Lecture 2:

Model Organisms

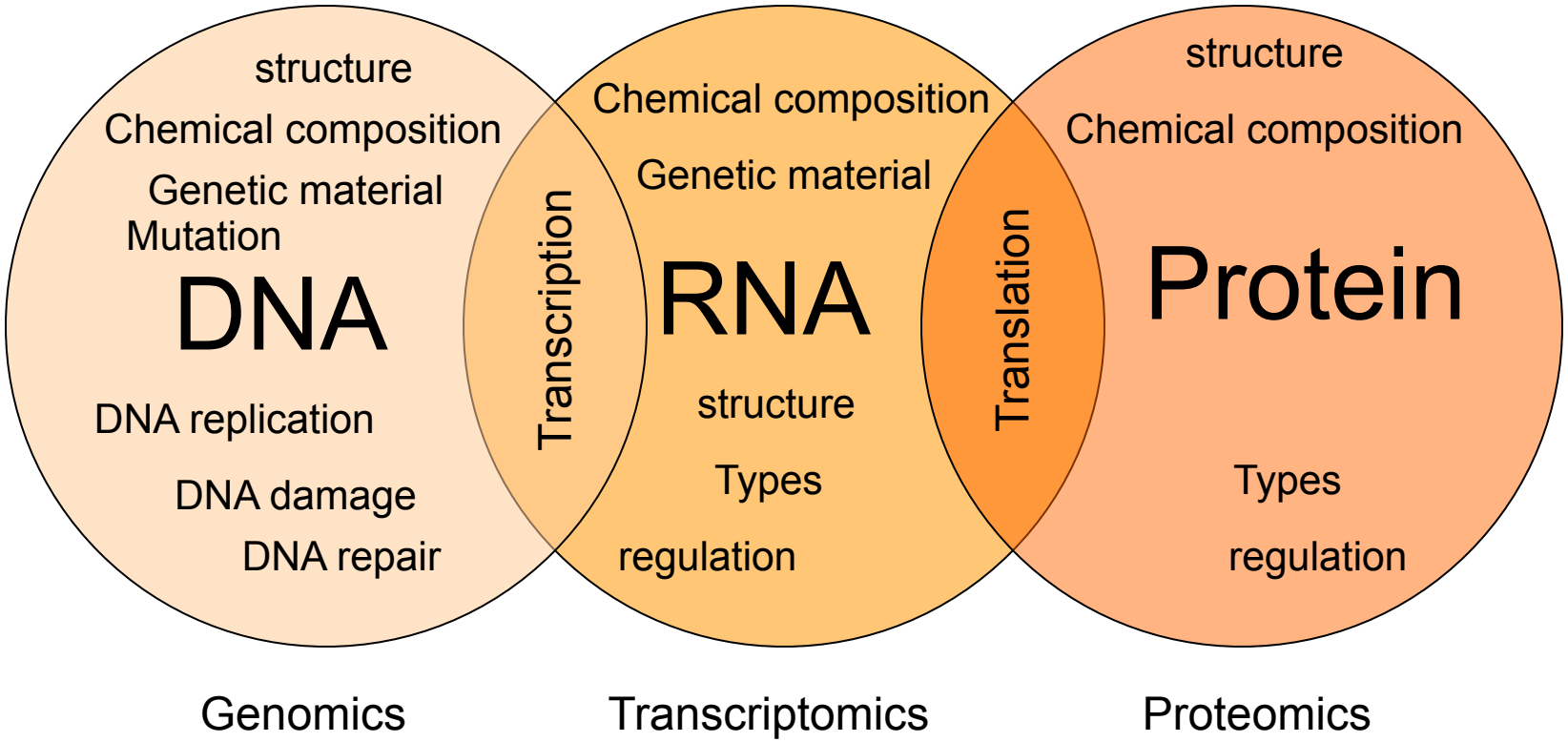
Course 371

AIMS

- Understand the importance of models in science.
- Introduce models organisms.
- Present the most important model organisms.
- Present the general characteristics of the model organisms.

The molecules

This class is about the molecules of life !



The molecules

How can we study these molecules?

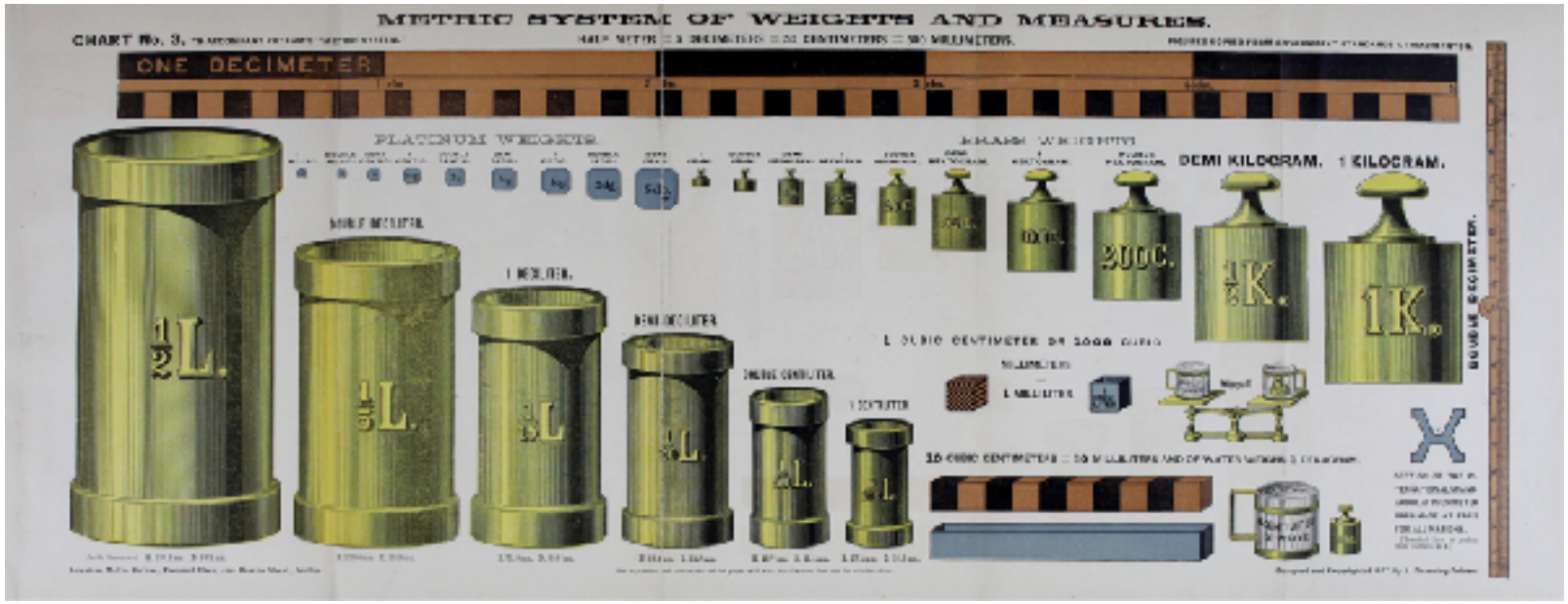
How can we learn about their biology?

We need models!

All branches of science need model systems

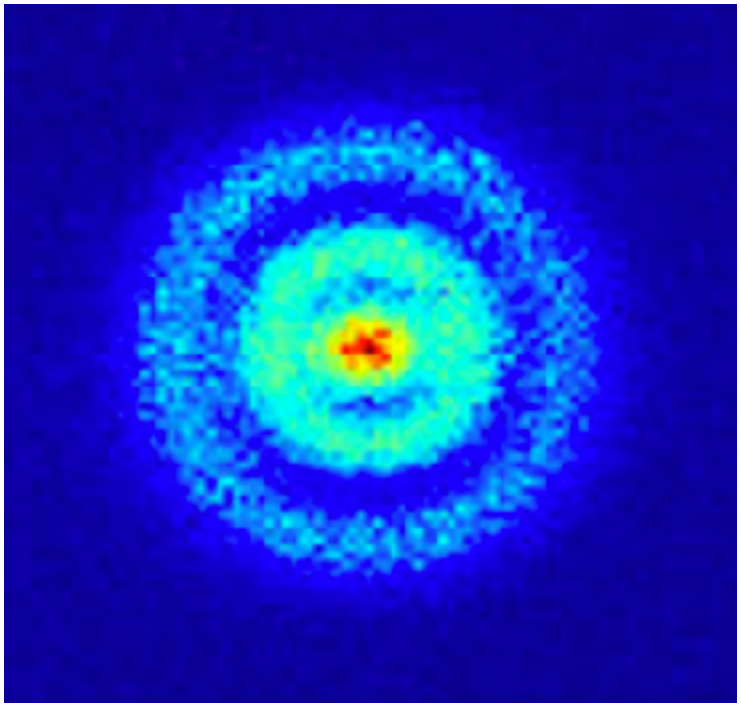
Science and models

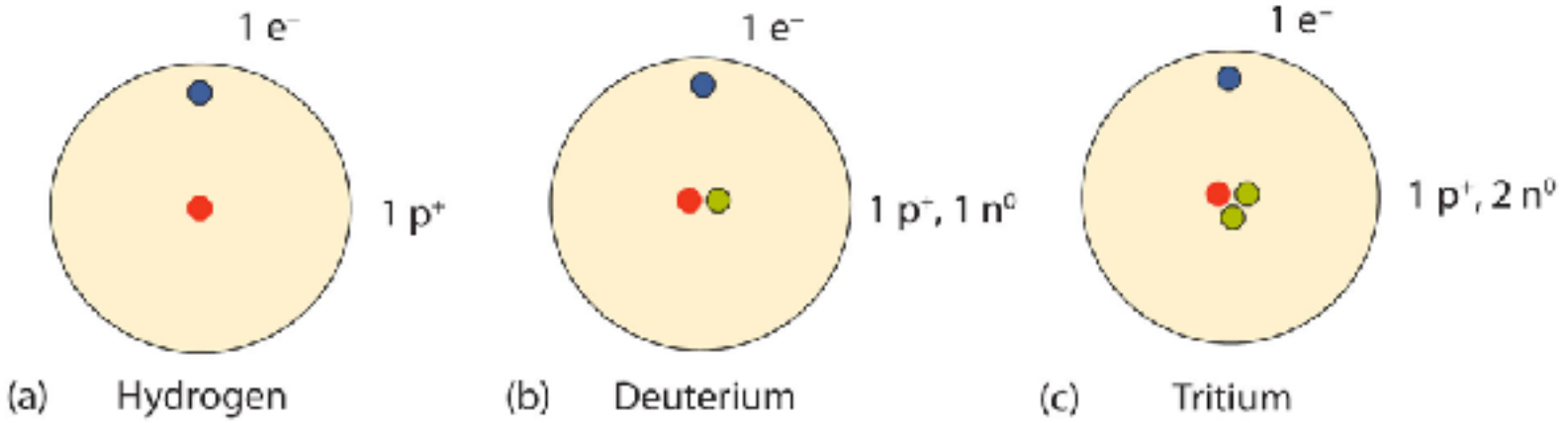
- To study physical sciences, we need a model of weights, volumes etc.
- We need a simple system that can be easily studied to gain the knowledge to understand more complicated systems.



Science and models

- The hydrogen atom is a model to study physics and chemistry.
- The hydrogen atom is simple (one proton and one electron).





Model organisms

How do we learn about all these molecules and mechanisms involved?

Bacteriophage (virus)

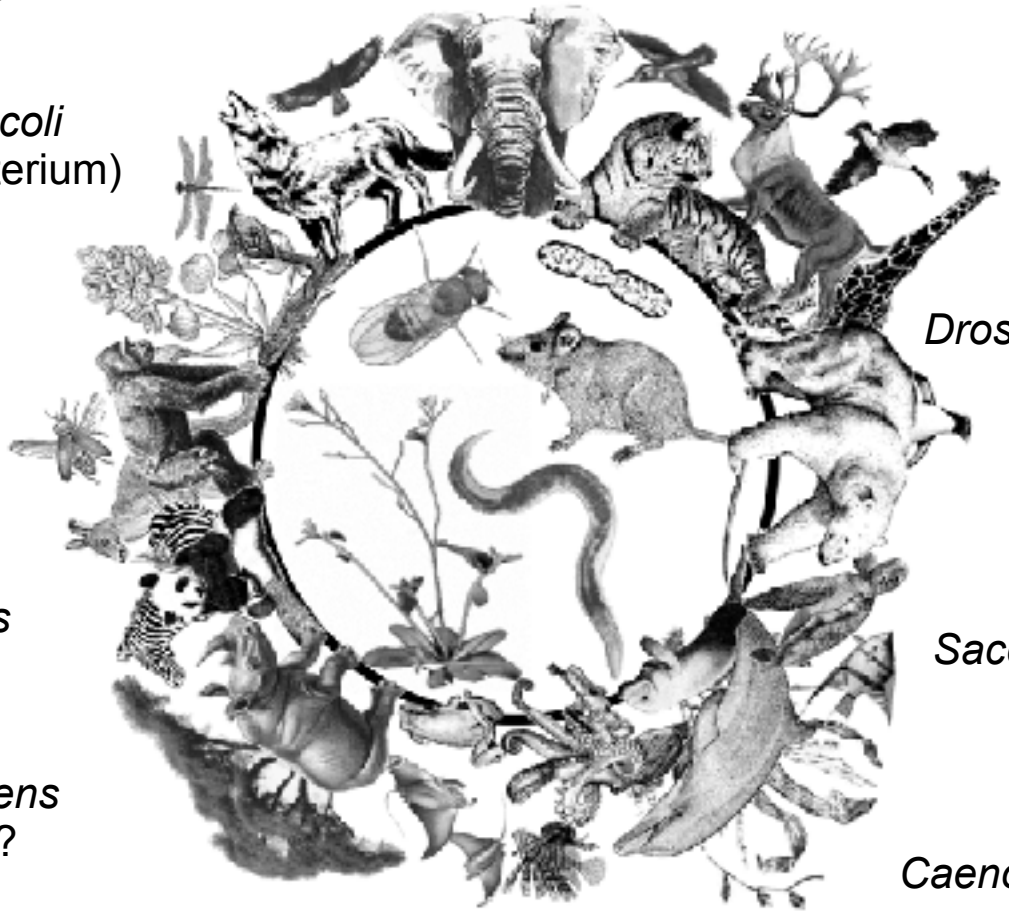
Escherichia coli (intestines' bacterium)

Zea mays (corn)

Danio rerio (zebrafish)

Mus musculus (mouse)

Homo sapiens (Human)?



Arabidopsis thaliana (plant)

Neurospora crassa (bread mold)

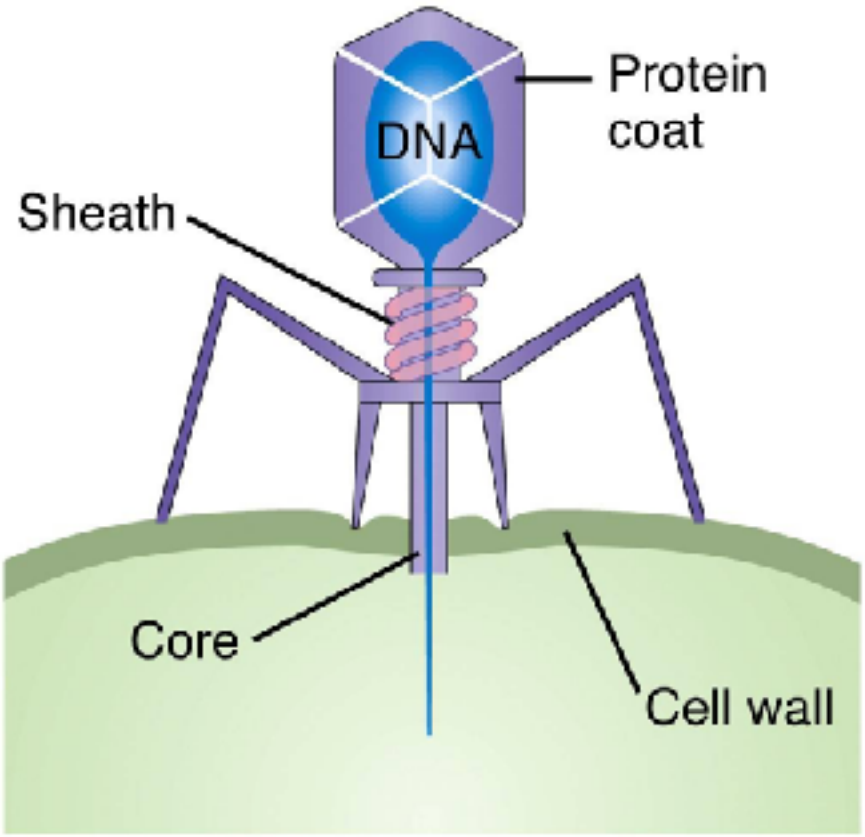
Drosophila melanogaster (fruit fly)

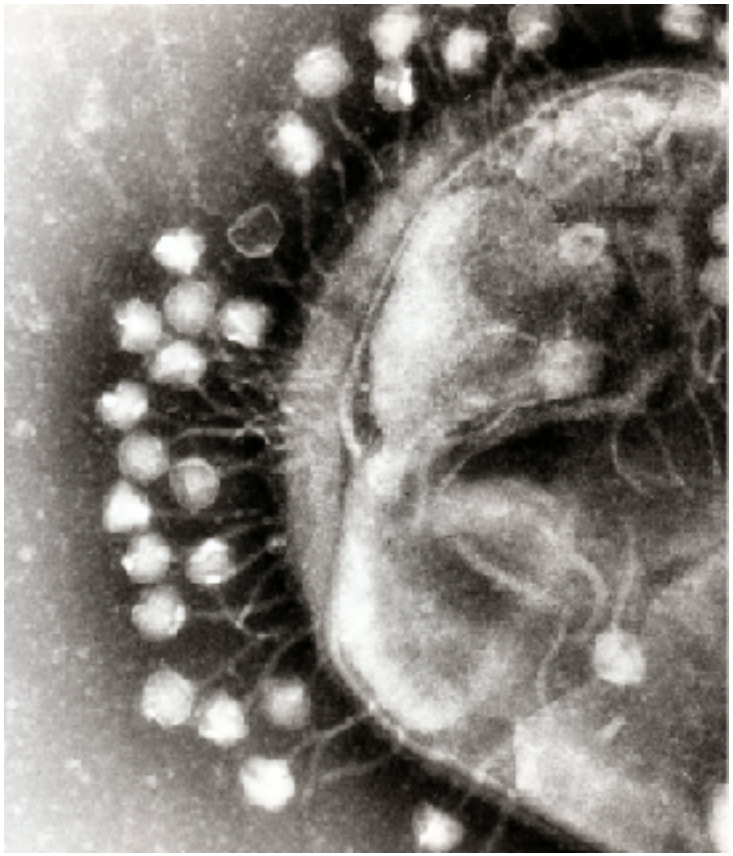
Pisum sativum (garden pea)

Saccharomyces cerevisiae (budding yeast)

Caenorhabditis elegans (worm)

Bacteriophage (virus)



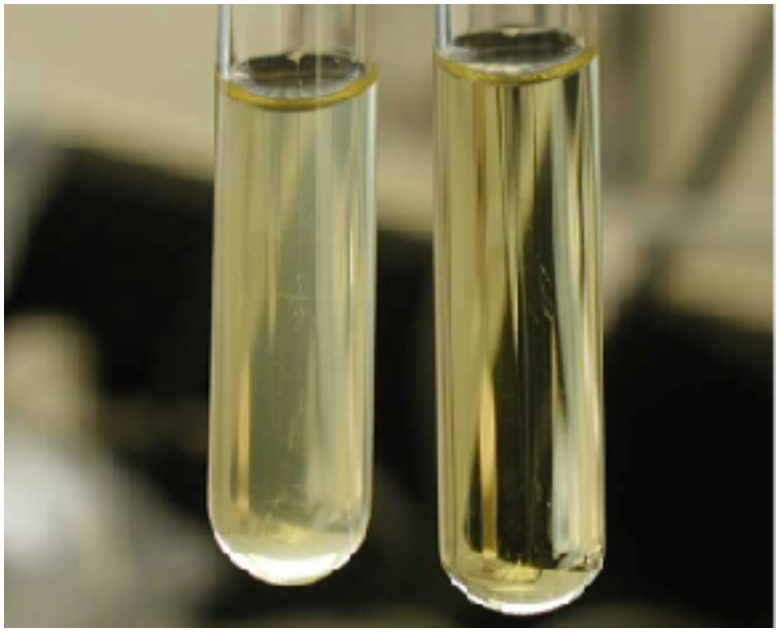


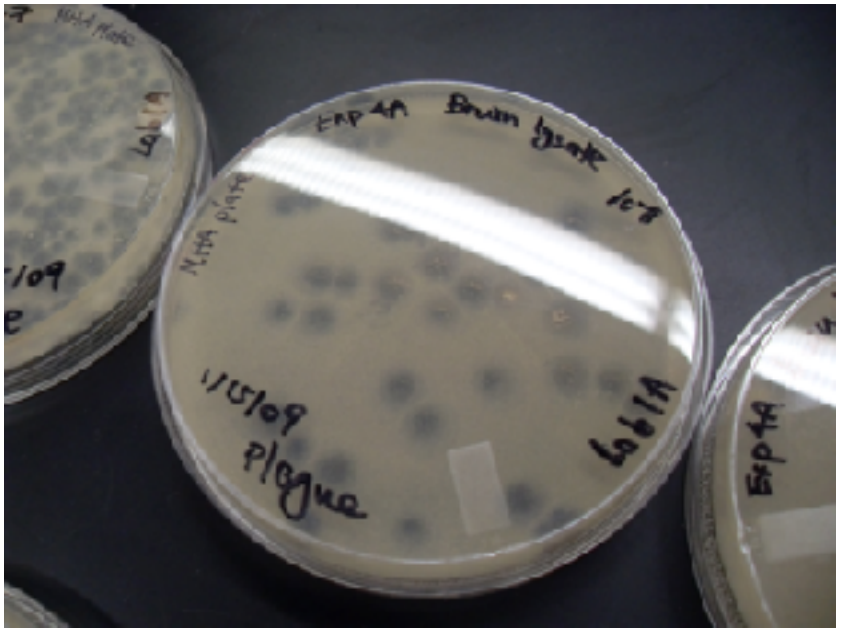
- Virus (living?)
- 24-200 nm in length

- single entities.
- Simple structure
- Haploid

Bacteriophage (virus)

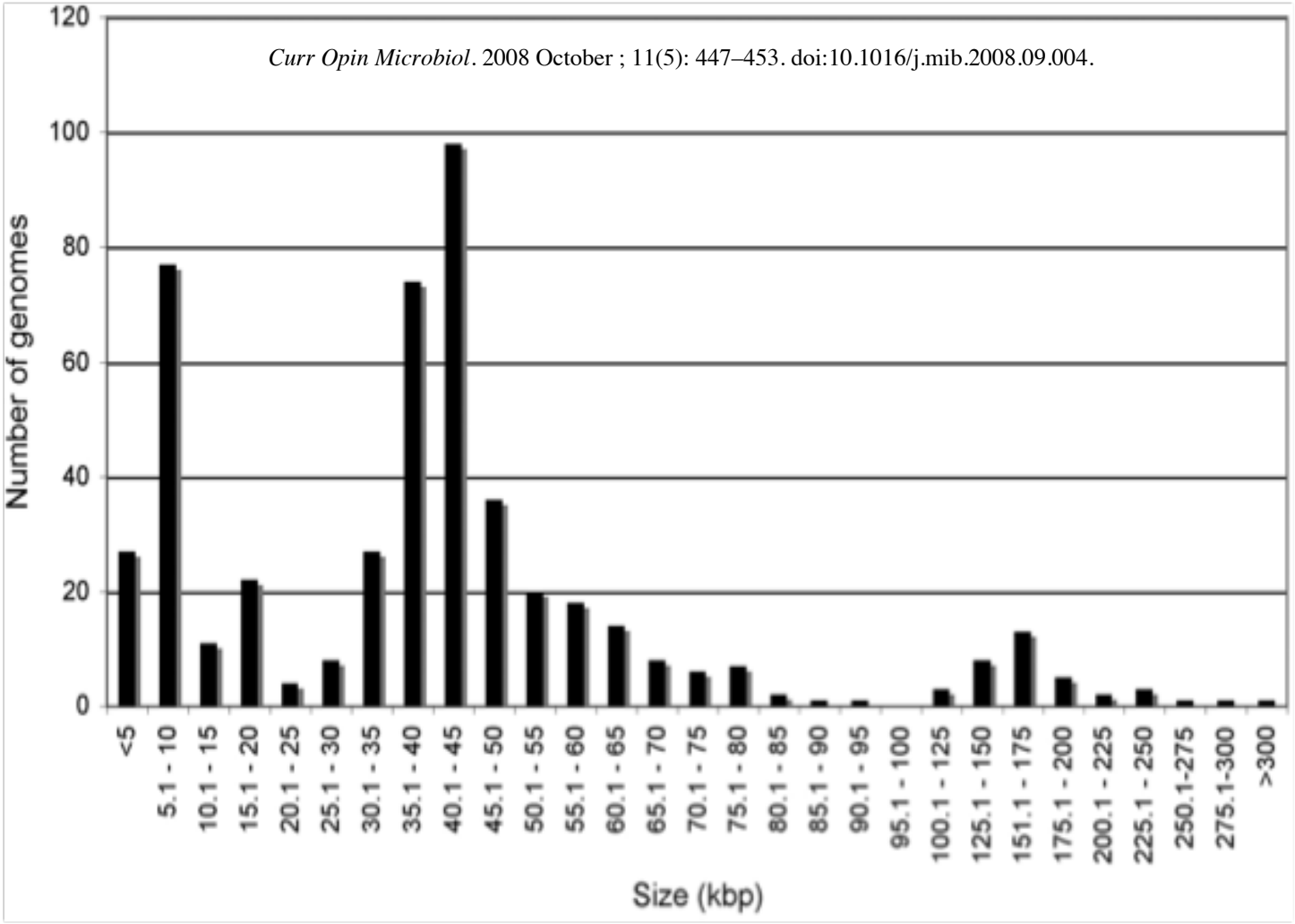
- Grows on/in bacteria
- Can be grown into millions of copies
- Fast growth
- Easy to culture, store, and manipulate genetically





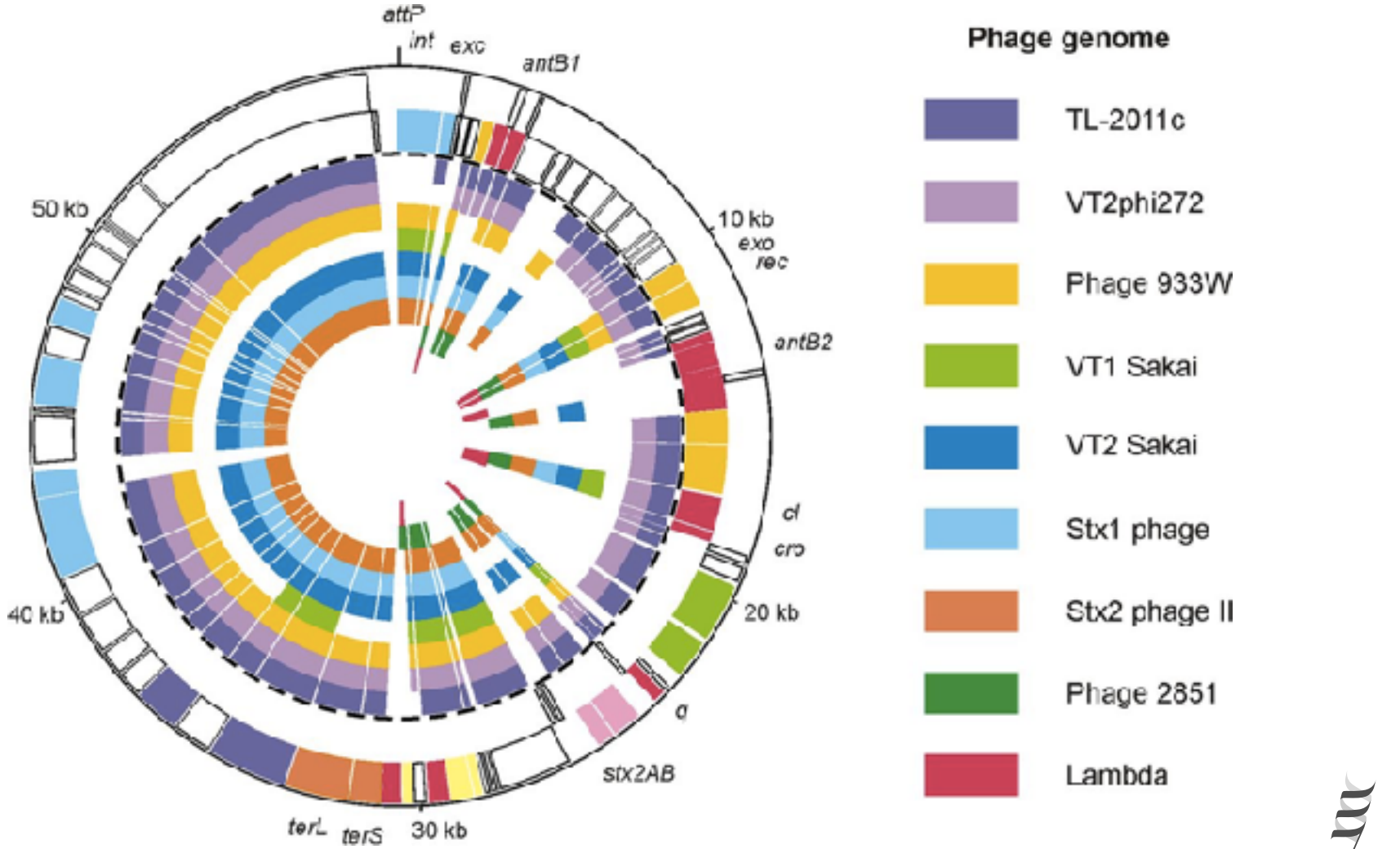
Bacteriophage (virus)

- Small genome (~5-300 Kb)



Bacteriophage (virus)

- Simple genome



Bacteriophage (virus)

- Research resources are available.

Nature 265, 687-695 (24 February 1977) | doi:10.1038/265687a0; Accepted

Nucleotide sequence of bacteriophage ΦX174 DNA

F. Sanger, G. M. Air*, B. G. Barrell, N. L. Brown†, A. R. Coulson, J. C. Fiddes, C. A. Hutchison, III‡, P. M. Slocombe§ & M. Smith¶

1. MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK
2. Present addresses: * John Curtin School of Medical Research, Microbiology Department, Canberra City ACT 2601, Australia; † Department of Biochemistry, University of Bristol, Bristol BS8 1TD, UK; ‡ Department of Bacteriology and Immunology, University of North Carolina, Chapel Hill, North Carolina 27514, § Max Planck Institut für Molekulare Genetik, 1 Berlin 33, FRG; ¶ Department of Biochemistry, University of British Columbia, Vancouver BC, Canada V6T 1W5,

A DNA sequence for the genome of bacteriophage ΦX174 of approximately 5,375 nucleotides has been determined using the rapid and simple 'plus and minus' method. The sequence identifies many of the features responsible for the production of the proteins of the nine known genes of the organism, including initiation and termination sites for the proteins and RNAs. Two pairs of genes are coded by the same region of DNA using different reading frames.

▲ Top

Journal of Molecular Biology
Volume 162, Issue 4, 25 December 1982, Pages 729-773

Nucleotide sequence of bacteriophage λ DNA

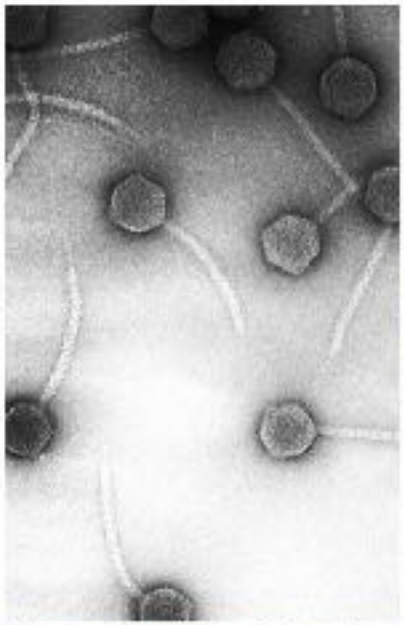
F. Sanger, A.R. Coulson, G.F. Hong, D.F. Hill, G.B. Petersen

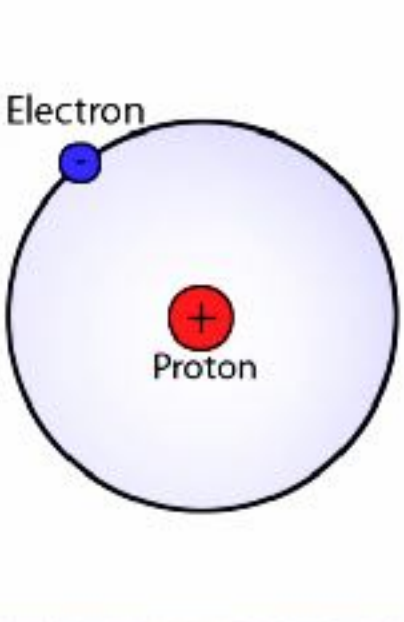
Abstract

The nucleotide sequence of the DNA of bacteriophage λ has been determined using the dideoxy chain termination method in conjunction with random cloning in M13 vectors. Various methods were studied for sequencing specific regions to complete the sequence, but all were much slower than the random approach. The DNA in its circular form contains 48,502 base pairs. Open reading frames were identified and, where possible, ascribed to genes by comparing with the previously determined genetic map. The reading frames for 46 genes were clearly identified, though in about 20 the position of the protein initiation site could not be rigorously established. Probable positions for the *Nil*, *cII* and *Am* genes are suggested but remain uncertain. There are about 20 other unidentified reading frames that may code for proteins.

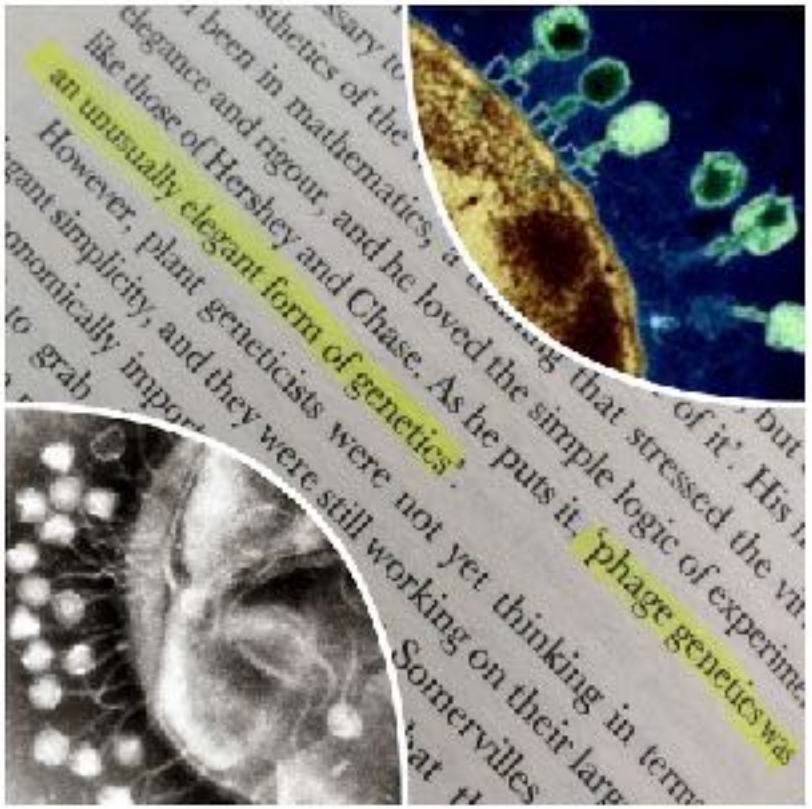
The genome is fairly compact with comparatively little non-coding DNA. In many cases the translation terminator and initiator overlap, particularly in the sequence A-T-G-A where the TGA terminates one gene and the ATG initiates the next. Such structures seem to be characterized by a purine-rich sequence, rather than by a specific "Shine and Dalgarno" sequence, before the initiator. In the whole of the left arm the codon CTA, which is normally read by a minor leucine tRNA, is absent. The distribution of other rare codons in the genes of the left arm suggests that they may have a controlling function on the relative amounts of the proteins produced.

Bacteriophage (virus)





that was as clean and simple as physics, which produced clear-cut, mathematical results. Phage seemed to Delbrück to have the potential to be biology's hydrogen atoms – the simplest possible example of life's ability to reproduce. And he decided to work with them until they provided the paradox he was after. 'This seemed



The hydrogen atom for biologists

Bacteriophage (virus)

word about phage. The year after *What is Life?* appeared, Delbrück began teaching a summer course at Cold Spring Harbor Laboratory to instruct newcomers in the basics of phage.

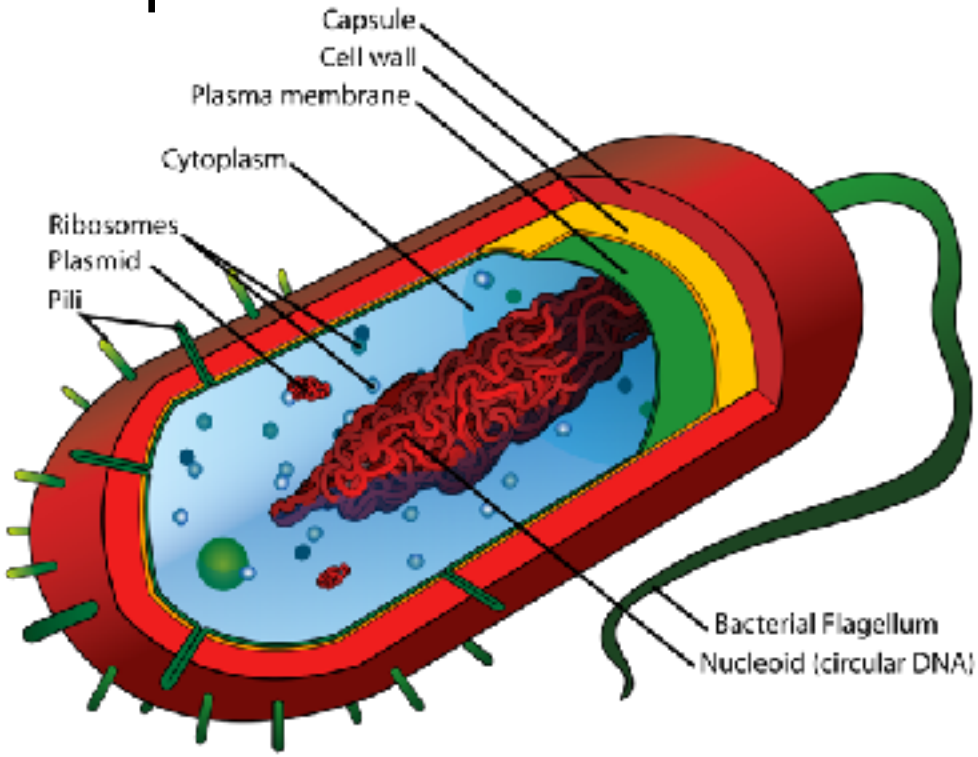
The phage church

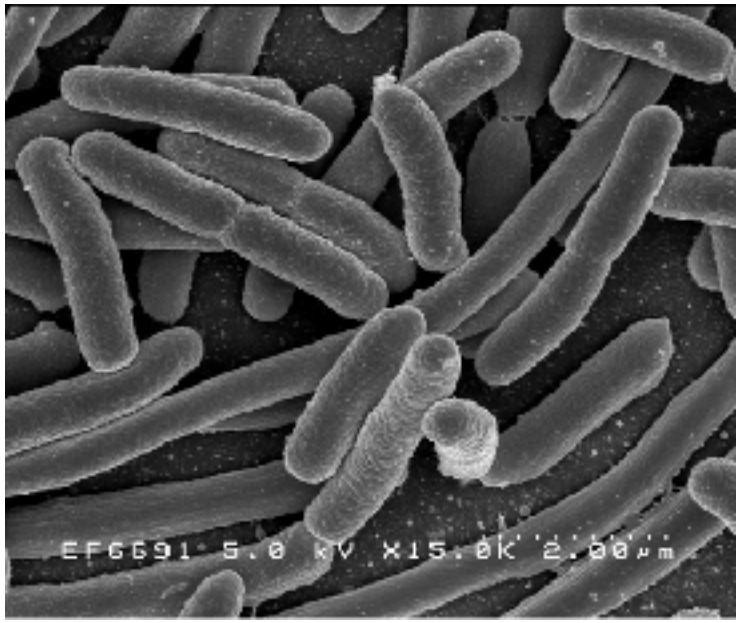
Max Delbrück (1906 - 1981) | Alfred D. Hershey (1908 - 1997) | Salvador E. Luria (1921 - 1991)

guanine with cytosine, so that each strand of the helix is a complementary copy of the other. Phage and the Phage Group had provided the crucial evidence that phage and bacteria actually possessed genes – thanks to Hershey and Delbrück's experiments on phage resistance; they then demonstrated that those genes were composed of DNA. None of that diminishes Watson and Crick's achievement, but it is a useful reminder that – like any brilliant theoreticians – they could not have done it on

Escherichia coli (intestines' bacterium)

- Prokaryote.
- Single celled organism.
- haploid

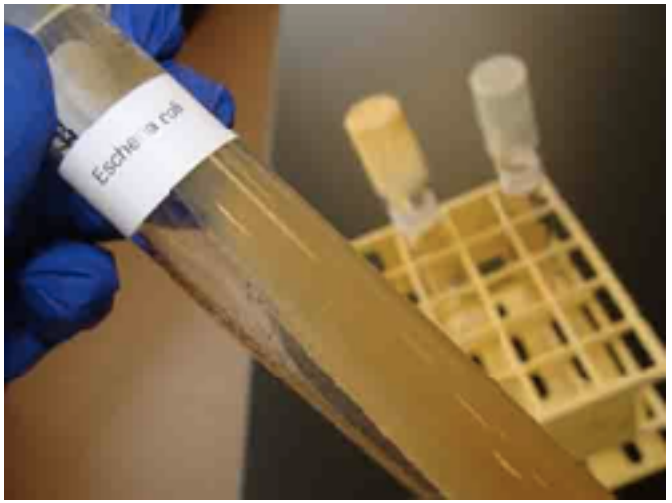




- Small in size
- ~ 2um in length
- ~ 0.5 um in width

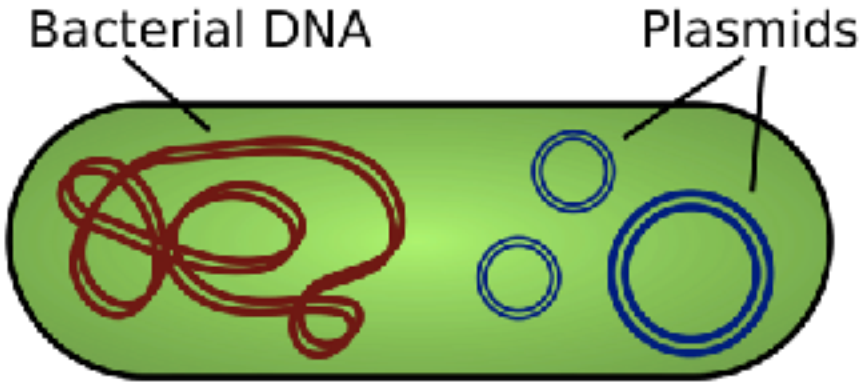
Escherichia coli (intestines' bacterium)

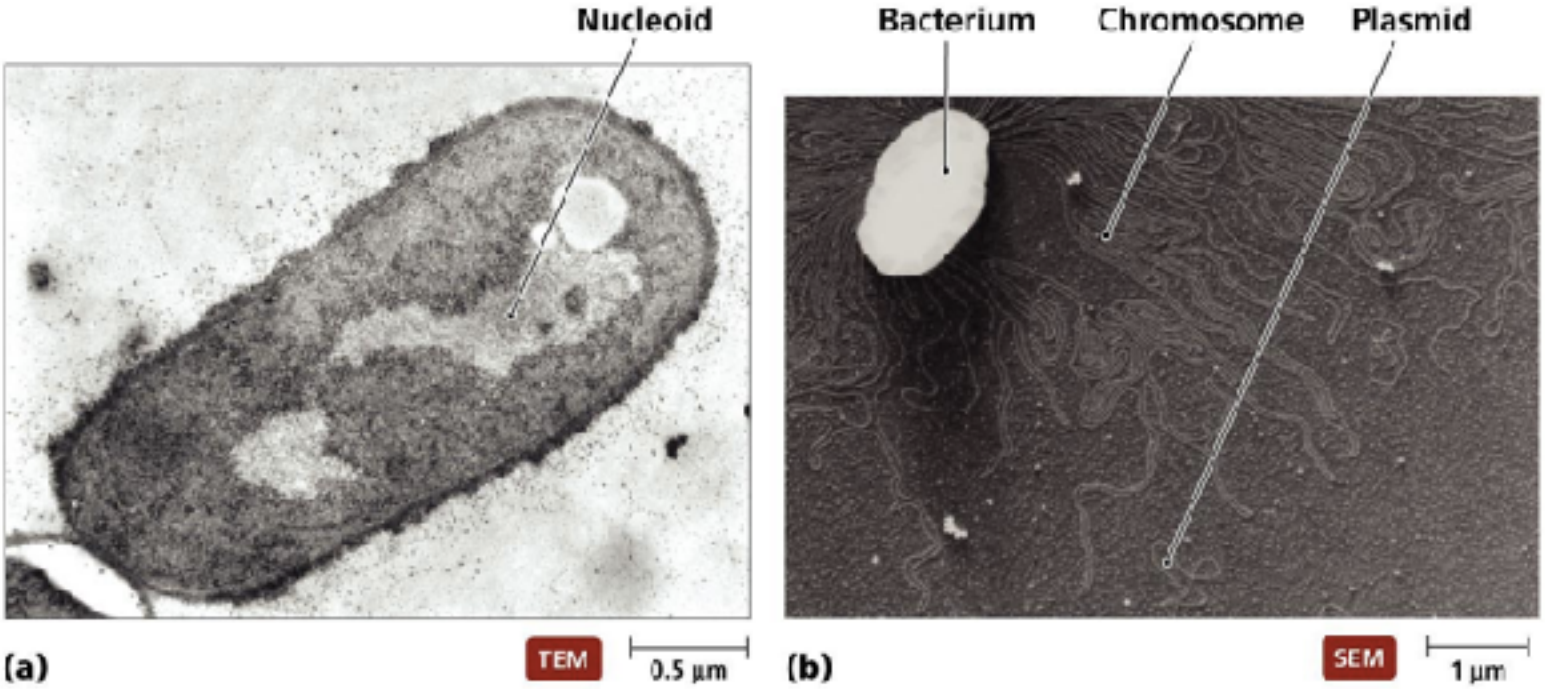
- Easy to grow in lab
- Can be grown into millions of copies
- Fast growth
- Easy to culture, store, and manipulate genetically



Escherichia coli (intestines' bacterium)

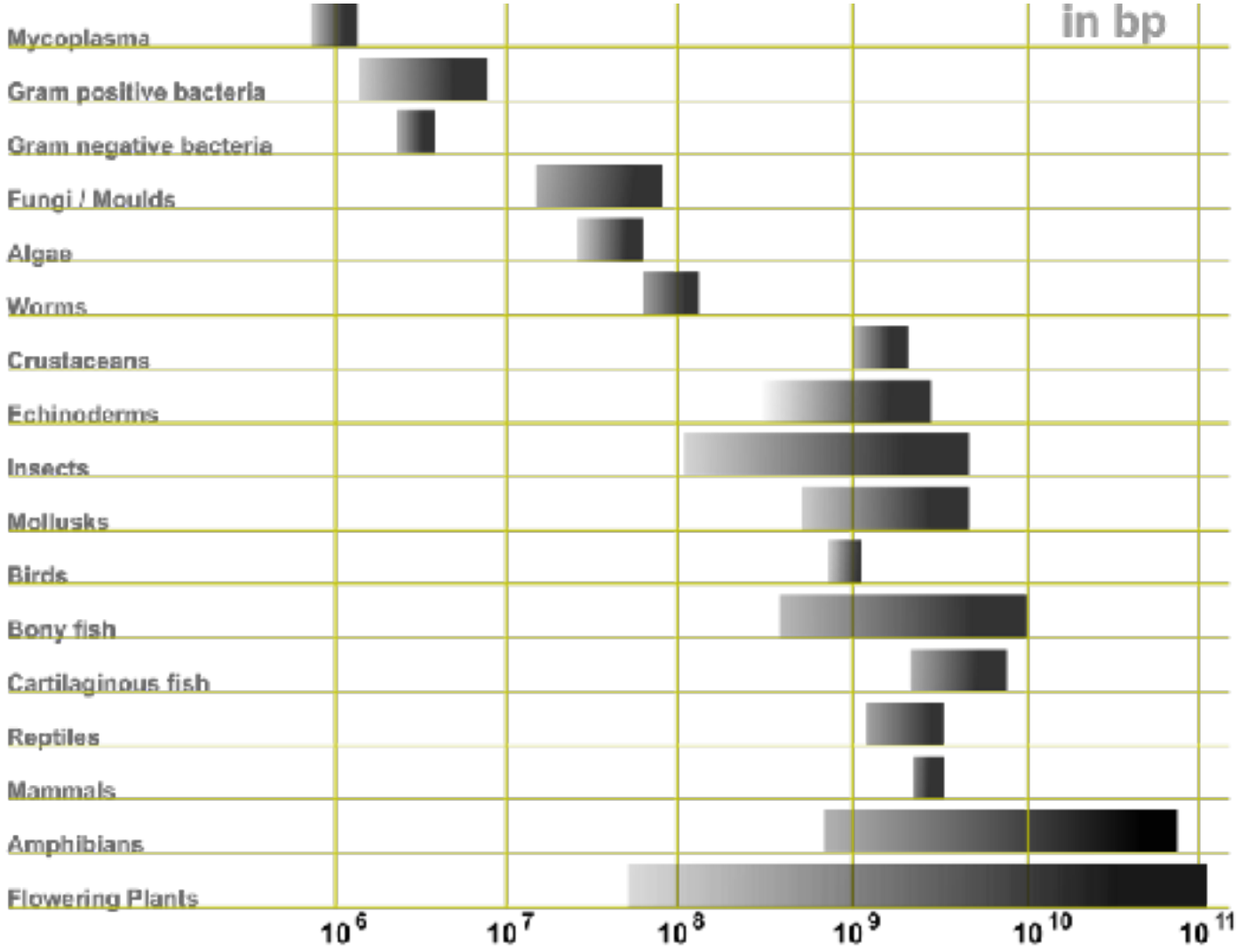
- Genetic material in:
  - •A single circular chromosome.
  - • Small plasmid





Escherichia coli (intestines' bacterium)

- Relatively small genome



Escherichia coli (intestines' bacterium)

- Research resources are available.

ARTICLE

The Complete Genome Sequence of *Escherichia coli* K-12

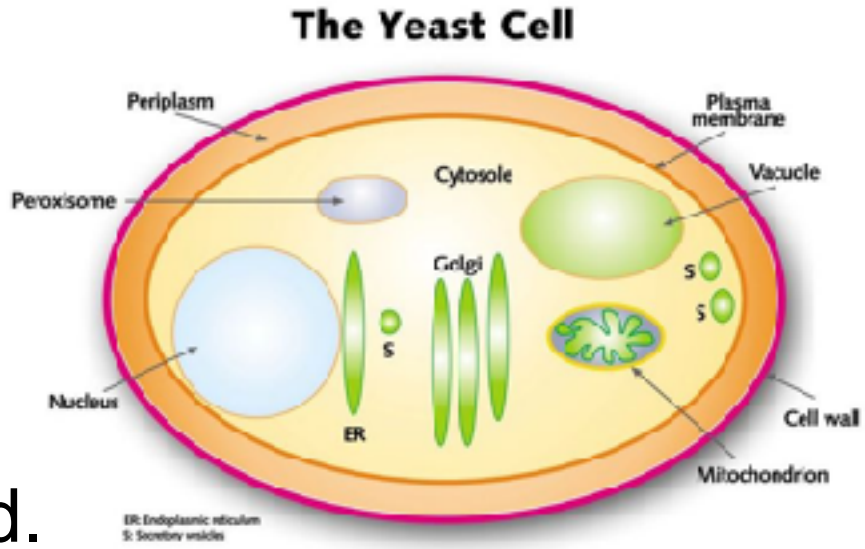
Frederick R. Blattner,* Guy Plunkett III,* Craig A. Bloch, Nicole T. Perna, Valerie Burland, Monica Riley, Julio Collado-Vides, Jeremy D. Glasner, Christopher K. Rode, George F. Mayhew, Jason Gregor, Nelson Wayne Davis, Heather A. Kirkpatrick, Michael A. Goeden, Debra J. Rose, Bob Mau, Ying Shao

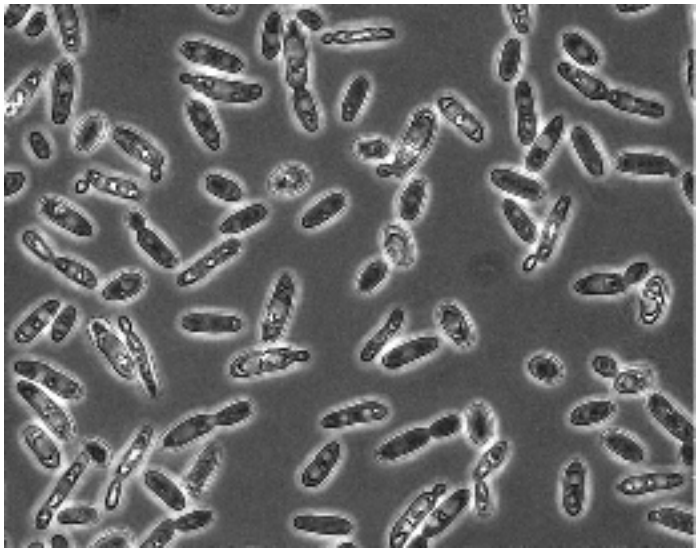
The 4,639,221–base pair sequence of *Escherichia coli* K-12 is presented. Of 4288 protein coding genes annotated, 38 percent have no attributed function. Comparison with five other sequenced microbes reveals ubiquitous as well as narrowly distributed gene families; many families of similar genes within *E. coli* are also evident. The largest family of paralogous proteins contains 80 ABC transporters. The genome as a whole is strikingly organized with respect to the local direction of replication; guanines, oligonucleotides possibly related to replication and recombination, and most genes are so oriented. The genome also contains insertion sequence (IS) elements, phage remnants, and many other patches of unusual composition indicating genome plasticity through horizontal transfer.

The first 1.92 Mb (*13*, *14*), positions 2,686,777 to 4,639,221 [in base pairs (bp)], was sequenced from our overlapping set of 15- to 20-kb MG1655 lambda clones (*15*) by means of radioactive chemistry and was deposited in GenBank between 1992 and 1995. Subsequently, we switched to dye-terminator fluorescence sequencing (Applied Biosystems). In addition to greater speed and lower cost, this new technology avoided electrophoretic compression arti-

Saccharomyces cerevisiae (budding yeast)

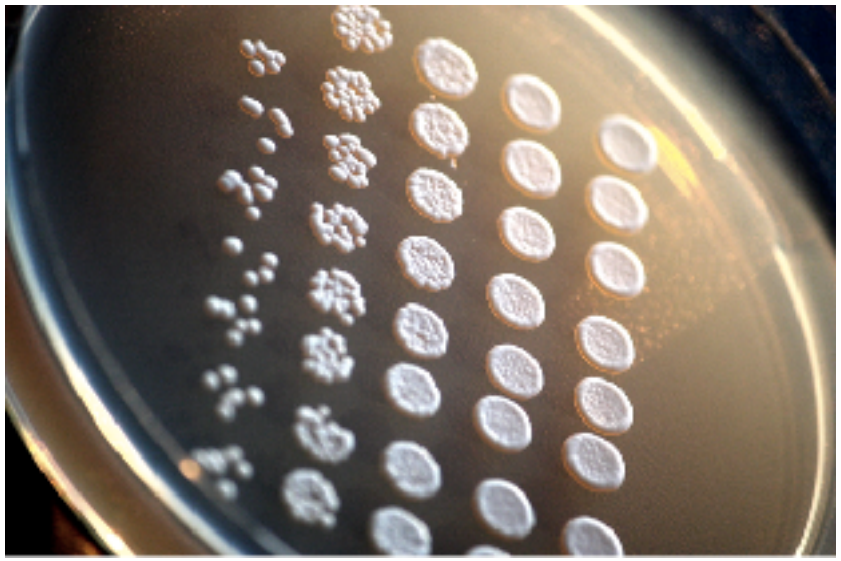
- Eukaryote.
- Fungi.
- Single celled organism.
- Grows haploid or diploid.
- Sexual and asexual life cycles.
- Small in size (~ 5-10 um in diameter).





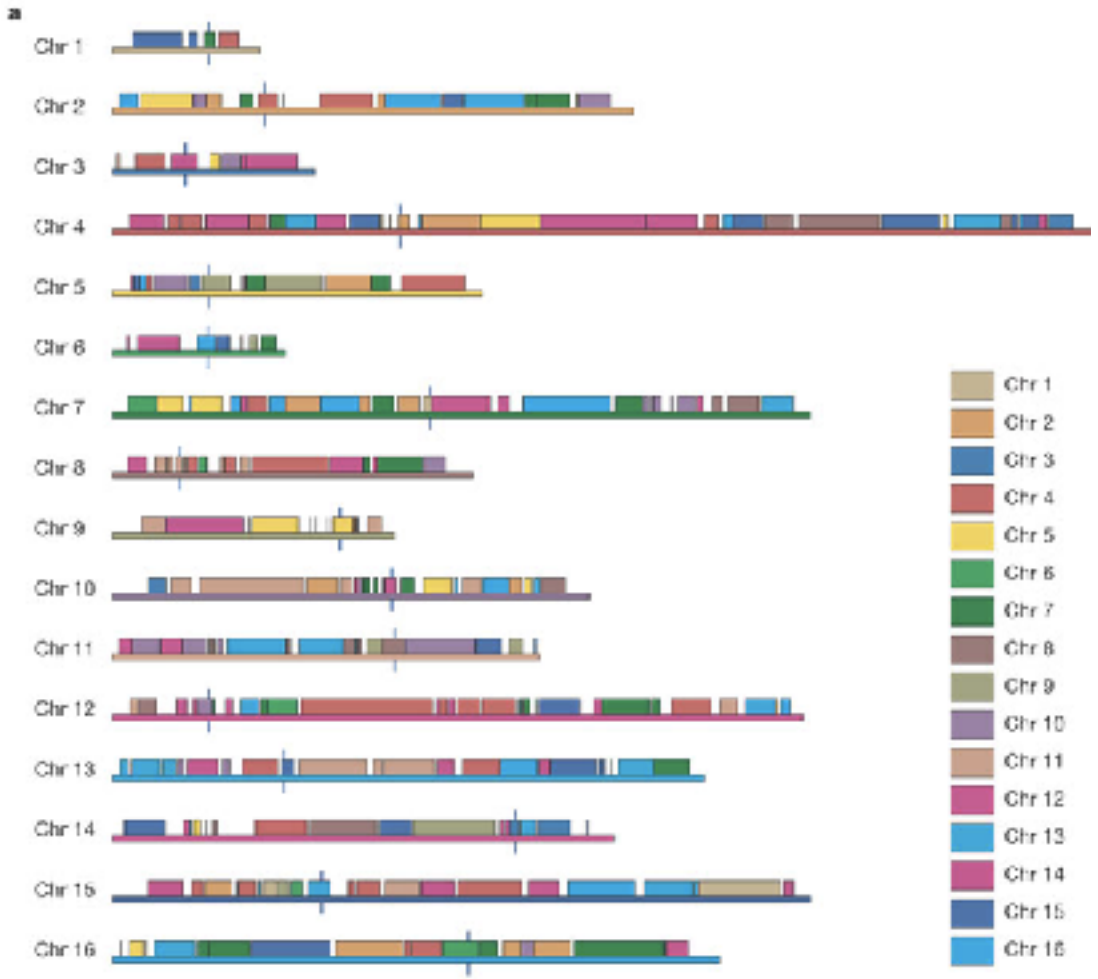
Saccharomyces cerevisiae (budding yeast)

- Easy to grow in lab
- Fast growth
- Easy to culture, store, and manipulate genetically



Saccharomyces cerevisiae (budding yeast)

- Relatively small genome (12Mb)
- Genome distributed over 16 chromosomes



Saccharomyces cerevisiae (budding yeast)

- Research resources are available.

Overview of the yeast genome

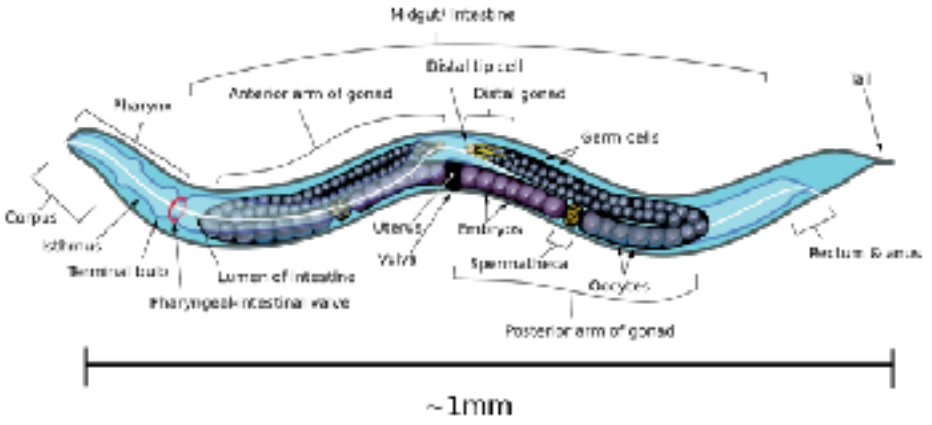
H. W. Mewes, K. Albermann, M. Bähr, D. Frishman, A. Gleissner, J. Hani, K. Heumann, K. Kleine, A. Maierl, S. G. Oliver†, F. Pfeiffer & A. Zollner

Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany
†University of Manchester Institute of Science And Technology (UMIST), Sackville Street, Manchester M60 1QD, UK

The collaboration of more than 600 scientists from over 100 laboratories to sequence the *Saccharomyces cerevisiae* genome was the largest decentralised experiment in modern molecular biology and resulted in a unique data resource representing the first complete set of genes from a eukaryotic organism. 12 million bases were sequenced in a truly international effort involving European, US, Canadian and Japanese laboratories. While the yeast genome represents only a small fraction of the information in today's public sequence databases, the complete, ordered and non-redundant sequence provides an invaluable resource for the detailed analysis of cellular gene function and genome architecture. In terms of throughput, completeness and information content, yeast has always been the lead eukaryotic organism in genomics; it is still the largest genome to be completely sequenced.

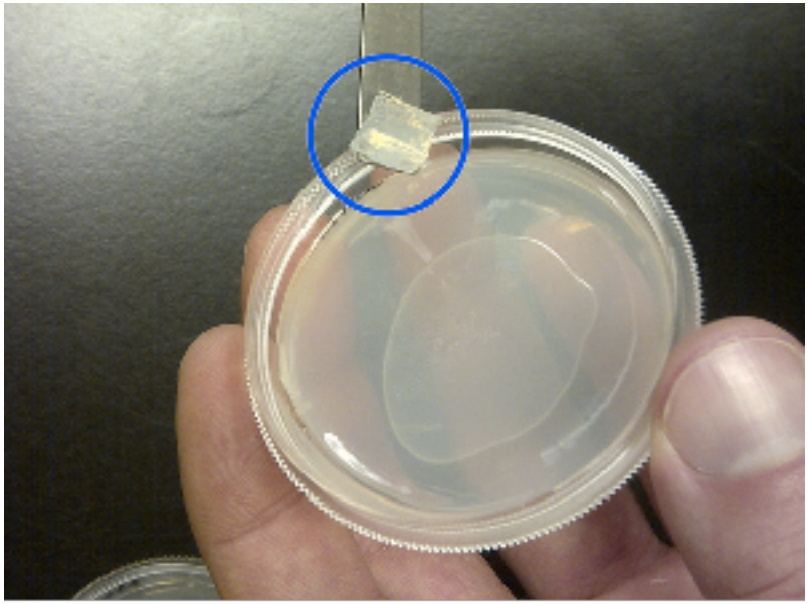
Caenorhabditis elegans (worm)

- Eukaryote.
- Animal - Nematode.
- Multicellular.
- Hermaphrodite.
- Sexual and asexual life cycles.
- Small in size (~ 1 mm in length).
- Diploid.



Caenorhabditis elegans (worm)

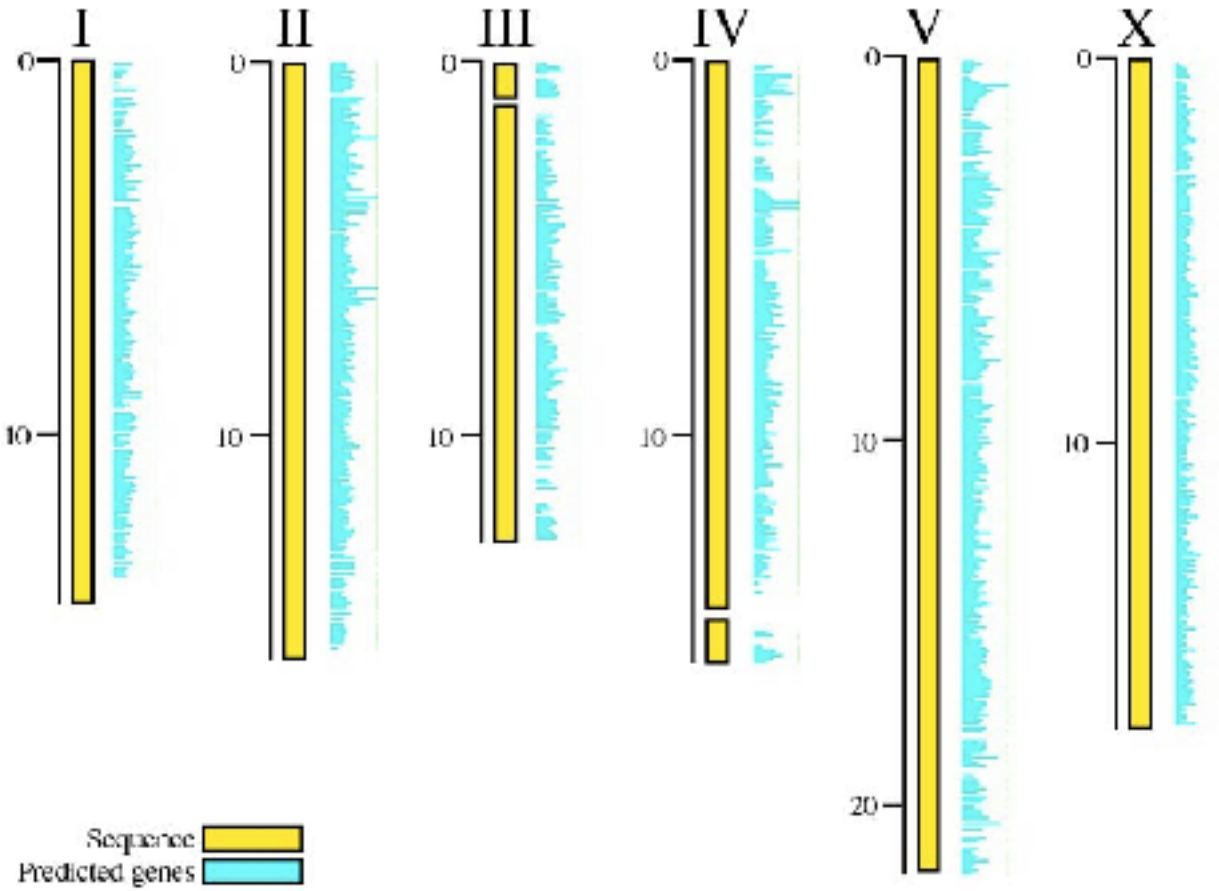
- Easy to grow in lab
- Fast growth
- Short life cycle
- Known number of cells



- Easy to culture, store, and manipulate genetically.
- Eggs can be stored.

Caenorhabditis elegans (worm)

- Genome size (~ 97Mb)
- Genome distributed over 6 chromosomes



Caenorhabditis elegans (worm)

- Research resources are available.

SPECIAL SECTION C. ELEGANS: SEQUENCE TO BIOLOGY

Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology

The *C. elegans* Sequencing Consortium*

REVIEW

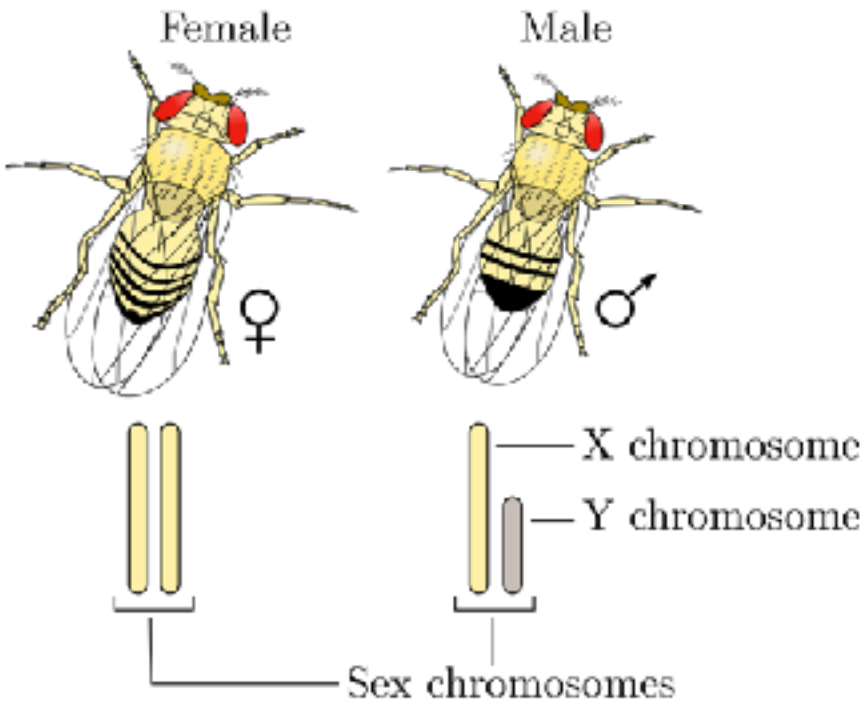
The 97-megabase genomic sequence of the nematode *Caenorhabditis elegans* reveals over 19,000 genes. More than 40 percent of the predicted protein products find significant matches in other organisms. There is a variety of repeated sequences, both local and dispersed. The distinctive distribution of some repeats and highly conserved genes provides evidence for a regional organization of the chromosomes.

represented only in YACs.

By 1989, it became apparent that, with the physical map in hand, complete sequencing of the genome might be both feasible and desirable. Joint funding [from the National Institutes of Health and the UK Medical Research Council (MRC)] for a pilot study was arranged, and in 1990, the first 3-megabase (Mb) sequence was undertaken. Success in this venture (*6*, *7*) resulted in full funding and the expansion of the two groups of the consortium in 1993.

Sequencing began in the centers of the chromosomes, where

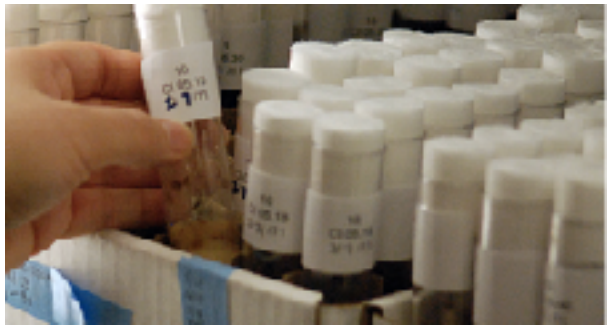
Drosophila melanogaster (fruit fly)



- Eukaryote.
- Animal - Insect.
- Multicellular.
- Diploid

- Sexual life cycle.
- Sexual dimorphism
- ~ 2.5 mm in length

Drosophila melanogaster (fruit fly)

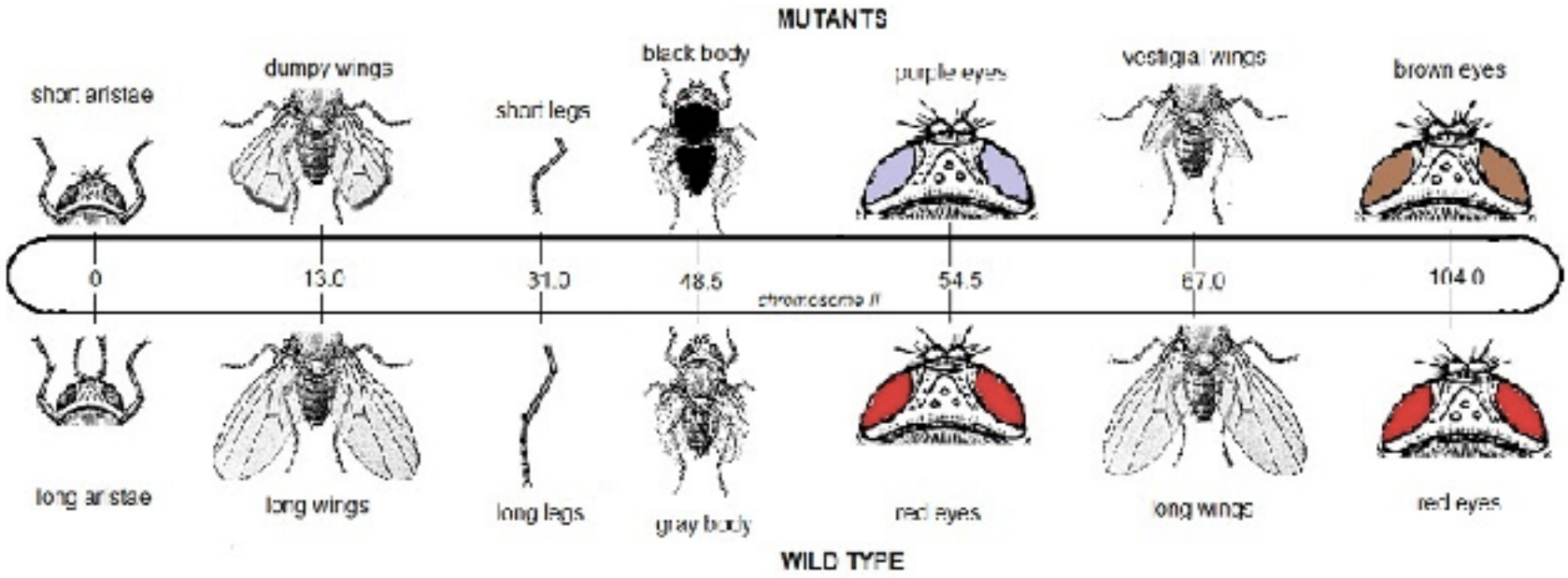


- Easy to grow in lab
- Occupies relatively a small space
- Short life cycle

- Easy to manipulate genetically.
- A living stock has to be maintained.

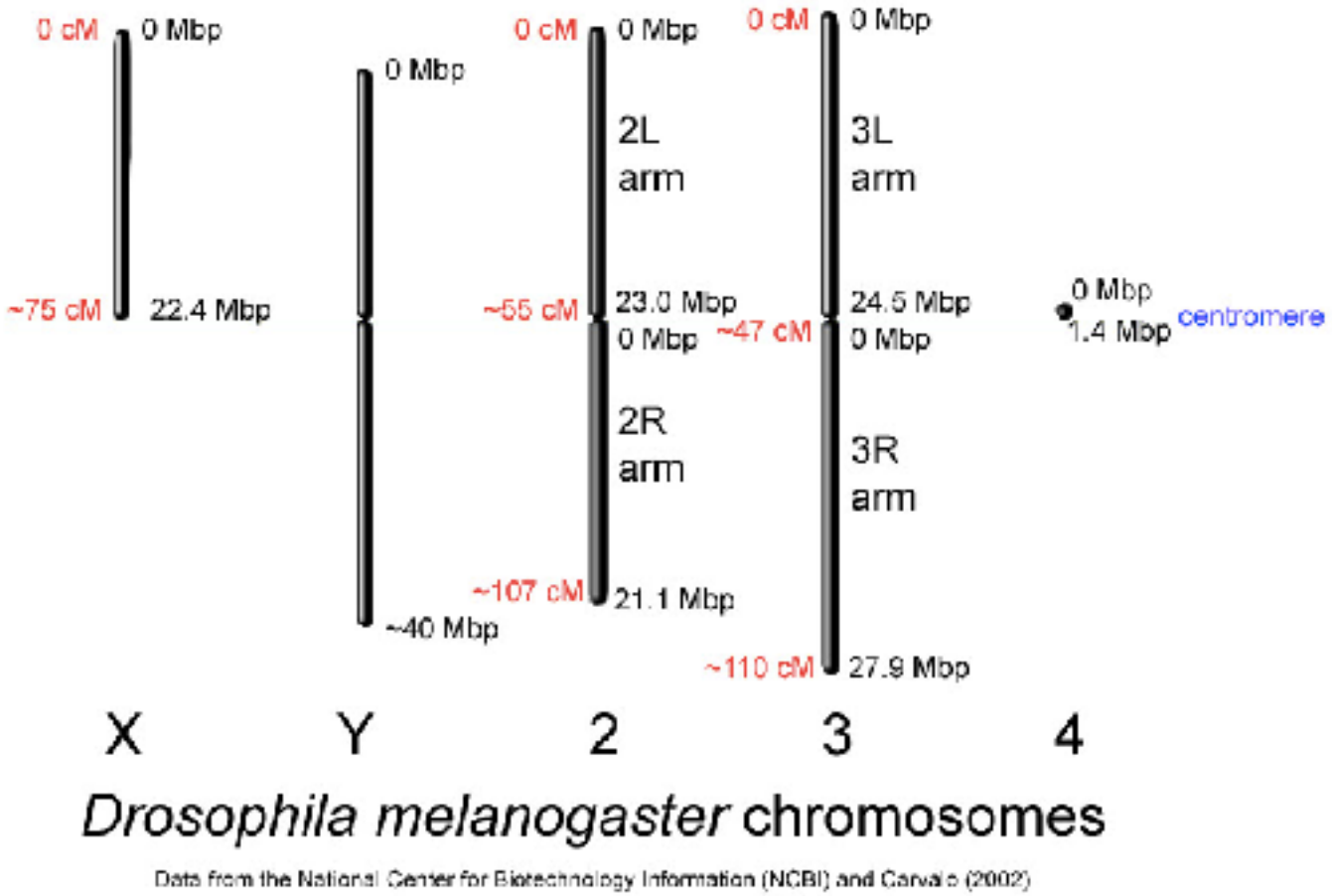
Drosophila melanogaster (fruit fly)

- Many phenotypic markers available to conduct genetic studies.



Drosophila melanogaster (fruit fly)

- Genome size (~ 120Mb)
- Genome distributed over 4 chromosomes



Drosophila melanogaster (fruit fly)

- Research resources are available.

THE DROSOPHILA GENOME

REVIEW

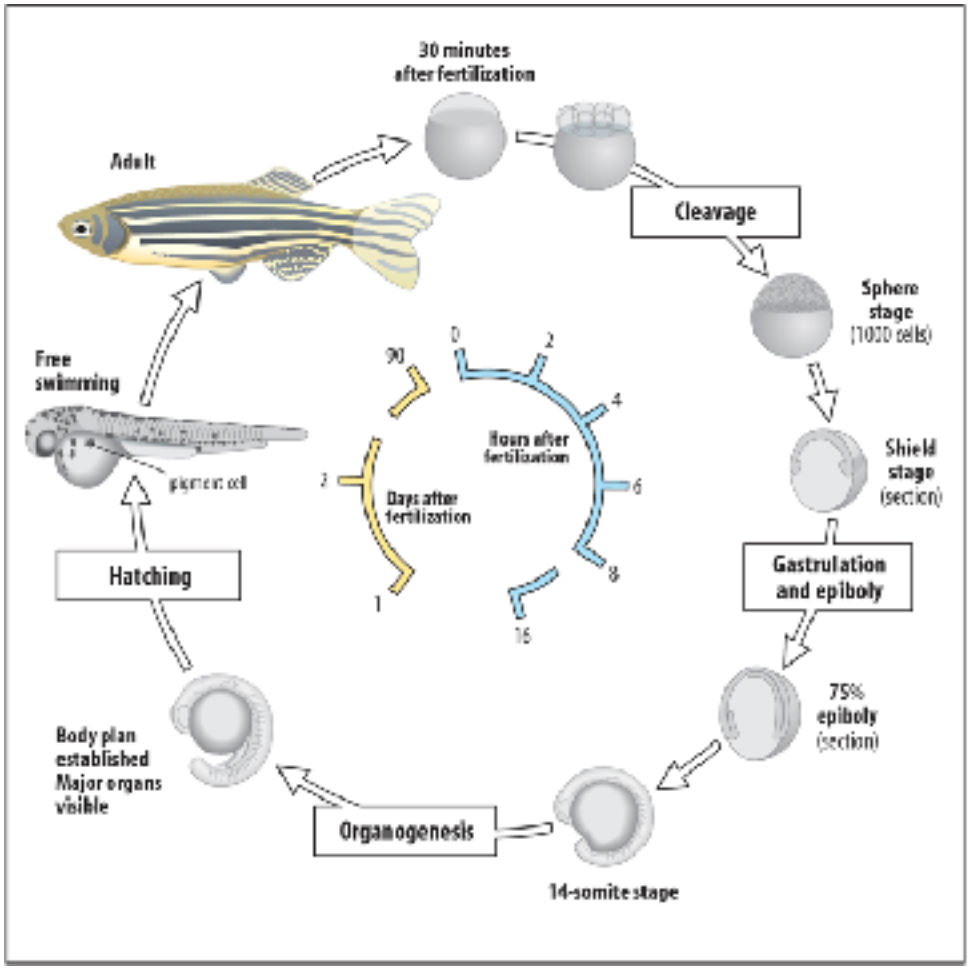
The Genome Sequence of *Drosophila melanogaster*

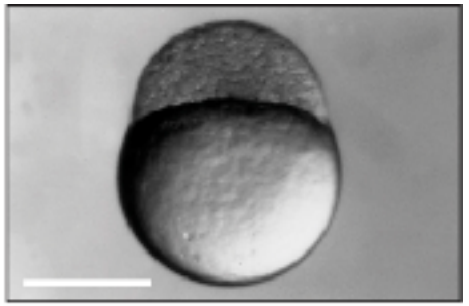
Mark D. Adams,[1*] Susan E. Celniker,[2] Robert A. Holt,[1] Cheryl A. Evans,[1] Jeannine D. Gocayne,[1] Peter G. Amanatides,[1] Steven E. Scherer,[1] Peter W. Li,[1] Roger A. Hoskins,[2] Richard F. Galle,[1] Reed A. George,[2] Suzanna E. Lewis,[2] Stephen Richards,[2] Michael Ashburner,[3] Scott N. Henderson,[1] Granger G. Sutton,[1] Jennifer R. Wortman,[1] Mark D. Yandell,[1] Qing Zhang,[1] Lin X. Chen,[1] Rhonda C. Brandon,[1] Yu-Hui C. Rogers,[1] Robert C. Blazej,[1] Mark Champe,[1] Barret D. Pfeiffer,[2] Kenneth H. Wan,[1] Clare Doyle,[1] Evan G. Baxter,[1] Gregg Helt,[4] Catherine R. Nelson,[4] George L. Gabor Miklos,[5] Josep F. Abril,[6] Anna Agbayani,[2] Hui-Jin An,[1] Cynthia Andrews-Pfannkoch,[1] Danita Baldwin,[1] Richard M. Ballew,[1] Anand Basu,[1] James Baxendale,[1] Leyla Bayraktaroglu,[1] Ellen M. Beasley,[1] Karen Y. Beeson,[1] P. V. Benos,[10] Benjamin P. Berman,[2] Deepali Bhandari,[1] Slava Bolshakov,[11] Dana Borkova,[12] Michael R. Botchan,[10] John Bouck,[3] Peter Brokstein,[1] Phillipe Brottier,[11] Kenneth C. Burtis,[13] Dana A. Busam,[1] Heather Butler,[11] Edouard Cadieu,[11] Angela Center,[1] Ishwar Chandra,[1] J. Michael Cherry,[14] Simon Cawley,[10] Carl Dahlke,[1] Lionel B. Davenport,[1] Peter Davies,[1] Beatriz de Pablos,[20] Arthur Delcher,[1] Zuoming Deng,[1] Anne Deslattes Mays,[1] Ian Dew,[1] Suzanne M. Dietz,[1] Kristina Dodson,[1] Lisa E. Doup,[1] Michael Downes,[21] Shannon Dugan-Rocha,[3] Boris C. Dunkov,[22] Patrick Dunn,[1] Kenneth J. Durbin,[1] Carlos C. Evangelista,[1] Concepcion Ferraz,[20] Steven Ferriera,[1] Wolfgang Fleischmann,[3] Carl Fosler,[1] Andrei E. Gabrielian,[1] Neha S. Garg,[1] William M. Gelbart,[9] Kan Glasser,[1] Anna Glodek,[1] Fangcheng Gong,[1] J. Harley Gorrell,[1] Zhiping Gu,[1] Ping Guan,[1] Mikhail Harris,[1] Nomi L. Harris,[2] Damon Harvey,[4] Thomas J. Heiman,[1] Judith R. Hernandez,[1] Jarrett Houck,[1] Damon Hostin,[1] Kathryn A. Houston,[1] Timothy J. Howland,[1] Ming-Hui Wei,[1] Chinyere Ibegwam,[1] Mena Jalali,[1] Francis Kalush,[1] Gary H. Karpen,[23] Zhaoxi Ke,[1] James A. Kennison,[24] Karen A. Ketchum,[1] Bruce E. Kimmel,[1] Chinnappa D. Kodira,[1] Cheryl Kraft,[1] Saul Kravitz,[1] David Kulp,[4] Zhongwu Lai,[1] Paul Lasko,[25] Yiding Lei,[1] Alexander A. Levitsky,[3] Jiayin Li,[1] Zhenya Li,[1] Yong Liang,[1] Xiaoying Lin,[26] Xiangjun Liu,[1] Bettina Mattei,[1] Tina C. McIntosh,[1] Michael P. McLeod,[3] Duncan McPherson,[1] Gennady Merkulov,[1] Natalia V. Milshina,[1] Clark Mobarry,[1] Joe Morris,[1] Ali Moshrefi,[2] Stephen M. Mount,[27] Mee Moy,[1] Brian Murphy,[1] Lee Murphy,[28] Donna M. Muzny,[1] David L. Nelson,[3] David R. Nelson,[29] Keith A. Nelson,[1] Katherine Nixon,[2] Deborah R. Nusskern,[1] Joanne M. Pacleb,[2] Michael Palazzolo,[1] Gjange S. Pittman,[1] Sue Pan,[1] John Pollard,[1] Vinita Puri,[1] Martin G. Reese,[4] Knut Reinert,[1] Karin Remington,[1] Robert D. C. Saunders,[30] Frederick Scheeler,[1] Hua Shen,[1] Baoling Christopher Shue,[1] Inga Siden-Kiamos,[31] Michael Simpson,[1] Marian P. Skupski,[1] Tom Smith,[1] Eugene Spier,[1] Allan C. Spradling,[31] Mark Stapleton,[2] Reuben Strong,[1] Eric Sun,[1] Robert Svirskas,[32] Cyndee Tector,[1] Russell Turner,[1] Eli Venter,[1] Alfred H. Wang,[1] Xin Wang,[1] Zhen-Yuan Wang,[1] David A. Wassarman,[33] George M. Weinstock,[3] Jean Weissenbach,[34] Sherita M. Williams,[1] Trevor Woodage,[1] Kim C. Worley,[3] Davis Wu,[1] Song Yang,[2] Q. Alison Yao,[1] Jane Ye,[1] Ru-Fang Yeh,[14] Jaydutse S. Zaveri,[1] Ming Zhan,[1] Guangren Zhang,[1] Qi Zhao,[1] Liansheng Zheng,[1] Xianquan H. Zheng,[1] Fei N. Zhong,[1] Wenyan Zhong,[1] Xiaojun Zhou,[3] Chiaping Zhu,[1] Xiaobong Zhu,[1] Hamilton O. Smith,[3] Richard A. Gibbs,[1] Eugene W. Myers,[1] Gerald M. Rubin,[3,4] J. Craig Venter[1]

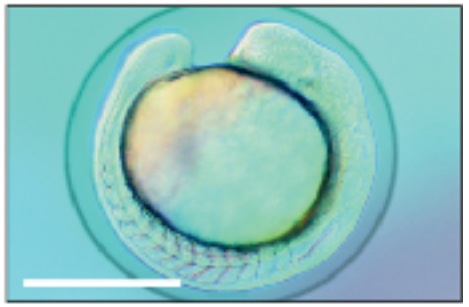
The fly *Drosophila melanogaster* is one of the most intensively studied organisms in biology and serves as a model system for the investigation of many developmental and cellular processes common to higher eukaryotes, including humans. We have determined the nucleotide sequence of nearly all of the ~120-megabase euchromatic portion of the *Drosophila* genome using a whole-genome shotgun sequencing strategy supported by extensive clone-based sequence and a high-quality bacterial artificial chromosome physical map. Efforts are under way to close the remaining gaps; however, the sequence is of sufficient accuracy and contiguity to be declared substantially complete and to support an initial analysis of genome structure and preliminary gene annotation and interpretation. The genome encodes ~13,600 genes, somewhat fewer than the number in the nematode *Caenorhabditis elegans* genome, but with comparable functional diversity.

Danio rerio (zebrafish)

- Eukaryote.
- Diploid.
- Animal - Vertebrate.
- 4-6 cm in length.

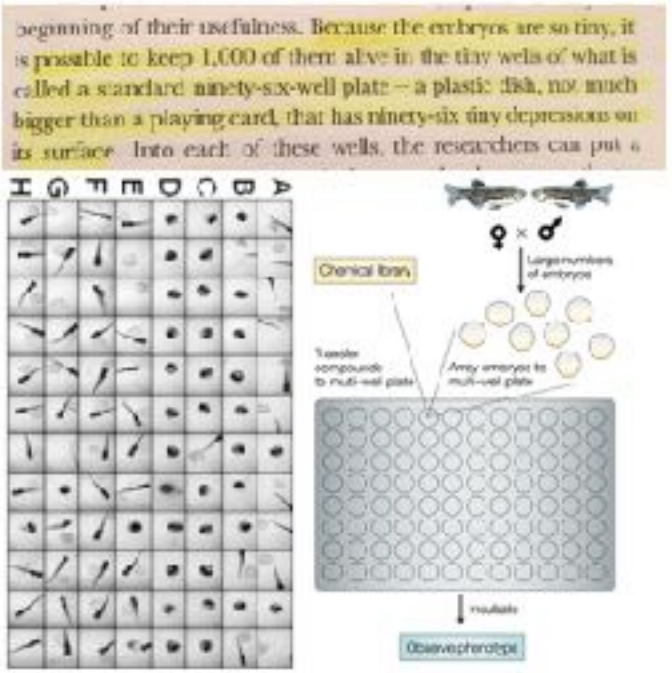






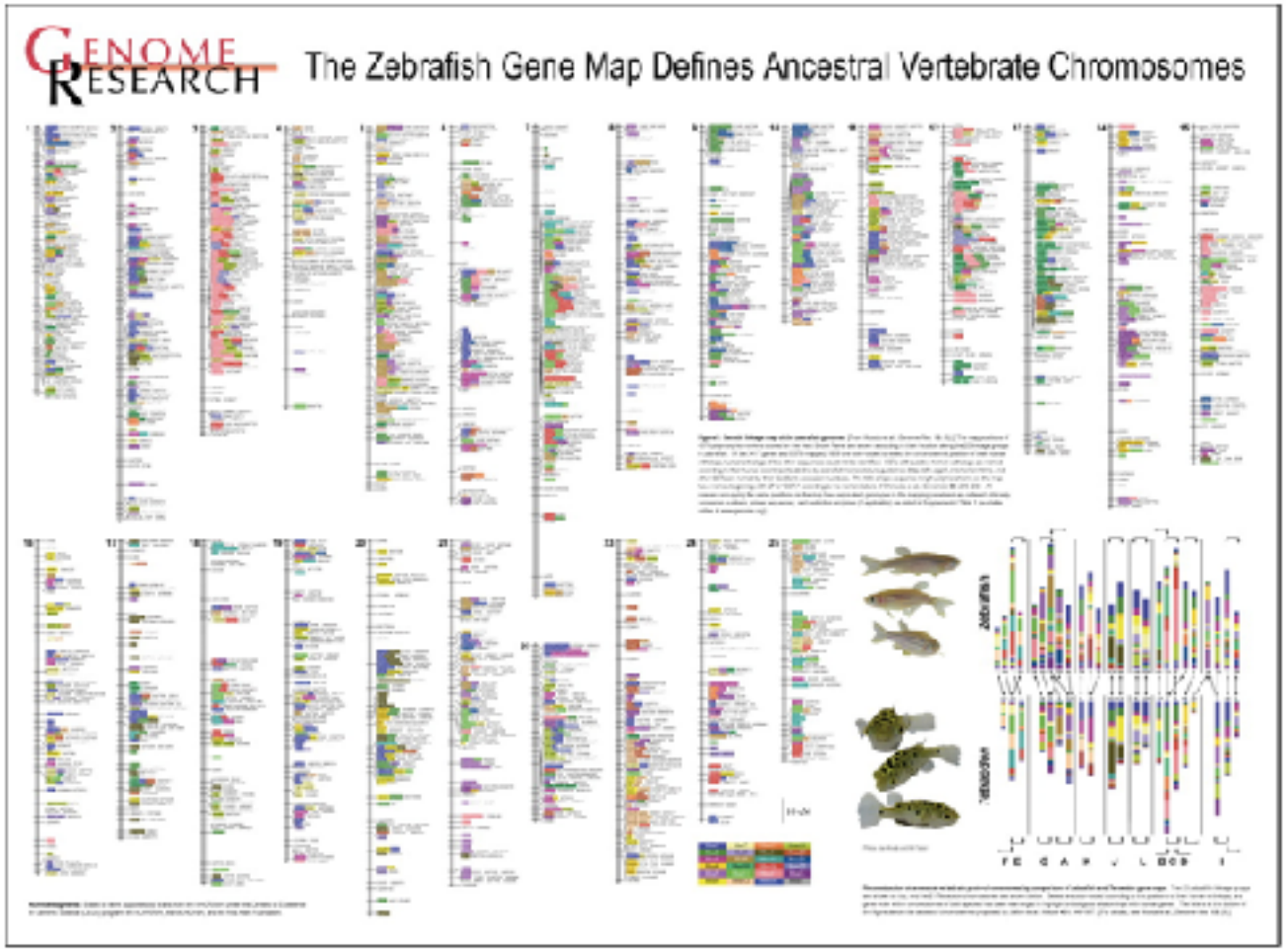
Danio rerio (zebrafish)

- Easy to grow in lab
- Occupies relatively a small space
- Short life cycle
- Good development model.



Danio rerio (zebrafish)

- Genome size (~ 1.5 Gb)
- Genome distributed over 25 chromosomes



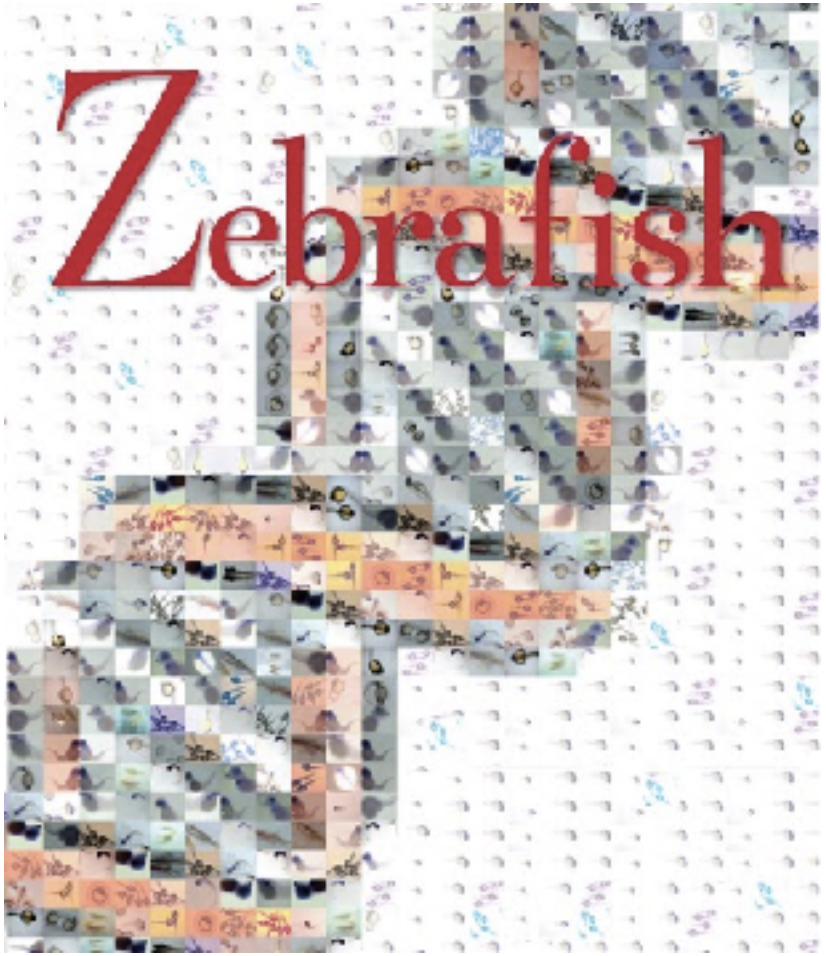
Danio rerio (zebrafish)

- Research resources are available.

LETTER

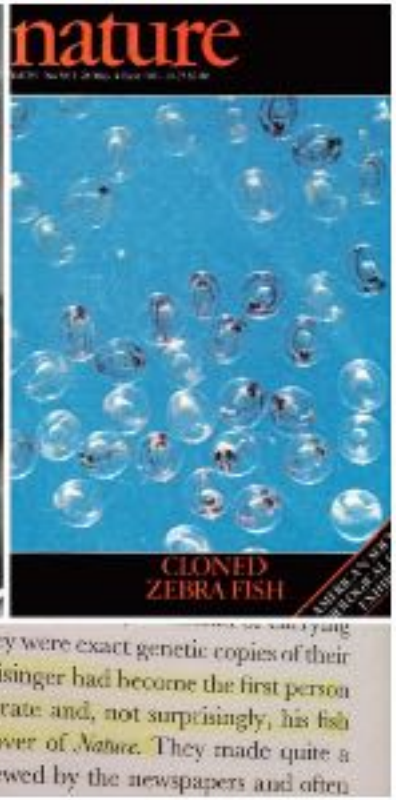
OPEN

The zebrafish reference genome sequence and its relationship to the human genome



Danio rerio (zebrafish)

- First cloned vertebrate!
- Fun scientific community.

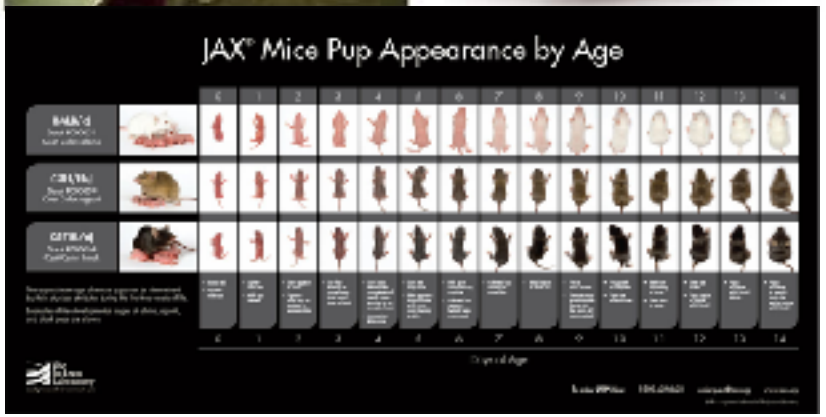


mothers – clones. In 1981 Streisinger had become the first person in the world to clone a vertebrate and, not surprisingly, his fish eggs appeared on the front cover of *Nature*. They made quite a splash: Streisinger was interviewed by the newspapers and often

Mus musculus (mouse)

- Eukaryote.
- Animal - mammal.
- Diploid
- Model for human.
- Small mammal.
- 7.5 - 10 cm in length.
- Long history as a model in biology and medicine.

Mus musculus (mouse)



- Small mammal.
- Can be grown in lab.
- Genome can be manipulated.
- Knockout mice.
- A variety of phenotypes can be studied.

Mus musculus (mouse)

- Genome size (~ 2.8 Gb)
- Genome distributed over 20 chromosomes

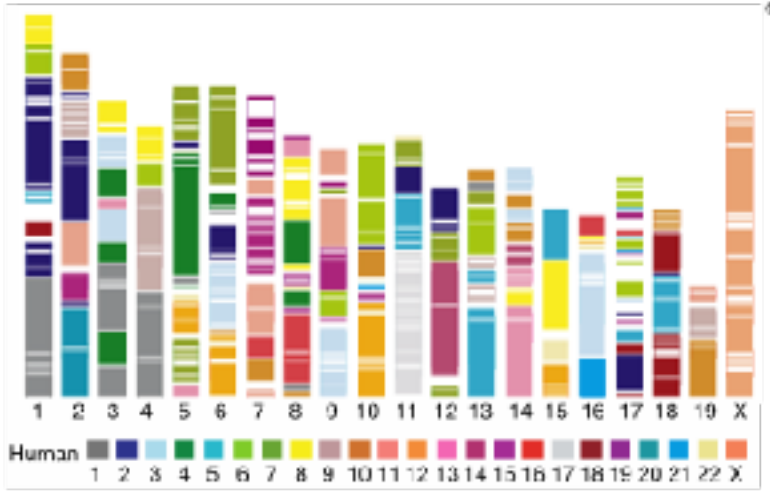
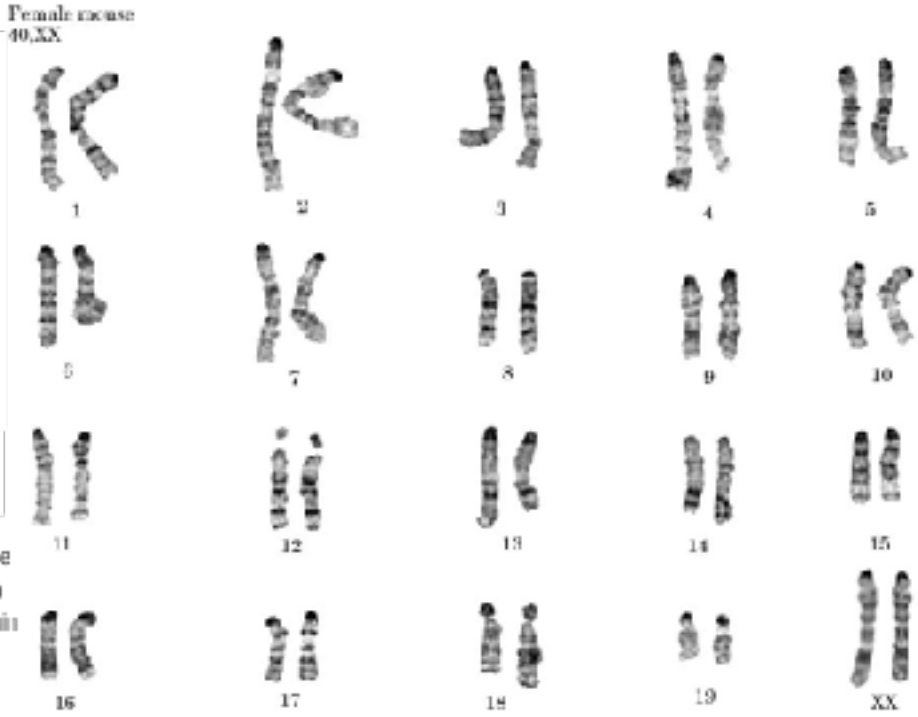


Figure 3 Segments and blocks >300kb in size with conserved synteny in human are superimposed on the mouse genome. Each colour corresponds to a particular human chromosome. The 342 segments are separated from each other by thin white lines within the 217 blocks of consistent colour.



Mus musculus (mouse)

- Research resources are available.

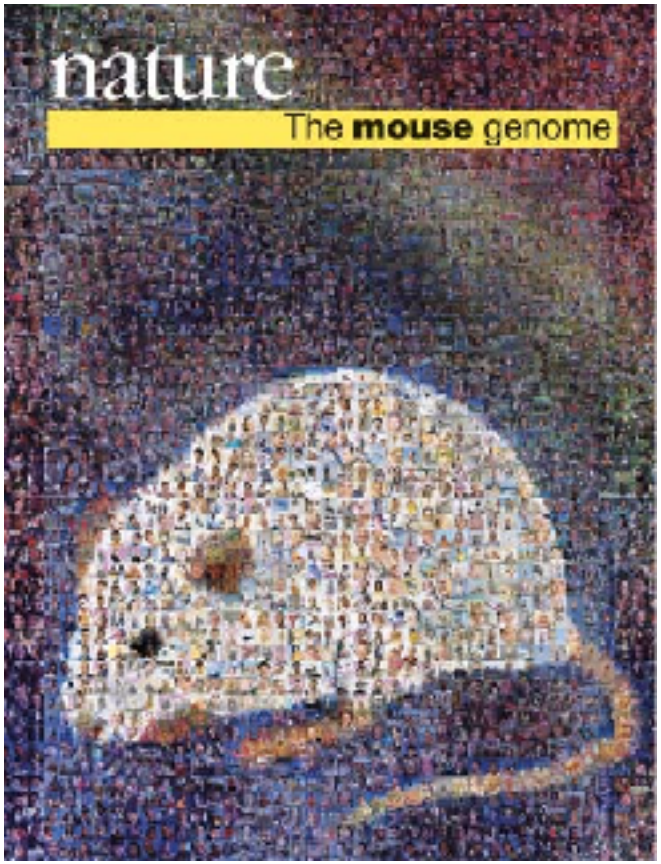
articles

Initial sequencing and comparative analysis of the mouse genome

Mouse Genome Sequencing Consortium*

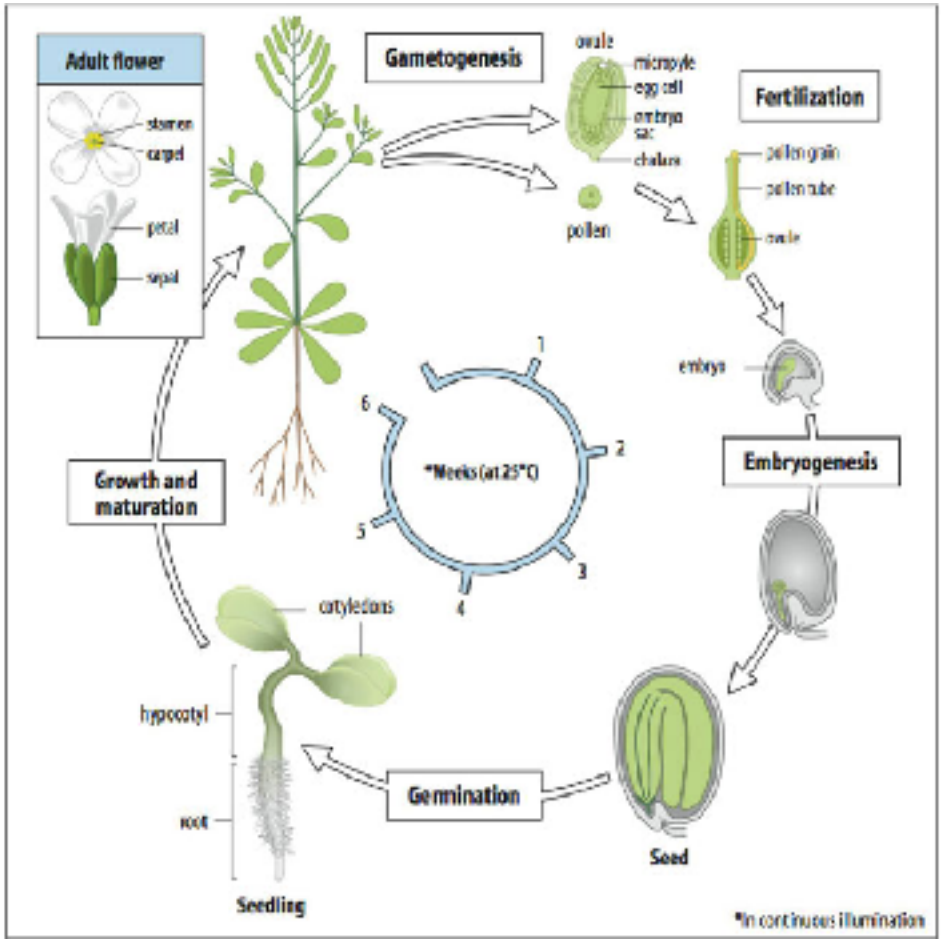
*A list of authors and their affiliations appears at the end of the paper.

The sequence of the mouse genome is a key informational tool for understanding the contents of the human genome and a key experimental tool for biomedical research. Here, we report the results of an international collaboration to produce a high-quality draft sequence of the mouse genome. We also present an initial comparative analysis of the mouse and human genomes, describing some of the insights that can be gleaned from the two sequences. We discuss topics including the analysis of the evolutionary forces shaping the size, structure and sequence of the genomes; the conservation of large-scale synteny across most of the genomes; the much lower extent of sequence orthology covering less than half of the genomes; the proportions of the genomes under selection; the number of protein-coding genes; the expansion of gene families related to reproduction and immunity; the evolution of proteins; and the identification of intraspecies polymorphism.



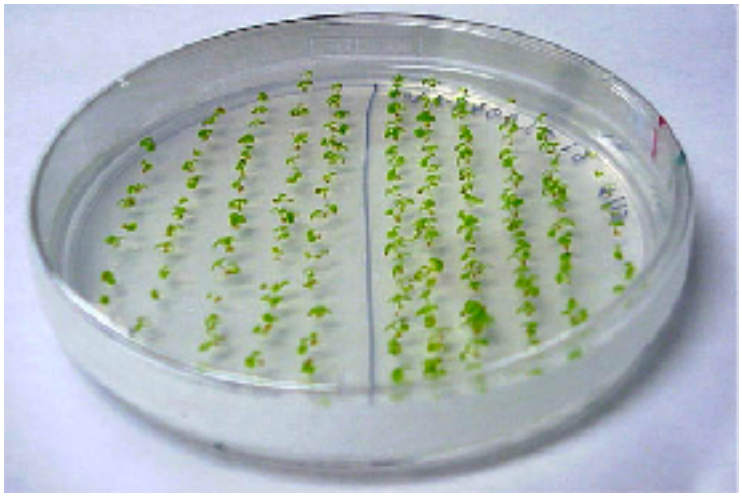
Arabidopsis thaliana (plant)

- Eukaryote.
- Plant - Dicot.
- Diploid.
- 20-25 cm in height



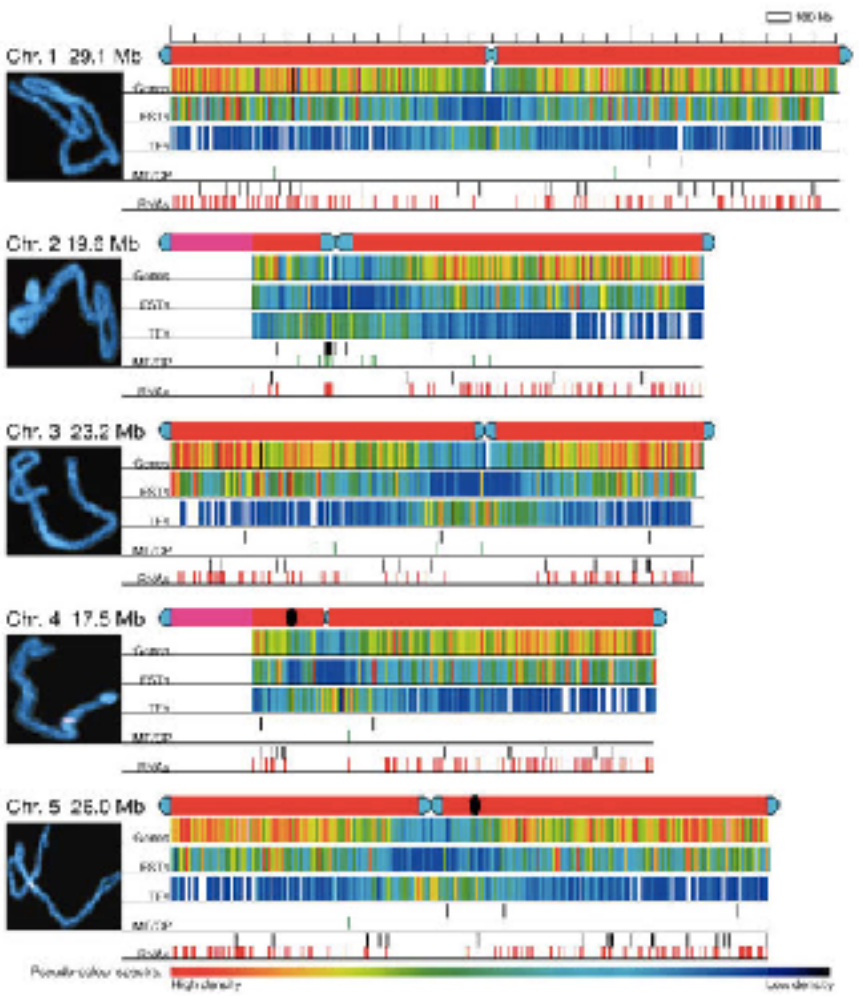
Arabidopsis thaliana (plant)

- Easy to grow in lab
- Occupies a small space
- Short life cycle
- Easy to cross
- Seeds can be stored.



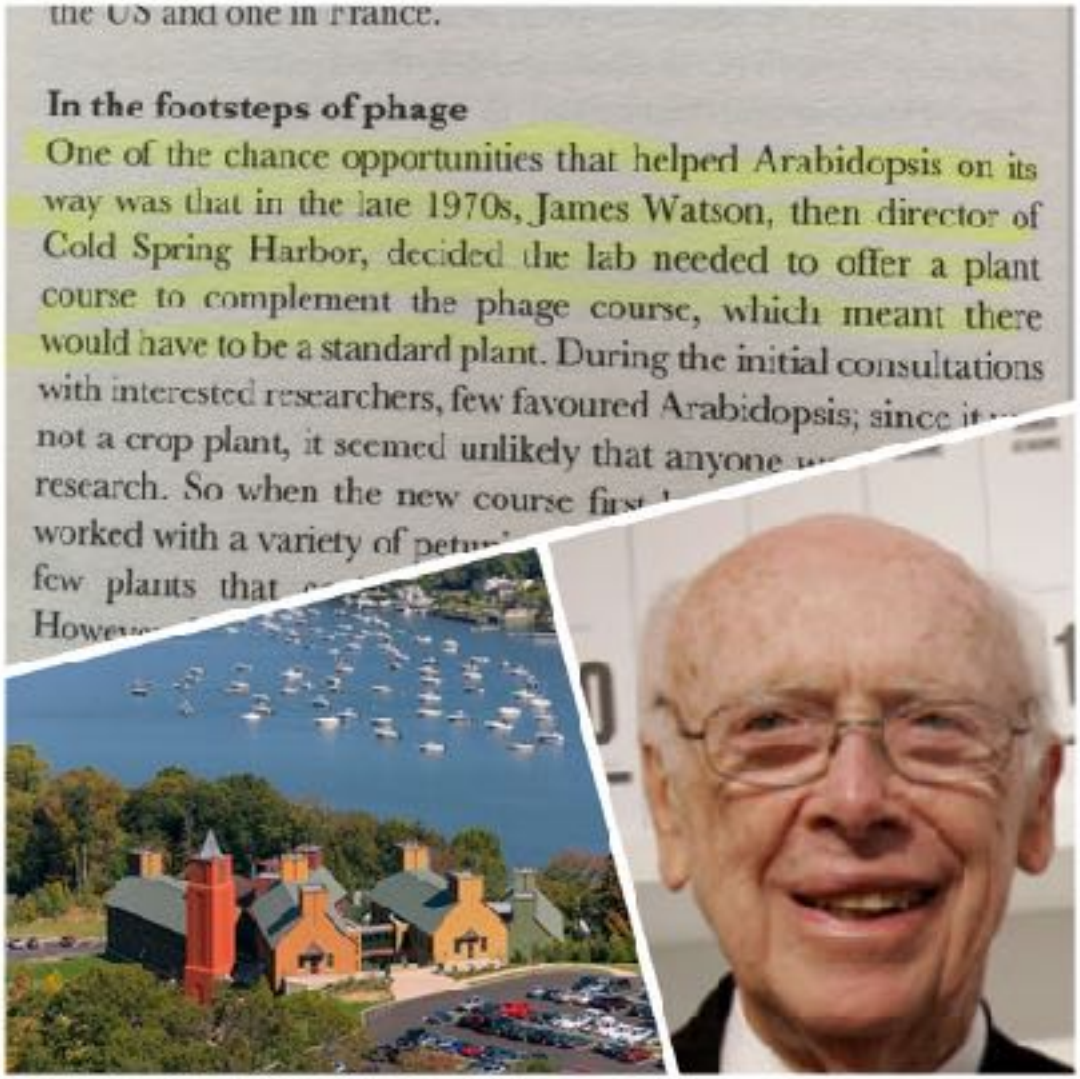
Arabidopsis thaliana (plant)

- Genome size (~ 135 Mb)
- Genome distributed over 5 chromosomes



Arabidopsis thaliana (plant)

- The model/standard for plants



the US and one in France.

In the footsteps of phage

One of the chance opportunities that helped Arabidopsis on its way was that in the late 1970s, James Watson, then director of Cold Spring Harbor, decided the lab needed to offer a plant course to complement the phage course, which meant there would have to be a standard plant. During the initial consultations with interested researchers, few favoured Arabidopsis; since it
not a crop plant, it seemed unlikely that anyone
research. So when the new course fir
worked with a variety of petu
few plants that
Howe

Arabidopsis thaliana (plant)

- Research resources are available.

articles

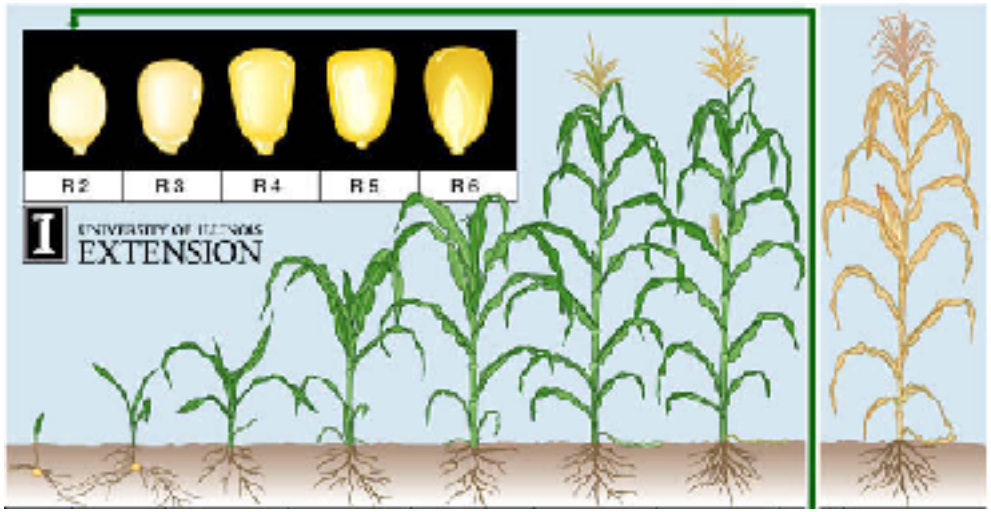
Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*

The Arabidopsis Genome Initiative*

The flowering plant *Arabidopsis thaliana* is an important model system for identifying genes and determining their functions. Here we report the analysis of the genomic sequence of *Arabidopsis*. The sequenced regions cover 115.4 megabases of the 125-megabase genome and extend into centromeric regions. The evolution of *Arabidopsis* involved a whole-genome duplication, followed by subsequent gene loss and extensive local gene duplications, giving rise to a dynamic genome enriched by lateral gene transfer from a cyanobacterial-like ancestor of the plastid. The genome contains 25,498 genes encoding proteins from 11,000 families, similar to the functional diversity of *Drosophila* and *Caenorhabditis elegans*— the other sequenced multicellular eukaryotes. *Arabidopsis* has many families of new proteins but also lacks several common protein families, indicating that the sets of common proteins have undergone differential expansion and contraction in the three multicellular eukaryotes. This is the first complete genome sequence of a plant and provides the foundations for more comprehensive comparison of conserved processes in all eukaryotes, identifying a wide range of plant-specific gene functions and establishing rapid systematic ways to identify genes for crop improvement.

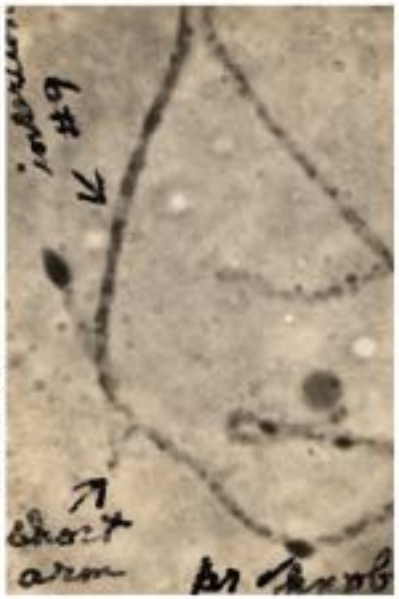
Zea mays (corn)

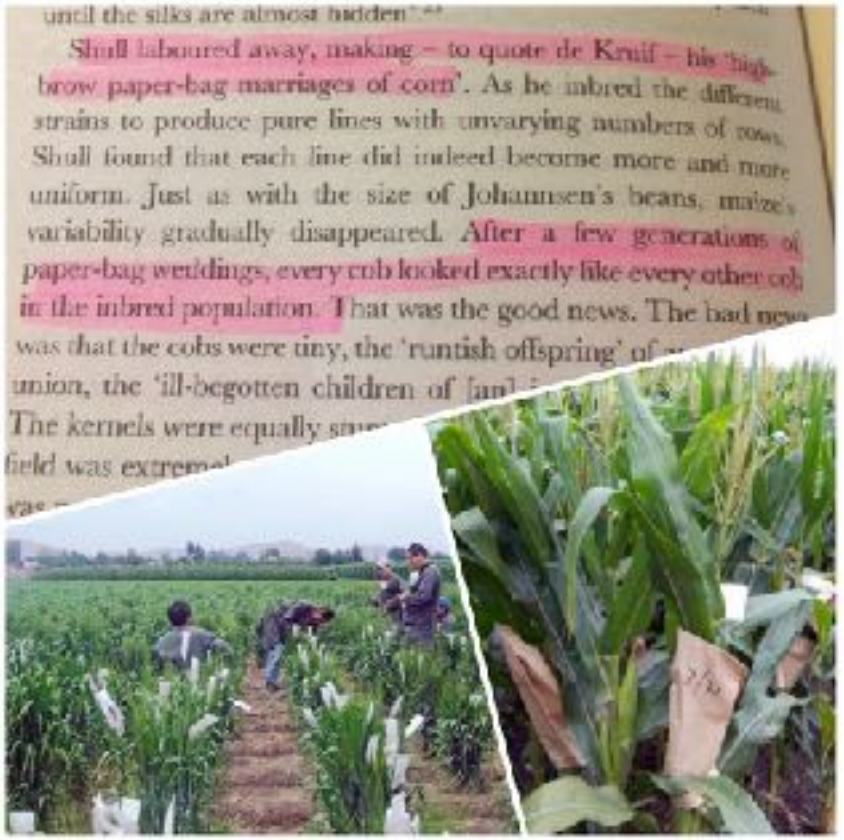
- Eukaryote.
- Plant - monocot.
- Diploid.
- Agricultural importance.
- ~ 2.5 m in height.



Zea mays (corn)

- Large plant.
- Can't be held in lab.
- Crosses must be conducted in the field.
- Long breeding history





- Model for cytogenetics.

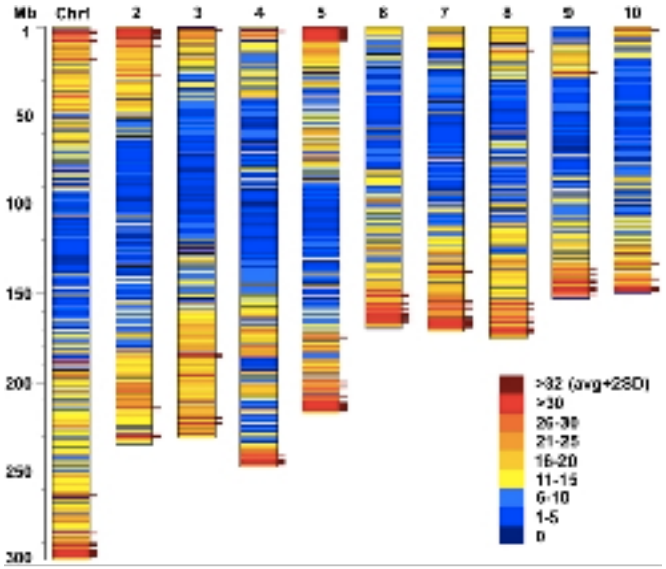
Zea mays (corn)

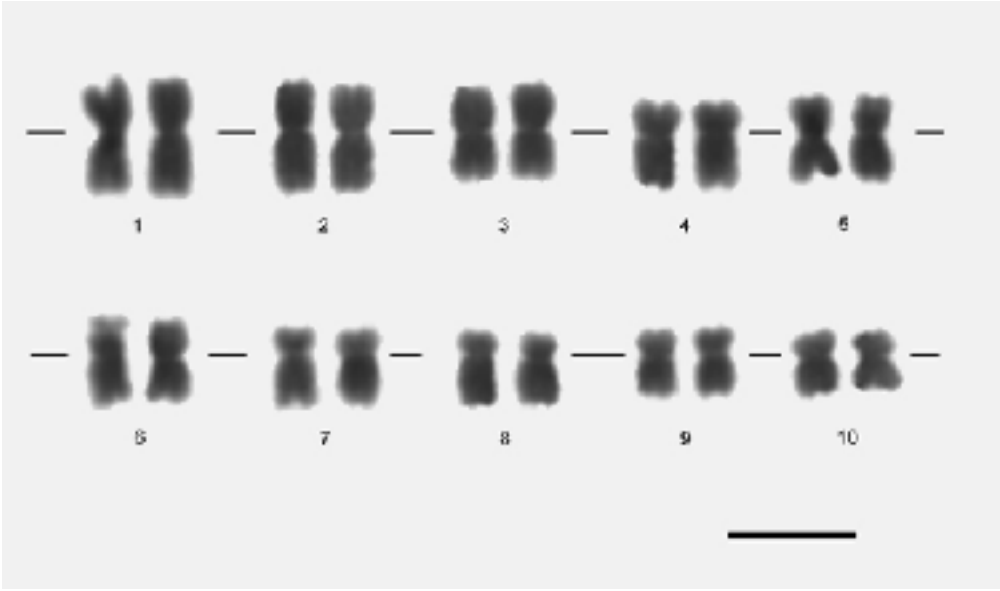
Barbara McClintock and jumping genes

Next to McClintock's maize fields were the labs where Delbrück, Luria and their colleagues were running their phage course every summer. She got to know them pretty well, and

Zea mays (corn)

- Genome size (~ 2.3 Gb)
- Genome distributed over 10 chromosomes





Zea mays (corn)

- Research resources are available.

The B73 Maize Genome: Complexity, Diversity, and Dynamics

Patrick S. Schnable,[1,2,3,4]* Doreen Ware,[5,6]* Robert S. Fulton,[7]† Joshua C. Stein,[5]† Fusheng Wei,[8]† Shiran Pasternak,[5] Chengzhi Liang,[5] Jianwei Zhang,[5] Lucinda Fulton,[7] Tina A. Graves,[7] Patrick Minx,[7] Amy Denise Reily,[7] Laura Courtney,[7] Scott S. Kruchowski,[7] Chad Tomlinson,[7] Cindy Strong,[7] Kim Delehaunty,[7] Catrina Fronick,[7] Bill Courtney,[7] Susan M. Rock,[7] Eddie Belter,[7] Feiyu Du,[7] Kyung Kim,[7] Rachel M. Abbott,[7] Marc Cotton,[7] Andy Levy,[7] Pamela Marchetto,[7] Kerri Ochoa,[7] Stephanie M. Jackson,[7] Barbara Gillam,[7] Weizu Chen,[7] Le Yan,[7] Jamey Higginbotham,[7] Marco Cardenas,[7] Jason Waligorski,[7] Elizabeth Applebaum,[7] Lindsey Phelps,[7] Jason Falcone,[7] Krishna Kanchi,[7] Thynn Thane,[7] Adam Scimone,[7] Nay Thane,[7] Jessica Henke,[7] Tom Wang,[7] Jessica Ruppert,[7] Neha Shah,[7] Kelsi Rotter,[7] Jennifer Hodges,[7] Elizabeth Ingenthron,[7] Matt Cordes,[7] Sara Kohlberg,[7] Jennifer Sgro,[7] Brandon Delgado,[7] Kelly Mead,[7] Asif Chinwalla,[7] Shawn Leonard,[7] Kevin Crouse,[7] Kristi Collura,[8] Dave Kudrna,[8] Jennifer Currie,[8] Ruifeng He,[8] Angelina Angelova,[8] Shanmugam Rajasekar,[8] Teri Mueller,[8] Rene Lomeli,[8] Gabriel Scara,[8] Ara Ko,[8] Krista Delaney,[8] Marina Wissotski,[8] Georgina Lopez,[8] David Campos,[8] Michele Braidotti,[1] Elizabeth Ashley,[9] Wolfgang Golser,[9] HyeRan Kim,[9] Seunghee Lee,[9] Jinke Lin,[9] Zeljko Dujmic,[9] Woojin Kim,[9] Jayson Talag,[9] Andrea Zuccolo,[9] Chuanzhu Fan,[9] Aswathy Sebastian,[9] Melissa Kramer,[4] Lori Spiegel,[4] Lidia Nascimento,[4] Theresa Zutavern,[4] Beth Miller,[4] Claude Ambroise,[4] Stephanie Muller,[4] Will Spooner,[4] Apurva Narechania,[4] Liya Ren,[4] Sharon Wei,[4] Sunita Kumari,[4] Ben Faga,[4] Michael J. Levy,[4] Linda McMahan,[4] Peter Van Buren,[4] Matthew W. Vaughn,[4] Kai Ying,[3] Cheng-Ting Yeh,[1,2] Scott J. Emrich,[9,10] Yi Jia,[3] Ananth Kalyanaraman,[9,11] An-Ping Hsia,[1,2] W. Brad Barbazuk,[12] Regina S. Baucom,[13] Thomas P. Brutnell,[14] Nicholas C. Carpita,[15] Cristian Chaparro,[16] Jer-Ming Chia,[4] Jean-Marc Deragon,[16] James C. Estill,[15,17] Yan Fu,[2,4] Jeffrey A. Jeddeloh,[18] Yujun Han,[15,17] Hyeran Lee,[19] Pinghua Li,[14] Damon R. Lisch,[20] Sanzhen Liu,[3] Zhijie Liu,[4] Dawn Holligan Nagel,[3,17] Maureen C. McCann,[21] Phillip SanMiguel,[22] Alan M. Myers,[23] Dan Nettleton,[24] John Nguyen,[25] Bryan W. Penning,[15,21] Lalit Ponnala,[16] Kevin L. Schneider,[27] David C. Schwartz,[28] Anupma Sharma,[27] Carol Soderlund,[29] Nathan M. Springer,[30] Qi Sun,[26] Hao Wang,[15,17] Michael Waterman,[31] Richard Westerman,[32] Thomas K. Wolfgruber,[27] Lixing Yang,[33] Yeisoo Yu,[29] Lifang Zhang,[4] Shiguo Zhou,[28] Qihui Zhu,[15,17] Jeffrey L. Bennetzen,[15] R. Kelly Dawe,[15,17] Jiming Jiang,[19] Ning Jiang,[15] Gernot G. Presting,[27] Susan R. Wessler,[15,17] Srinivas Aluru,[3,4,26] Robert A. Martienssen,[+] Sandra W. Clifton,[7] W. Richard McCombie,[6] Rod A. Wing,[8] Richard K. Wilson[7,15]‡

We report an improved draft nucleotide sequence of the 2.3-gigabase genome of maize, an important crop plant and model for biological research. Over 32,000 genes were predicted, of which 99.8% were placed on reference chromosomes. Nearly 85% of the genome is composed of hundreds of families of transposable elements, dispersed nonuniformly across the genome. These were responsible for the capture and amplification of numerous gene fragments and affect the composition, sizes, and positions of centromeres. We also report on the correlation of methylation-poor regions with *Mu* transposon insertions and recombination, and copy number variants with insertions and/or deletions, as well as how uneven gene losses between duplicated regions were involved in returning an ancient allotetraploid to a genetically diploid state. These analyses inform and set the stage for further investigations to improve our understanding of the domestication and agricultural improvements of maize.

Homo sapiens

Humans: A model organism?

Model organisms. Why?

- Genome can be manipulated experimentally.
- Short life-cycle.
- Minimal living requirements.
- Small genome (some of them)!
- Easy to grow in lab.
- Small in size.
- Accumulated knowledge about the organism.
- Organism does NOT need to be BEAUTIFUL!!

Molecular Biology

The biology of molecules

Sub-cellular biology

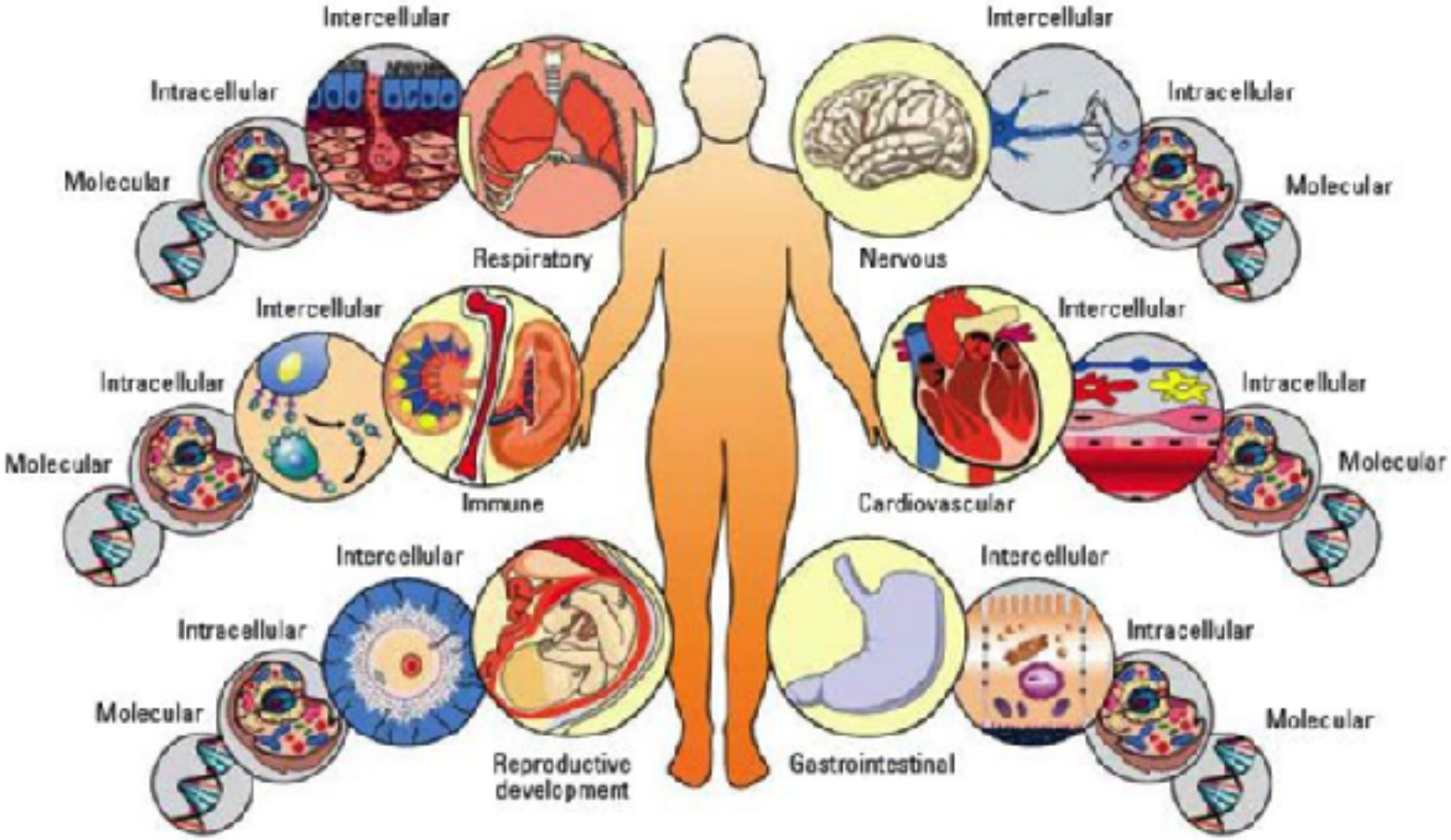
Molecular Biology

ray crystallography, and in 1938 it acquired a new name, 'molecular biology'. The name was coined by Warren Weaver, director of Rockefeller's Natural Sciences Division. He defined the field as the 'biology of molecules' or as 'sub-cellular biology', shifting from the cell itself as the object of study to a more fundamental level of analysis. Weaver made an explicit analogy with the sub-atomic world of the quantum physicists; to make

Warren Weaver

Why molecular biology is fun?

- Physical characters start with a molecule.
- Cognitive and emotional characters also start with a molecule (I think ☺).



To study

sexual life cycle

diploid

Eukaryotes

Arabidopsis thaliana
(plant)

Bacteriophage
(virus)

short life cycle

Zea mays
(corn)

Hermaphrodite

Danio rerio
(zebrafish)

prokaryotes

Saccharomyces cerevisiae
(budding yeast)

Mus musculus
(mouse)

Escherichia coli
(intestines' bacterium)

haploid

Homo sapiens
(Human)?

Asexual life cycle

Caenorhabditis elegans
(worm)

Drosophila melanogaster (fruit fly)

Expectations

- You know the importance of models in science.
- You know the most important model organisms.
- You know general characteristics of the model organisms.
- You know the taxonomic representation of each model organism.

For a smile


escherichia
CONJUGATION
16 NEW
Antibiotic
Resistances
for this summer!
special
does your host
really love you?
SEXY FLAGELLA
in only 10 days!
TOP10
Animal Guts
to colonize!
VELIC'I_2009