

I'm an athletic mature, independent, 33 yrs. old, with a youthful appearance of 23. I'm a single white male and a working professional in the research industry. I'm 5'10" tall, 175 lbs, with light brown hair and hazel eyes. I'm a non-drinker and non-smoker. I've never been married and have no dependants. Currently seeking a female companion who enjoys the outdoors and understands DNA replication. If interested in finding out more, call Box 1044.

Vybrané experimentální postupy pro manipulaci s DNA a proteiny

segmentace enzymy a rekombinace DNA

transformace organismů cizí DNA

klonování DNA

cDNA knihovny

elektroforéza DNA

hybridizace se značenými sondami

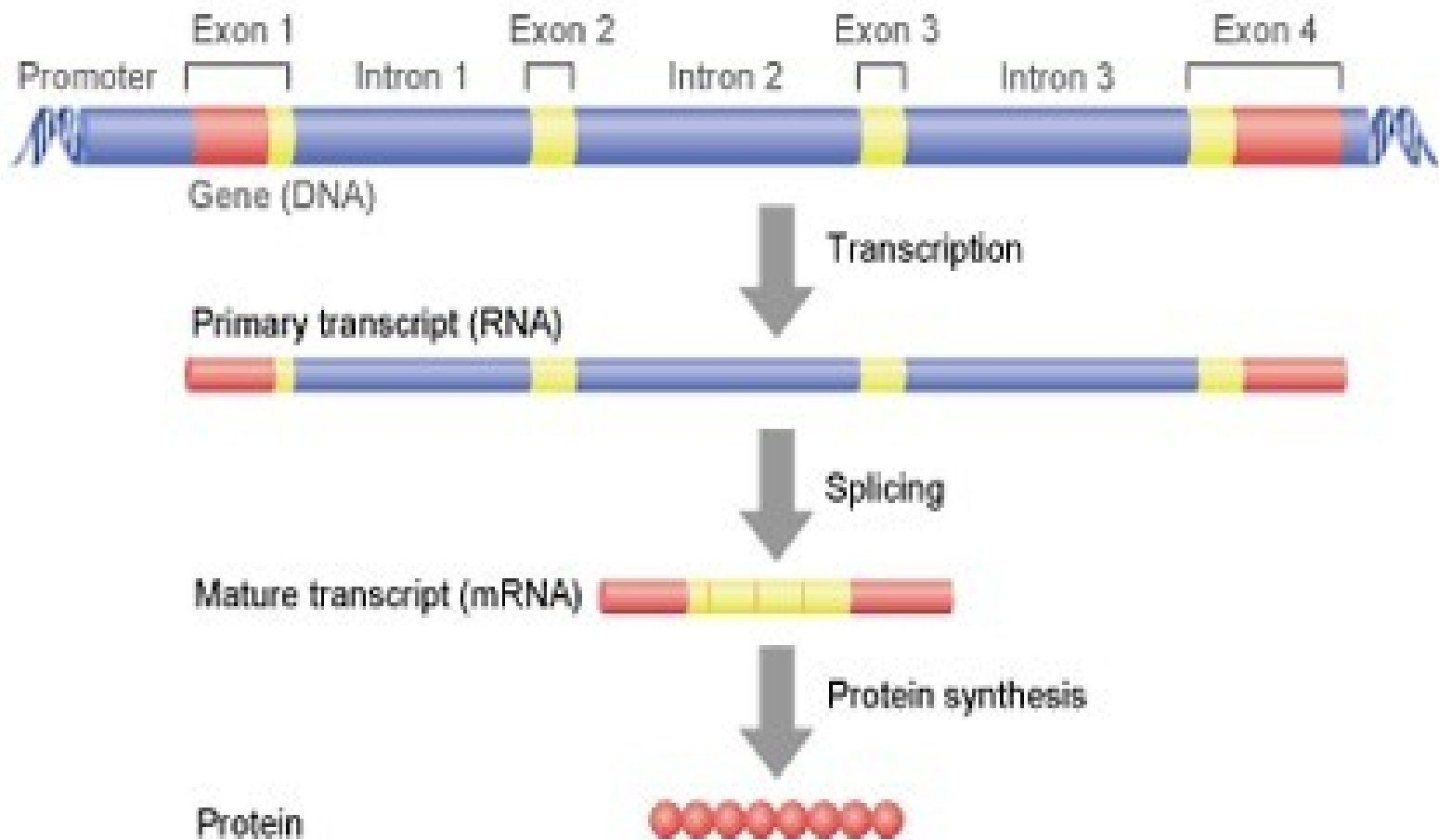
PCR (Polymerase Chain Reaction)

určování sekvence DNA

produkce rekombinantních proteinů

určování struktury proteinů (X-ray, NMR)

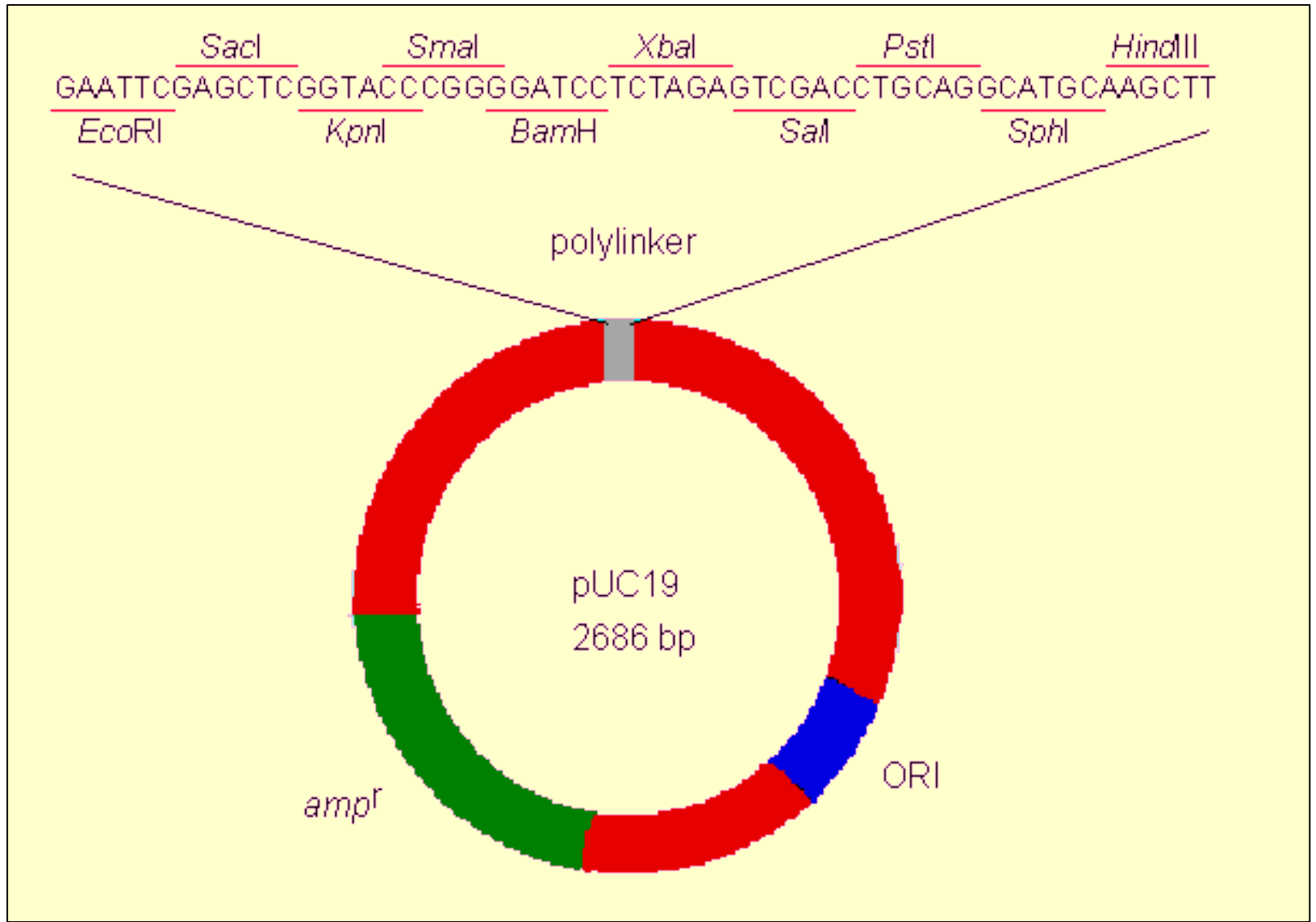
Structure of a Gene



Restrikcni endonukleazy

Enzyme	Source	Recognition site	Average cleaved size (kb)
<i>AluI</i>	<i>Arthrobacter luteus</i>	AG↓CT TCTGA	0.3
<i>BamHI</i>	<i>Bacillus amyloliquefaciens H</i>	G↓GATC C C CTAG↑G	7.0
<i>EcoRI</i>	<i>Escherichia coli R factor</i>	G↓AATT C C TTAA↑G	3.1
<i>HaeIII</i>	<i>Hemophilus aegyptus</i>	GG↓CC CC↑GG	0.6
<i>HindIII</i>	<i>Hemophilus influenzae Rd</i>	A↓AGCT T T TCGA↑A	3.1
<i>NofI</i>	<i>Norcadia otitidis-caviarum</i>	GC↓GGCC GC CG CCGG↑CG	< 9700
<i>PstI</i>	<i>Providencia stuartii</i>	C TGCA↓G G↑ACGT C	7.0
<i>TaqI</i>	<i>Thermus aquaticus</i>	T↓CG A A GCT	1.4

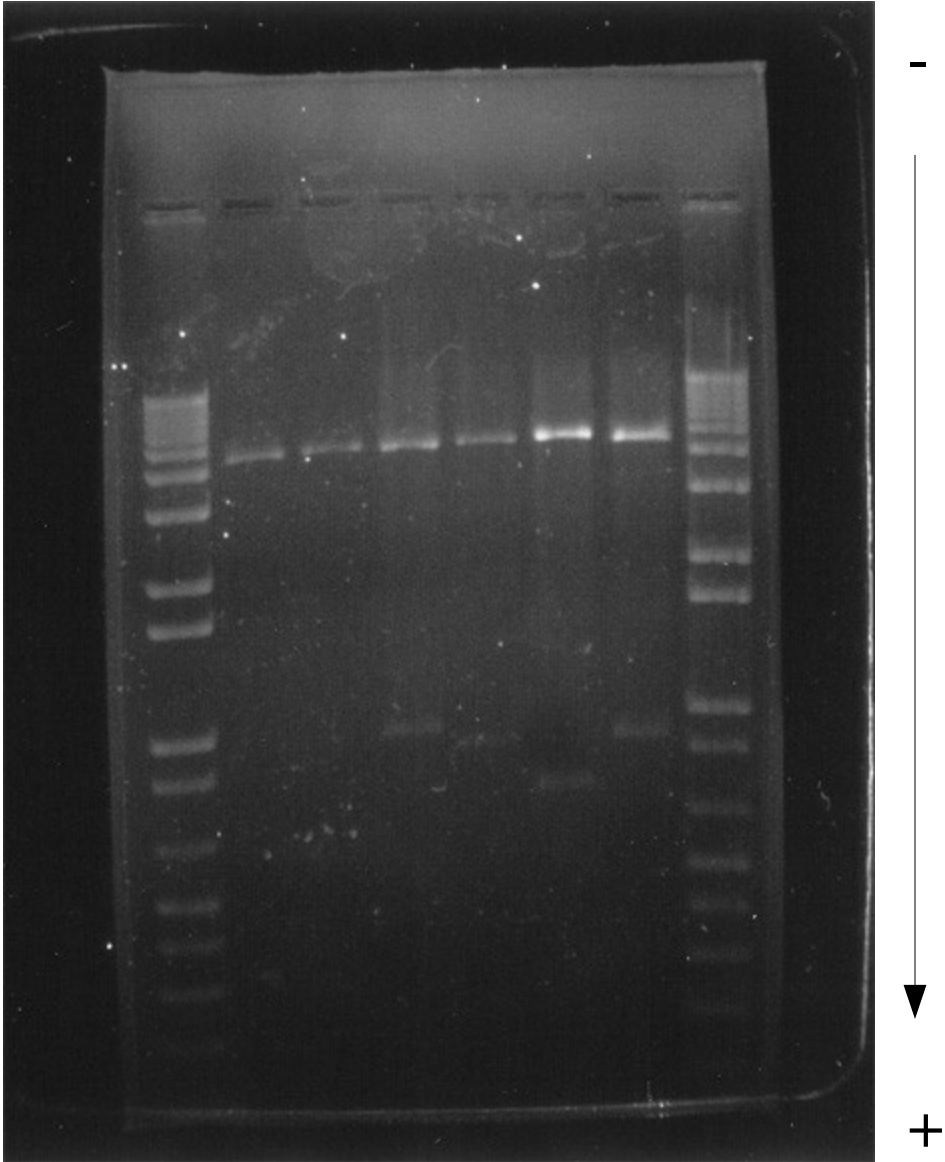
Vektor pro rekombinantni DNA (klonovani DNA)



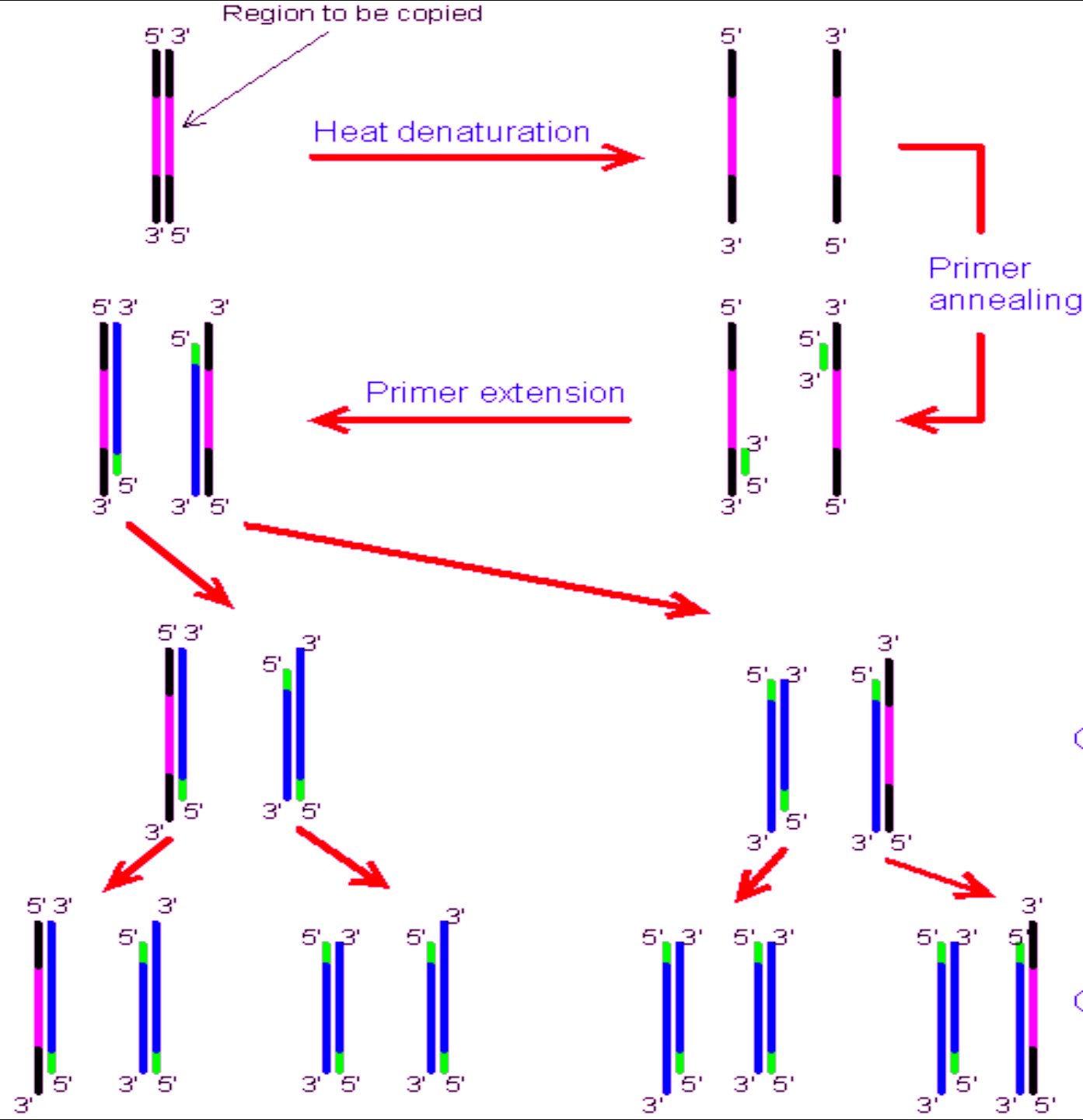
Manipulate DNA



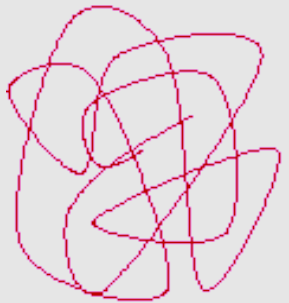
Elektroforeza DNA



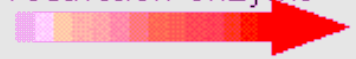
PCR



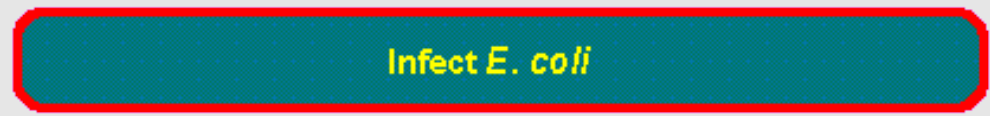
Target DNA

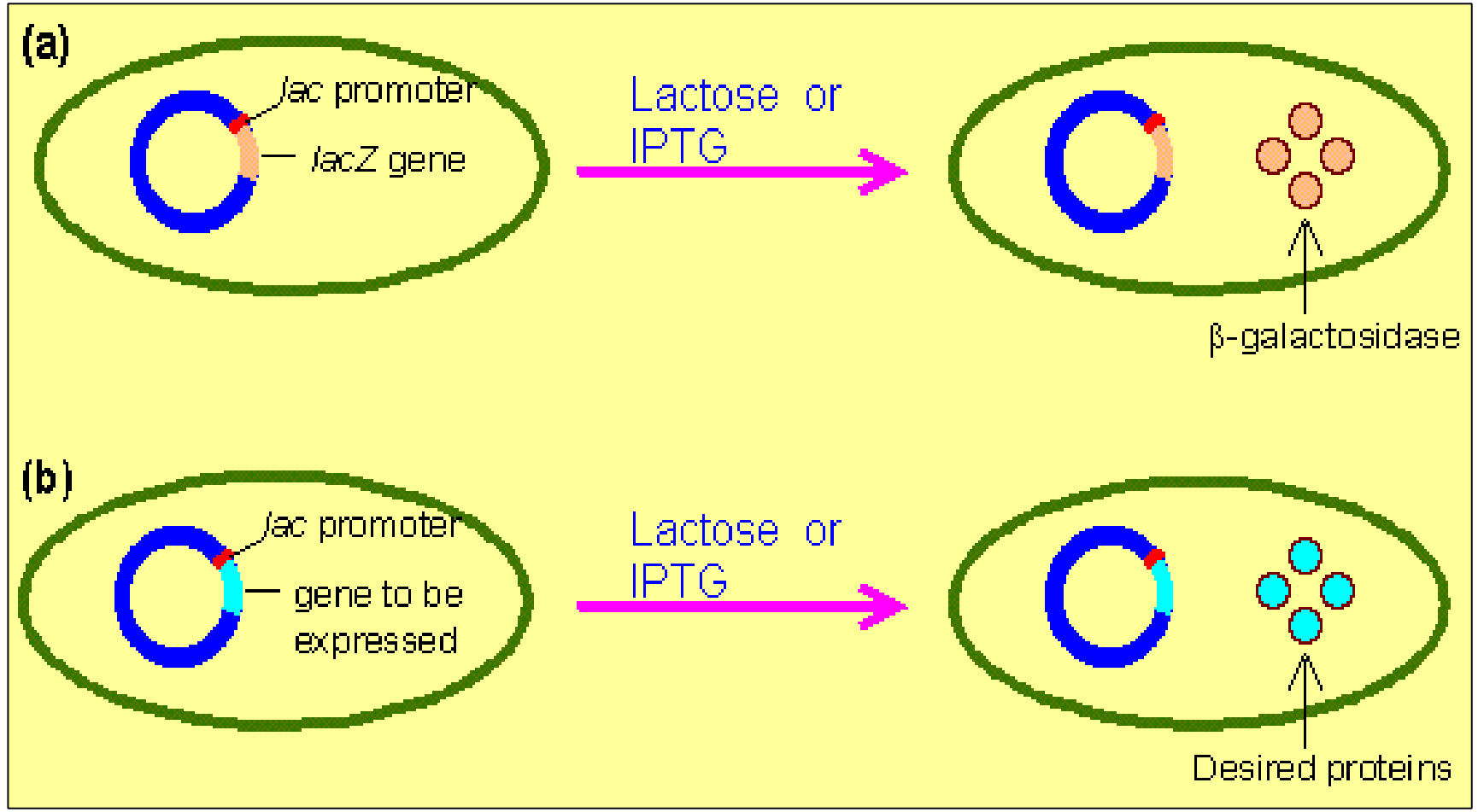


Digest with restriction enzyme



DNA recombination, packaging, assembly





Sekvence DNA

- Pořadí nukleotidů ve směru 5' k 3' uloženo pro budoucí generace jako dlouhý řetězec symbolů A, C, G a T
- Dobrá správa pro budoucnost bioinformatiků: zatím je osekvenováno jenom asi 150 organismů, data budou přibývat
- Polymorfismus: Rozdíl mezi jedinci Homo sapiens sapiens 1/1000 bazí

Sekvence DNA

- Sekvenace se dělá ve dvou fázích
 - Za použití DNA polymerázy a směsi normálních a modifikovaných deoxyribonukleotidů se syntetizují vlákna DNA, templátem je sekvenovaná molekula DNA
 - Produkty polymerizace se v automatických strojích chromatograficky dělí podle velikosti a vypovídají o sekvenci
- Jedna reakce podá informaci o 500-1000 bazích

DNA Polymerase reads the template strand and synthesizes a new second strand to match:



IF 5% of the T nucleotides are actually dideoxy T, then each strand will terminate when it gets a ddT on its growing end:

5' - TACGCGGTACGGTATGTTTCGACCGTTTAGCTACCGAT•

5' - TACGCGGTACGGTATGTTTCGACCGTTTAGCT•

5' - TACGCGGTACGGTATGTTTCGACCGTTT•

5' - TACGCGGTACGGTATGTTTCGACCGTT•

5' - TACGCGGTACGGTATGTTTCGACCGT•

5' - TACGCGGTACGGTATGTT•

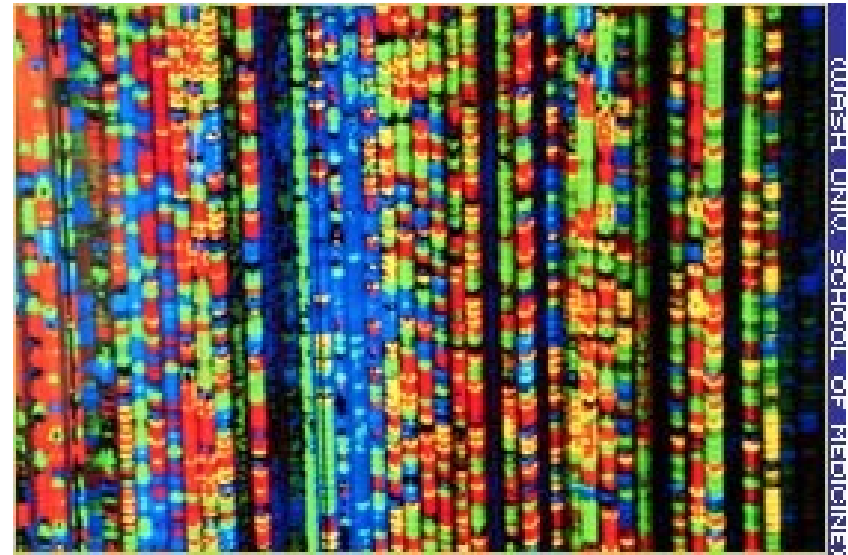
5' - TACGCGGTACGGTATGT•

5' - TACGCGGTACGGTAT•

5' - TACGCGGTACGGT•

5' - TACGCGGT•

Sekvence DNA



Hierarchical shotgun sequencing

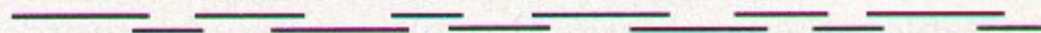
Genomic DNA



BAC library



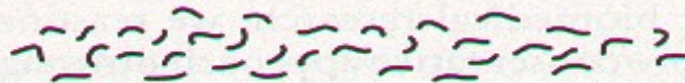
Organized mapped large clone contigs



BAC to be sequenced



Shotgun clones



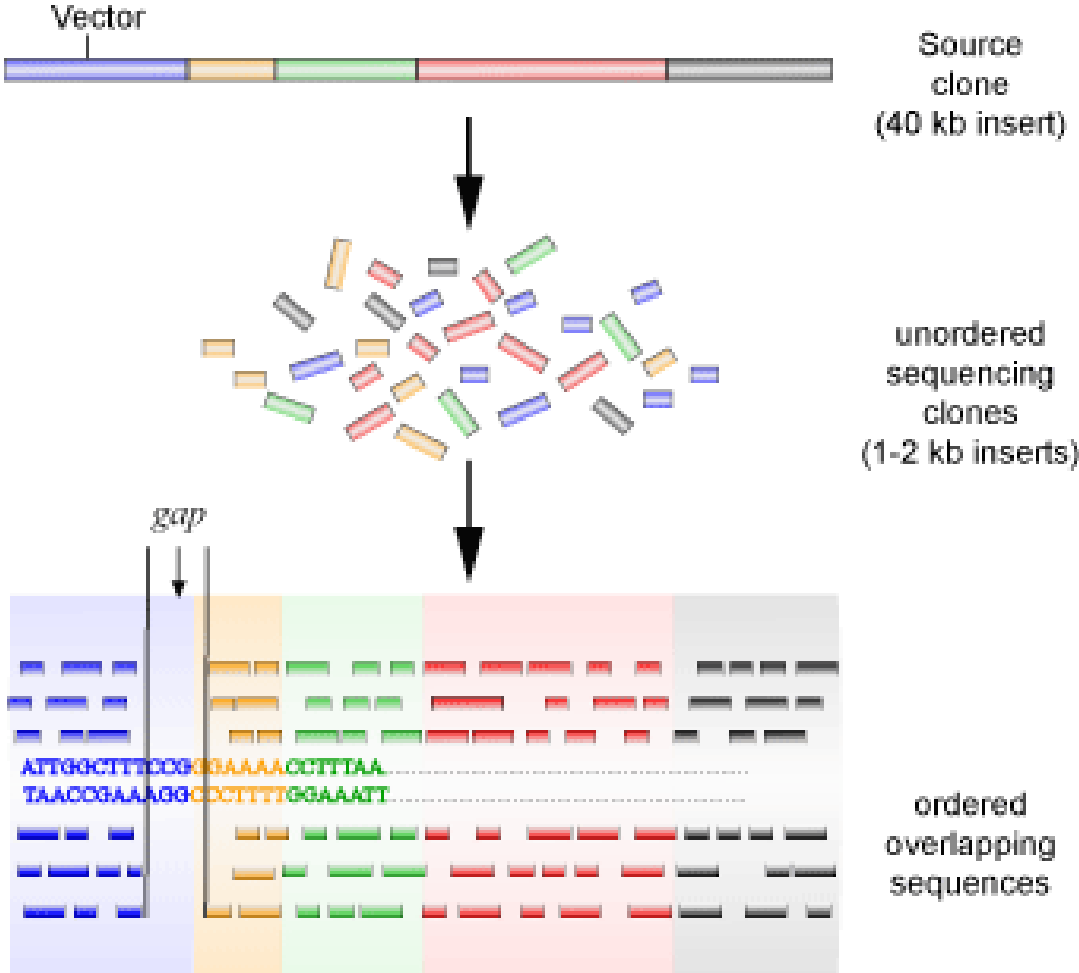
Shotgun sequence

```
...ACCGTAAATGGGCTGATCATGCTTAAA  
TGATCATGCTTAAACCCTGTGCATCCTACTG...
```

Assembly

```
...ACCGTAAATGGGCTGATCATGCTTAAACCCTGTGCATCCTACTG...
```

Shotgun Sequencing



Sekvence DNA

1975	bakteriofag MS2	
1977	bakteriofag PhiX174	5,375 bp
1982	bakteriofag Lambda	
1984	HIV-1	
1990	virus HCMV	230 Kbp
1995	H. influenzae	1,83 Mbp
1996	E. coli	4,60 Mbp
	S. cerevisiae	12,00 Mbp
1998	C. elegans	96 Mbp
2000	D. melanogaster	120 Mbp
	A. thaliana	130 Mbp
2001	H. sapiens	3 000 Mbp
	M. musculus	3 000 Mbp
2004-5	R. norvegicus, A. melifera, G. gallus, C. familiaris, B. taurus	



RESEARCH ARTICLE

Environmental Genome Shotgun Sequencing of the Sargasso Sea

J. Craig Venter,^{1*} Karin Remington,¹ John F. Heidelberg,³
Aaron L. Halpern,² Doug Rusch,² Jonathan A. Eisen,³
Dongying Wu,³ Ian Paulsen,³ Karen E. Nelson,³ William Nelson,³
Derrick E. Fouts,³ Samuel Levy,² Anthony H. Knap,⁶
Michael W. Lomas,⁶ Ken Nealson,⁵ Owen White,³
Jeremy Peterson,³ Jeff Hoffman,¹ Rachel Parsons,⁶
Holly Baden-Tillson,¹ Cynthia Pfannkoch,¹ Yu-Hui Rogers,⁴
Hamilton O. Smith¹

We have applied "whole-genome shotgun sequencing" to microbial populations collected en masse on tangential flow and impact filters from seawater samples collected from the Sargasso Sea near Bermuda. A total of 1.045 billion base pairs of nonredundant sequence was generated, annotated, and analyzed to elucidate the gene content, diversity, and relative abundance of the organisms within these environmental samples. These data are estimated to derive from at least 1800 genomic species based on sequence relatedness, including 148 previously unknown bacterial phylotypes. We have identified over 1.2 million previously unknown genes represented in these samples, including more than 782 new rhodopsin-like photoreceptors. Variation in species present and stoichiometry suggests substantial oceanic microbial diversity.

Microorganisms are responsible for most of the biogeochemical cycles that shape the environment of Earth and its oceans. Yet, these organisms are the least well understood on Earth, as the ability to study and understand the metabolic potential of microorganisms has been hampered by the inability to generate pure cultures. Recent studies have begun to explore environ-

mental microbial diversity and to begin to characterize. To help ensure a tractable pilot study, we sampled in the Sargasso Sea, a nutrient-limited, open ocean environment. Further, we concentrated on the genetic material captured on filters sized to isolate primarily microbial inhabitants of the environment, leaving detailed analysis of dissolved DNA and viral particles on one end of the size spectrum and eukaryotic inhabitants on

the other. We found that *chlorococcus*, that numerically dominate the photosynthetic biomass in the Sargasso Sea.

Surface water samples (170 to 200 liters) were collected aboard the RV Weatherbird II from three sites off the coast of Bermuda in February 2003. Additional samples were collected aboard the SV Sorcerer II from "Hydrostation S" in May 2003. Sample site locations are indicated on Fig. 1 and described in table S1; sampling protocols were fine-tuned from one expedition to the next (5). Genomic DNA was extracted from filters of 0.1 to 3.0 μm , and genomic libraries with insert sizes ranging from 2 to 6 kb were made as described (5). The prepared plasmid clones were sequenced from both ends to provide paired-end reads at the J. Craig Venter Science Foundation Joint Technology Center on ABI 3730XL DNA sequencers (Applied Biosystems, Foster City, CA). Whole-genome random shotgun sequencing of the Weatherbird II samples (table S1, samples 1 to 4) produced 1.66 million reads averaging 818 bp in length, for a total of approximately 1.36 Gbp of microbial DNA sequence. An additional 325,561 sequences were generated from the Sorcerer II samples (table S1, samples 5 to 7), yielding approximately 265 Mbp of DNA sequence.

Environmental genome shotgun assembly. Whole-genome shotgun sequencing projects have traditionally been applied to identify the genome sequence(s) from one particular organism, whereas the approach taken here is intended to capture representative sequence from many diverse organisms simultaneously. Variation in genome size and relative abundance determines the depth of coverage of any

- ★ DNA Microarray
- ★ DNA Chip

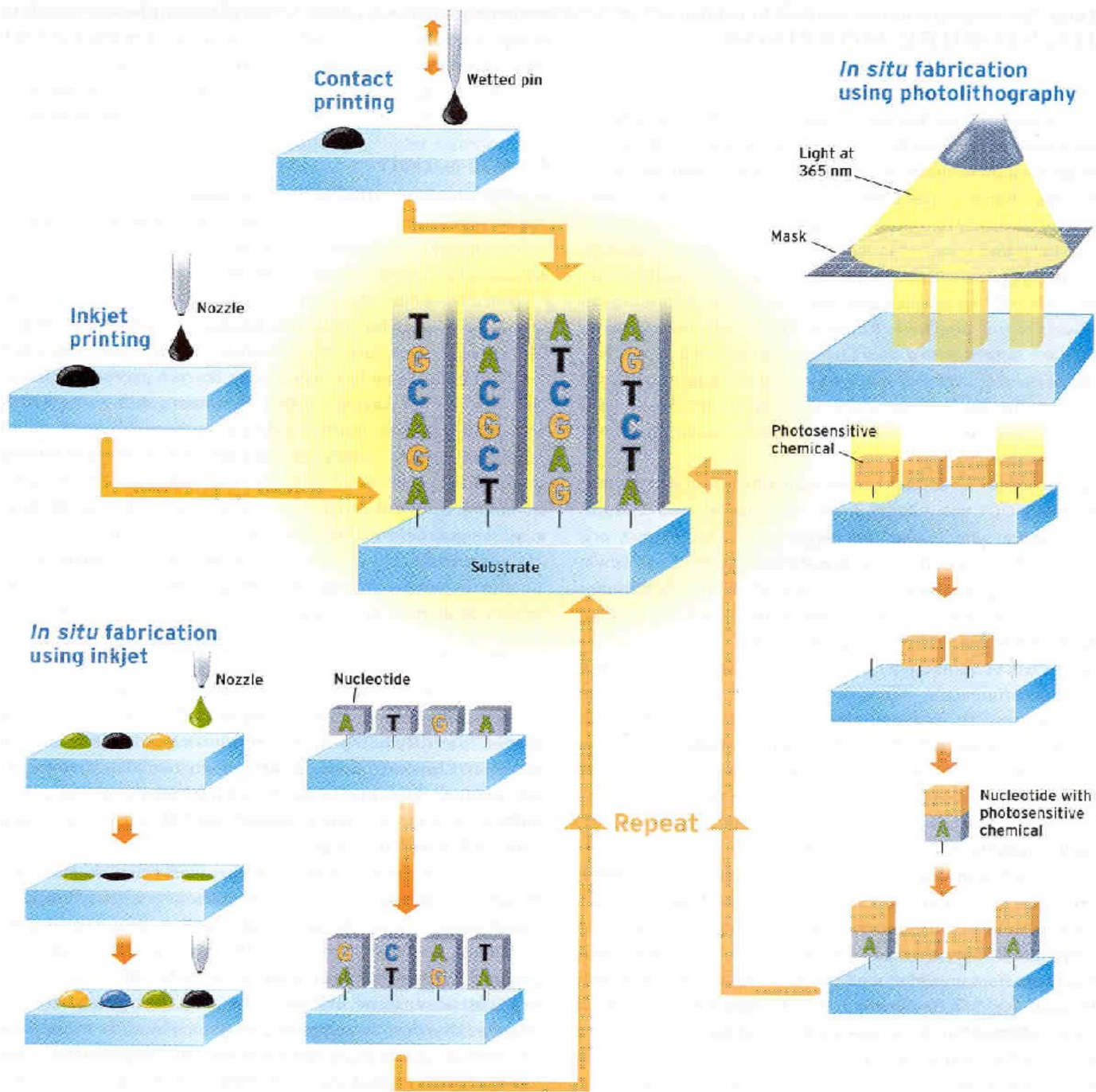


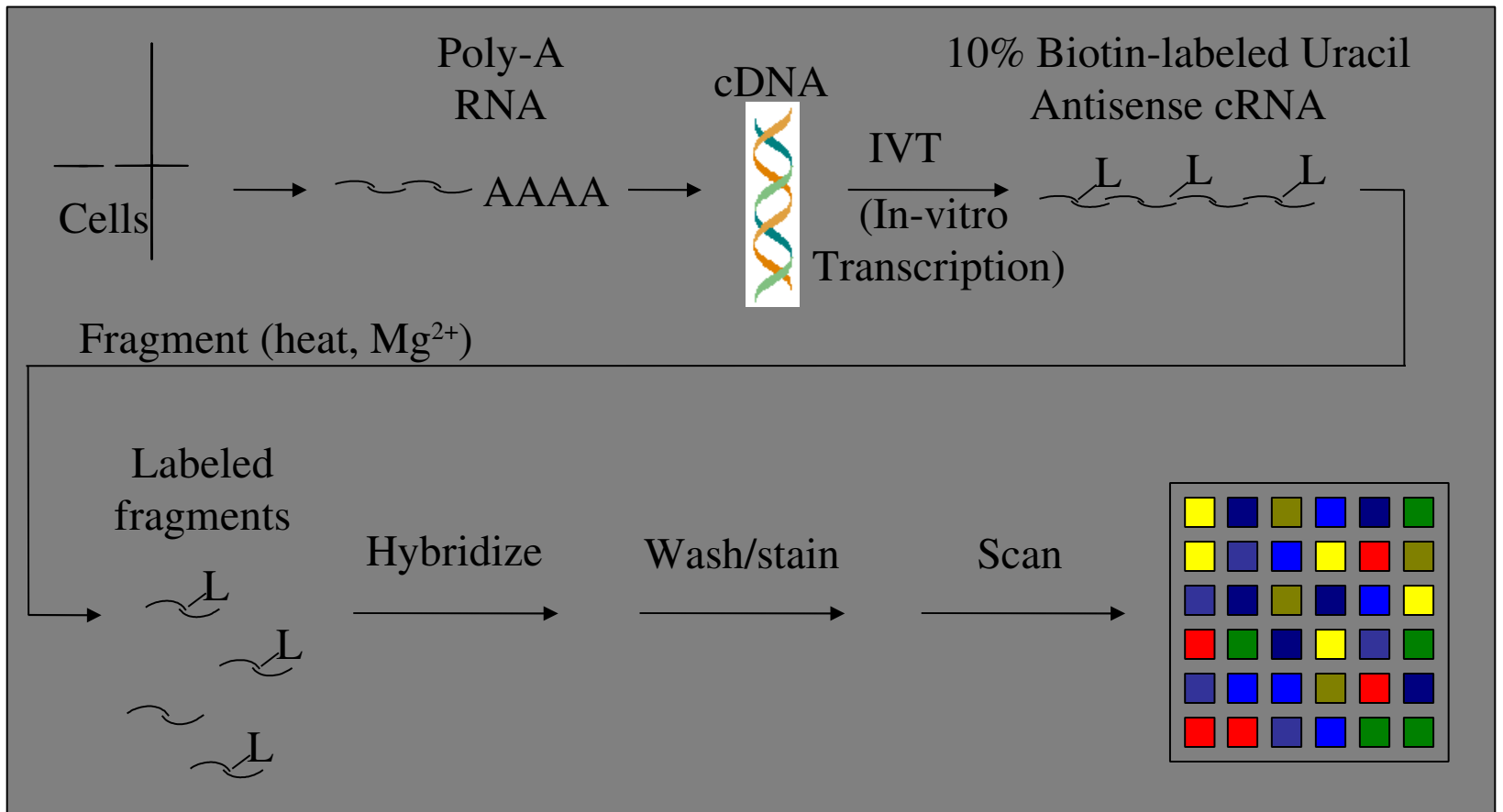
Každé pole (sonda) nese tisíce oligonukleotidů

Destička o rozměrech 1.28 cm x 1.28+ cm, sklo, silikon

Destičky s nanášenými oligonukleotidy mají pole asi 100 mikrometrů, řádově 10000 oligonukleotidových sond o 60 bází

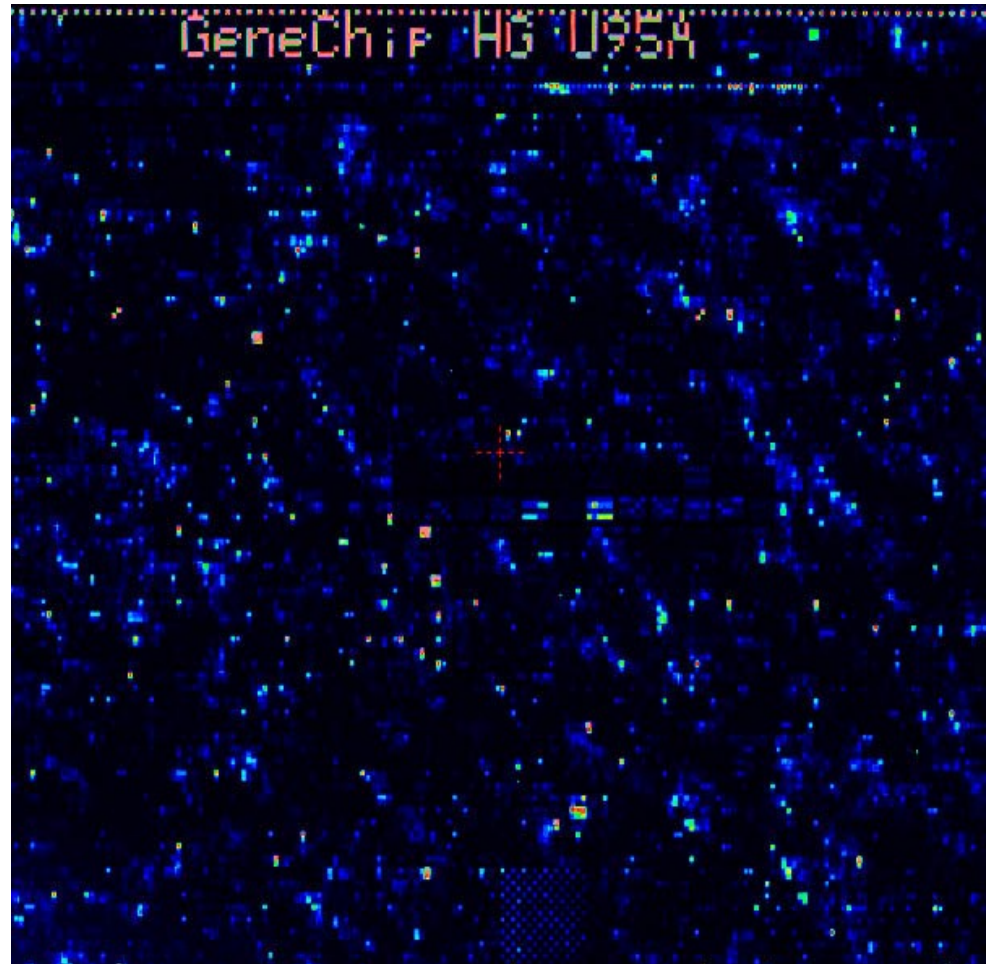
Destičky s oligonukleotidy syntetizovanými na místě (DNA čipy Affymetrix) mají pole 25 mikrometrů, řádově 100000 sond 25 bází

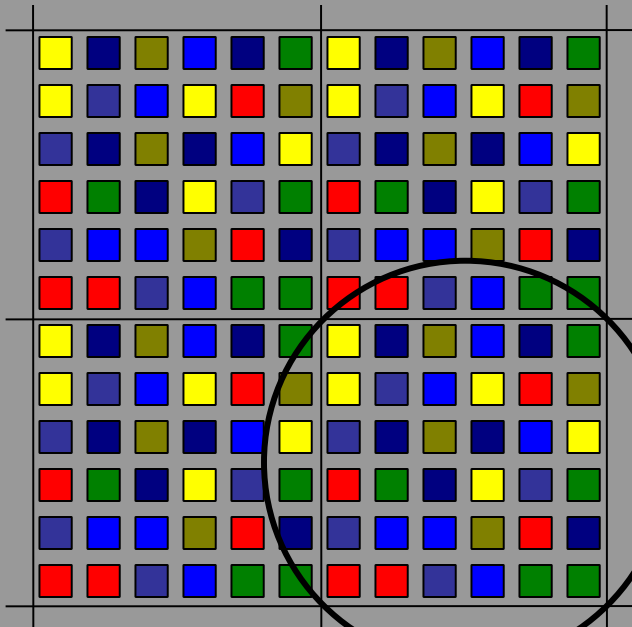






GENEARRAY™ SCANNER





	83	112	96	32	
	47	382	165	87	
	55	246	140	93	
	104	552	187	65	

S jakou frekvencí se bude v dostatečně dlouhé náhodné sekvenci DNA vyskytovat sekvence AAGT?

S jakou frekvencí se bude v dostatečně dlouhé molekule DNA vyskytovat sekvence AAGT, pokud víme, že obsah C + G je $\frac{2}{3}$ (cca 66%)?

S jakou frekvencí se bude v dostatečně velkém genomu vyskytovat sekvence AAGT, pokud víme, že obsah C+G je $2/3$ (cca 66%)?

central dogma
genome
DNA
nucleotide
3' / 5'
hybridization
replication
DNA polymerase
vector
plasmid
sequence alignment
PCR
DNA sequence
contig

gene
ORF
gene structure
promoter
intron
exon
prokaryote
ATG
GC content
eukaryote
TATA
enhancer
silencer

RNA
transcription
mRNA
gene expression
microarray
probe
EST
translation
codon
protein
proteome
mass spectrometry
signal transduction