Genomics

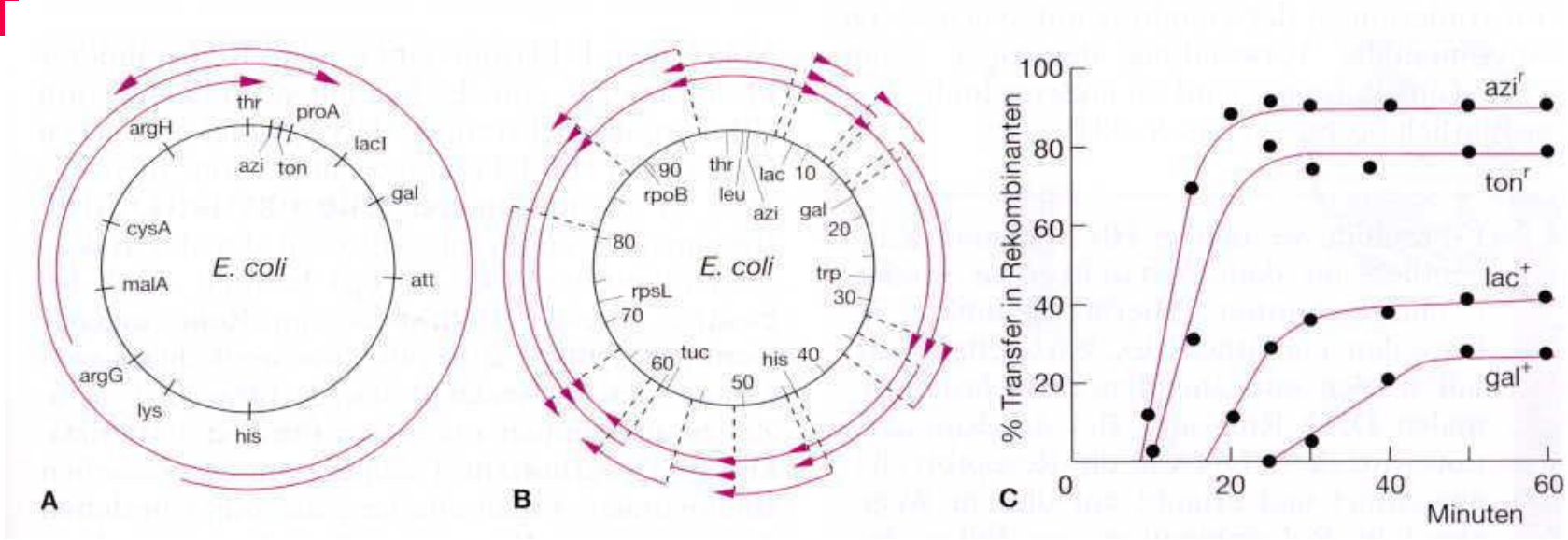
Genome:
total of all genetic information

Gene:
Section on DNA, which encodes a specific feature,

→ nucleotide sequence comprising the coding region for one protein

Locus:
Specific position within a genome encoding a specific trait

Allele:
Variant of a specific gene, differing in at least one bp



Mapping of the *E.coli* genome by F-duction. A Chromosome map of *E.coli* with important marker genes. Different Hfr strains start the transfer of the *E.coli* genome not only from different insertion sites but also in different directions (arrows). **B** Detailed *E.coli* map showing the different plasmid insertion sites (dashed lines) but also the direction of the transfer (arrows). The numbers in the inner circle indicate the time (in minutes) that it takes to transfer the respective gene into the recipient cell (starting from an arbitrarily fixed zero point). **C** Frequency and duration of the transfer of certain *E.coli* marker genes (azi, ton, lac, gal) by F-duction with a certain *Hfr* strain (for the position in the genome compare map A). (**B**: according to Bachmann and Low 1980, **C**: according to Jacob and Wollman 1961)

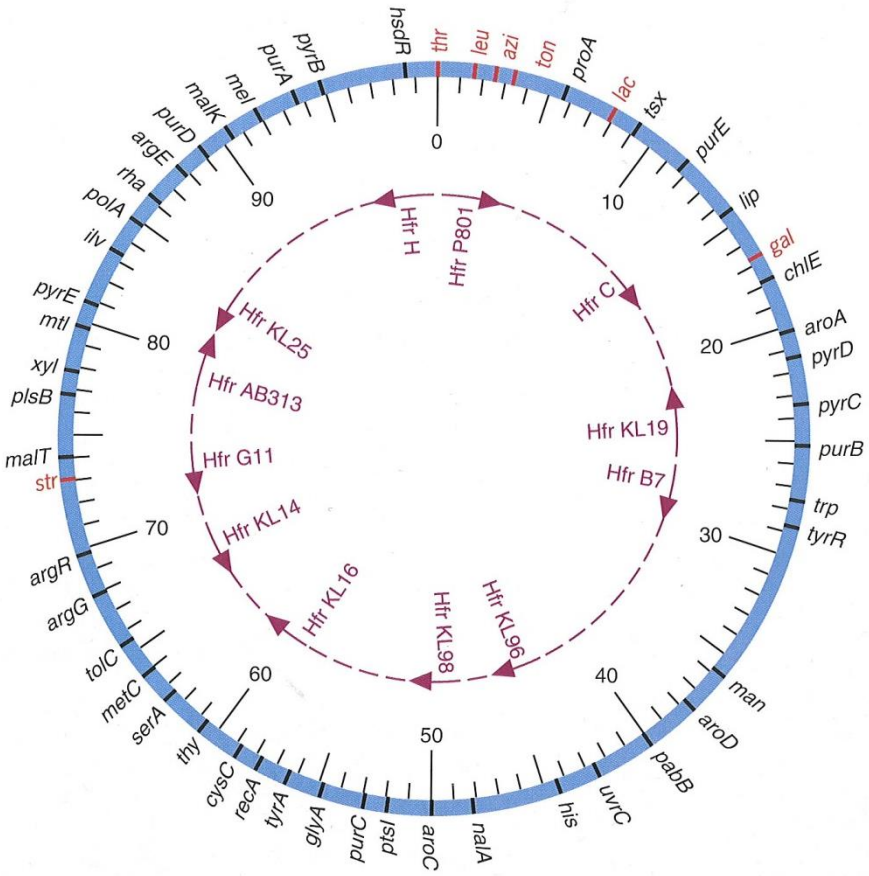
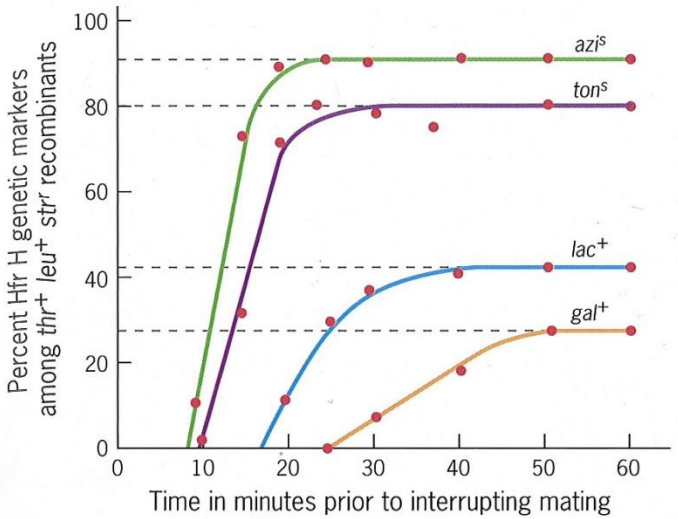


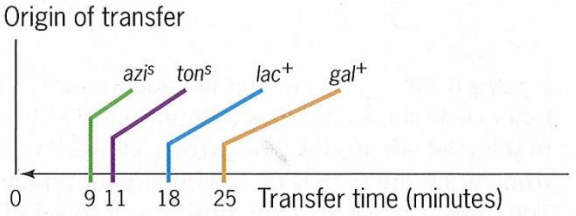
Figure 8.27 ▶ The circular linkage map of *E. coli*. The inner circle shows the sites of integration of the F factor in selected Hfr strains. The arrows indicate whether transfer by the Hfr's is clockwise or counterclockwise. The outer circle shows the position of selected genes. The map is divided into 100 units, where each unit is the length of DNA transferred during one minute of conjugation. The genes shown in red were used in Wollman and Jacob's famous interrupted mating experiment (see Figures 8.25 and 8.26).

Summary of the results



(*a*)

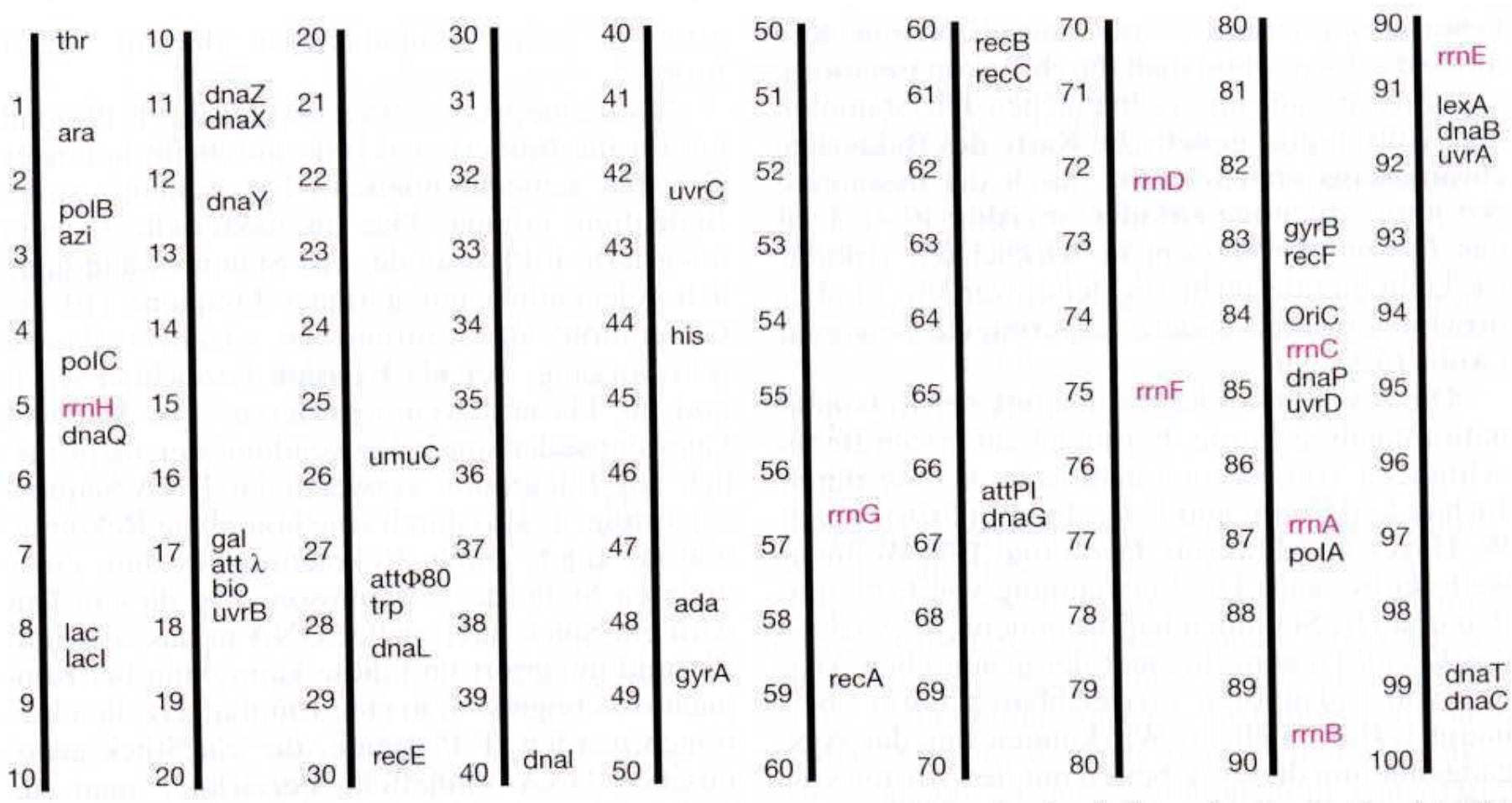
Interpretation of the results



(*b*)

Figure 8.25 ▶ Wollman and Jacob's classic interrupted mating experiment. (*a*) The frequencies of the unselected donor alleles present in thr^+ leu^+ str^r recombinants are shown as a function of the time at which mating was interrupted. (*b*) Interpretation of the results based on the linear transfer of genes from the Hfr cell to the F^- cell. Transfer is initiated at the origin on the F factor, and the time at which a gene is transferred to the F^- cell depends on its distance from the F factor.

Genetic maps

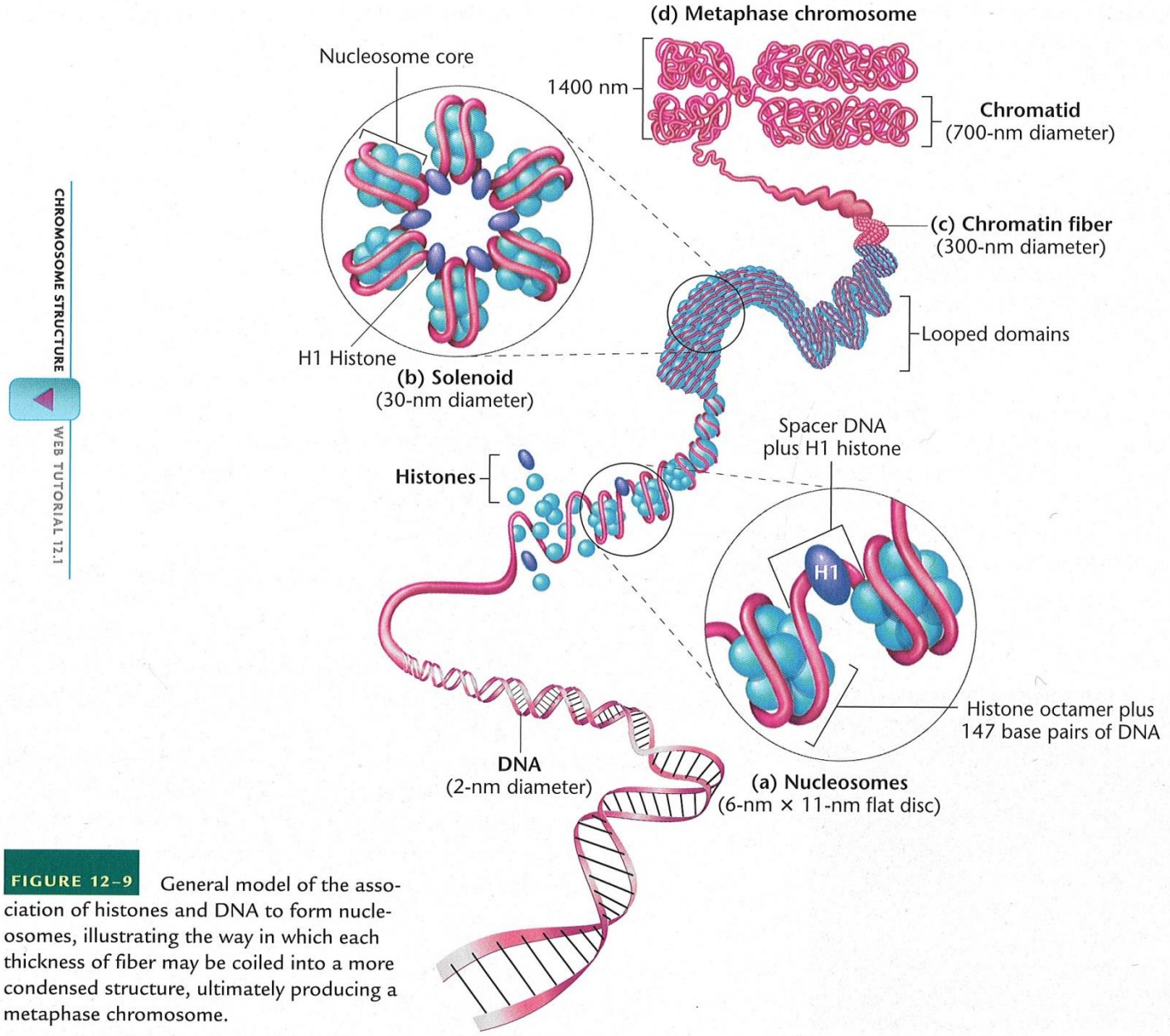


Genetic map of *E.coli*. Groups of rDNA genes are highlighted in red. „dna“: loci involved in replication (according to Bachman and Low 1980).

Nomenclature – *E.coli*

	Genotype	Phenotype	Protein
Genetic Locus	*lac*	Lac	Lac
Wild-type	lac^{+}, $lacZ^{+}$	Lac^{+}	LacZ
Mutant	lac^{-}, *lacZ*	Lac^{-}	LacZ-W123A*
Alleles	$lacZ^{+}$, *lacZ123*		LacZ123

*) in case exact change of amino acids are known the exact coordinates are provided: e.g.: exchange of aa tryptopane (W) at position 123 to aa alanine (A)



CHROMOSOME STRUCTURE
WEB TUTORIAL 12.1

FIGURE 12–9 General model of the association of histones and DNA to form nucleosomes, illustrating the way in which each thickness of fiber may be coiled into a more condensed structure, ultimately producing a metaphase chromosome.

Taken from: W.S. Klug, M. R. Cummings, C.A. Spencer, M.A. Palladino; Concepts of Genetics, 9th Edition; Pearson Benjamin Cummings

Nomenclature *E.coli*

Specific elements:		
Plasmids:	general:	pAB123
		specific: e.g. F, R1, RP4, ColE1 etc.
Phages		specific: λ, M13, T4 etc
Insertion elements		IS*1*
Transposons		Tn*3*

Structural changes:		
Deletions	Δ	e.g. Δ(*lacZ-galE*)123
Inversions	IN	e.g. IN (*lacZ-galE*)123
Transposition	TP	e.g. TP (*lacZ-galE*)123
Fusions	Φ	e.g. Φ (*lacZ-galE*)123
Insertion	::	e.g. *lacZ*::Tn*10*

Strains

Escherichia coli → *E.coli*

E.coli K12(F^-, *lacZ1*, *lacY*::IS*10*, *galE*)

Nomenclature – *Yeast*

	Genotype	Phenotype	Protein
Genetic Locus	*ade5, ADE5*	Ade5	Ade5
Wild-type	*ADE, ADE5*	Ade^{+}	
Mutant	*ade^{-}, ade5*	Ade^{-}	
Alleles	*ADE5*, *ade5-123*, *cup1^{+}, CUP1*		Ade5-123

cup1^{+} → Genotype: Wild-type allele, Phenotype: sensitive against Cu^{++}
CUP1 → Genotype: Mutant allele, Phenotype: resistent against Cu++

Genome sequencing projects

Sequencing

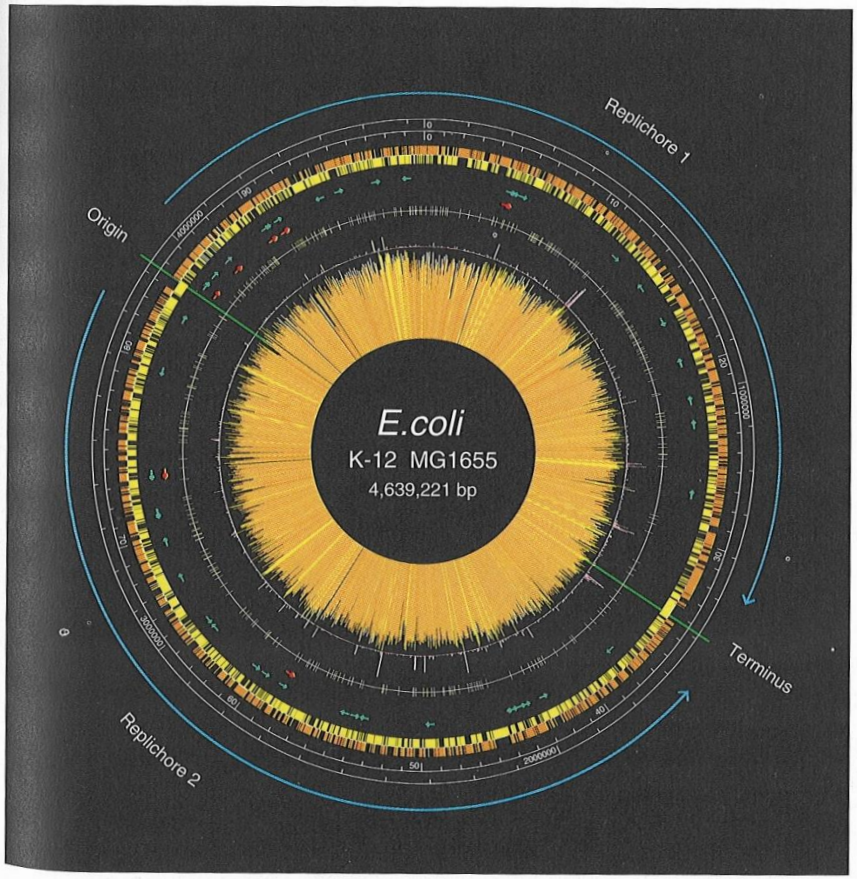
Single Sequence
Assembling
Contigs
Genome Sequence

Annotation

Detection of ORFs
Sequence comparison
Assignment of genes
Assignment of functional elements

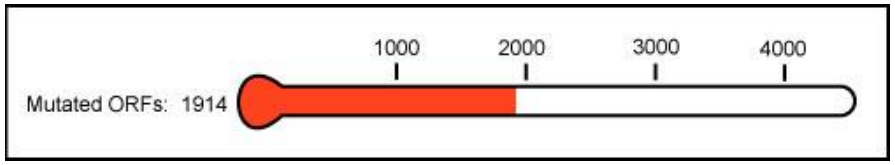
Functional analysis – Functional Genomics

Transcriptomics
Proteomics
Metabolomics
Other „omics“
Knock out mutations



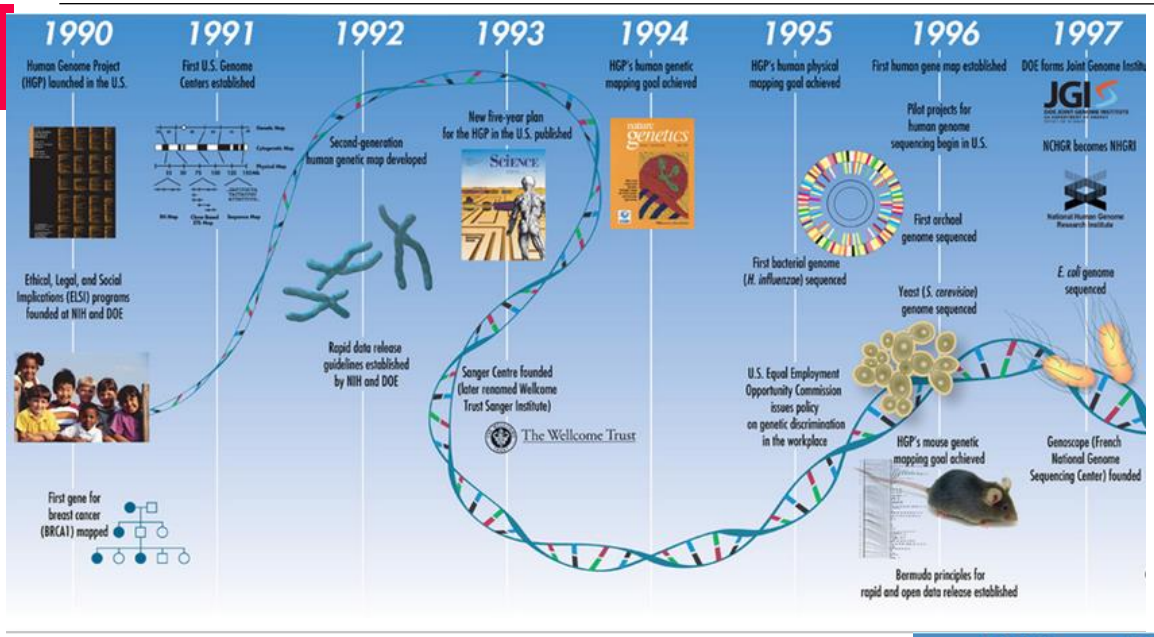
***Escherichia coli* K12** Chromosome
Strain -MG1655
completely sequenced
4,639.221 bp
4.288 Protein-coding Genes
38 % unknown function

Diagram of the DNA sequence organization of *Escherichia coli* strain K-12. The coordinates are given in base pairs as well as in minutes on the genetic map. The coding sequences are shown as gold and yellow bars, which are transcribed in a clockwise (gold) or counterclockwise (yellow) direction. Green and red arrows denote genes for transfer RNAs or for ribosomal RNAs, respectively. The gold rays of the "sunburst" are proportional to the degree of randomness of codon usage in the coding sequences. Genes with the longest rays use the codons in the genetic code almost randomly. The origin and terminus of DNA replication are indicated. Bidirectional replication creates two "replichores." The peaks on the circle immediately outside the sunburst indicate coding sequences with high similarity to previously described bacteriophage proteins. [Courtesy of Frederick R. Blattner and Guy Plunkett III. From F. R. Blattner et al. 1997. *Science* 277: 1453.]

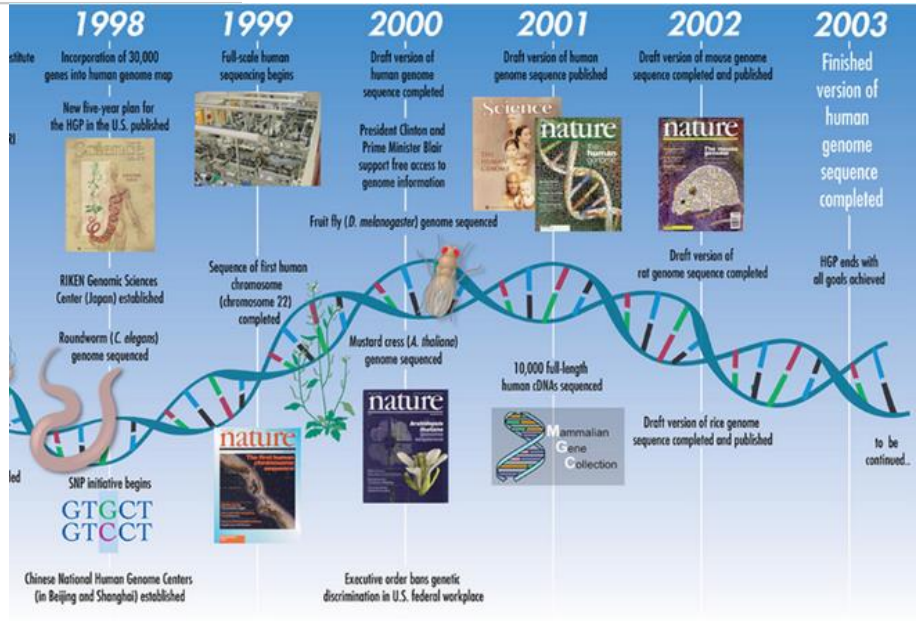


http://www.ecogene.org/

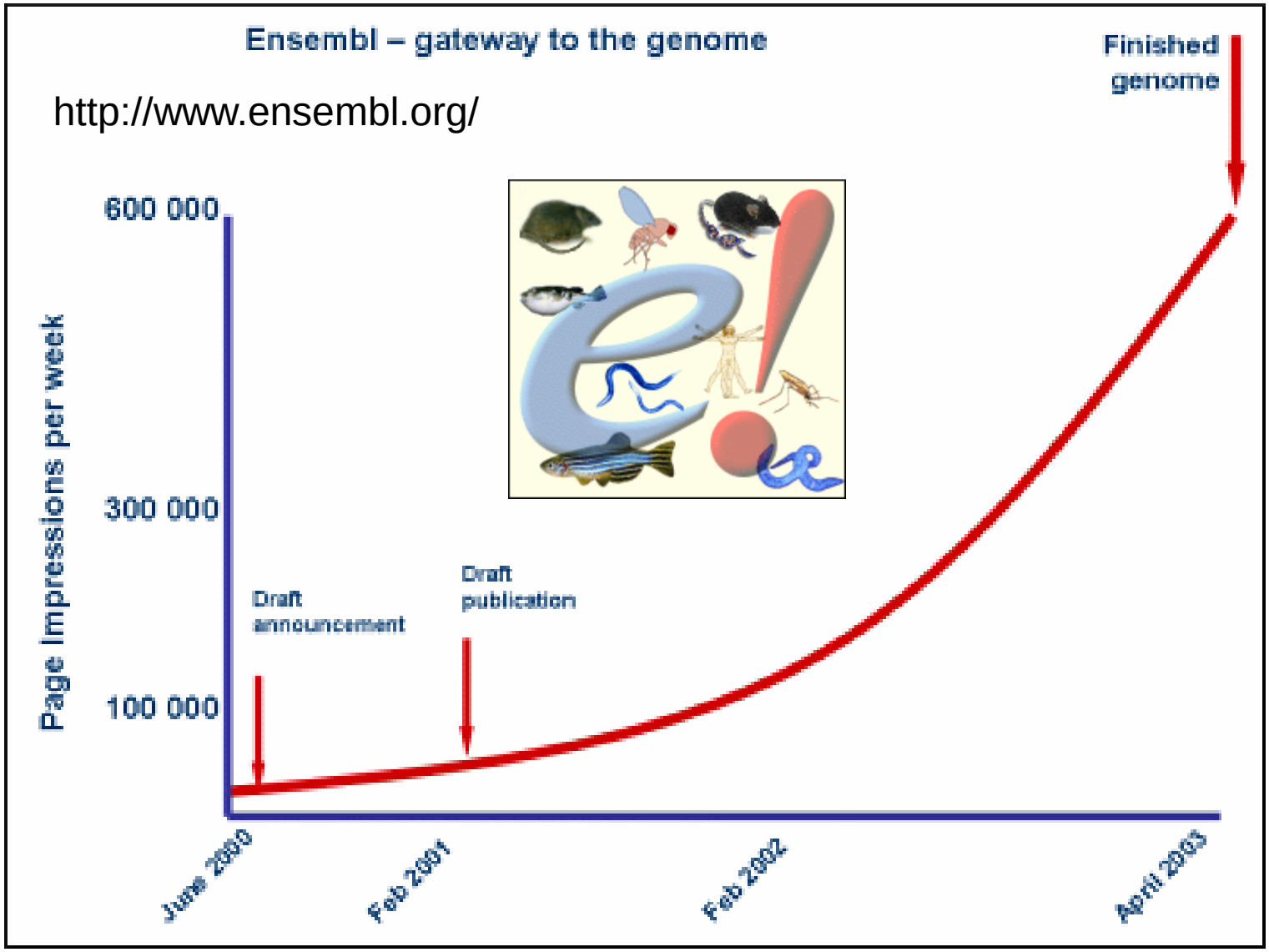
The human genome project



Development of human genome project

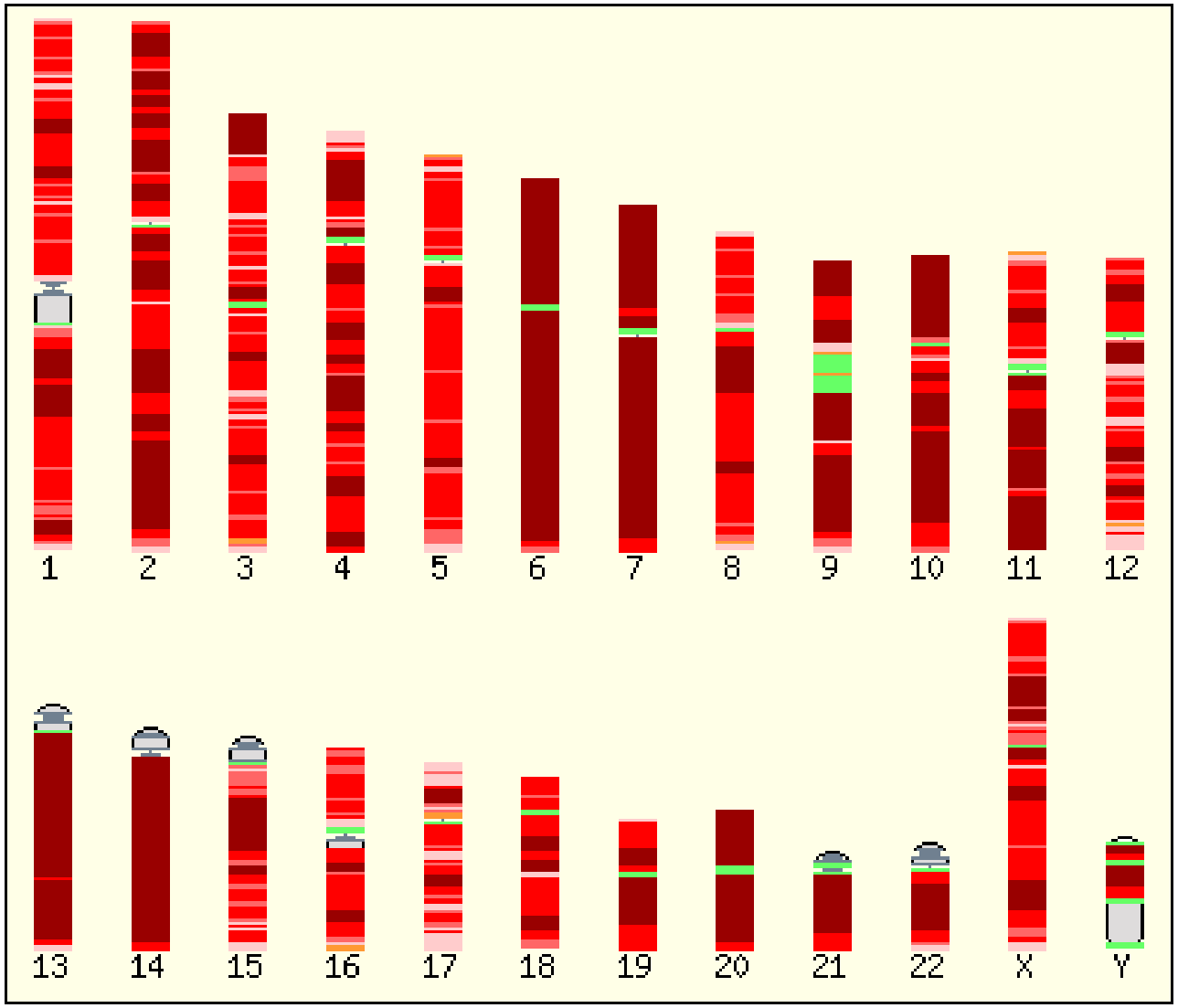


The human genome project



The human genome - Assembly

Nov. 2002



N50 length > 5mb

N50 length > 1mb

N50 length > 500kb

N50 length > 100kb

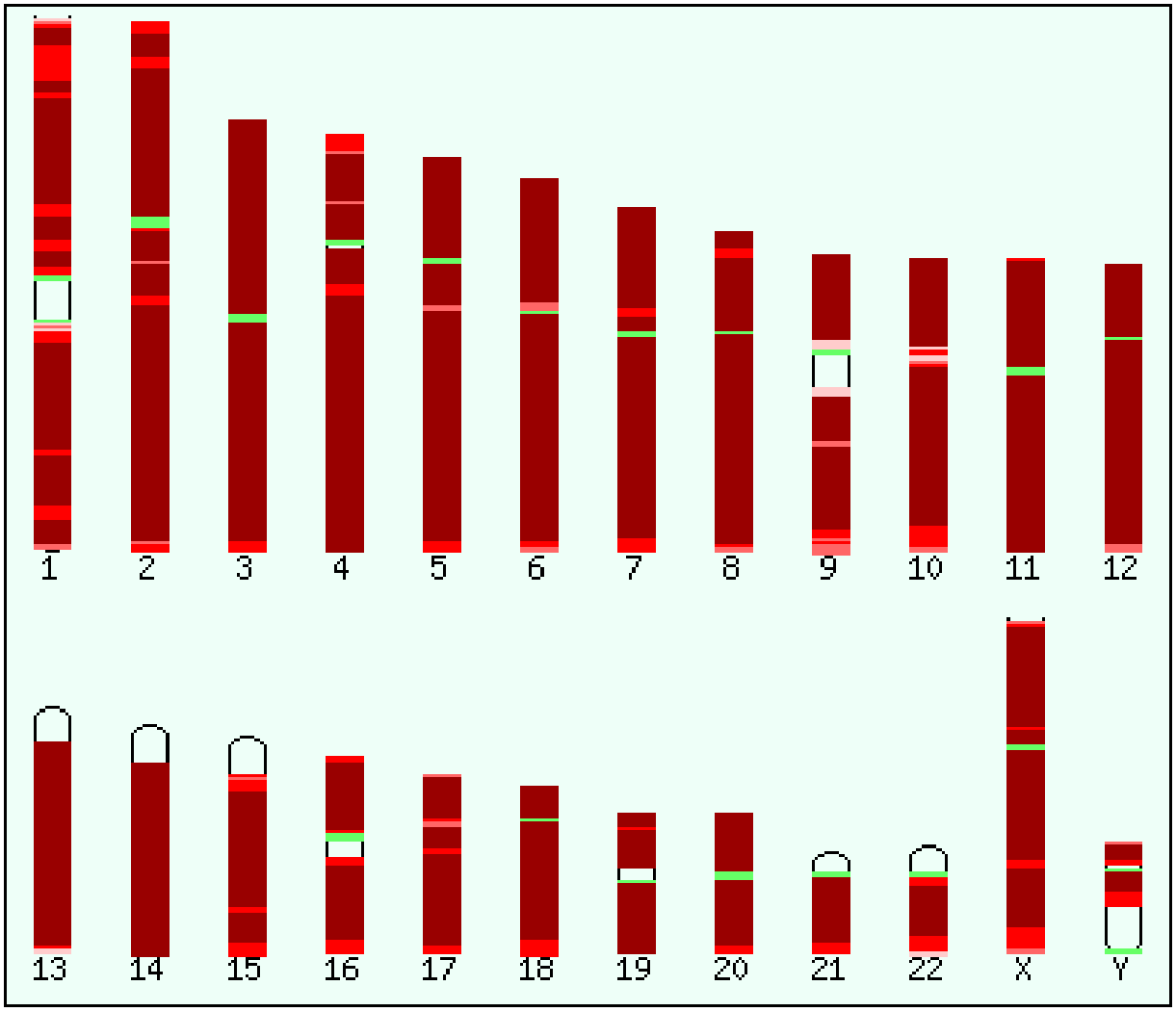
N50 length > 10kb

More than 50% of bin in golden path

Less than 50% of bin in golden path

Clear-

None of bin in golden path



Stats

Freeze date:

July 2003

Estimated size:

3 069.43 Mb

Total mapped:

2 843.41 Mb (92.64%)

No. of supercontigs:

350

Super contig N50 length:

29 104 799 bps

In super contigs > 10Mb

2 307.65 Mb (76 s'ctgs)

In super contigs > 1Mb

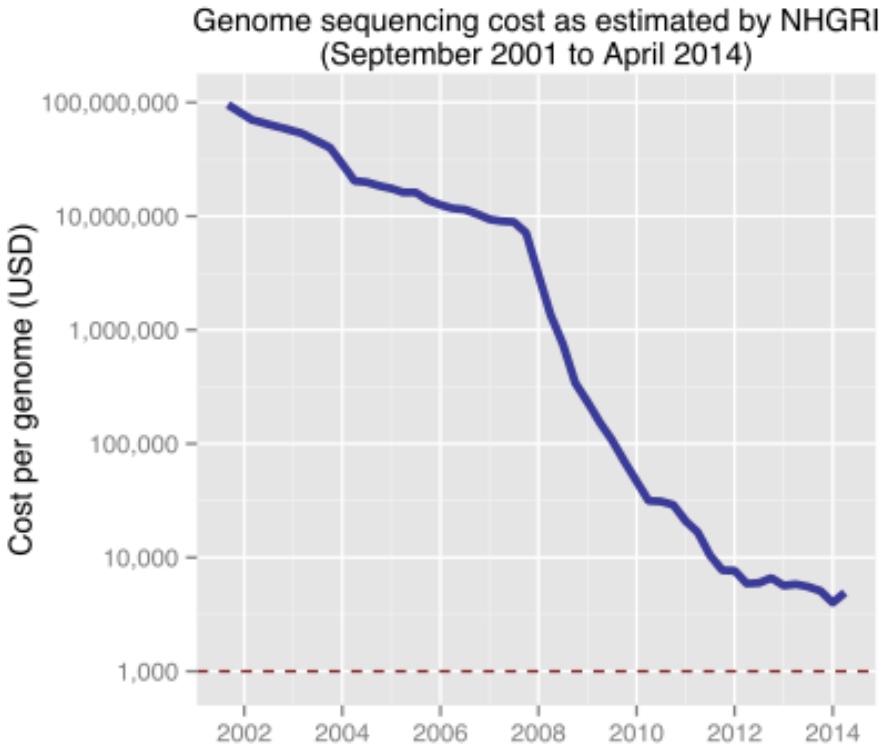
2 789.20 Mb (199 s'ctgs)

In super contigs > 100Kb

Genome sequencing: present situation

Estimates 2015

Bacterial genomes:	> 10 E3
Fungal genomes:	> 10E2
Plant genomes:	~ 10E2
Animal genomes	~ 10E2
Human genomes	> 10
Metagenomes	< 10 E3



https://en.wikipedia.org/wiki/DNA_sequencing