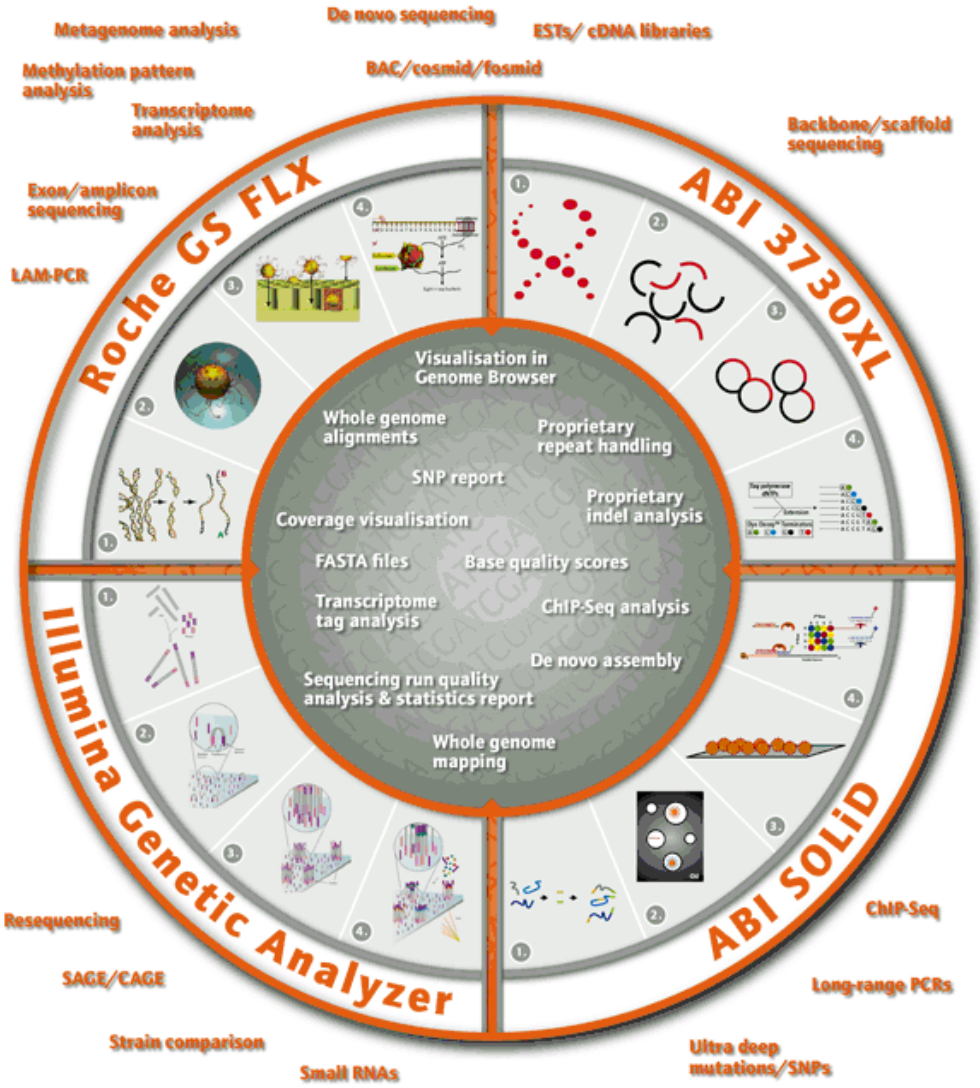
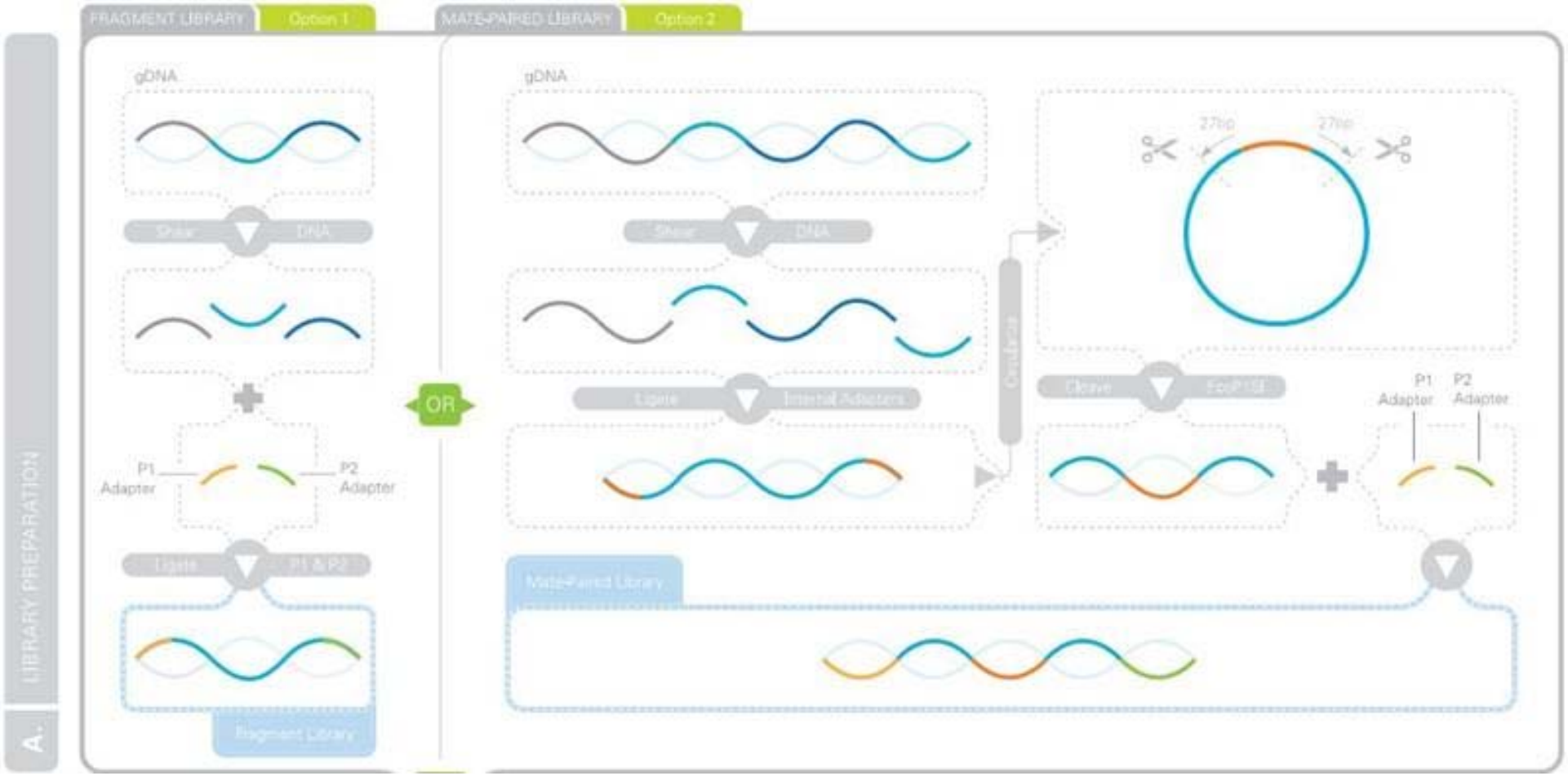
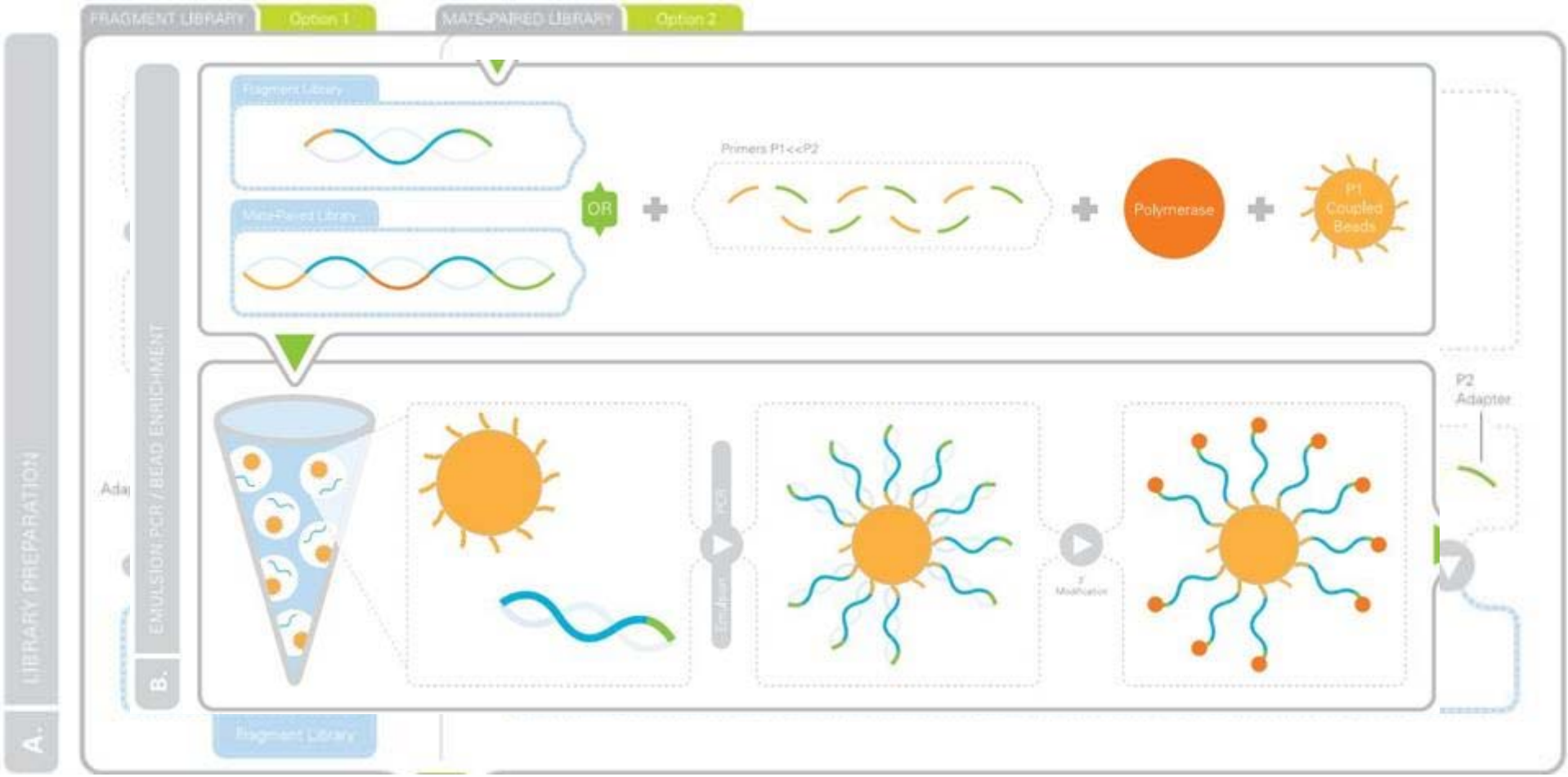
Einführung in die Gentechnologie
Erwin R. Schmidt

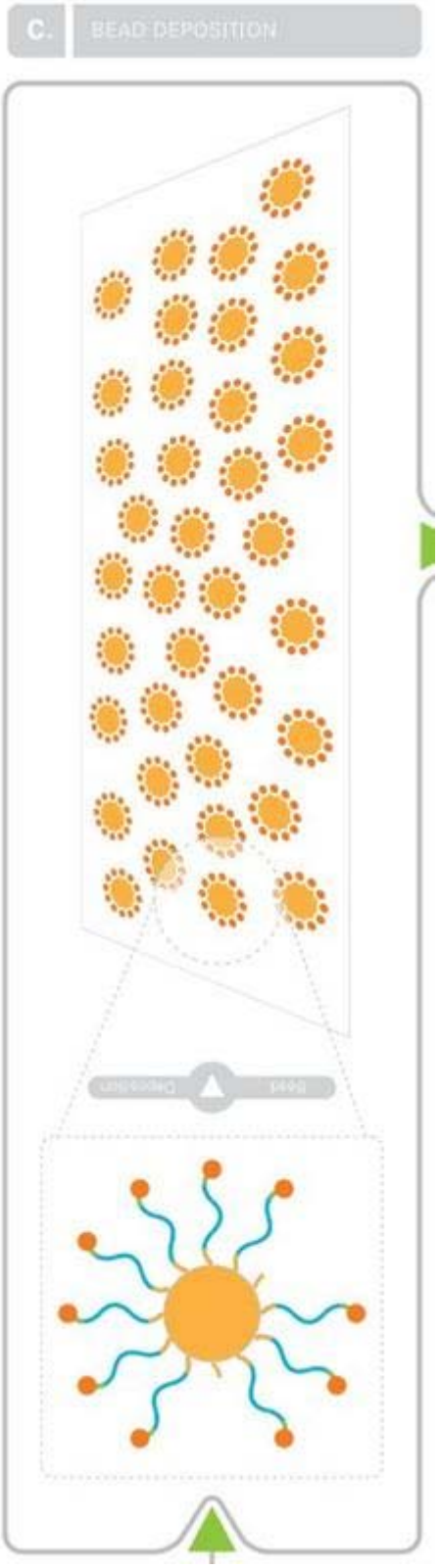
- Vorlesung # 11
- 30. 06. 2009

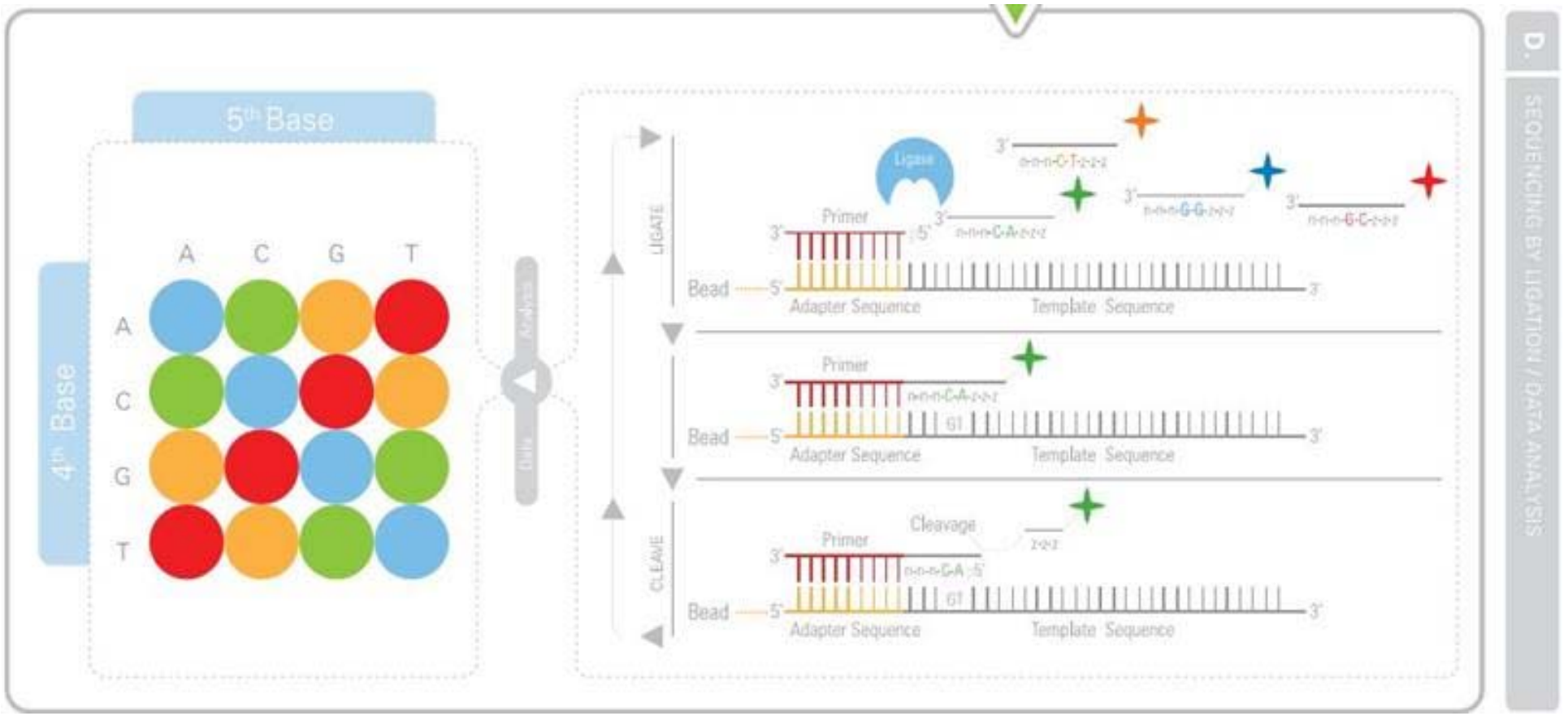
Next generation sequencing



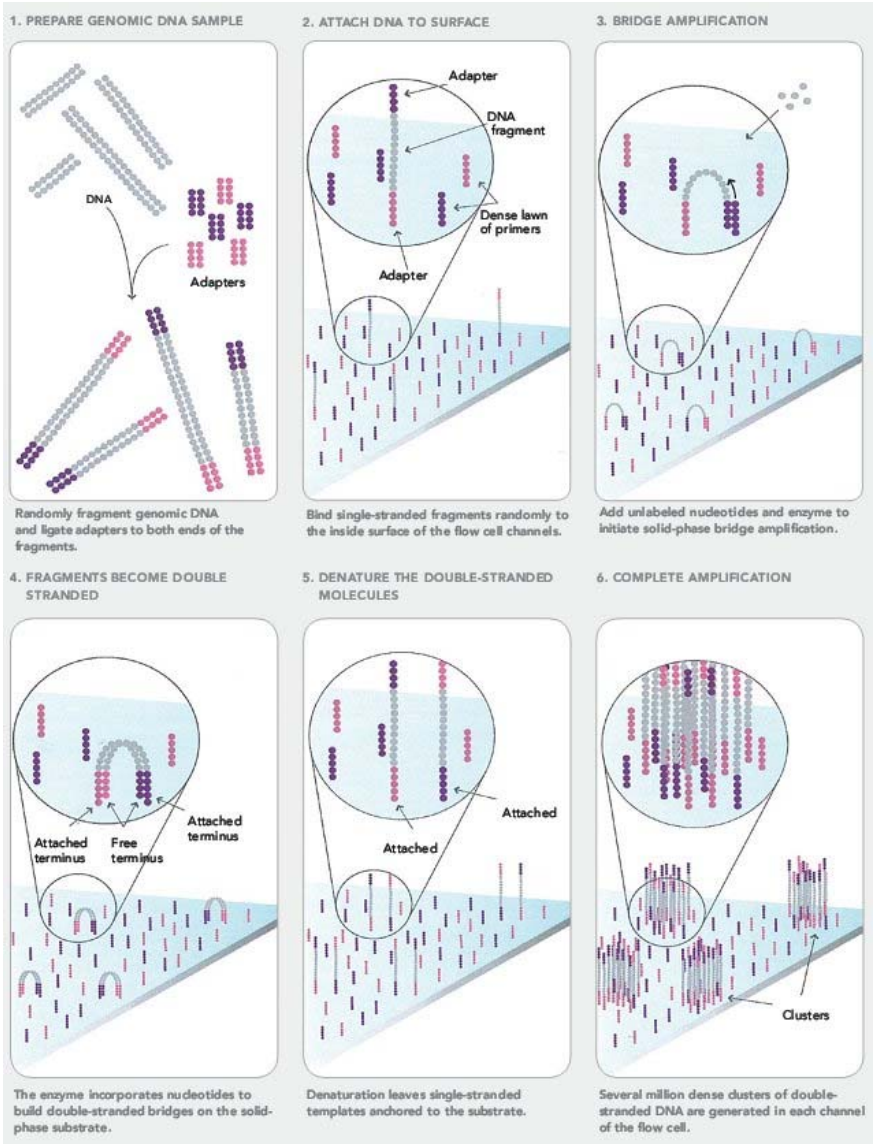

A. LIBRARY PREPARATION
FRAGMENT LIBRARY
Option 1
MATE-PAIRED LIBRARY
Option 2
gDNA
Shear DNA
P1 Adapter
P2 Adapter
Ligate P1 & P2
Fragment Library
OR
gDNA
Shear DNA
Ligate Internal Adapters
Circularize
27bp
27bp
Cleave EcoP15I
P1 Adapter
P2 Adapter
Mate-Paired Library


A.
LIBRARY PREPARATION
B.
EMULSION PCR / BEAD ENRICHMENT
FRAGMENT LIBRARY
Option 1
MATE-PAIRED LIBRARY
Option 2
Fragment Library
Mate-Paired Library
OR
Primers P1<<P2
Polymerase
P1
Coupled
Beads
P2
Adapter
Fragment Library

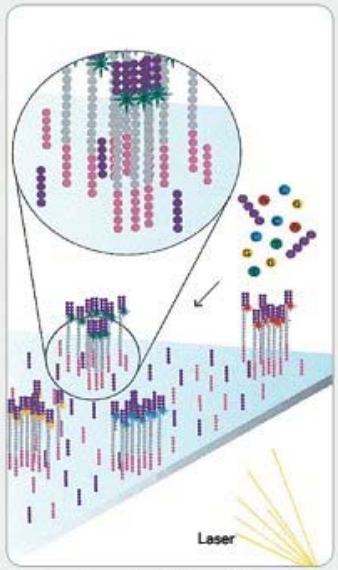

C.
BEAD DEPOSITION


D.
SEQUENCING BY LIGATION / DATA ANALYSIS
5th Base
4th Base
A
C
G
T
Image Analysis
Data Analysis
LIGATE
CLEAVE
Ligase
Primer
Bead
Adapter Sequence
Template Sequence
Cleavage

„paired end sequencing“ (Fa. Illumina

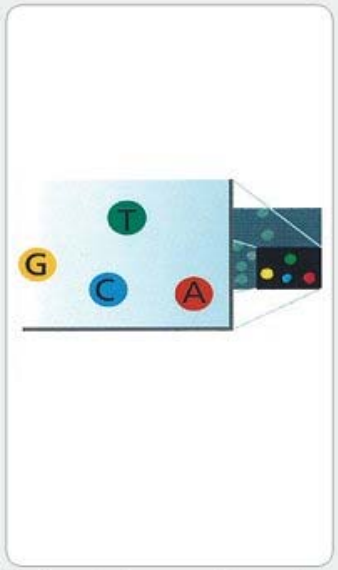


7. DETERMINE FIRST BASE



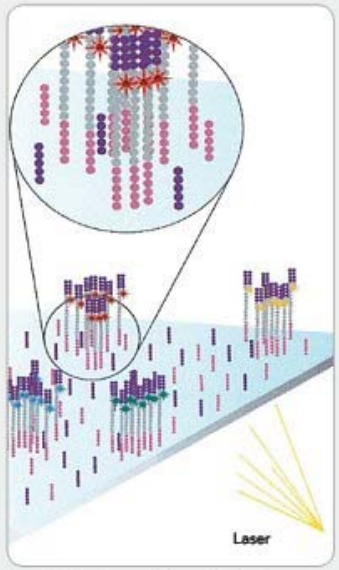
First chemistry cycle: to initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.

8. IMAGE FIRST BASE



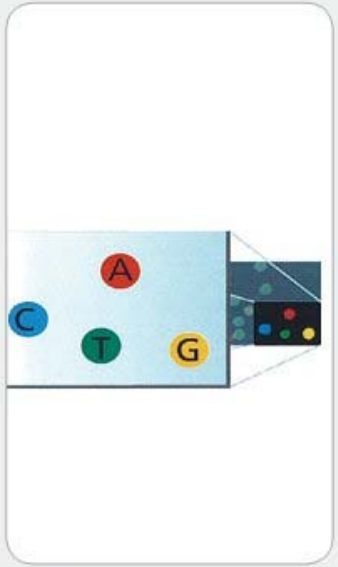
After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

9. DETERMINE SECOND BASE



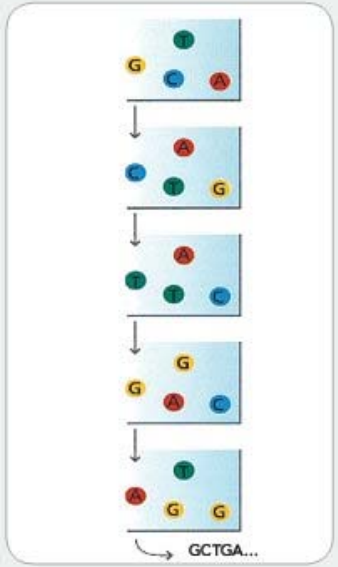
Second chemistry cycle: to initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.

10. IMAGE SECOND CHEMISTRY CYCLE



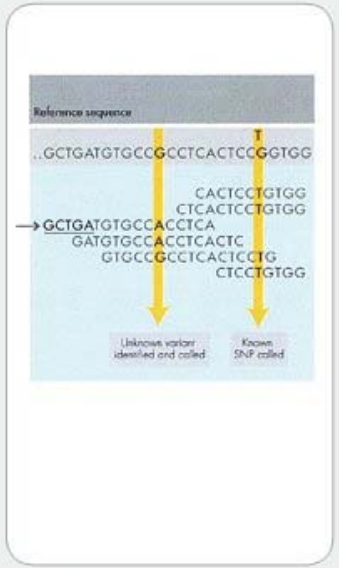
After laser excitation, collect the image data as before. Record the identity of the second base for each cluster.

11. SEQUENCE READS OVER MULTIPLE CHEMISTRY CYCLES



Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at time.

12. ALIGN DATA



Align data, compare to a reference, and identify sequence differences.

Die Ära der Genomik

„Even the smallest functional DNA varieties seen, those occurring in small phages, must have something like 5000 nucleotides in a row. We may, therefore, leave the task of reading the complete nucleotide sequence of a DNA for the next century, which will, however, have other worries.

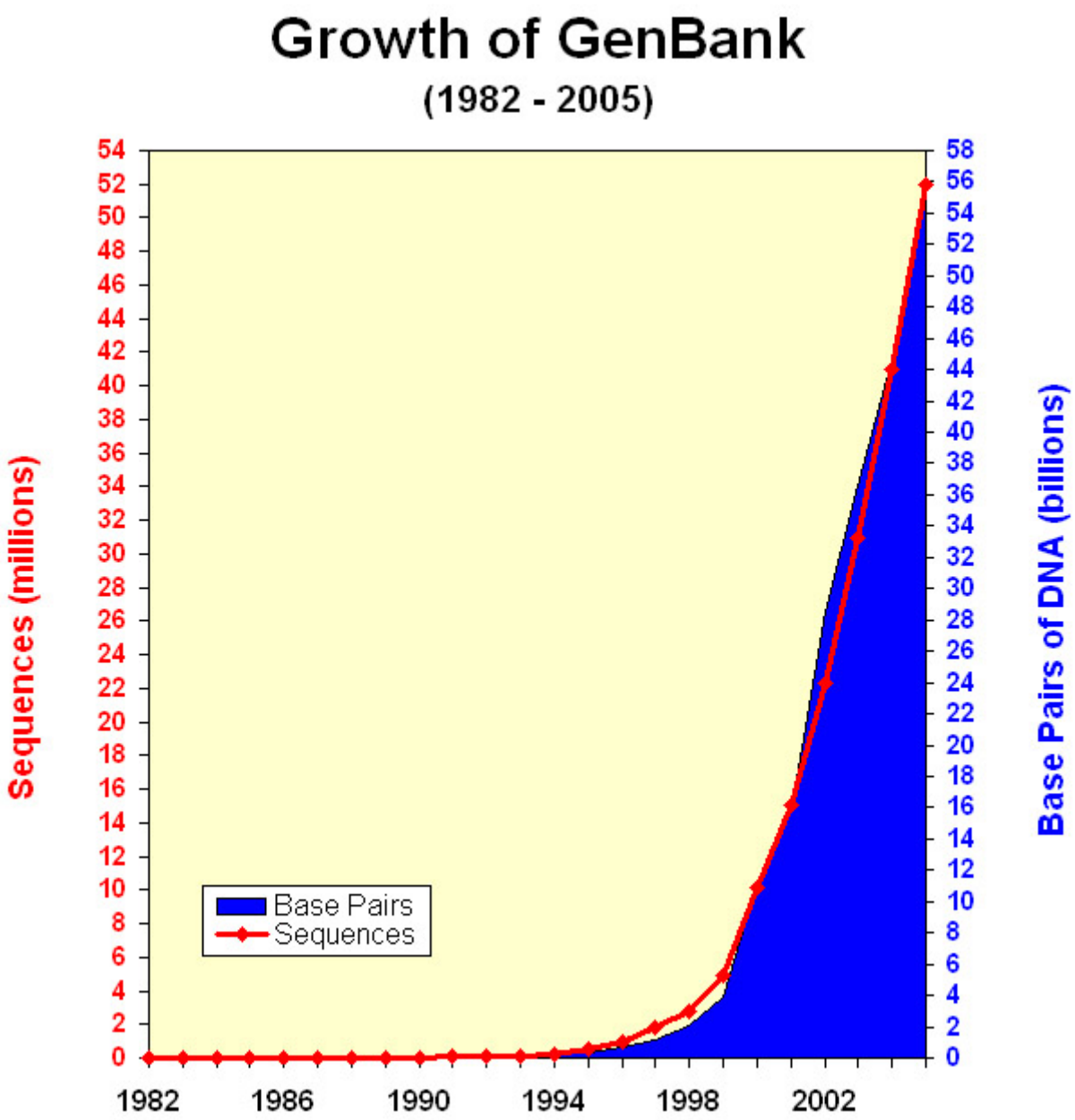
Progress in Nucleic Acid Research and Molecular Biology, 1968

Phi-X 174 sequenced, Nature 1977

Die Geschwindigkeit von Genomanalysen ist exponentiell gewachsen

Die Ära der Genomforschung:

Phi X 174	1977	5.386 bp
λ– Phage	1982	48.502 bp
M. genitalium	1995	580.000 bp
H. influenzae	1995	1.830.000 bp
M. jannaschii	1996	1.660.000 bp
S. cerevisae	1997	12.500.000 bp
E. coli	1997	4.654.000 bp
C. elegans	1998	97.000.000 bp
D. melanog.	1999	116.000.000 bp
A. thaliana	2000	115.000.000 bp
H. sapiens	2001	2.693.000.000 bp
Oryza sativa	2002	420.000.000 bp


Growth of GenBank
(1982 - 2005)
Sequences (millions)
Base Pairs of DNA (billions)
Base Pairs
Sequences
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
42
44
46
48
50
52
54
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
42
44
46
48
50
52
54
56
58
1982
1986
1990
1994
1998
2002

Stand der Genomik 25. 06. 09

() Stand vor einem Jahr

- 1035 (815) komplette Genome sequenziert und publiziert

  861 (672) Eubakterien

  64 (27) Archebakterien

  110 (94) Eukaryoten

- 4932 (2874) aktuelle Genomprojekte

  2605 (1848) Bakterien

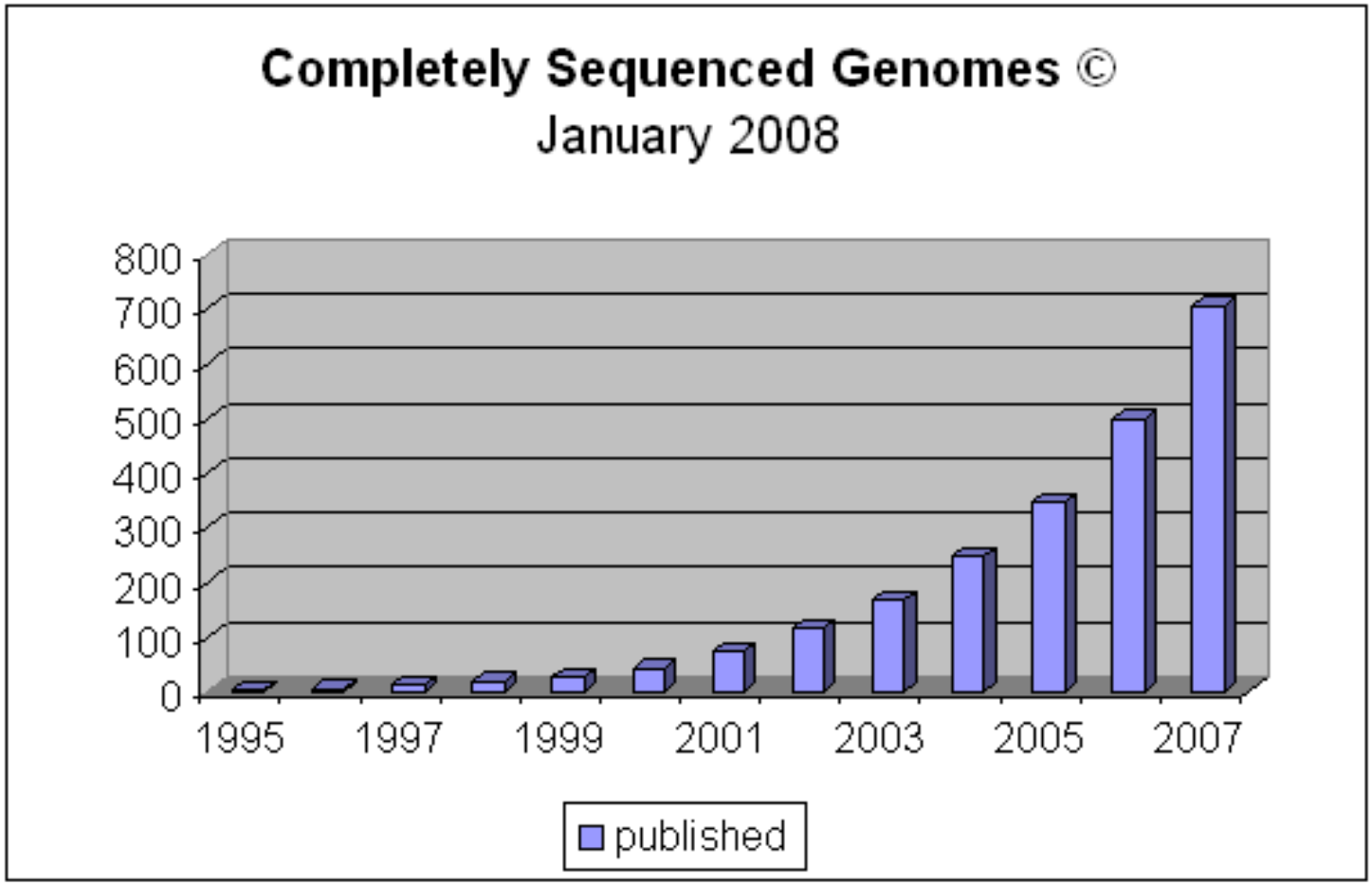
  96 (90) Archebakterien

  1029 (936) Eukaryoten

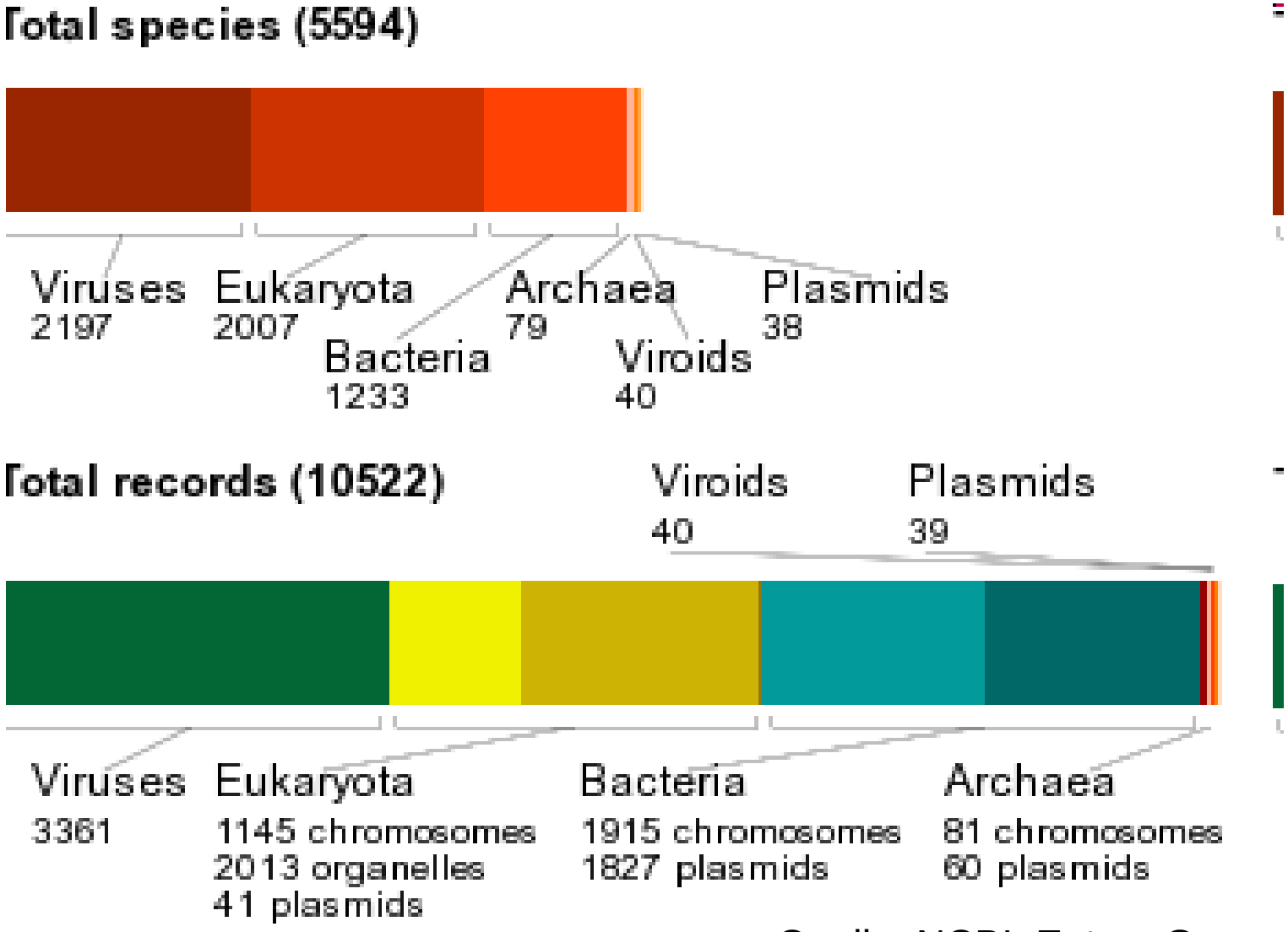
  167 (130) Metagenome

http://www.genomesonline.org/gold.cgi

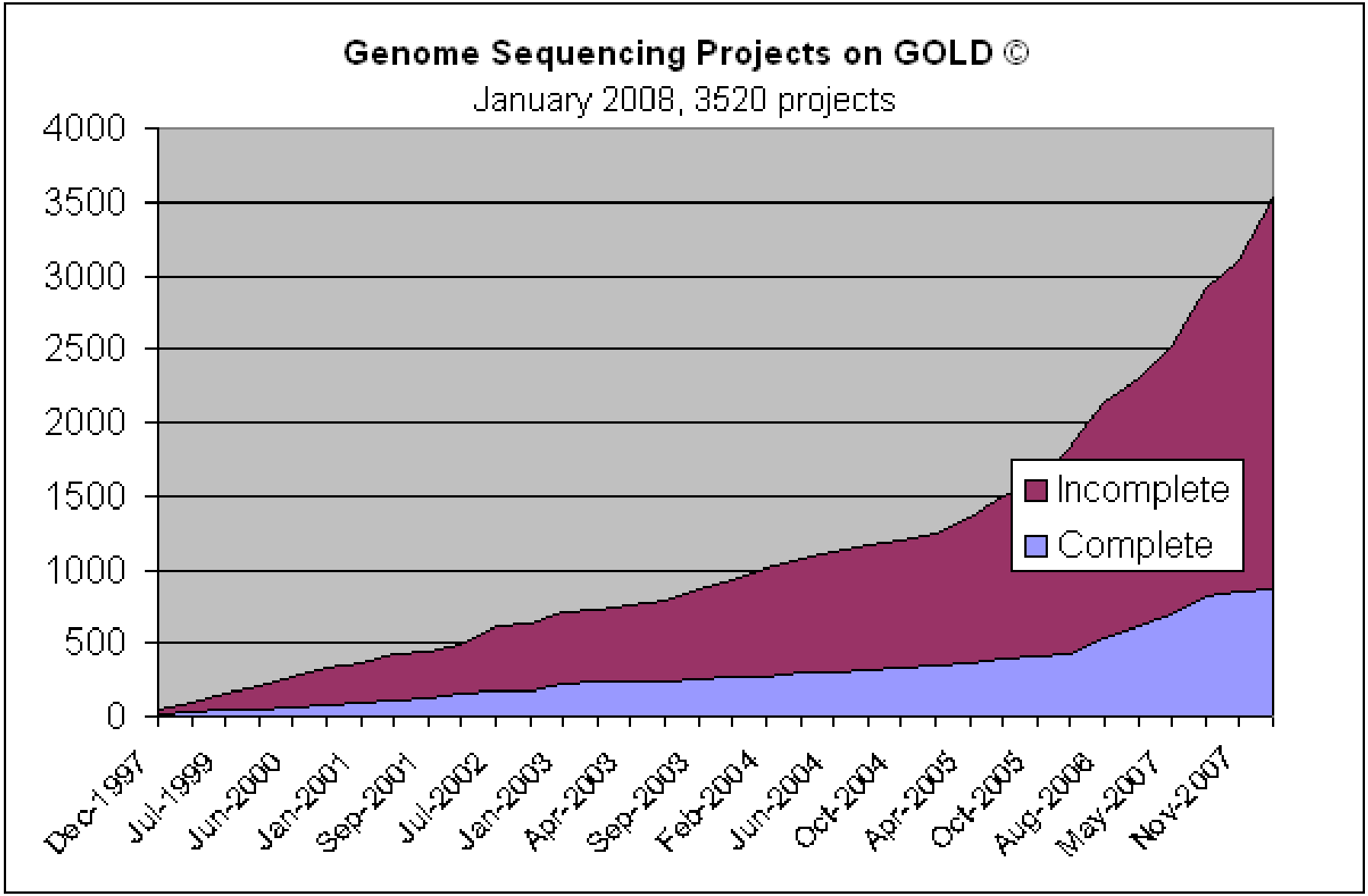
Vollständig sequenzierte Genome

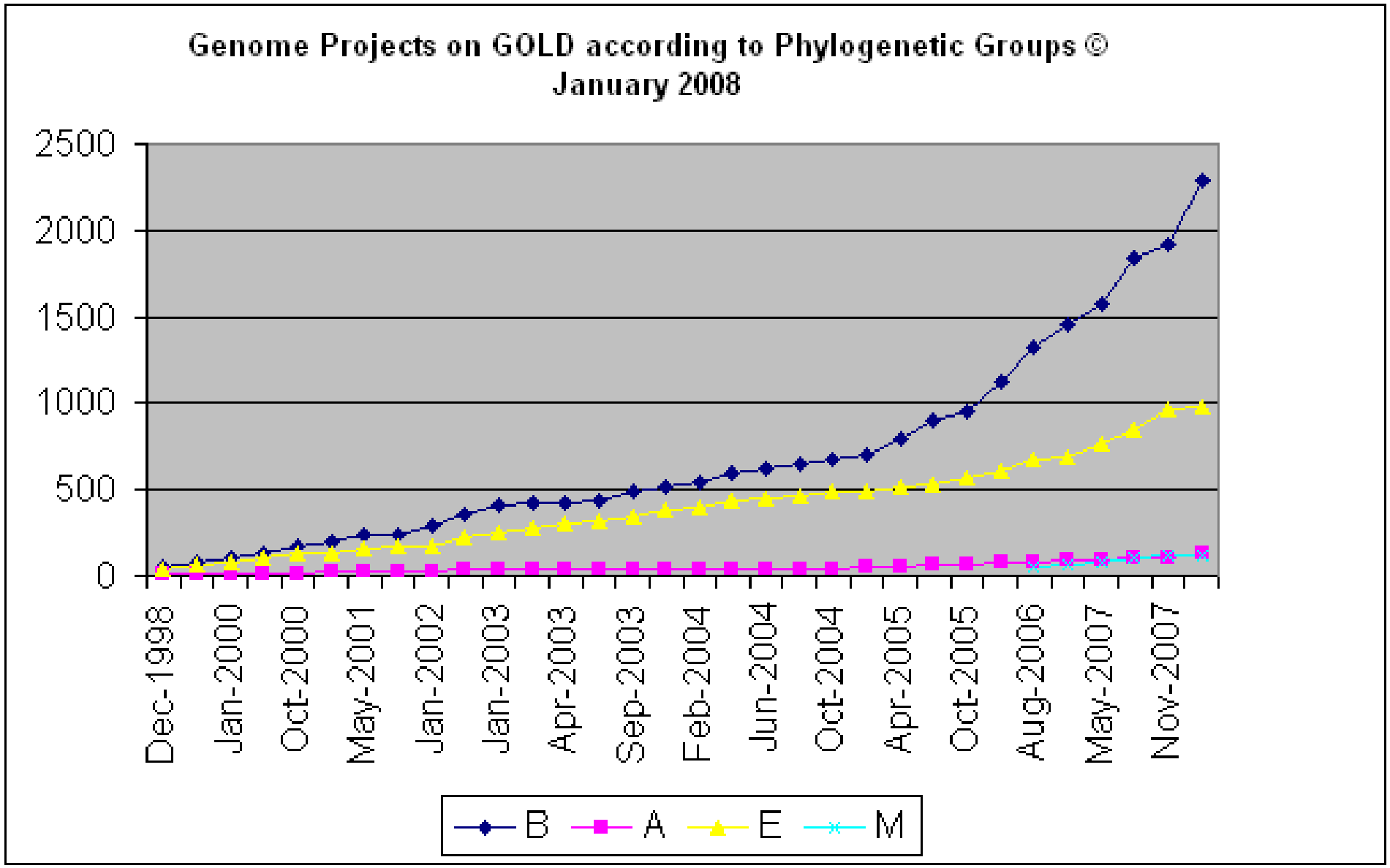


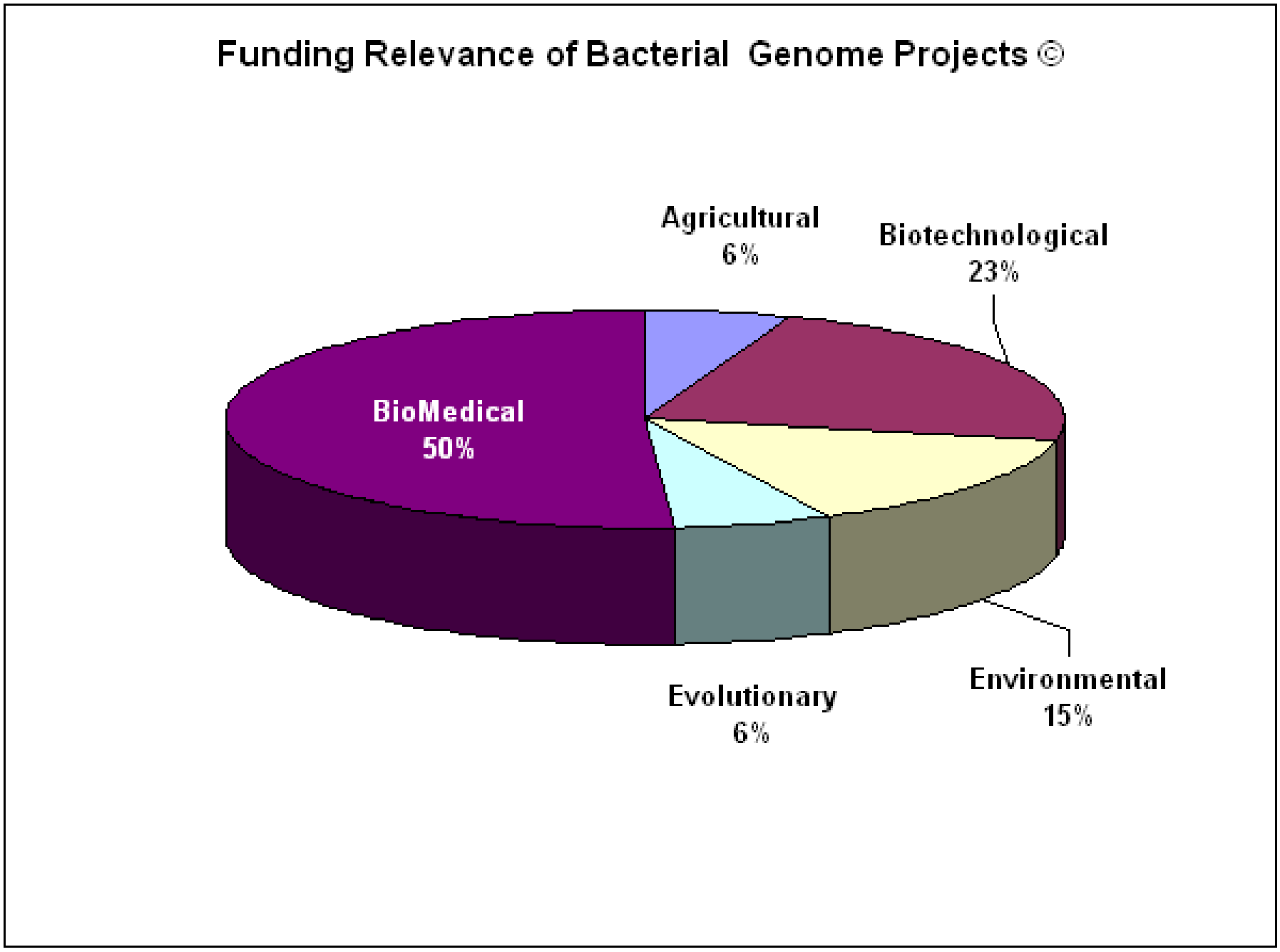
Vollständig sequenzierte Genome

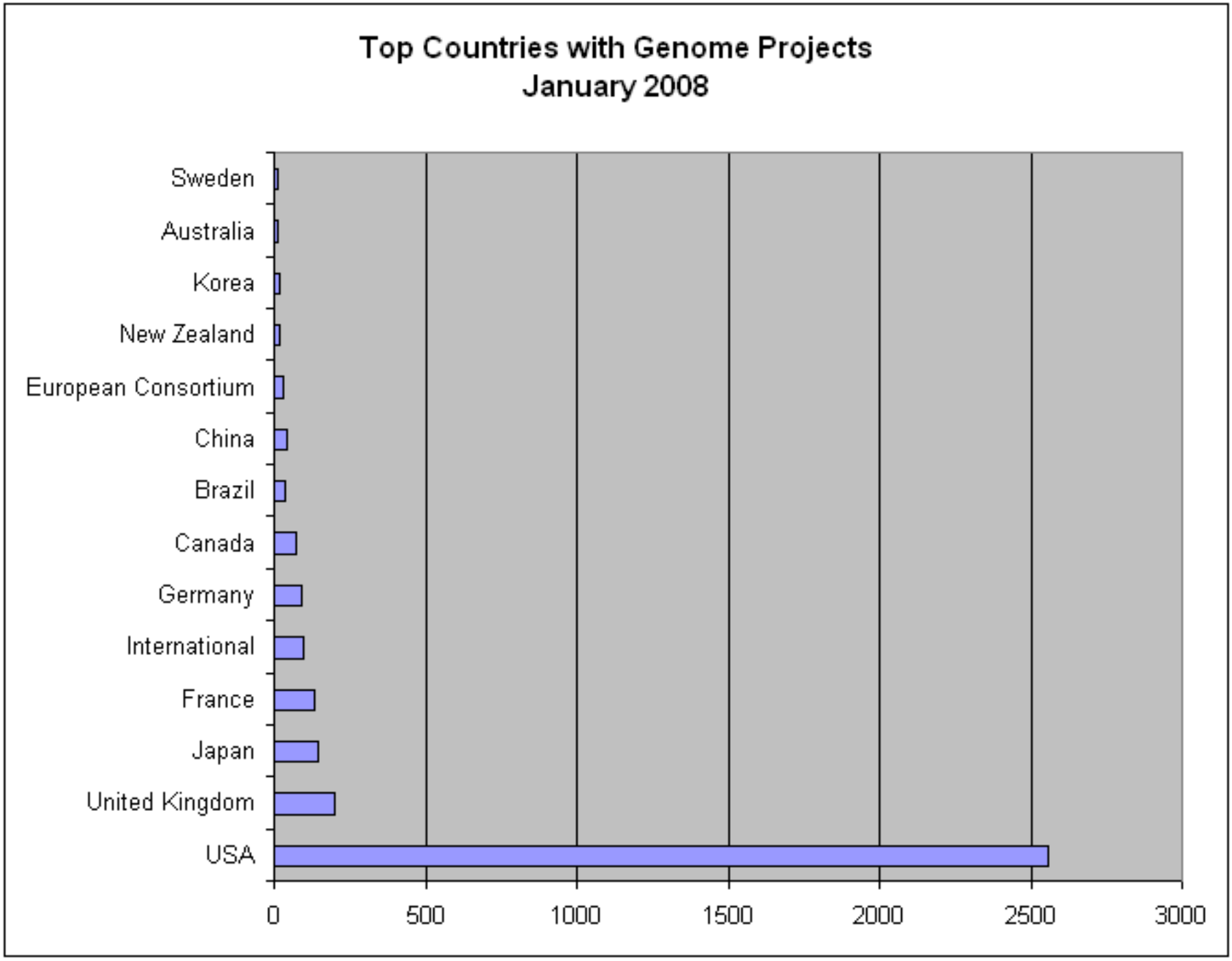


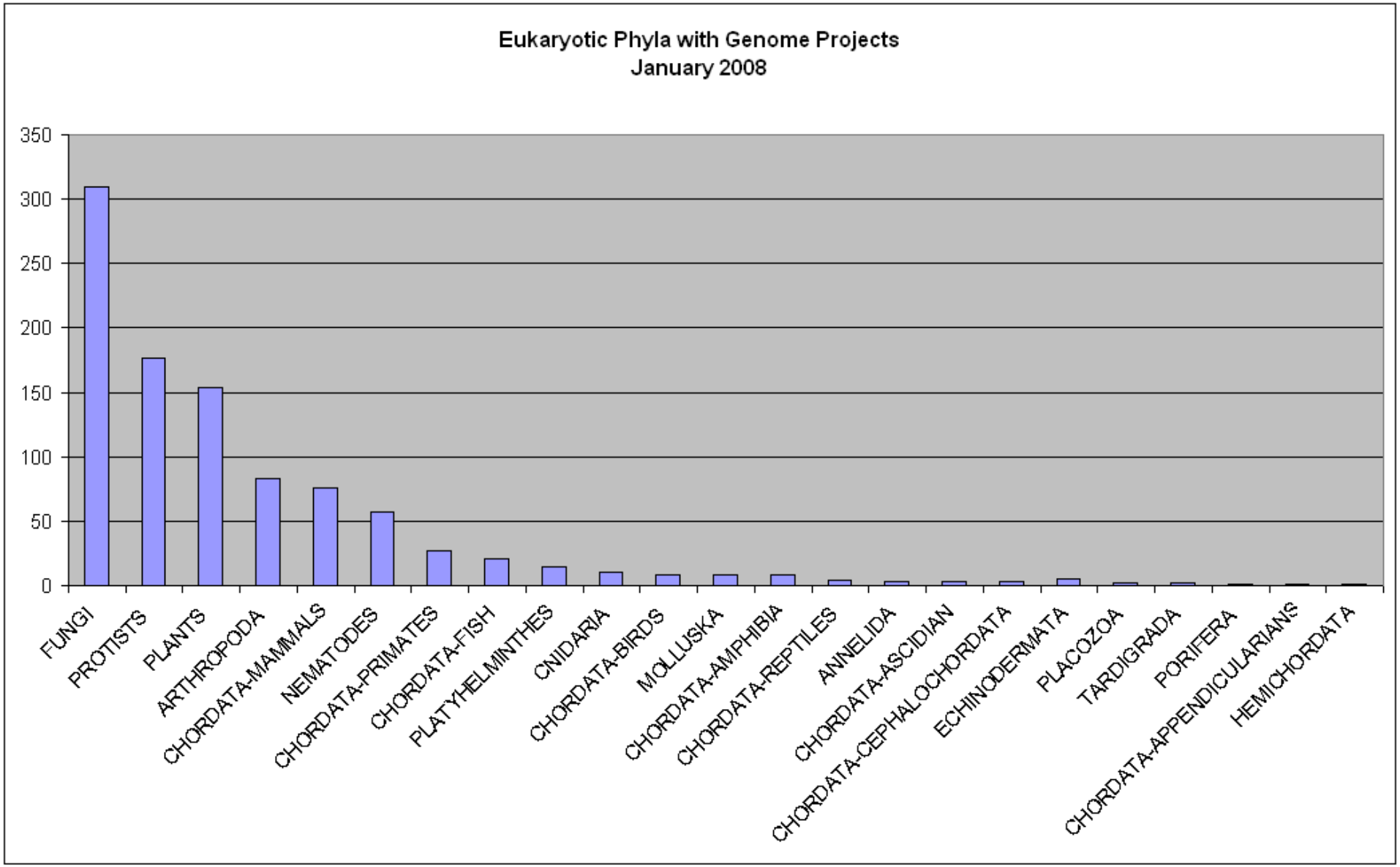
Quelle: NCBI; Entrez Genome

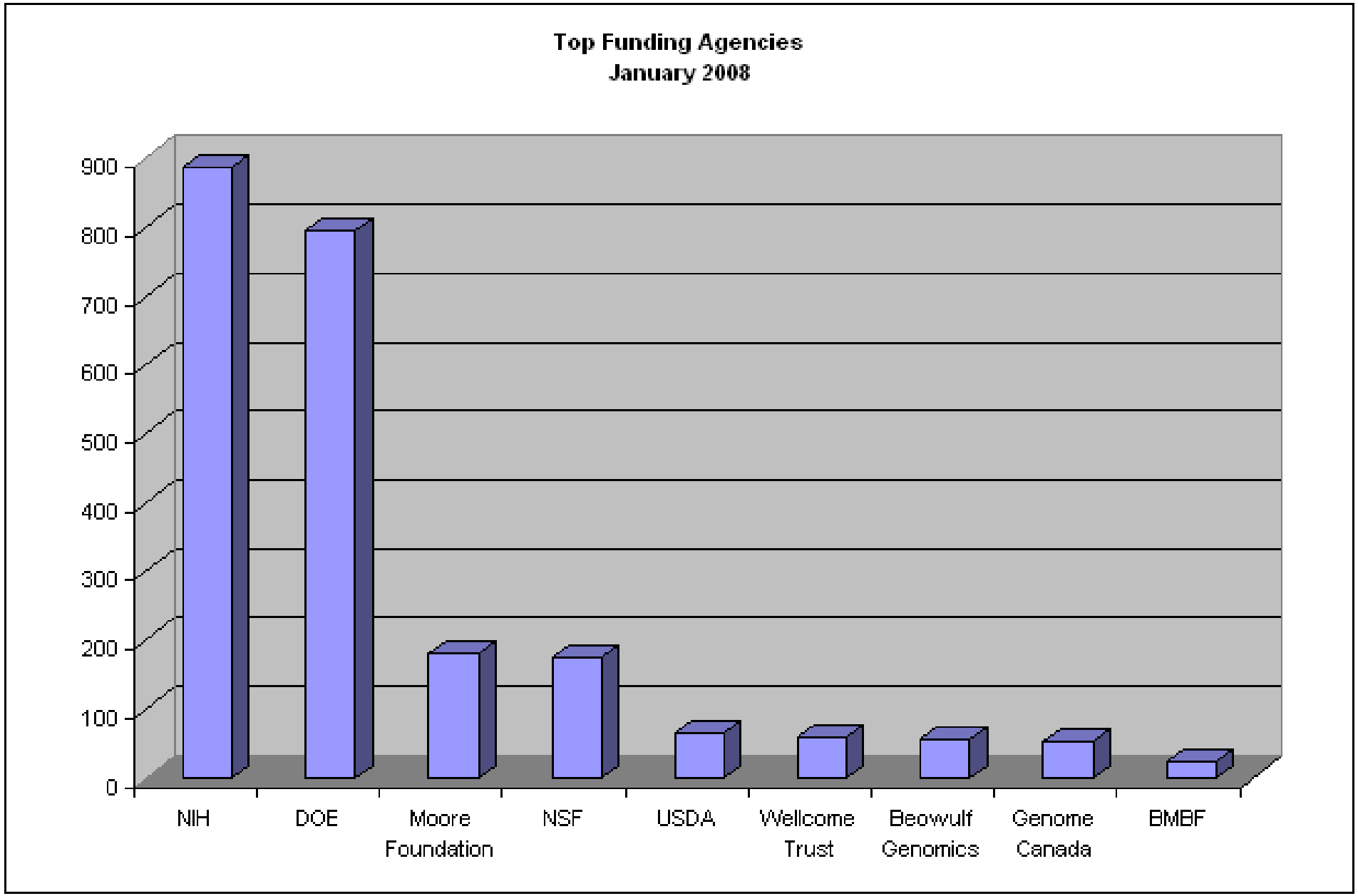

Genome Sequencing Projects on GOLD ©
January 2008, 3520 projects
4000
3500
3000
2500
2000
1500
1000
500
0
Incomplete
Complete
Dec-1997
Jul-1999
Jun-2000
Jan-2001
Sep-2001
Jul-2002
Jan-2003
Apr-2003
Sep-2003
Feb-2004
Jun-2004
Oct-2004
Apr-2005
Oct-2005
Aug-2006
May-2007
Nov-2007

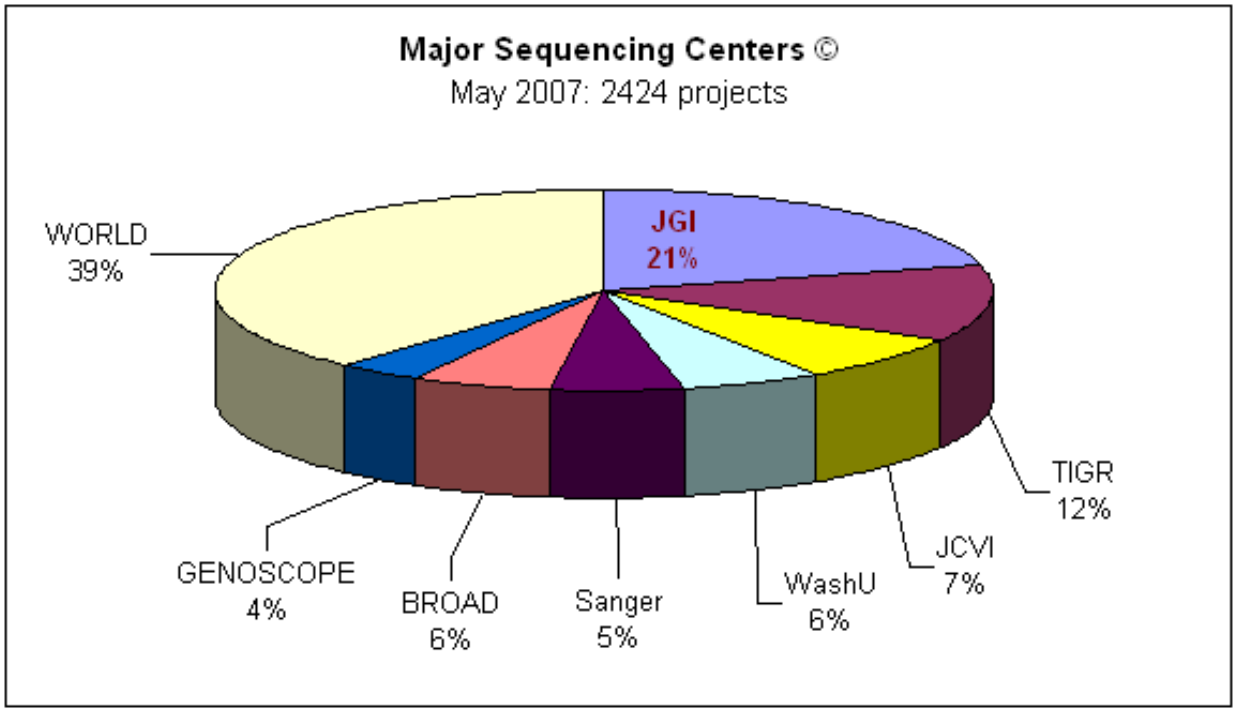

Genome Projects on GOLD according to Phylogenetic Groups ©
January 2008
2500
2000
1500
1000
500
0
Dec-1998
Jan-2000
Oct-2000
May-2001
Jan-2002
Jan-2003
Apr-2003
Sep-2003
Feb-2004
Jun-2004
Oct-2004
Apr-2005
Oct-2005
Aug-2006
May-2007
Nov-2007
B
A
E
M


Funding Relevance of Bacterial Genome Projects ©
Agricultural
6%
Biotechnological
23%
BioMedical
50%
Evolutionary
6%
Environmental
15%

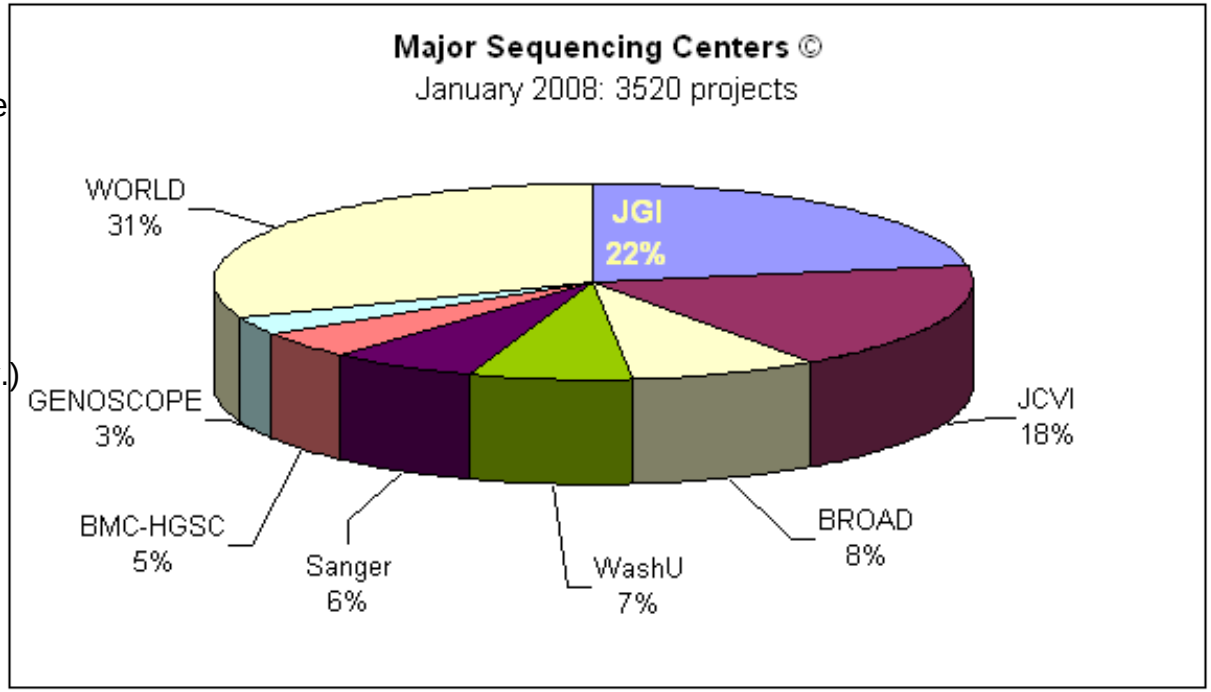

Top Countries with Genome Projects
January 2008
Sweden
Australia
Korea
New Zealand
European Consortium
China
Brazil
Canada
Germany
International
France
Japan
United Kingdom
USA
0
500
1000
1500
2000
2500
3000


Eukaryotic Phyla with Genome Projects
January 2008
350
300
250
200
150
100
50
0
FUNGI
PROTISTS
PLANTS
ARTHROPODA
CHORDATA-MAMMALS
NEMATODES
CHORDATA-PRIMATES
CHORDATA-FISH
PLATYHELMINTHES
CNIDARIA
CHORDATA-BIRDS
MOLLUSKA
CHORDATA-AMPHIBIA
CHORDATA-REPTILES
ANNELIDA
CHORDATA-ASCIDIAN
CHORDATA-CEPHALOCHORDATA
ECHINODERMATA
PLACOZOA
TARDIGRADA
PORIFERA
CHORDATA-APPENDICULARIANS
HEMICHORDATA


Top Funding Agencies
January 2008
900
800
700
600
500
400
300
200
100
0
NIH
DOE
Moore Foundation
NSF
USDA
Wellcome Trust
Beowulf Genomics
Genome Canada
BMBF



BMC: Baylor College of Medicine
Genoscope: Centre National de ´sequencage
Sanger: ThéWellcome Trust Sanger Institute
WashU: Washington University Genome Sequencing Center
JGI: Joint Genome Institute, Departmant of Energy, USA
BROAD: Eli & Edythe Broad Institute (MIT, Whitehead, Harvard Univ.)
JCVI: J. Craig Venter Institute



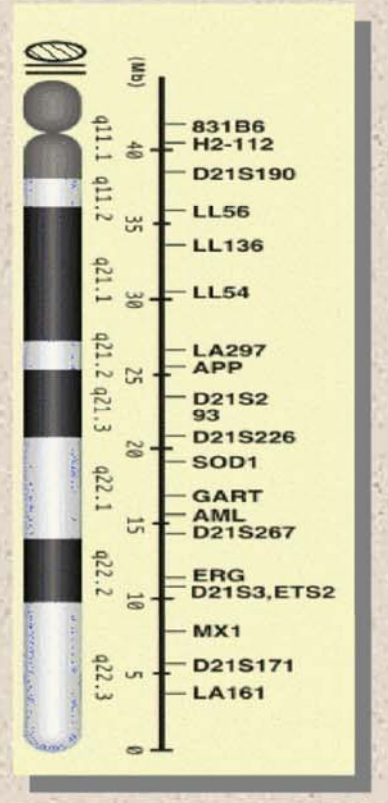
Nature vom 2. Dez. 1999:

„The DNA sequence of human chromosome 22“

- 33 464 000 Basenpaare
- noch 11 Lücken
- 545 Gene (davon 298 = 55% neu)
- 134 Pseudogene
- 39% der DNA von Genen besetzt (Exons + Introns)
- Genlängen: 1 kb min. bis 583 kb max
- 3% der DNA kodiert für Proteine
- 42% besteht aus repetitiver„junk“-DNA

Nature **vom 10. Apr. 2000:**

„The DNA sequence of human chromosome 21“



- **das kleinste menschliche Chromosom (1% der DNA)**
- **33 546 361 Basenpaare (3 Lücken, <100kb)**
- **225 Gene, 127 davon bekannt**
- **59 Pseudogene**
- **7 Mbp Abschnitt ohne einziges Gen!**

Maximal 40 000 Gene ??

(zum Vergl.: MHC Genkomplex auf Chr. 6 hat 128 Gene in nur 3,6 Mb)