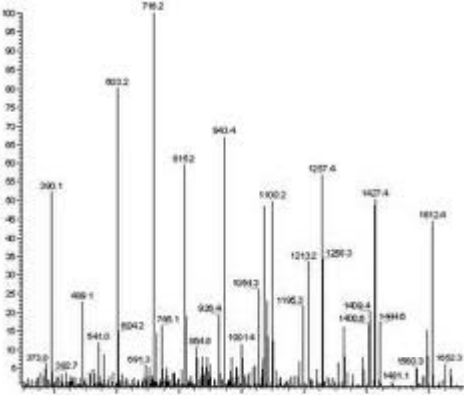
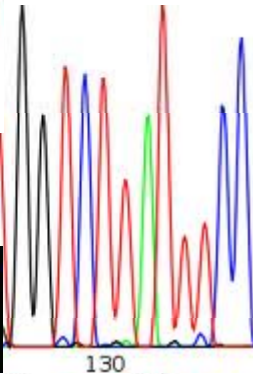
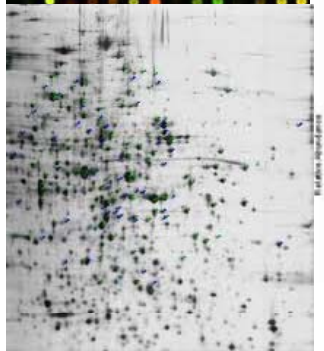
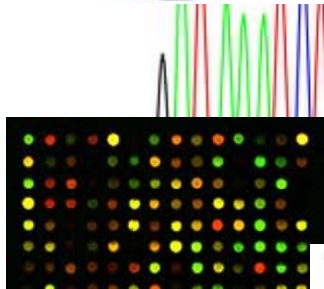
Curso de Posgrado: Alimentos Funcionales

BIONFORMATICA EN EL AMBITO DE LA NUTRIGENOMICA

Norma Paniego

Instituto de Biotecnología, CICVyA, CNIA INTA Castelar





CONCEPTOS GENERALES

NUTRIGENOMICA

NUTRIGENETICA

GENOMICA

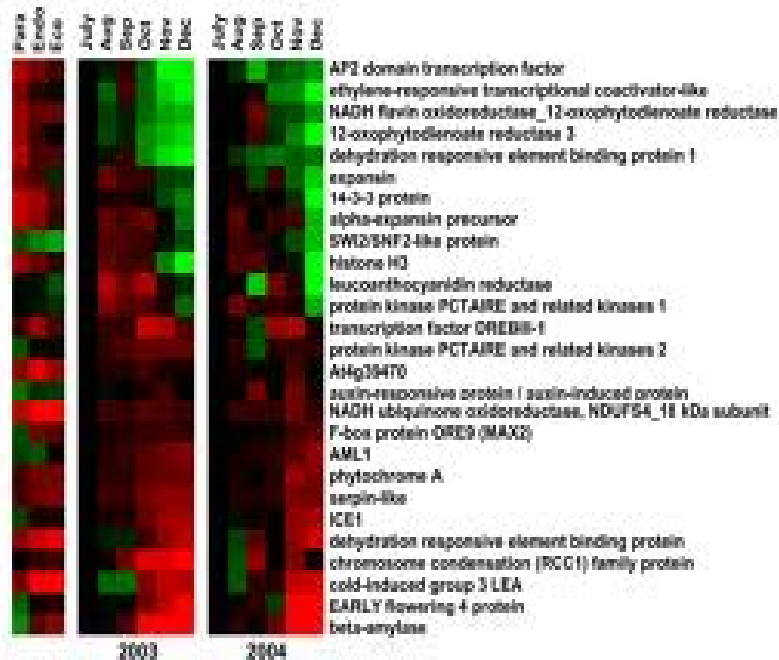
POST-GENOMICA

BASES DE DATOS

EXTRACCIÓN DEL CONOCIMIENTO

Nutrigenómica & Nutrigenética

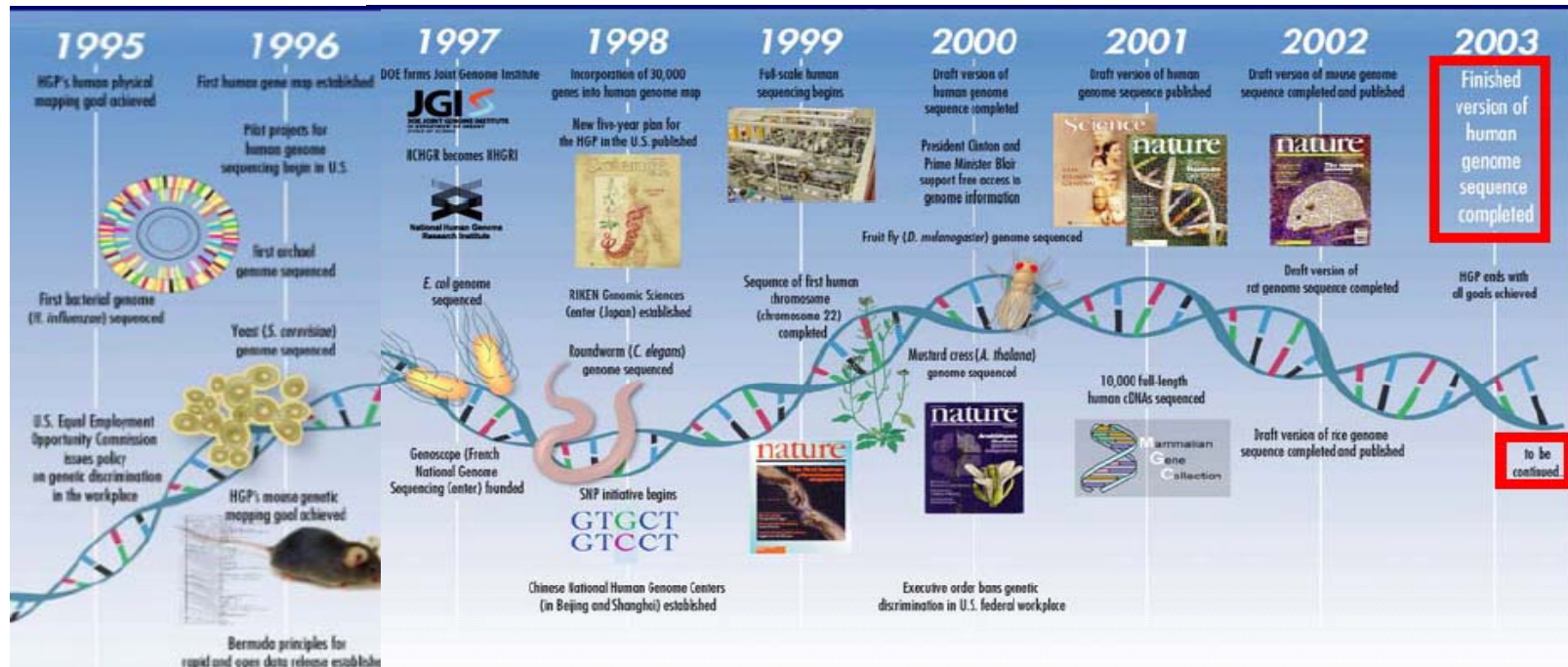
La **nutrigenómica** es la ciencia que estudia la **expresión de los genes** en relación con la nutrición y el desarrollo de enfermedades asociadas a dicha expresión.



La **nutrigenética** es la ciencia que estudia como el **genotipo** del individuo interacciona con ingesta de nutrientes.



Proyectos genómicos



NCBI **ENTREZ** Genome Project connection information discovery

Entrez Genome Gene Nucleotide Protein PopSet PubMed Taxonomy

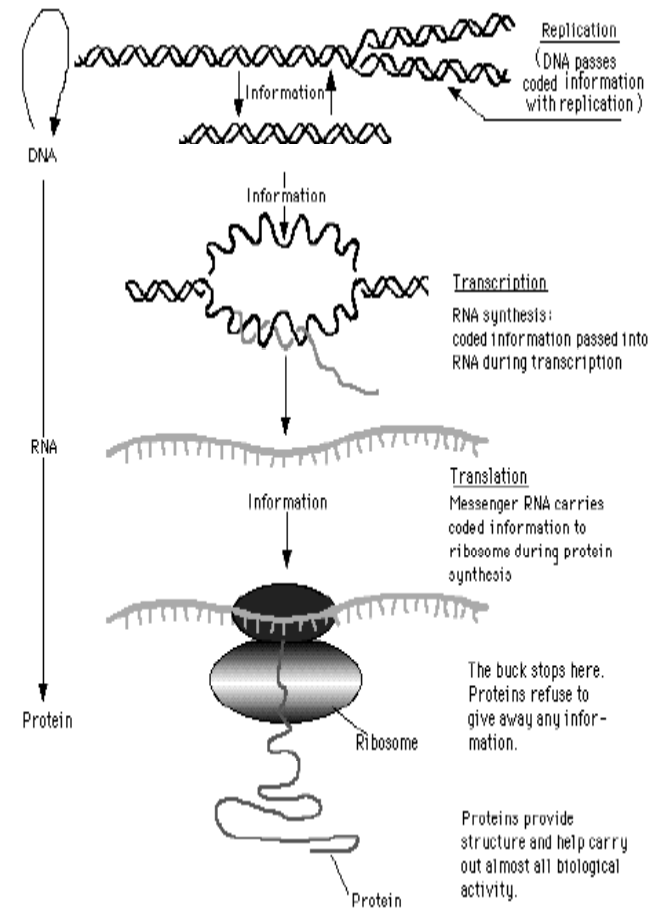
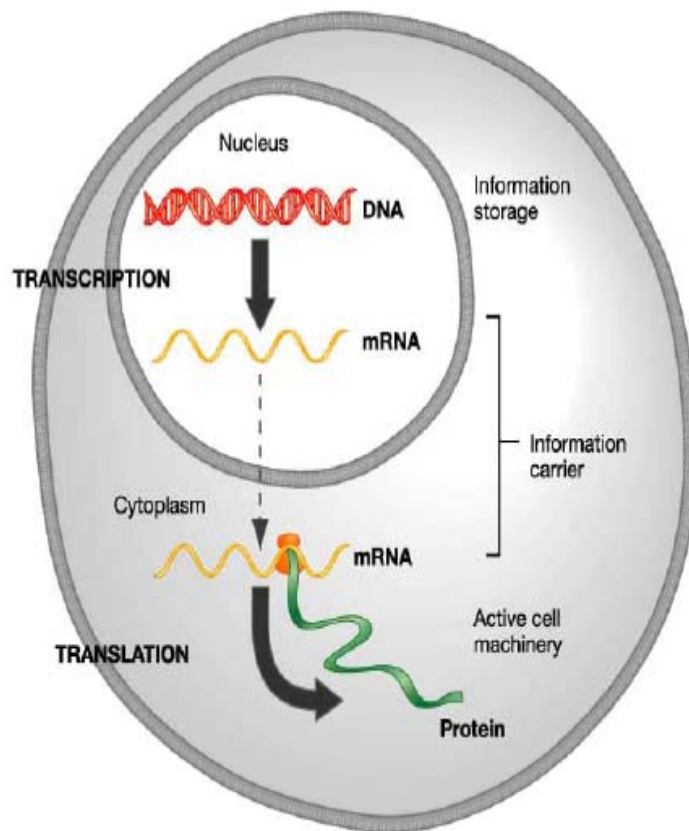
Search

Properties of Eukaryotic Genome Sequencing Projects

Organism Group: Sequencing Status: Sequencing Method:

Abbreviations: **GB** - GenBank Accessions; **PM** - PubMed; **R** - RefSeq Accessions; **G** - Entrez Gene; **T** - Trace Archive; **B** - BLAST; **M** - Map Viewer; **F** - FTP Sites

El Dogma Central de la Biología Molecular



Niveles básicos de Información biológica

- **Genoma:** es la totalidad de la información genética que posee un organismo en particular.
- **Transcriptoma:** la parte del genoma que se expresa en una célula en una etapa específica de su desarrollo.
- **Proteoma:** la totalidad de las proteínas codificadas por un genoma.
- **Metaboloma:** la totalidad de pequeñas moléculas o metabolitos que se pueden encontrar en una muestra biológica, tal como un organismo.

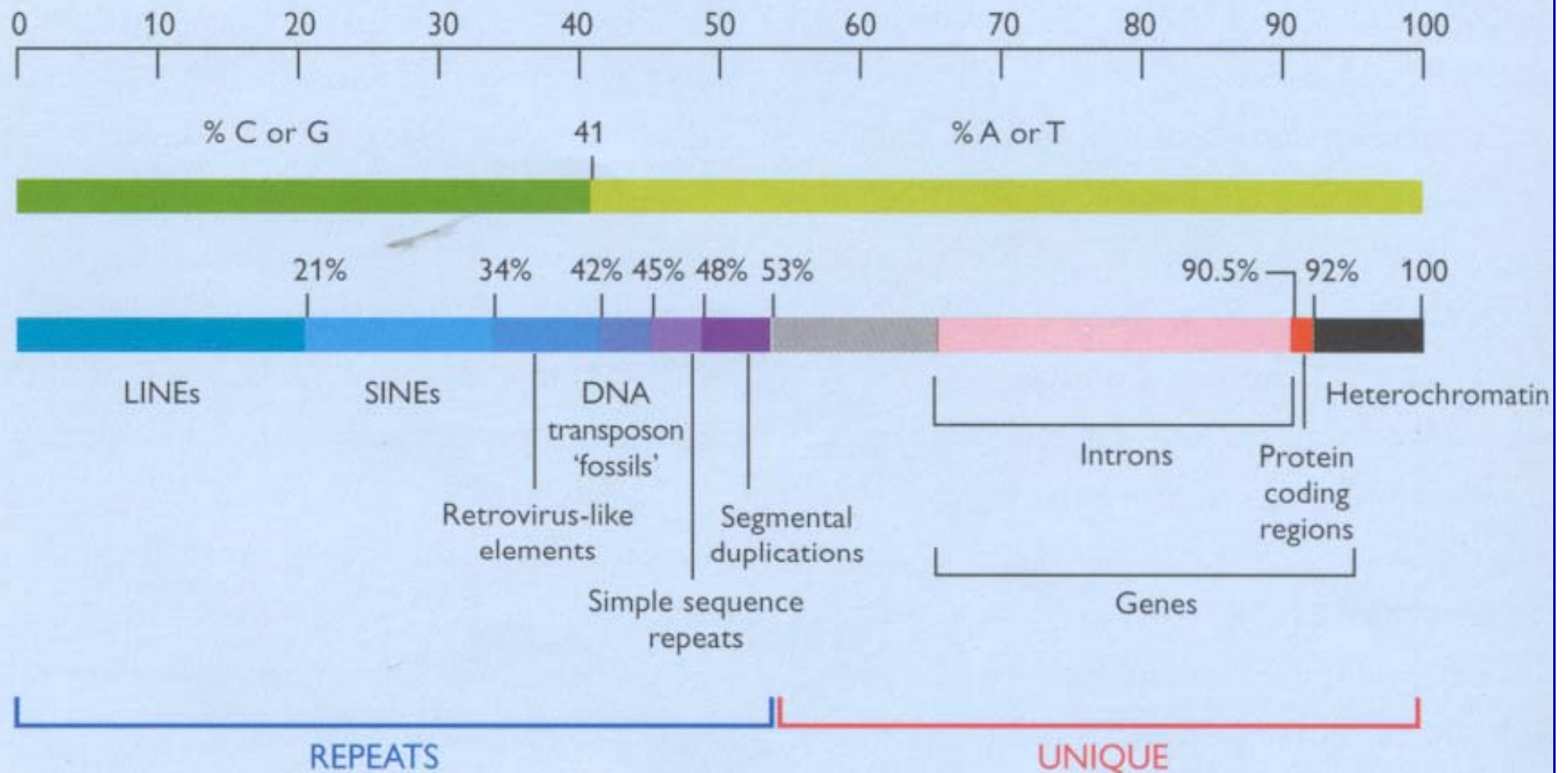
La era de la genómica

La genómica se ha desarrollado como consecuencia de los avances en Biología Molecular e Informática.

La introducción y popularización de las tecnologías de alta procesividad ha cambiado drásticamente la manera en que se abordan los problemas biológicos y se prueban las hipótesis.

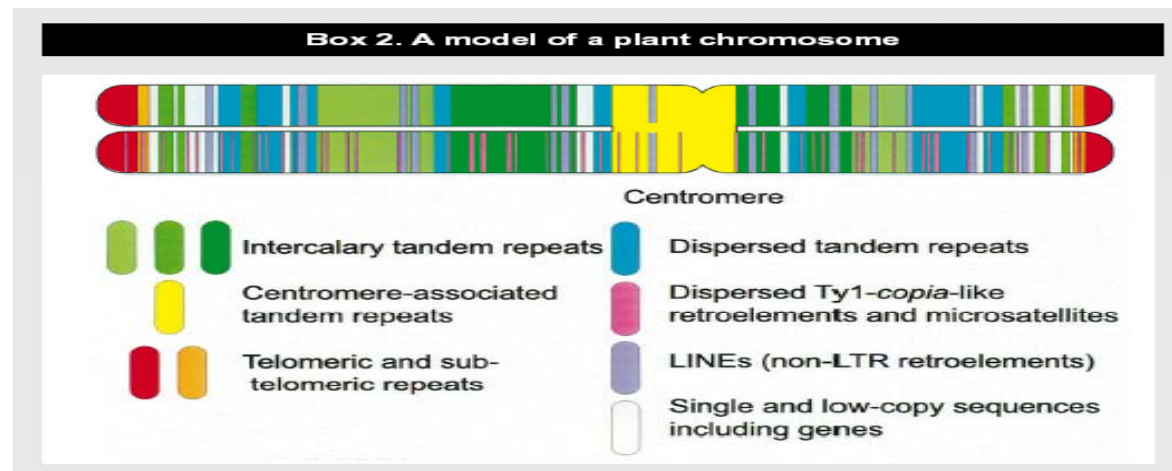
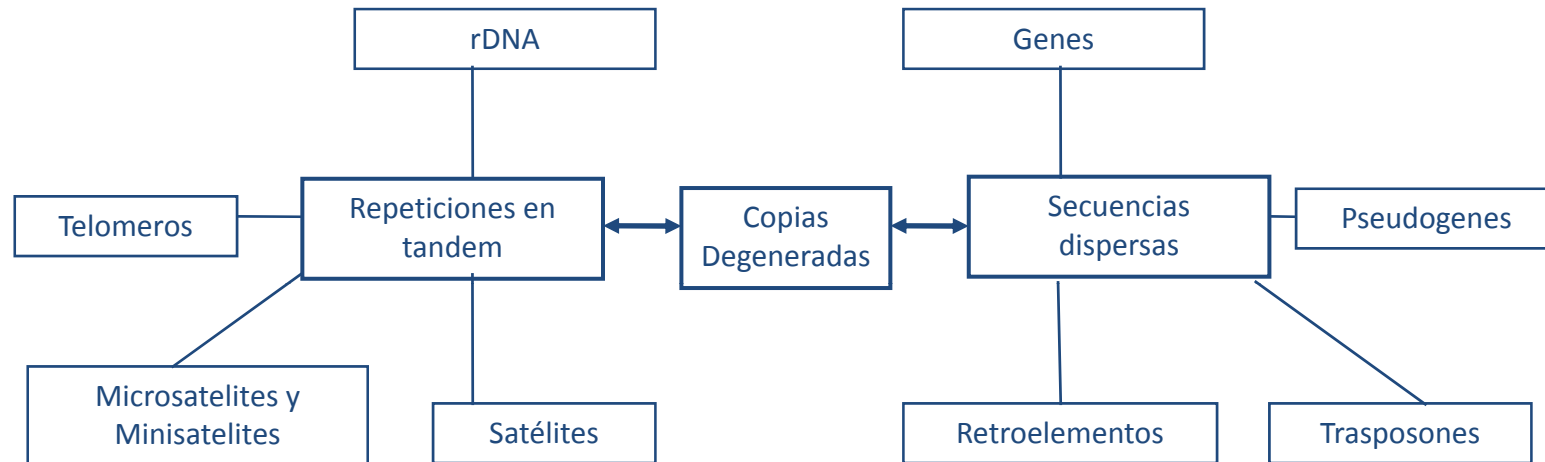
Inicialmente los proyectos genómicos se focalizaban en la secuencia completa del genoma, actualmente incluyen el análisis de la expresión y función de genes, proteínas, metabolitos entre otros.

Figure 1: The genome by numbers



[The Genome by Numbers](#), the Wellcome Trust.

Contenido de secuencias de ADN en el genoma vegetal

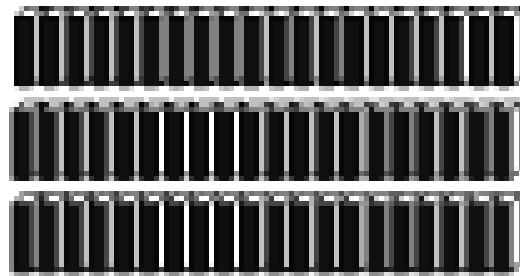


Single nucleotide variant	ATTGGCCTTAACCCCGATTATCAGGAT ATTGGCCTTAACCTCCGATTATCAGGAT	Structural variants
Insertion–deletion variant	ATTGGCCTTAACCCGATCCGATTATCAGGAT ATTGGCCTTAACCC---CCGATTATCAGGAT	
Block substitution	ATTGGCCTTAACCCCGATTATCAGGAT ATTGGCCTTAACAGTGATTATCAGGAT	
Inversion variant	ATTGGCCTTAACCCCGATTATCAGGAT ATTGGCCTTCGGGGGTTATTATCAGGAT	
Copy number variant	ATTGGCCTTAGGCCTTAACCCCGATTATCAGGAT ATTGGCCTTA-----ACCTCCGATTATCAGGAT	

Figure 1 | **Classes of human genetic variants.** The nomenclature used to describe the various types of structural variants is not yet standard¹²¹. Here, the terminology used aims to describe the nucleotide composition of the variant and distinguish it from other types of variants. Single nucleotide variants are DNA sequence variations in which a single nucleotide (A, T, G or C) is altered. Insertion–deletion variants (indels) occur when one or more base pairs are present in some genomes but absent in others. They are generally composed of only a few bases but can be greater than 80 kb in length¹¹. Block substitutions describe cases in which a string of adjacent nucleotides varies between two genomes. An inversion variant is one in which the order of the base pairs is reversed in a defined section of a chromosome. A well-characterized inversion variant that has been described in humans involves a section of chromosome 17 in which a ~900 kb interval is in the reverse order in approximately 20% of individuals with Northern European ancestry¹²². Copy number variants occur when identical or nearly identical sequences are repeated in some chromosomes but not others. The largest copy number variant identified in the Venter genome¹¹ was almost 2 Mb in length.

Genome Sizes

Human
Mouse



~3,000,000,000 bp

Fruit Fly



~160,000,000 bp

Nematode



~100,000,000 bp

Yeast



~15,000,000 bp

E. coli



~5,000,000 bp

YAC



~1,000,000 bp

BAC



~100,000 bp

Cosmid



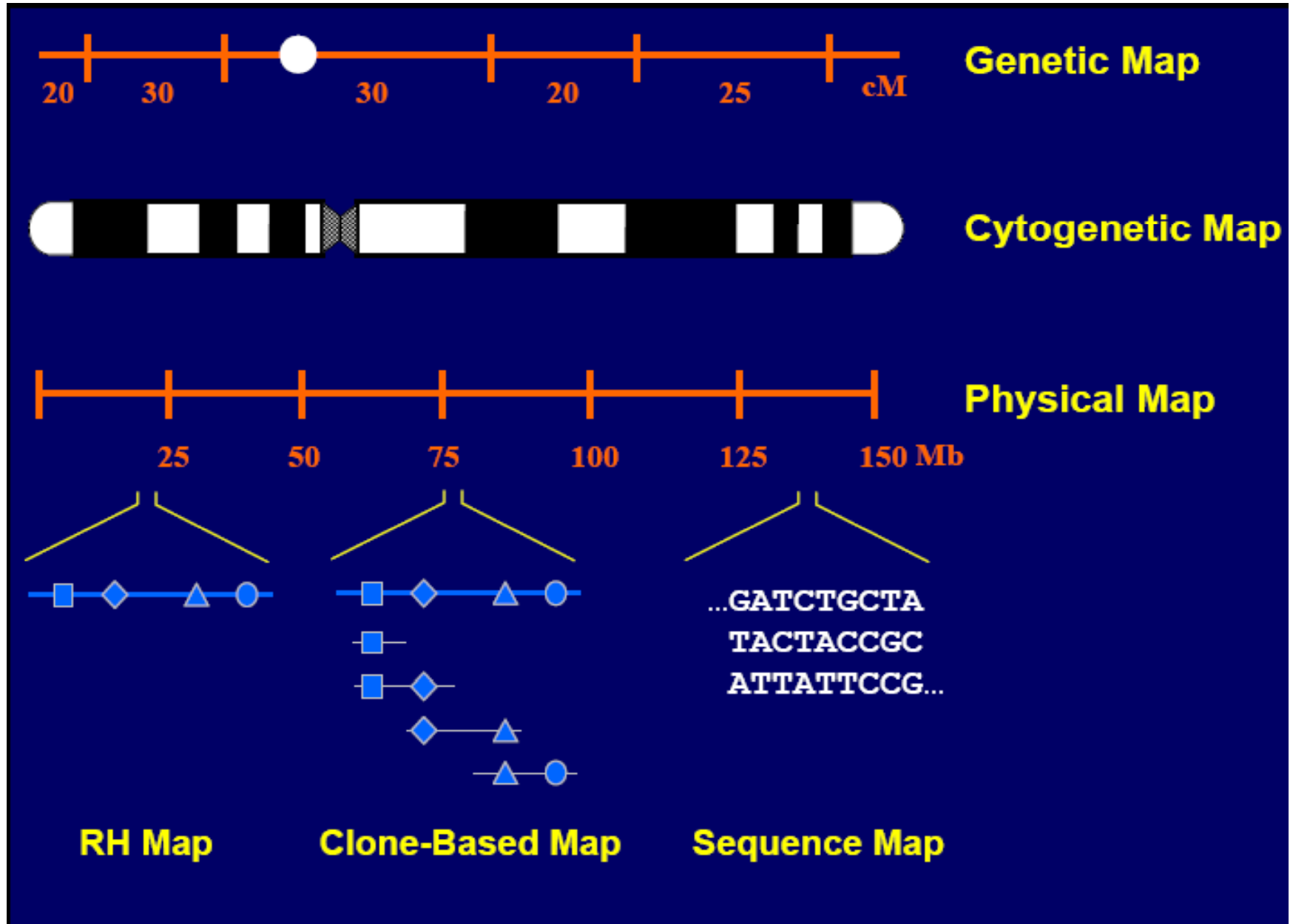
~45,000 bp

Bacteriophage



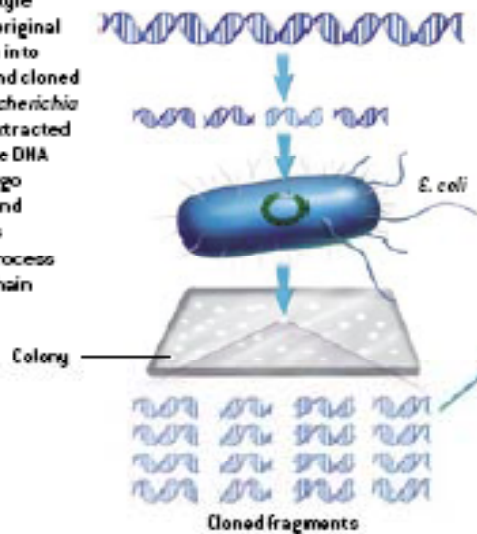
~25,000 bp

Como se estudian los genomas: técnicas de mapeo



Método de secuenciación

1 Before Sanger-style sequencing, an original DNA strand is broken into smaller fragments and cloned within colonies of *Escherichia coli* bacteria. Once extracted from the bacteria, the DNA fragments will undergo another massive round of copying, known as amplification, by a process called polymerase chain reaction (PCR).



2 During PCR, fragments are heated so they will separate into single strands. A short nucleotide sequence called a primer is then annealed to each original template. Starting at the primer, polymerase links free-floating nucleotides (called dNTPs) into new complementary strands. The process is repeated over and over to generate millions of copies of each fragment.

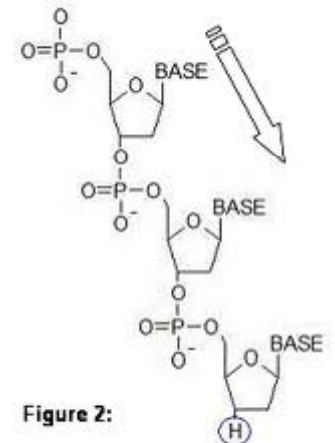
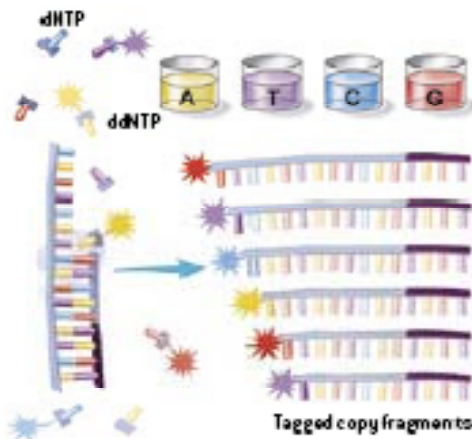


Figure 2:

3 Single-stranded fragments are next tagged in a process similar to PCR but with fluorescently labeled terminator nucleotides (ddNTPs) added to the mixture of primers, polymerase and dNTPs. Complementary strands are built until by chance a ddNTP is incorporated, halting synthesis. The resulting copy fragments have varying lengths and a tagged nucleotide at one end.



4 Capillary electrophoresis separates the fragments, which are negatively charged, drawing them toward a positively charged pole. Because the shortest fragment travels fastest, their order reflects their size and their ddNTP terminators can thus be read in the template's base sequence. Laser light activates the fluorescent tags as the fragments pass through a detection window, producing a color readout that is translated into a sequence.

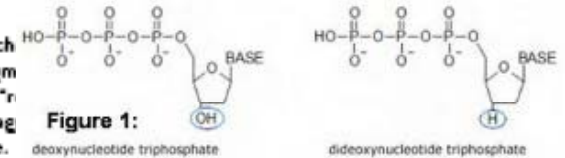
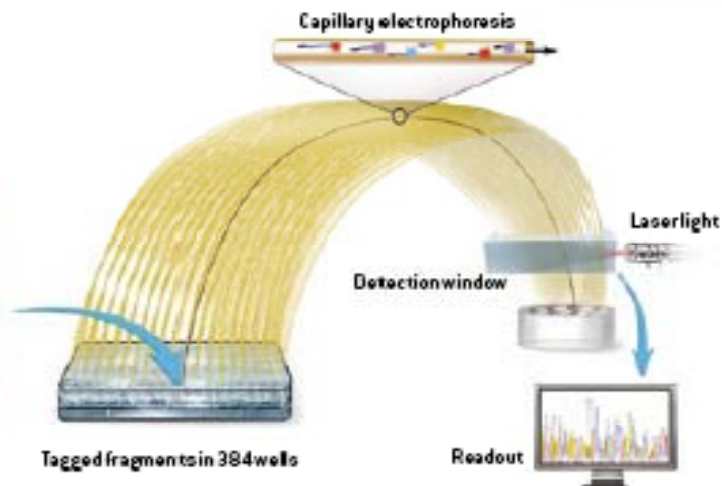
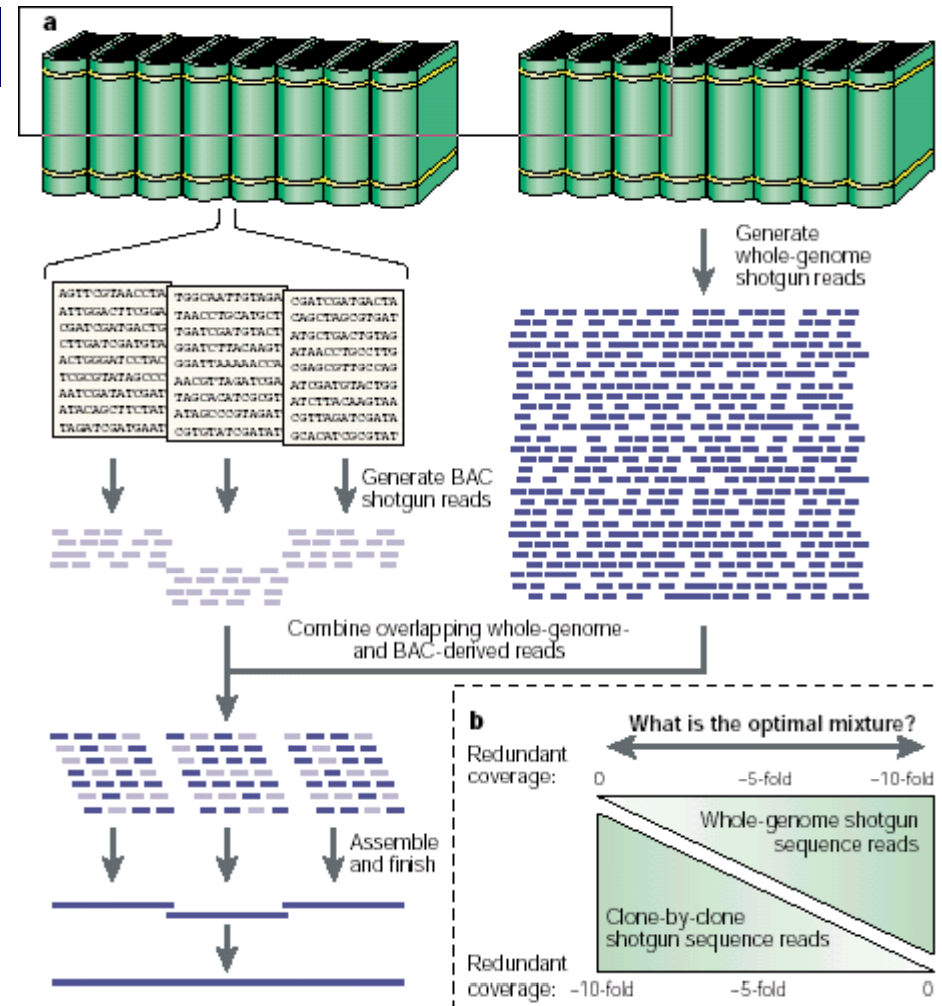
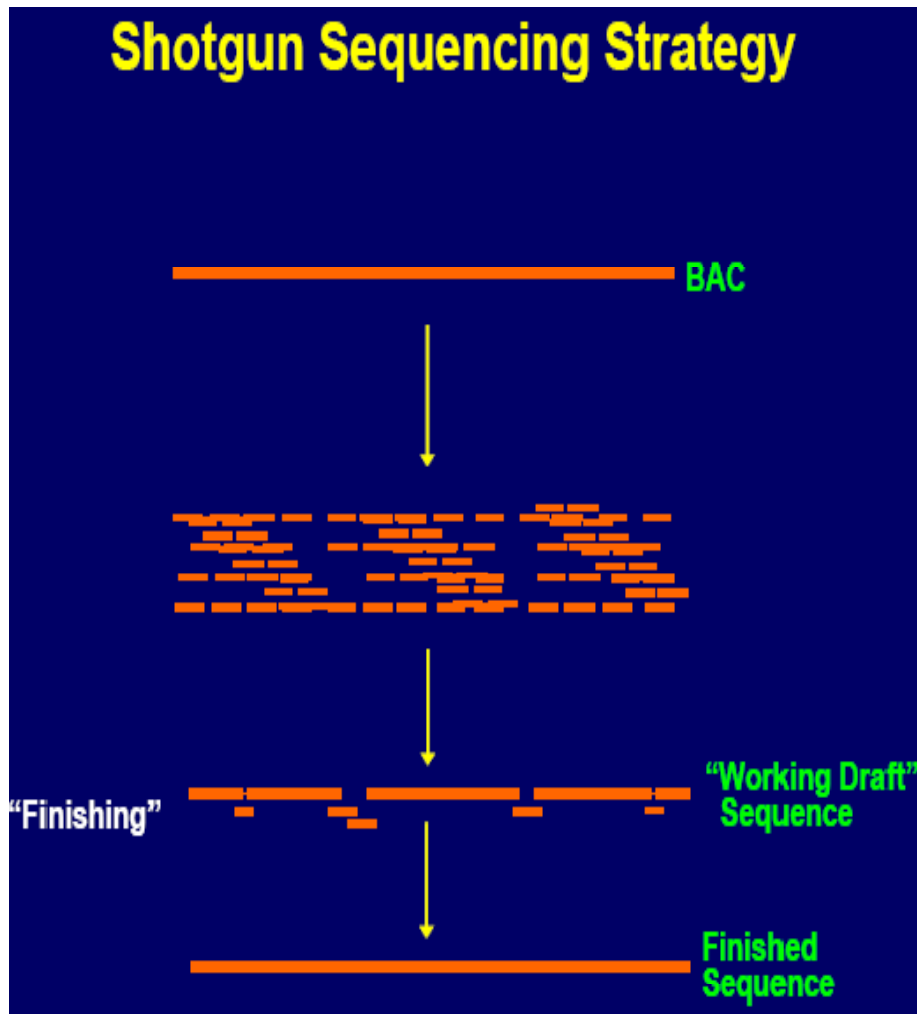


Figure 1:

Estrategias alternativas de secuenciación



The Genomic Landscape: *circa 2010, Eric Green*

Equipamiento automatizado

Human Genome Sequencing Centers



The Genomic Landscape: *circa 2010, Eric Green*

a) Sequence reads

Read 1 CACATACACATGG

Read 2 TCAATGGGGCTAA

Read 3 AGCAGGGAATTGTCACATACACATG

Read 4 ACACATGGAAATA

Read 5 GGGCTAATGATTGTCAC

Read 6 TGATTGTCACATA

Read 7 ATTCATGAAGCAGGA

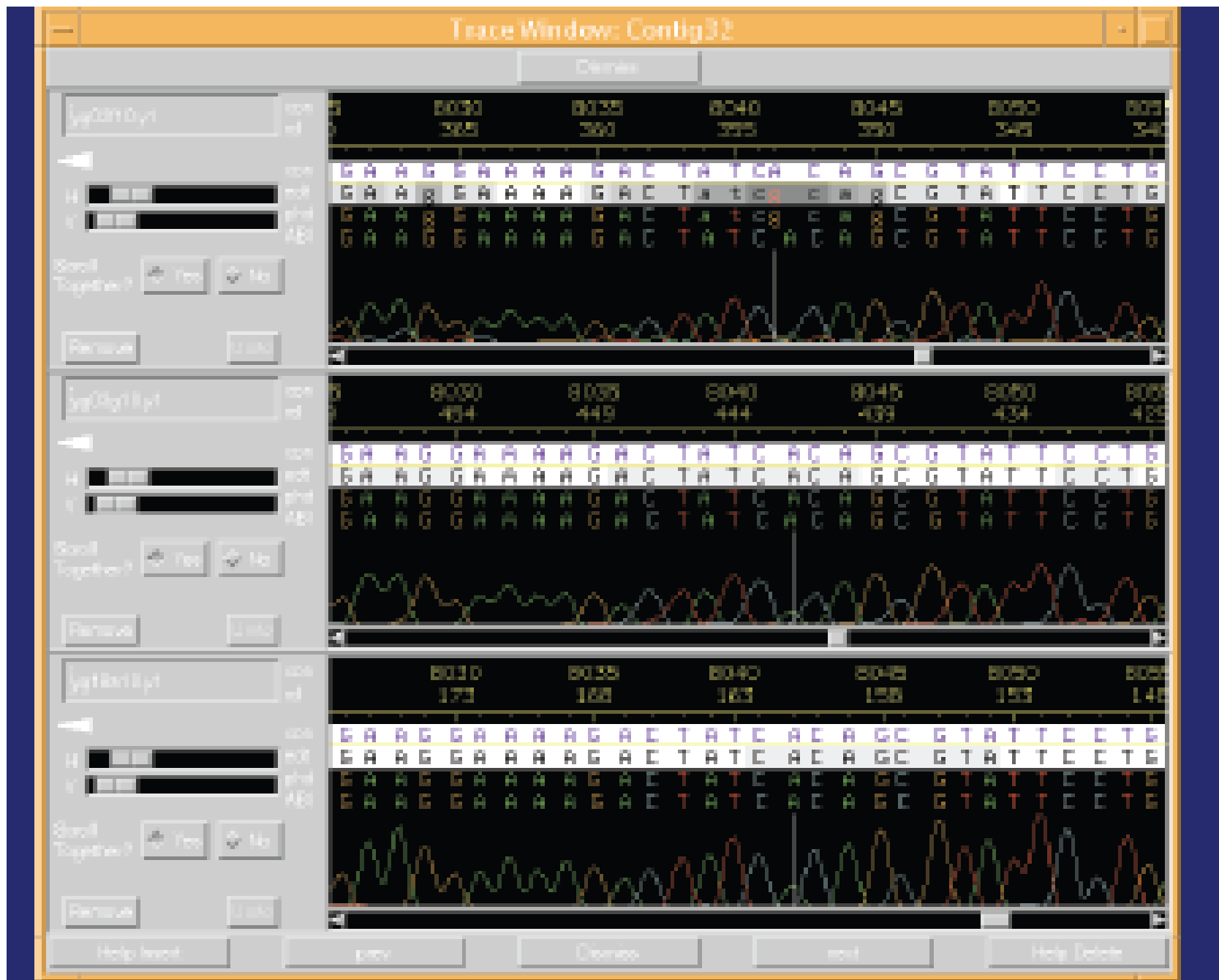
Read 8 GTCACATACACATGATCAATGGGG

Use computer to assemble sequence reads



Assembled sequence

c) ATTCATGAAGCAGGGAATTGTCACATACACATGATCAATGGGGCTAATGATTGTCACATACACATGGAAATA



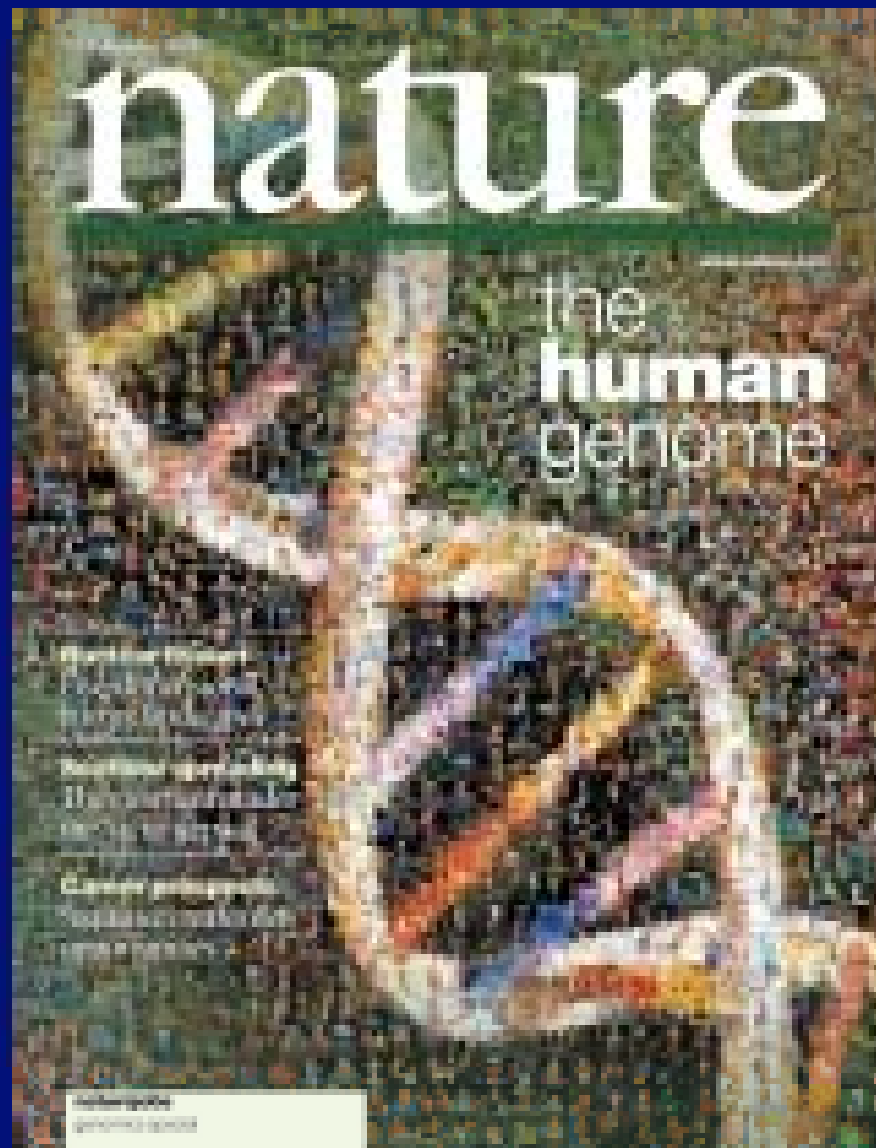
The Genomic Landscape: *circa 2010, Eric Green*

Sequence Finishing: Resolving Ambiguities



*** Sequence Finishing: Remains Relatively Expensive ***

February, 2001 Draft Sequence

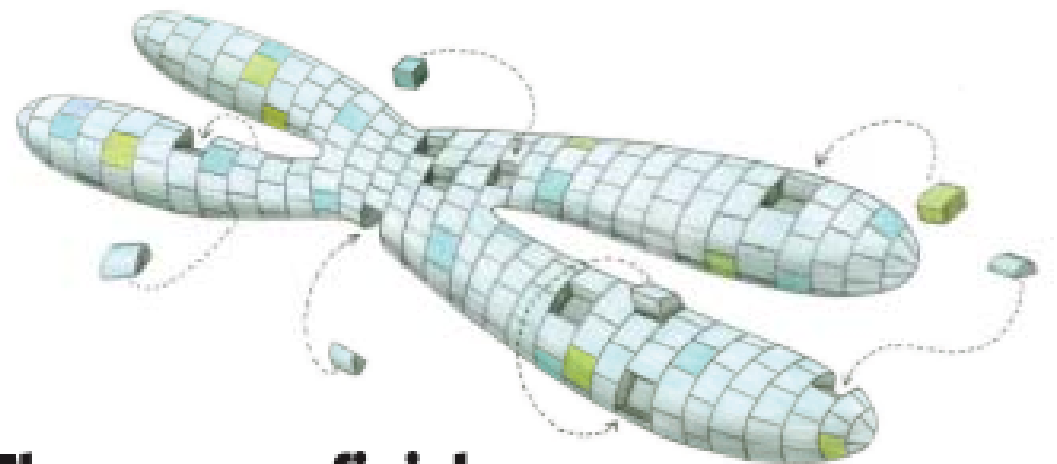
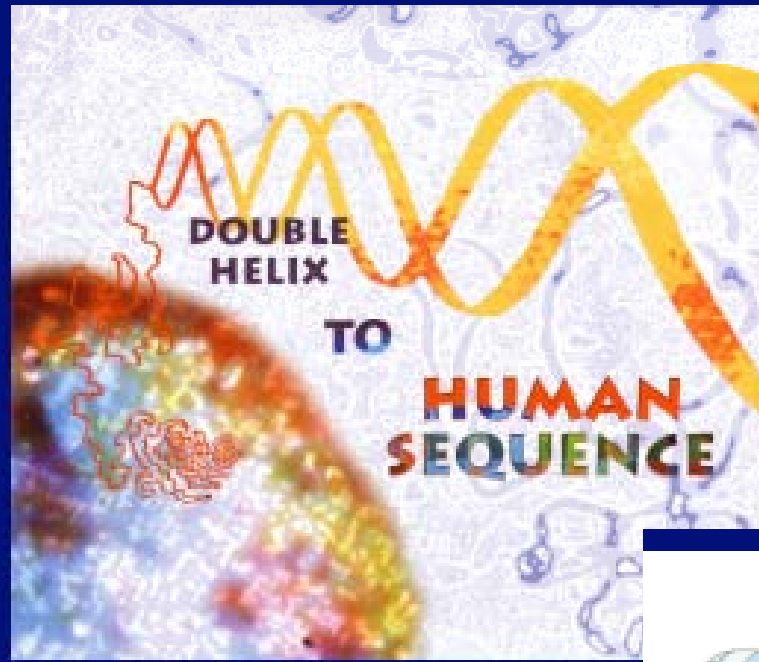


International Human Genome Sequencing Consortium (2001)



Venter et al. (2001)
The Genomic Landscape: *circa 2010*, Eric Green

April, 2003 Completion



The genome finishers

Dedicated scientists are working hard to close the gaps, fix the errors and finally complete the human genome sequence. **Elie Dolgin** looks at how close they are.

Nature (2009)

The Genomic Landscape: *circa 2010, Eric Green*

Mapping
the Human Genome

~1990 to ~2000

Sequencing
the Human Genome

~1998 to ~2003

The Human
Genome Project

Interpreting
the Human Genome
Sequence

~2003 to ???

Beyond
The Human
Genome Project

The Human Genome... by the Numbers

~5% of Human Genome Sequence is Constrained Across Mammals (and Presumed Functional)

5% of 3B Bases = ~150M Bases

Do NOT Yet Know the Position of these ~150M Functional Bases
Lower Bound for the Amount that is Functional

~1.5% Encodes for Protein (Genes)

Corresponds to ~18-22K Genes

Many More than ~22K Different Proteins

Good Inventory at Present

~3.5% Functional But Non-Coding

Gene Regulatory Elements

Chromosomal Functional Elements

Undiscovered Functional Elements (NOT Yet in Textbooks!)

Poor Inventory at Present

The Genomic Landscape: *circa 2010, Eric Green*

Comparative Sequence Analysis

Using the 'Experiments of Evolution' to Decode the Human Genome



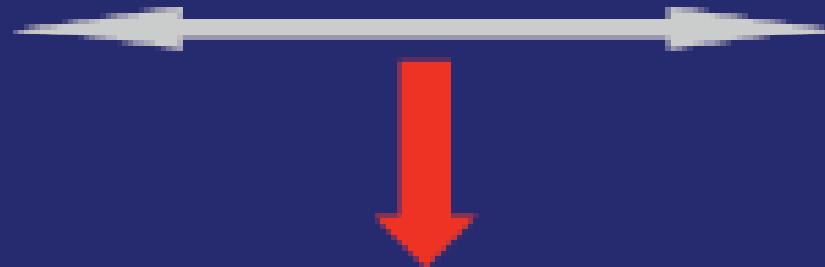
Species A

1. The first step is to identify the problem or goal. This involves understanding the current situation and what needs to be achieved.

Species B

[illegible]

Compare

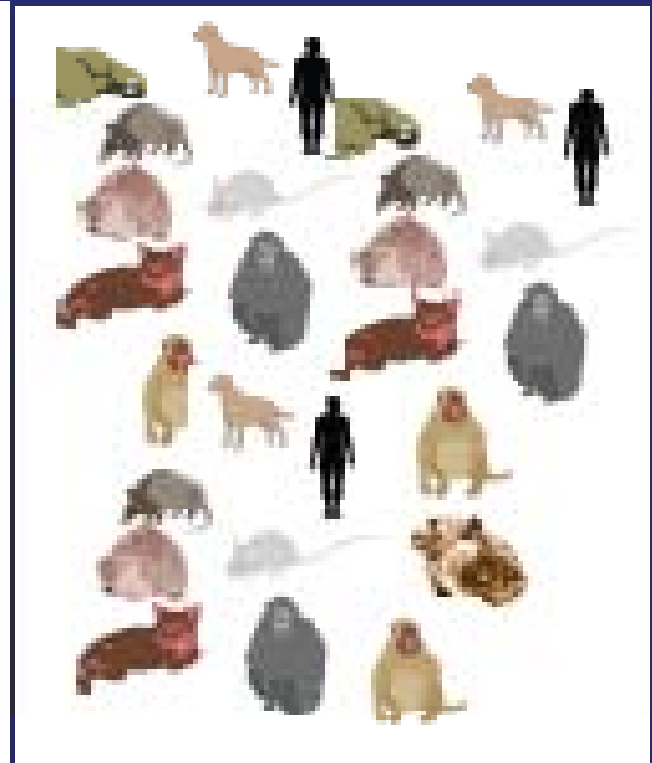
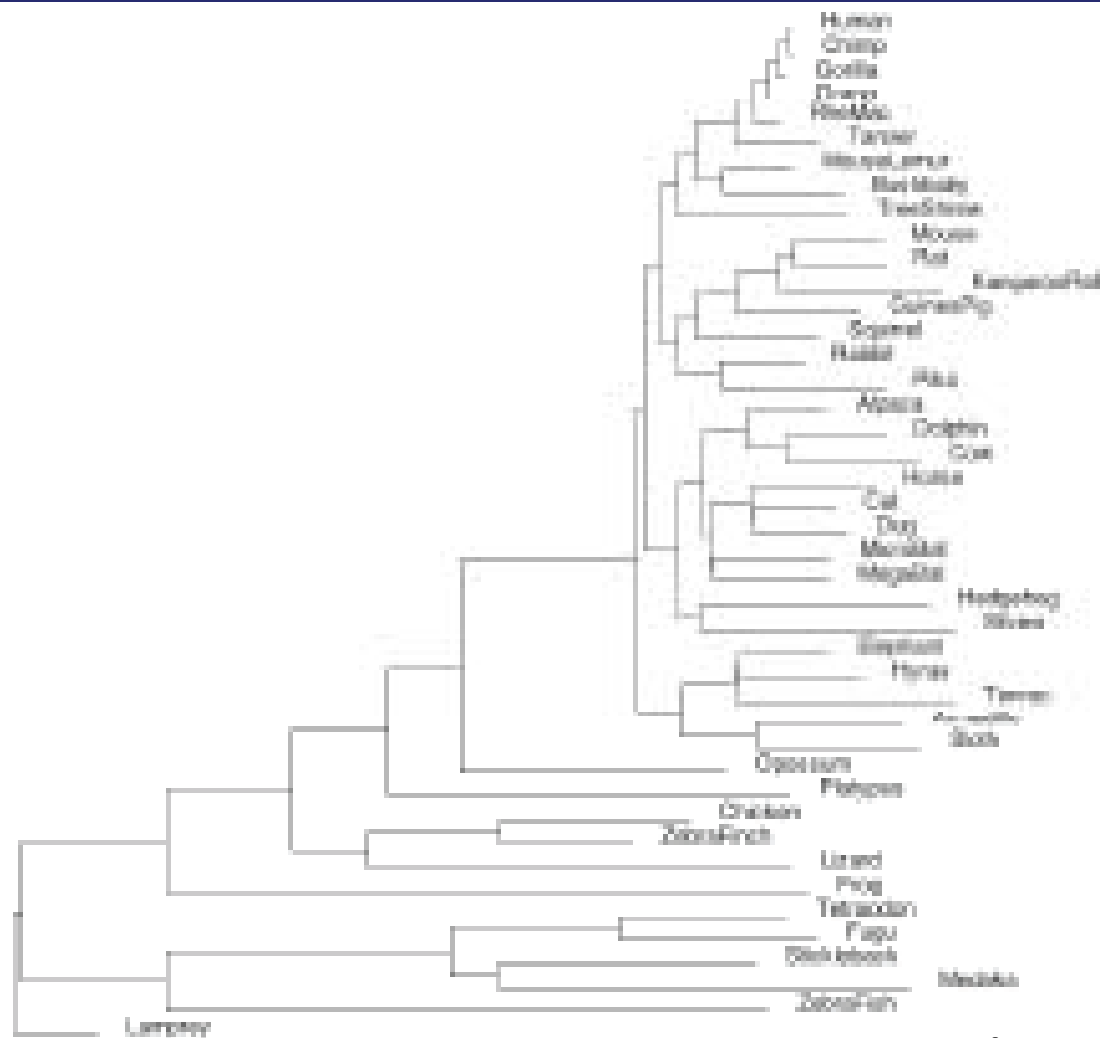


1. **Identify the main idea**
 2. **Identify the supporting details**
 3. **Identify the conclusion**
 4. **Identify the evidence**
 5. **Identify the counter-evidence**
 6. **Identify the author's purpose**
 7. **Identify the author's bias**
 8. **Identify the author's tone**
 9. **Identify the author's style**
 10. **Identify the author's audience**
 11. **Identify the author's context**
 12. **Identify the author's background**
 13. **Identify the author's education**
 14. **Identify the author's experience**
 15. **Identify the author's beliefs**
 16. **Identify the author's values**
 17. **Identify the author's goals**
 18. **Identify the author's interests**
 19. **Identify the author's hobbies**
 20. **Identify the author's family**
 21. **Identify the author's friends**
 22. **Identify the author's enemies**
 23. **Identify the author's mentors**
 24. **Identify the author's teachers**
 25. **Identify the author's employers**
 26. **Identify the author's colleagues**
 27. **Identify the author's subordinates**
 28. **Identify the author's superiors**
 29. **Identify the author's advisors**
 30. **Identify the author's critics**
 31. **Identify the author's supporters**
 32. **Identify the author's opponents**
 33. **Identify the author's allies**
 34. **Identify the author's enemies**
 35. **Identify the author's friends**
 36. **Identify the author's family**
 37. **Identify the author's education**
 38. **Identify the author's experience**
 39. **Identify the author's beliefs**
 40. **Identify the author's values**
 41. **Identify the author's goals**
 42. **Identify the author's interests**
 43. **Identify the author's hobbies**
 44. **Identify the author's family**
 45. **Identify the author's friends**
 46. **Identify the author's enemies**
 47. **Identify the author's mentors**
 48. **Identify the author's teachers**
 49. **Identify the author's employers**
 50. **Identify the author's colleagues**
 51. **Identify the author's subordinates**
 52. **Identify the author's superiors**
 53. **Identify the author's advisors**
 54. **Identify the author's critics**
 55. **Identify the author's supporters**
 56. **Identify the author's opponents**
 57. **Identify the author's allies**
 58. **Identify the author's enemies**
 59. **Identify the author's friends**
 60. **Identify the author's family**
 61. **Identify the author's education**
 62. **Identify the author's experience**
 63. **Identify the author's beliefs**
 64. **Identify the author's values**
 65. **Identify the author's goals**
 66. **Identify the author's interests**
 67. **Identify the author's hobbies**
 68. **Identify the author's family**
 69. **Identify the author's friends**
 70. **Identify the author's enemies**
 71. **Identify the author's mentors**
 72. **Identify the author's teachers**
 73. **Identify the author's employers**
 74. **Identify the author's colleagues**
 75. **Identify the author's subordinates**
 76. **Identify the author's superiors**
 77. **Identify the author's advisors**
 78. **Identify the author's critics**
 79. **Identify the author's supporters**
 80. **Identify the author's opponents**
 81. **Identify the author's allies**
 82. **Identify the author's enemies**
 83. **Identify the author's friends**
 84. **Identify the author's family**
 85. **Identify the author's education**
 86. **Identify the author's experience**
 87. **Identify the author's beliefs**
 88. **Identify the author's values**
 89. **Identify the author's goals**
 90. **Identify the author's interests**
 91. **Identify the author's hobbies**
 92. **Identify the author's family**
 93. **Identify the author's friends**
 94. **Identify the author's enemies**
 95. **Identify the author's mentors**
 96. **Identify the author's teachers**
 97. **Identify the author's employers**
 98. **Identify the author's colleagues**
 99. **Identify the author's subordinates**
 100. **Identify the author's superiors**
 101. **Identify the author's advisors**
 102. **Identify the author's critics**
 103. **Identify the author's supporters**
 104. **Identify the author's opponents**
 105. **Identify the author's allies**
 106. **Identify the author's enemies**
 107. **Identify the author's friends**
 108. **Identify the author's family**
 109. **Identify the author's education**
 110. **Identify the author's experience**
 111. **Identify the author's beliefs**
 112. **Identify the author's values**
 113. **Identify the author's goals**
 114. **Identify the author's interests**
 115. **Identify the author's hobbies**
 116. **Identify the author's family**
 117. **Identify the author's friends**
 118. **Identify the author's enemies**
 119. **Identify the author's mentors**
 120. **Identify the author's teachers**
 121. **Identify the author's employers**
 122. **Identify the author's colleagues**
 123. **Identify the author's subordinates**
 124. **Identify the author's superiors**
 125. **Identify the author's advisors**
 126. **Identify the author's critics**
 127. **Identify the author's supporters**
 128. **Identify the author's opponents**
 129. **Identify the author's allies**
 130. **Identify the author's enemies**
 131. **Identify the author's friends**
 132. **Identify the author's family**
 133. **Identify the author's education**
 134. **Identify the author's experience**
 135. **Identify the author's beliefs**
 136. **Identify the author's values**
 137. **Identify the author's goals**
 138. **Identify the author's interests**
 139. **Identify the author's hobbies**
 140. **Identify the author's family**
 141. **Identify the author's friends**
 142. **Identify the author's enemies**
 143. **Identify the author's mentors**
 144. **Identify the author's teachers**
 145. **Identify the author's employers**
 146. **Identify the author's colleagues**
 147. **Identify the author's subordinates**
 148. **Identify the author's superiors**
 149. **Identify the author's advisors**
 150. **Identify the author's critics**
 151. **Identify the author's supporters**
 152. **Identify the author's opponents**
 153. **Identify the author's allies**
 154. **Identify the author's enemies**
 155. **Identify the author's friends**
 156. **Identify the author's family**
 157. **Identify the author's education**
 158. **Identify the author's experience**
 159. **Identify the author's beliefs**
 160. **Identify the author's values**
 161. **Identify the author's goals**
 162. **Identify the author's interests**
 163. **Identify the author's hobbies**
 164. **Identify the author's family**
 165. **Identify the author's friends**
 166. **Identify the author's enemies**
 167. **Identify the author's mentors**
 168. **Identify the author's teachers**
 169. **Identify the author's employers**
 170. **Identify the author's colleagues**
 171. **Identify the author's subordinates**
 172. **Identify the author's superiors**
 173. **Identify the author's advisors**
 174. **Identify the author's critics**
 175. **Identify the author's supporters**
 176. **Identify the author's opponents**
 177. **Identify the author's allies**
 178. **Identify the author's enemies**
 179. **Identify the author's friends**
 180. **Identify the author's family**
 181. **Identify the author's education**
 182. **Identify the author's experience**
 183. **Identify the author's beliefs**
 184. **Identify the author's values**
 185. **Identify the author's goals**
 186. **Identify the author's interests**
 187. **Identify the author's hobbies**
 188. **Identify the author's family**
 189. **Identify the author's friends**
 190. **Identify the author's enemies**
 191. **Identify the author's mentors**
 192. **Identify the author's teachers**
 193. **Identify the author's employers**
 194. **Identify the author's colleagues**
 195. **Identify the author's subordinates**
 196. **Identify the author's superiors**
 197. **Identify the author's advisors**
 198. **Identify the author's critics**
 199. **Identify the author's supporters**
 200. **Identify the**

Sequences in Common (i.e., 'Conserved' or 'Constrained')

The Genomic Landscape: circa 2010, Eric Green

22 Additional Mammalian Genome Sequences (@ Low Redundancy)



The Genomic Landscape: *circa 2010, Eric Green*

All humans are ~99.7% identical at the DNA sequence level, and yet...

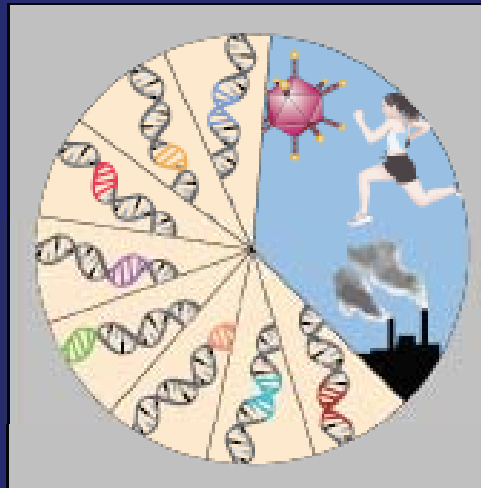


all of us carry a significant number of 'glitches' in our genomes.

Genomic Architecture of Genetic Diseases



Rare, Simple, Monogenic,
Mendelian...



Common, Complex, Multigenic,
Non-Mendelian...

The Genomic Landscape: *circa 2010, Eric Green*

International
HapMap
Project



International HapMap Project

[Home](#) | [About the Project](#) | [Data](#) | [Publications](#) | [Tutorial](#)



THE HUMAN VARIOME PROJECT
sharing data · reducing disease

[Home](#) [About](#) [Activities](#) [Publications](#) [Recommendations](#) [Meetings](#) [News](#) [Links](#) [Blogs](#) [Members](#)

Nuevas tecnologías de secuenciación de ADN

Eric D. Green, M.D., Ph.D.

Techniques for Genome Mapping & Sequencing

Genome sequencing in microfabricated high-density picolitre reactors

[illegible]

Margulies M et al. (2005)

Solexa Ltd

NewsSource PM/ Torana Labs developing an inkjet-based system, based on a reusable single-molecule sequencing technology. In a March 2012 release stated that it expected to show

Bennett (2004)

Toward the \$1000 human genome

<i>Steve J. Brown</i> <i>John Brown</i> <i>Andrew J. Yu</i> <i>Joe Flacco</i> <i>Cyril Jones</i>	Revolutionary new technologies, capable of transforming the economics of sequencing, are providing an unparalleled opportunity to analyze human genetic variation comprehensively at the whole-genome level within a realistic timeframe and at affordable costs. Current estimates suggest that it is all but inevitable in the region of \$500 million to sequence
---	---

Bennett et al. (2005)

Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome

Jay Shenders,^{1a} Gregory J. Porreca,^{1a} Nikos S. Reppas,¹
Xiaoxia Lin,² John F. McCutcheon,^{2,3} Abraham M. Rosenbaum,¹
Michael D. Wong,¹ Kun Zhang,¹ Rabi D. Nitro,² George M. Church¹

Shendure et al. (2005)

Perspective

Emerging technologies in DNA sequencing

Michael L. Metzker

Human Genome Sequencing Center and Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA

Demand for DNA sequence information has never been greater, yet current Sanger technology is too costly, time consuming, and labor intensive to meet this ongoing demand. Application spans numerous research interests, including sequence variation studies, comparative genomics and evolution, forensics, and diagnostic and applied therapeutics. Several emerging technologies show promise of delivering next-generation solutions for fast and affordable genome sequencing. In this review article, the DNA polymerase-dependent strategies of Sanger sequencing, single nucleotide addition, and cyclic reversible immobilization are discussed. The strengths and potential challenges these technologies face in their development for ultrafast sequencing are also discussed. **Metzker (2008)**

Metzker (2005)

SCIENTIFIC
AMERICAN

Alternatives to Toxic Tests on Animals

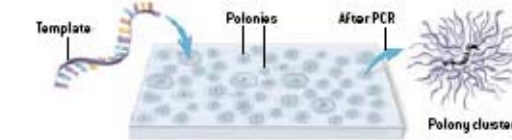
Church (2006)

Know Your DNA

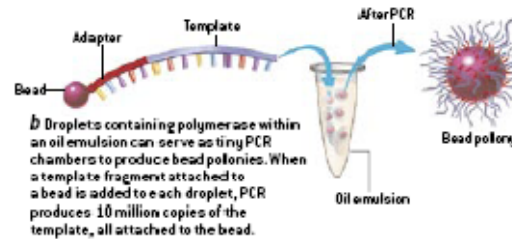
Inexpensive gene readers will soon unlock the secrets in your personal double helix

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.

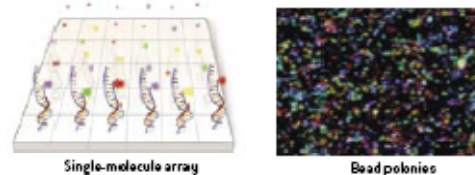


Q Polonies—polymerase colonies—created directly on the surface of a slide or gel each contain a primer, which a template fragment can find and bind to. PCR within each polony produces a cluster containing millions of template copies.



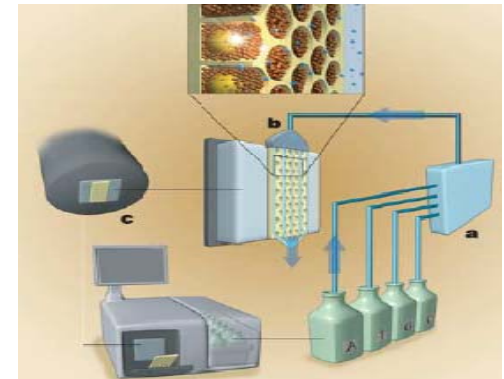
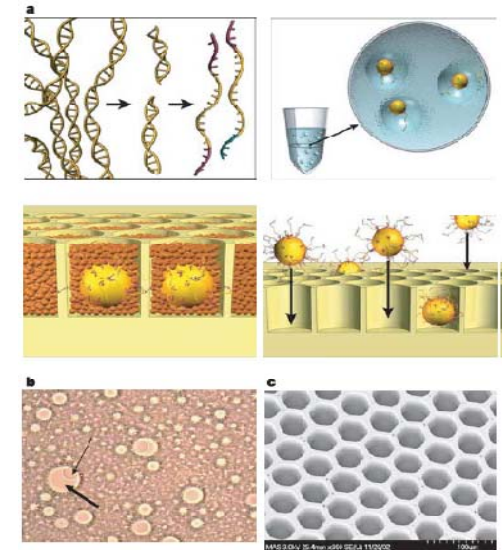
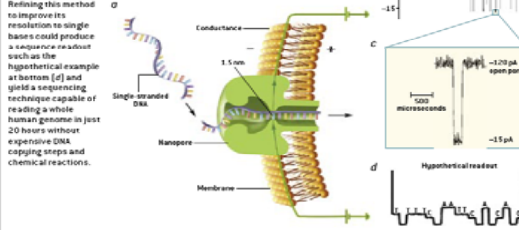
MULTIPLEXING

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 colonies 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.



NANOPORE SEQUENCING

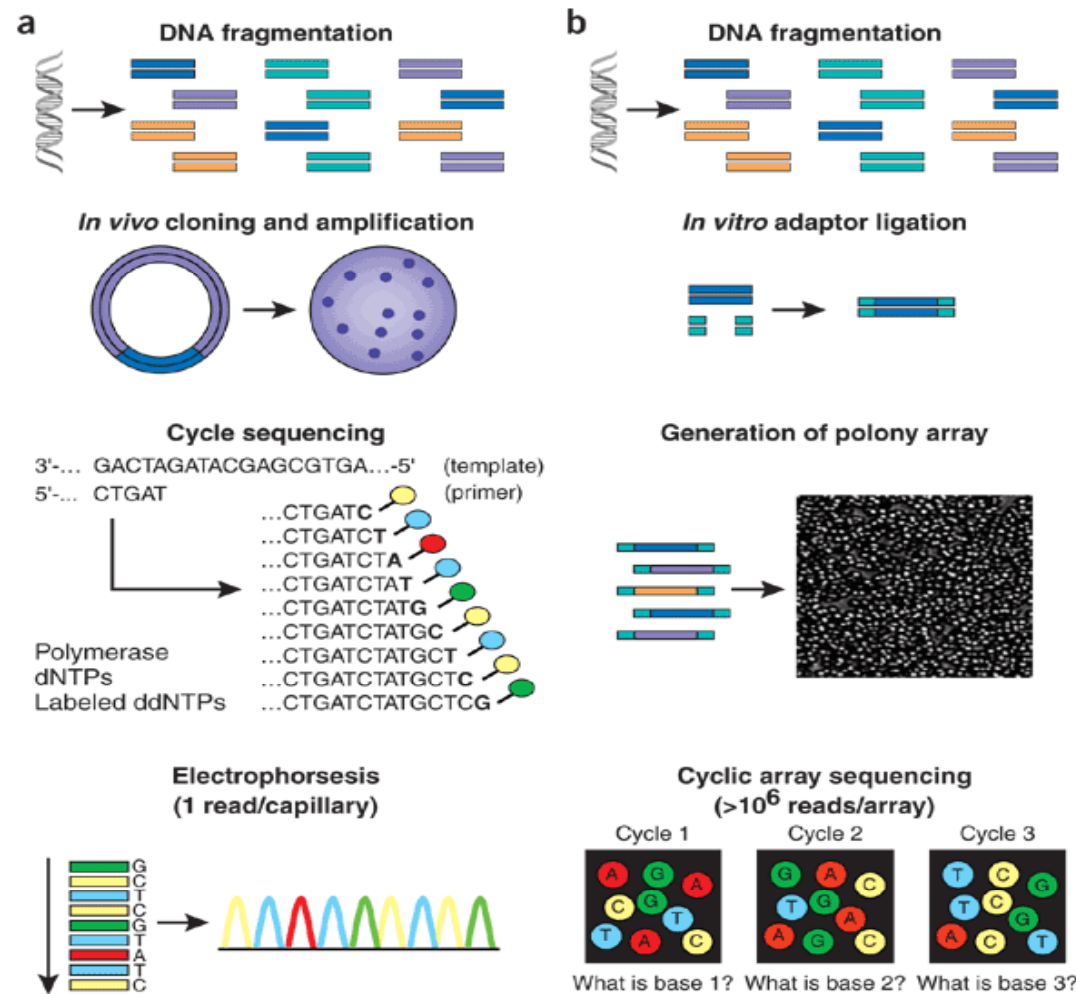
Like electrophoresis, this technique draws DNA toward a positive charge. To get there, the molecule must cross a membrane by going through a pore whose narrowest diameter of 1.5 nanometers will allow only single-stranded DNA to pass [a]. As the strand transits the pore, nucleotides block the opening momentarily, altering the membrane's electrical conductance, measured in picoamperes (pA). Physical differences between the four base types produce blockades of different degrees and durations [b]. A close-up of a blockade event measurement shows a conductance change when a 150-nucleotide strand of a single base type passes through the pore [c].



DNA Sequencing Technologies

<u>Method</u>	<u>Feasibility</u>	<u>Read Length</u>	<u>Data Quality</u>	<u>Raw Data Production</u>
Sanger Sequencing	Well Established	Long (800-1200 bases)	+++	+
Stepwise Synthesis	Becoming Established	Short (25-100 bases)	+	+++++
Single Molecule	Far from Established	Long (>1000 bases ?)	???	+++++ ?

Nueva generación de tecnologías de secuenciación



(a) With high-throughput shotgun Sanger sequencing, genomic DNA is fragmented, then cloned to a plasmid vector and used to transform *E. coli*. For each sequencing reaction, a single bacterial colony is picked and plasmid DNA isolated. Each cycle sequencing reaction takes place within a microliter-scale volume, generating a ladder of ddNTP-terminated, dye-labeled products, which are subjected to high-resolution electrophoretic separation within one of 96 or 384 capillaries in one run of a sequencing instrument. As fluorescently labeled fragments of discrete sizes pass a detector, the four-channel emission spectrum is used to generate a sequencing trace. (b) In shotgun sequencing with cyclic-array methods, common adaptors are ligated to fragmented genomic DNA, which is then subjected to one of several protocols that results in an array of millions of spatially immobilized PCR colonies or 'polonies'¹⁵. Each polony consists of many copies of a single shotgun library fragment. As all polonies are tethered to a planar array, a single microliter-scale reagent volume (e.g., for primer hybridization and then for enzymatic extension reactions) can be applied to manipulate all array features in parallel. Similarly, imaging-based detection of fluorescent labels incorporated with each extension can be used to acquire sequencing data on all features in parallel. Successive iterations of enzymatic interrogation and imaging are used to build up a contiguous sequencing

read for each array.

Nature Biotechnology 26, 1135 - 1145 (2008)

Table 1 Second-generation DNA sequencing technologies

	Feature generation	Sequencing by synthesis	Cost per megabase	Cost per instrument	Paired ends?	1° error modality	Read-length	References
454	Emulsion PCR	Polymerase (pyrosequencing)	~\$60	\$500,000	Yes	Indel	250 bp	14,20
Solexa	Bridge PCR	Polymerase (reversible terminators)	~\$2	\$430,000	Yes	Subst.	36 bp	17,22
SOLiD	Emulsion PCR	Ligase (octamers with two-base encoding)	~\$2	\$591,000	Yes	Subst.	35 bp	13,26
Polonator	Emulsion PCR	Ligase (nonamers)	~\$1	\$155,000	Yes	Subst.	13 bp	13,20
HeliScope	Single molecule	Polymerase (asynchronous extensions)	~\$1	\$1,350,000	Yes	Del	30 bp	18,30

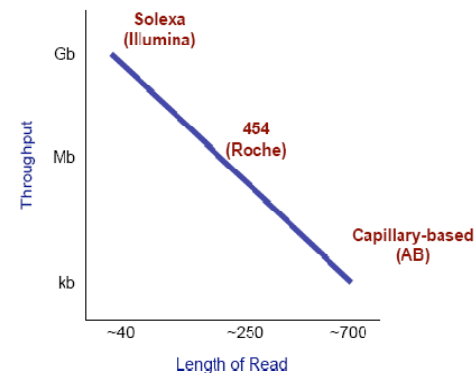
The pace with which the field is moving makes it likely that estimates for costs and read-lengths will be quickly outdated. Vendors including Roche Applied Science, Illumina, and Applied Biosystems have major upgrade releases currently in progress. Estimated costs-per-megabase are approximate and inclusive only of reagents. Read-lengths are for single tags. Subst., substitutions; indel, insertions or deletions; del, deletions.

Nature Biotechnology 26, 1135 - 1145 (2008)

Advances in DNA sequencing technologies

Technology	Approach	Read length	Bp per run	Company name and Web site
Automated Sanger sequencer ABI3730xl	Synthesis in the presence of dye terminators	Up to 900 bp	96 kb	Applied Biosystems www.appliedbiosystems.com
454/Roche FLX system	Pyrosequencing on solid support	200–300 bp	80–120 Mb	Roche Applied Science www.roche-applied-science.com
Illumina/Solexa	Sequencing by synthesis with reversible terminators	30–40 bp	1 Gb	Illumina, Inc. http://www.illumina.com/
ABI/SOLiD	Massively parallel sequencing by ligation	35 bp	1–3 Gb	Applied Biosystems www.appliedbiosystems.com

Trade-offs with Newer Sequencing Technologies



Throughput = $\frac{\text{Amount of Sequence Generated}}{\text{Unit of Time or Cost}}$

R McCombie, 2008. Next Generation Sequencing Symposium. Santa Fe

Genomics 92 (2008) 255–264

Procesamiento de secuencias en proyectos genómicos

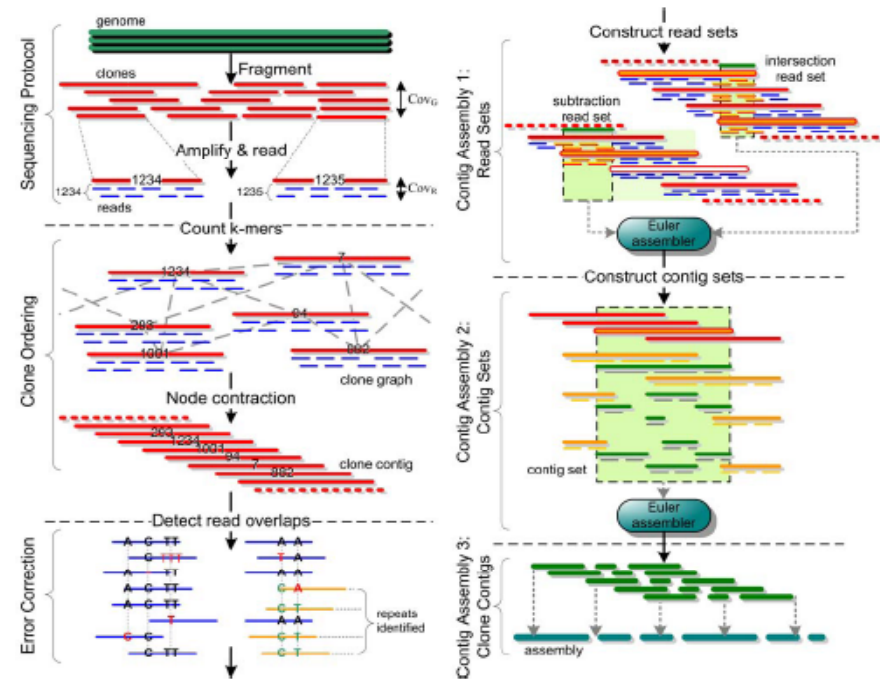
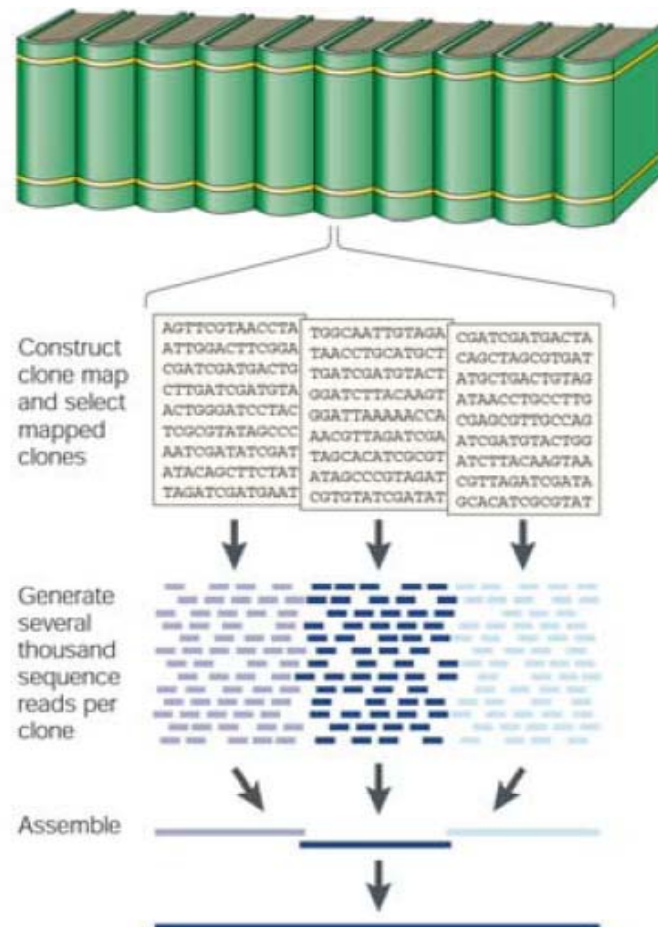


Figure 1. Sequencing protocol and assembly methodology. Reads are obtained in a hierarchical sequencing protocol with high genome-done coverage and low done-read coverage. From the k -mer content of each done we construct a done graph whose edge weights reflect the likely done proximities, and from this our clone ordering algorithm determines the clone contigs. Next, we find all putative read overlaps by only looking in nearby clones and perform error correction. In three stages of contig assembly we 1) create read sets via set operations that consist of reads from multiple overlapping clones within small clone subregions and assemble using Euler, 2) combine contigs resulting from the previous stage in clone-sized contig sets for assembly, and 3) use a scalable assembler to merge entire clone contigs.

PLoS ONE 2(5): e484.
doi:10.1371/journal.pone.0000484



<http://demo.decode.me.com/research-catalog>

Human Genome Sequence

>\$1,000,000,000



~\$1,000

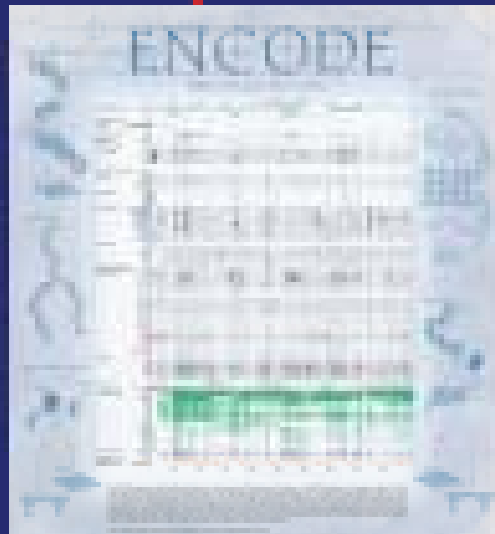
The Pathway to Genomic Medicine

Interpreting
the Human
Genome Sequence

Implicating
Genetic Variants
with Human Disease



HGP



Realization of
Genomic Medicine