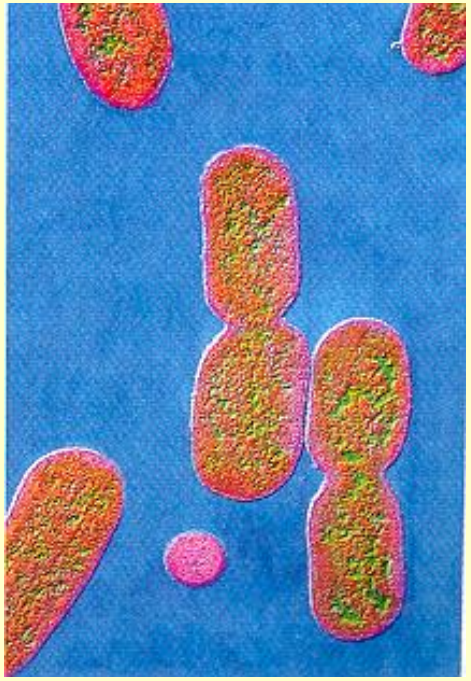
Molecular microbial diagnosis

Dr John Ferguson

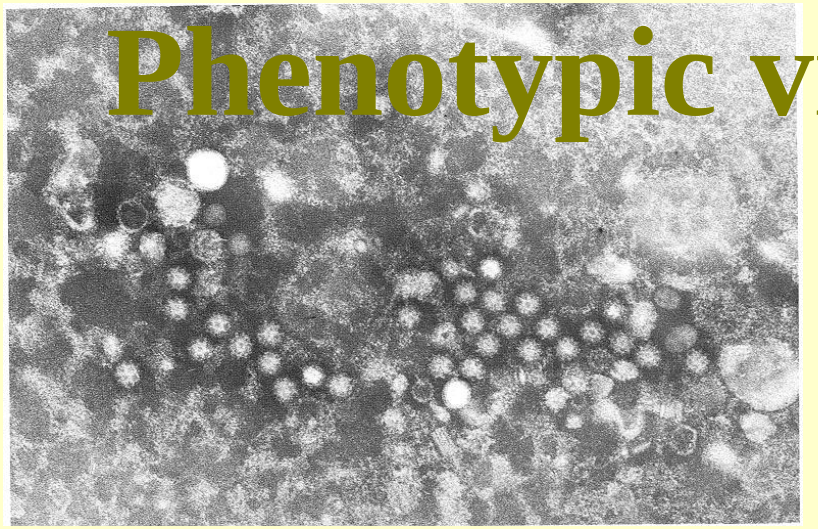
Microbiology and Infectious Diseases,

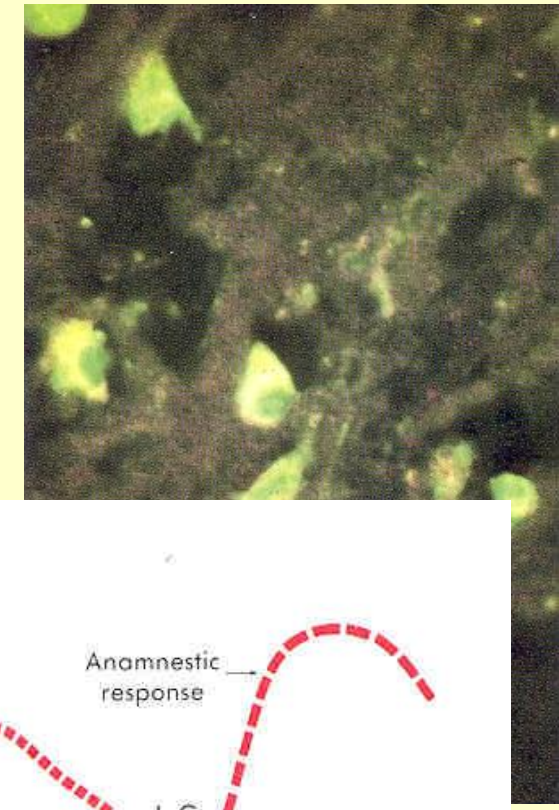
Newcastle, NSW, Australia

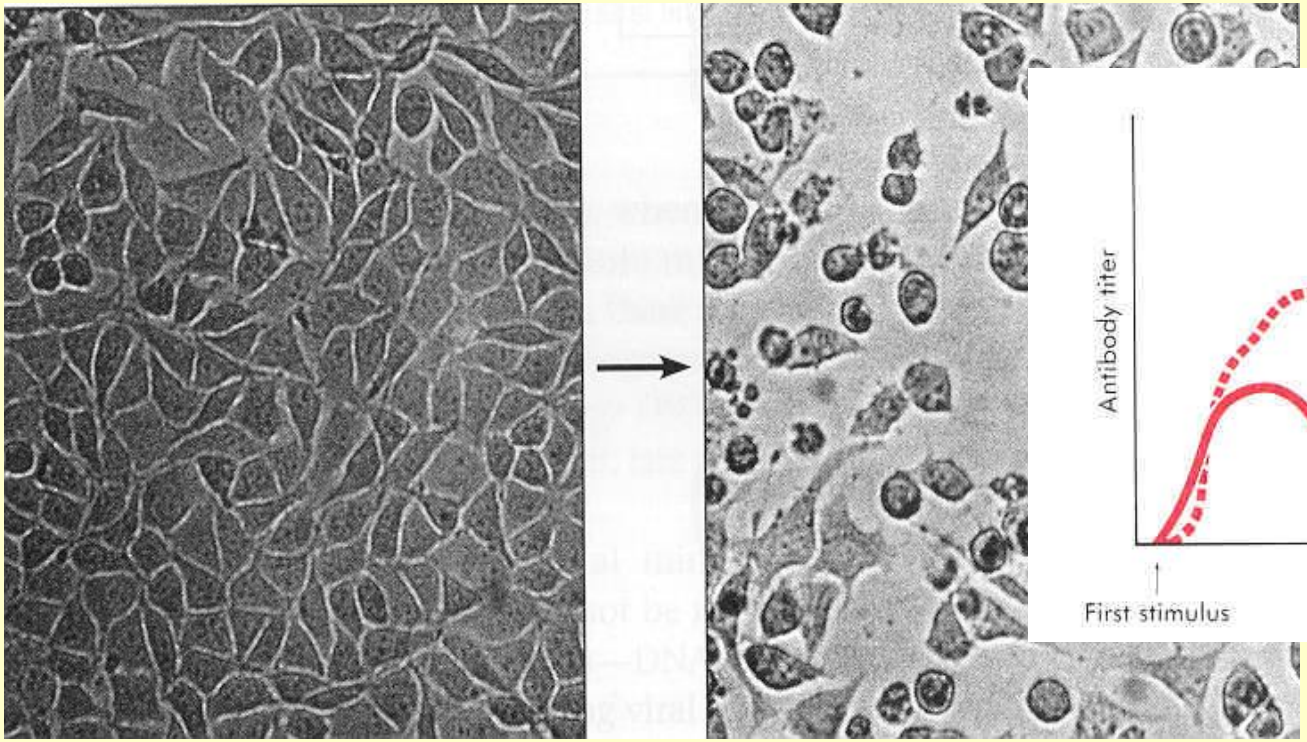
Updated 2016

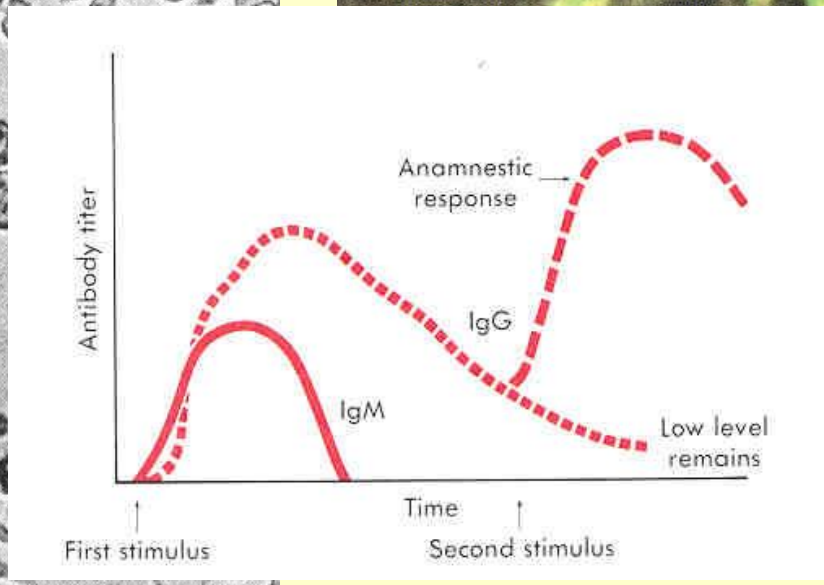


Phenotypic virology methods





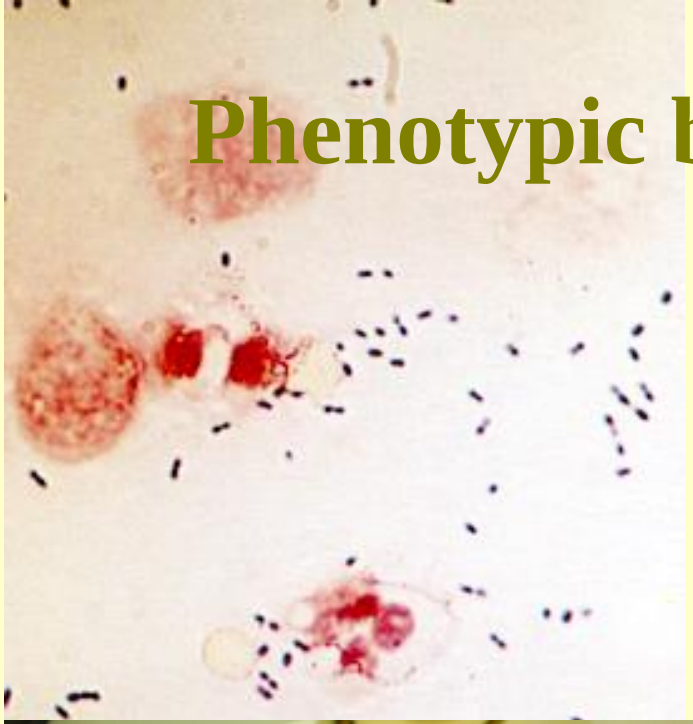


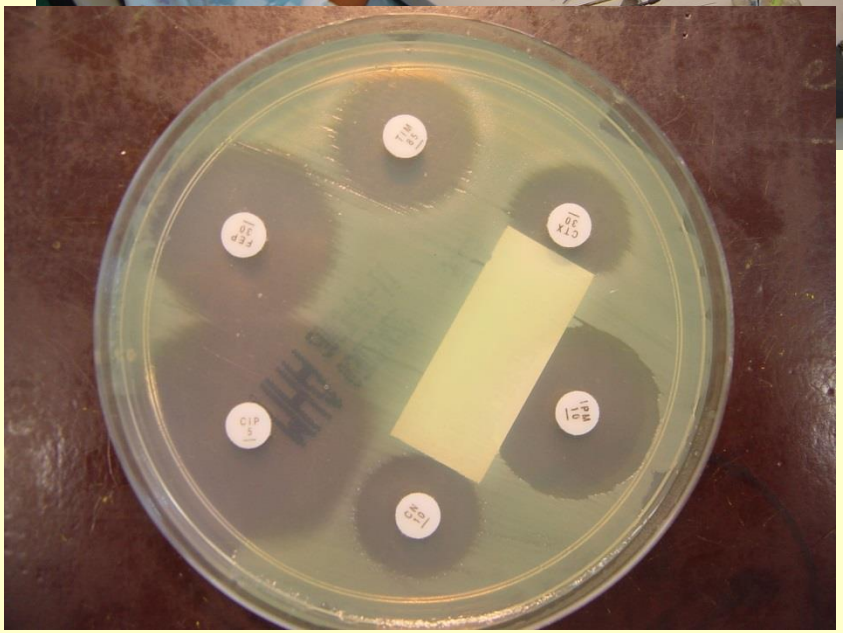


Molecular viral diagnosis

- Sensitive, specific, rapid detection
- Monitor antiviral therapy (viral loads)
- Distinguish active from latent infection (e.g. herpes virus family)
- Directly determine drug resistance genotypes (HIV, herpes simplex, cytomegalovirus)

Phenotypic bacteriological methods





Molecular bacterial diagnosis

Rapid quantitative detection of pathogens that are non-cultivatable or fastidious, slow growing organisms or indistinguishable from normal flora

Genotypic identification of bacteria/fungi for speciation or epidemiological typing

Detection of **virulence or antibiotic genetic determinants**

Molecular Approaches

1. Nucleic acid **extraction** (RNA or DNA)
   +/- reverse transcriptase process (viral RNA -> cDNA)
2. +/- nucleic acid **amplication** (duplication) (e.g. Polymerase chain reaction (PCR)
3. **Detection** of the nucleic acid product by-
   - agarose gel electrophoresis
   - nucleic acid probes
   - real-time assays
   - sequencing

1. Nucleic acid extraction

- Breaking open the cell/ nucleus

- Separation/ purification (DNA or RNA)
  - ethanol precipitation
  - spin columns
- +/- Assay

The Unusual Origin of the Polymerase Chain Reaction

A surprisingly simple method for making unlimited copies of DNA fragments was conceived under unlikely circumstances during a moonlit drive through the mountains of California

2. Amplification (PCR-nucleic acid amplification

by Kary B. Mullis

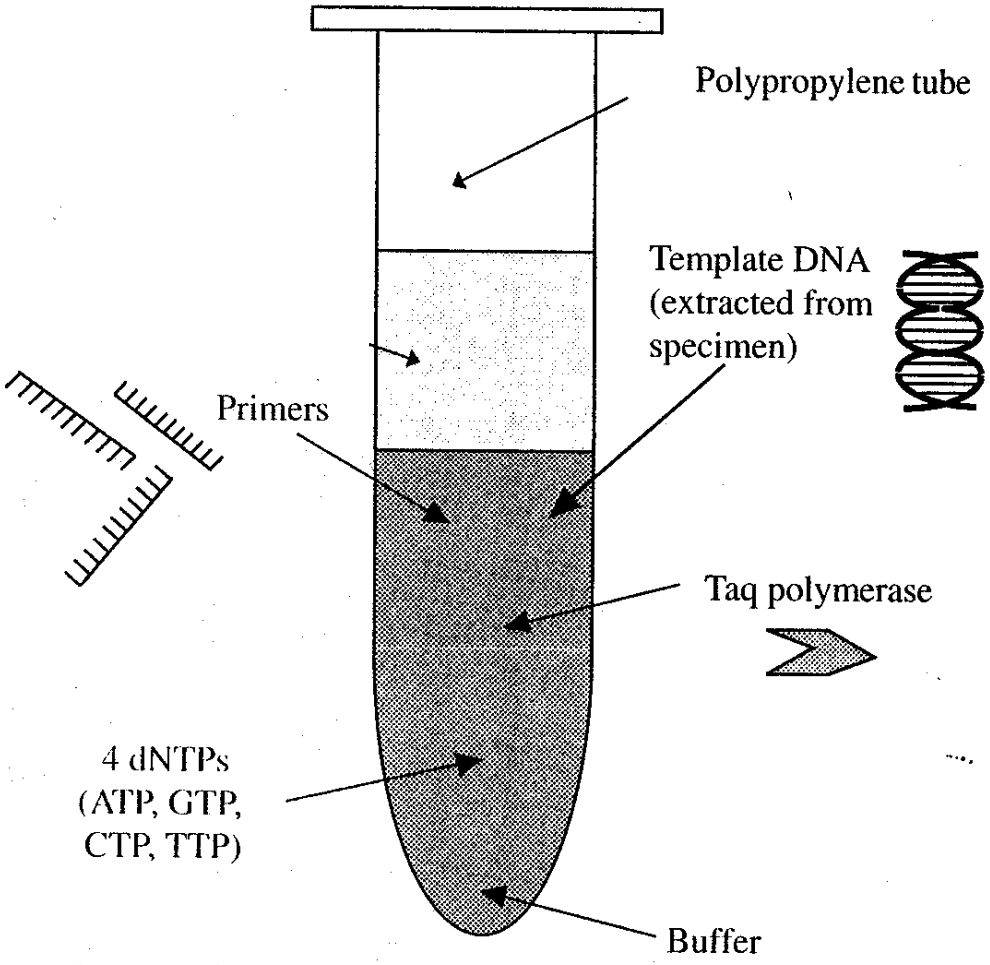
Sometimes a good idea comes to you when you are not looking for it. Through an improbable combination of coincidences, naiveté and lucky mistakes, such a revelation came to me one Friday night in April, 1983, as I gripped the steering wheel of my car and snaked along a moonlit mountain road into northern California's redwood country. That was how I stumbled across a process that could make unlimited numbers of copies of genes, a process now known as the polymerase chain reaction (PCR).

tal tissue specimen, from a single human hair, from a drop of dried blood at the scene of a crime, from the tissues of a mummified brain or from a 40,000-year-old woolly mammoth frozen in a glacier.

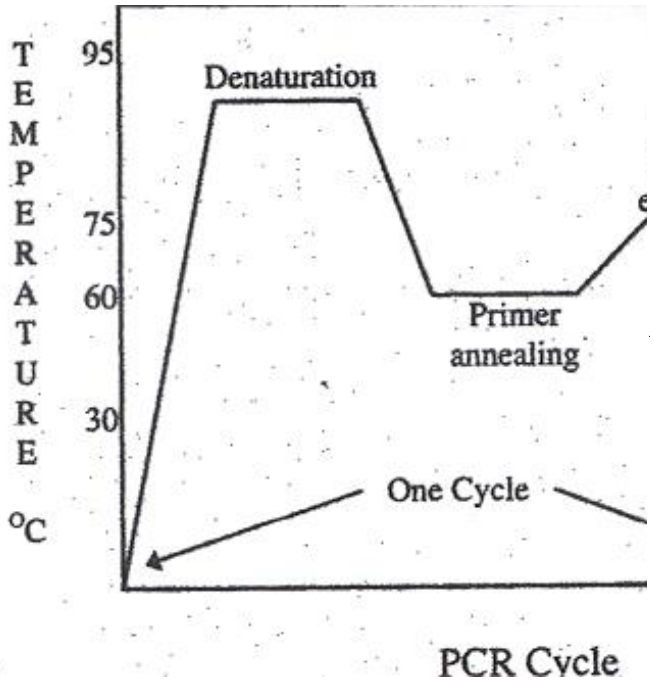
In the seven years since that night, applications for the PCR have spread throughout the biological sciences: more than 1,000 reports of its use have been published. Given the impact of the PCR on biological research and its conceptual simplicity, the fact that it lay unrecognized for more than 15

it at random points along its length. Consequently, if the DNA is removed from 1,000 identical cells, there will be 1,000 copies of any given gene, but each copy will be on a DNA fragment of differing length.

For years this problem made it difficult to study genes. Then in the 1970's enzymes known as restriction endonucleases were discovered: these enzymes snipped strands of DNA at specific points. The endonucleases made it possible to cut DNA into smaller, sturdier, more identifiable pieces and

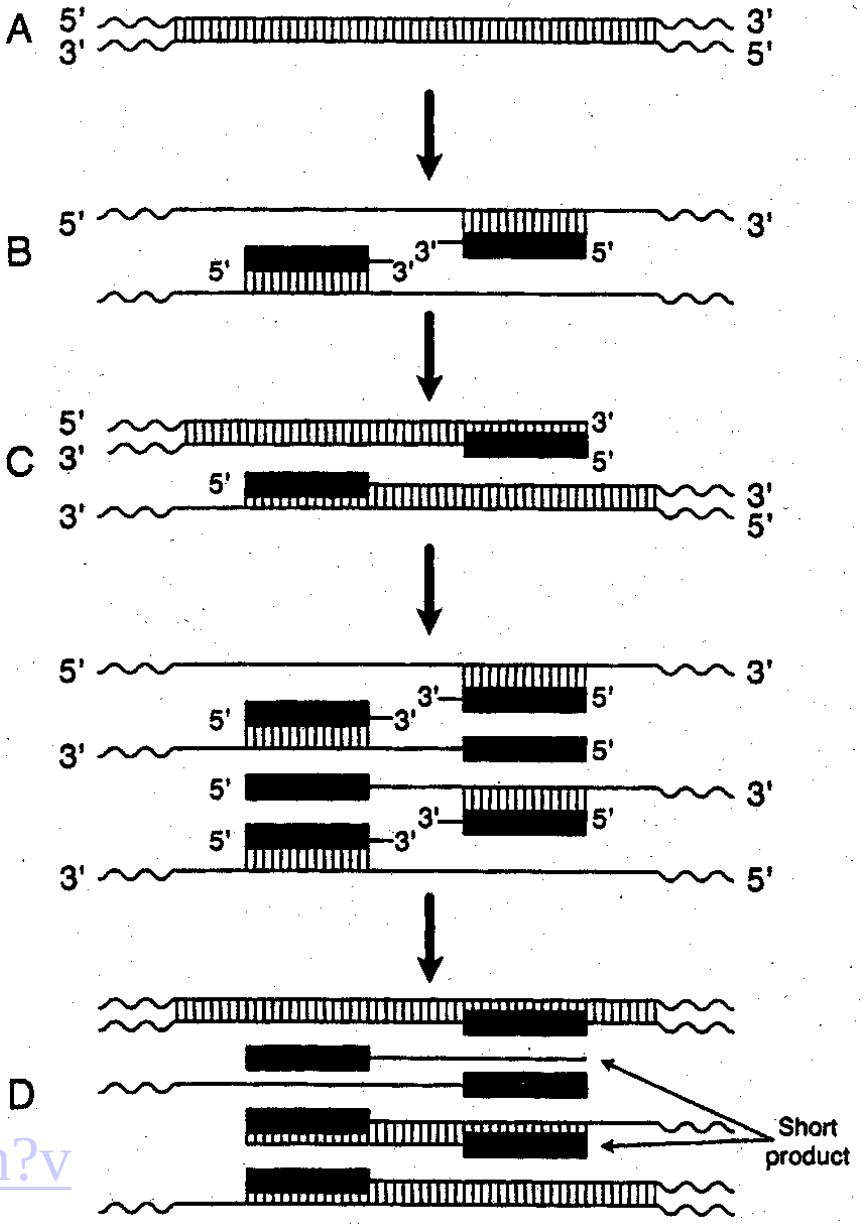


Constituents of PCR



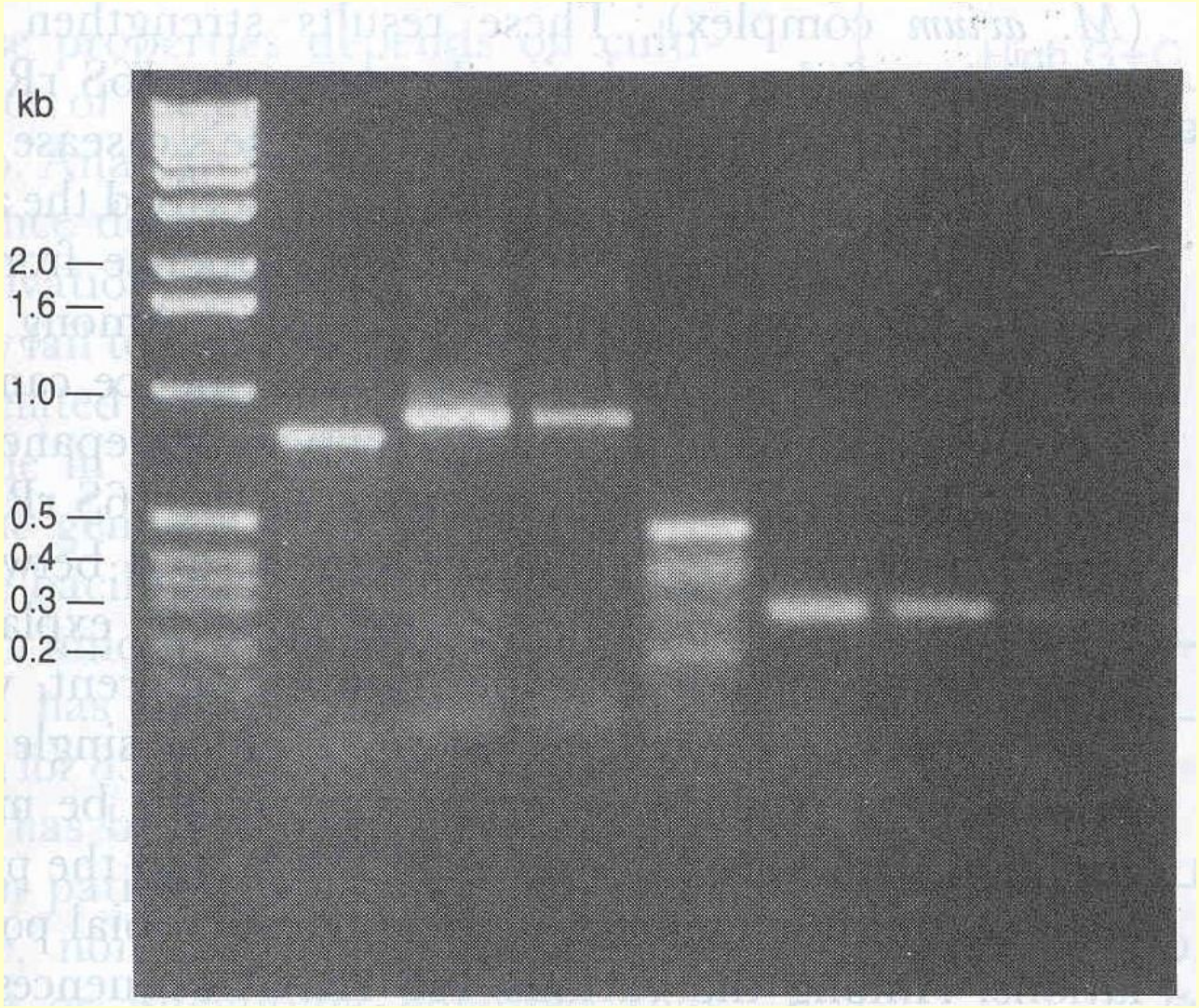
Polymerase chain reaction

Mullis-K 1985

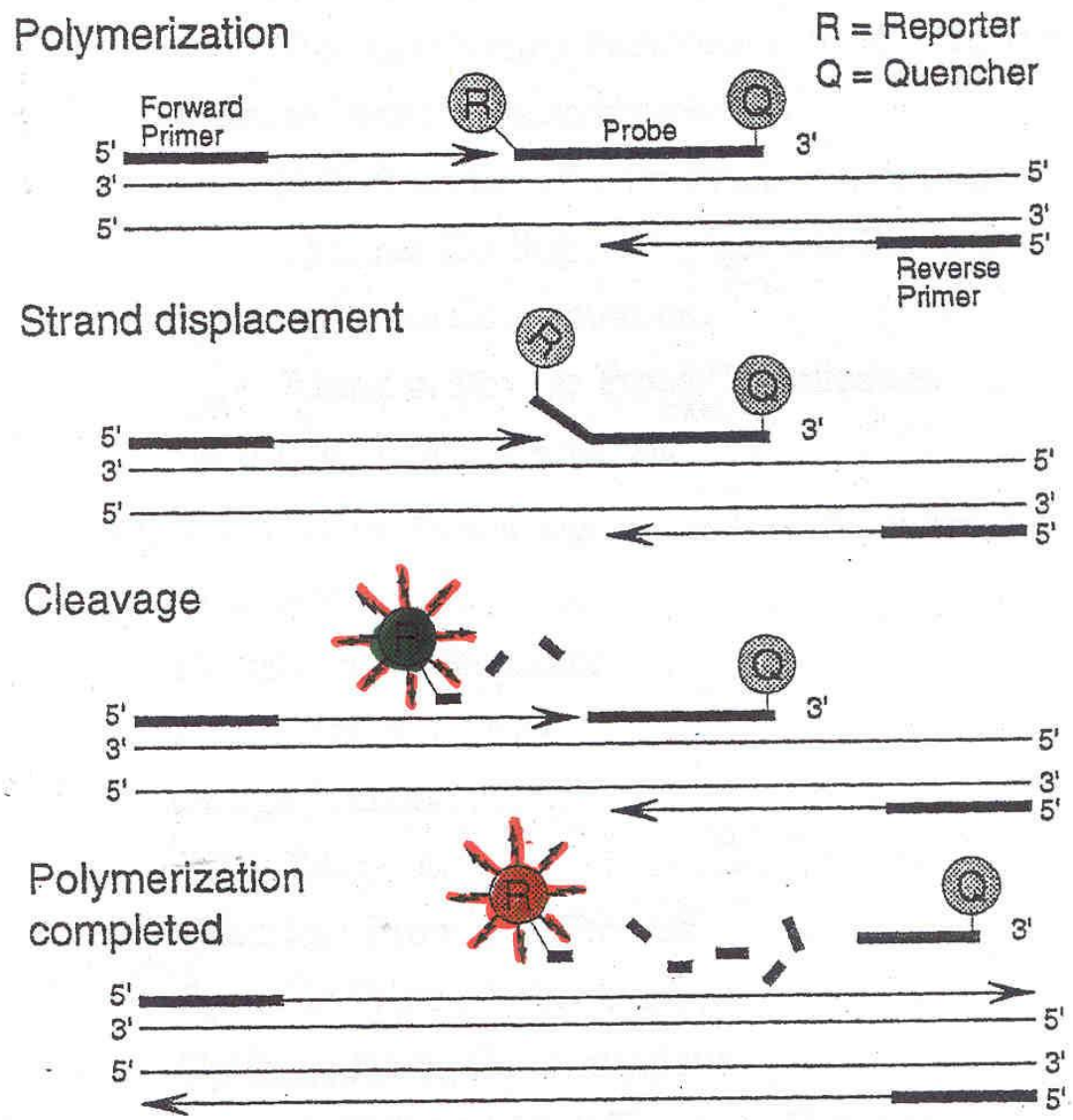


https://www.youtube.com/watch?v=JRAA4C2OPwg

3. Detection of product



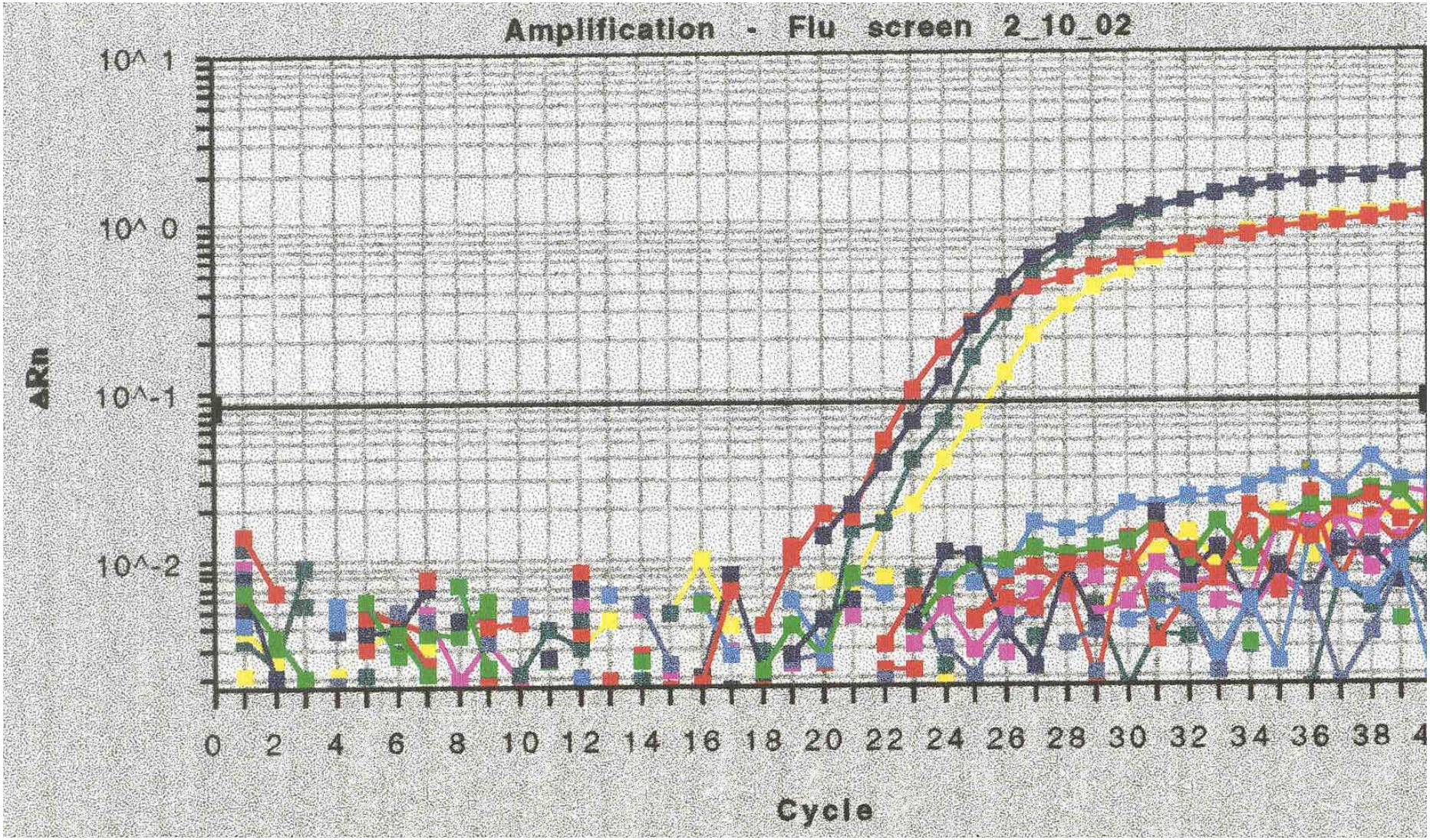
The old way…

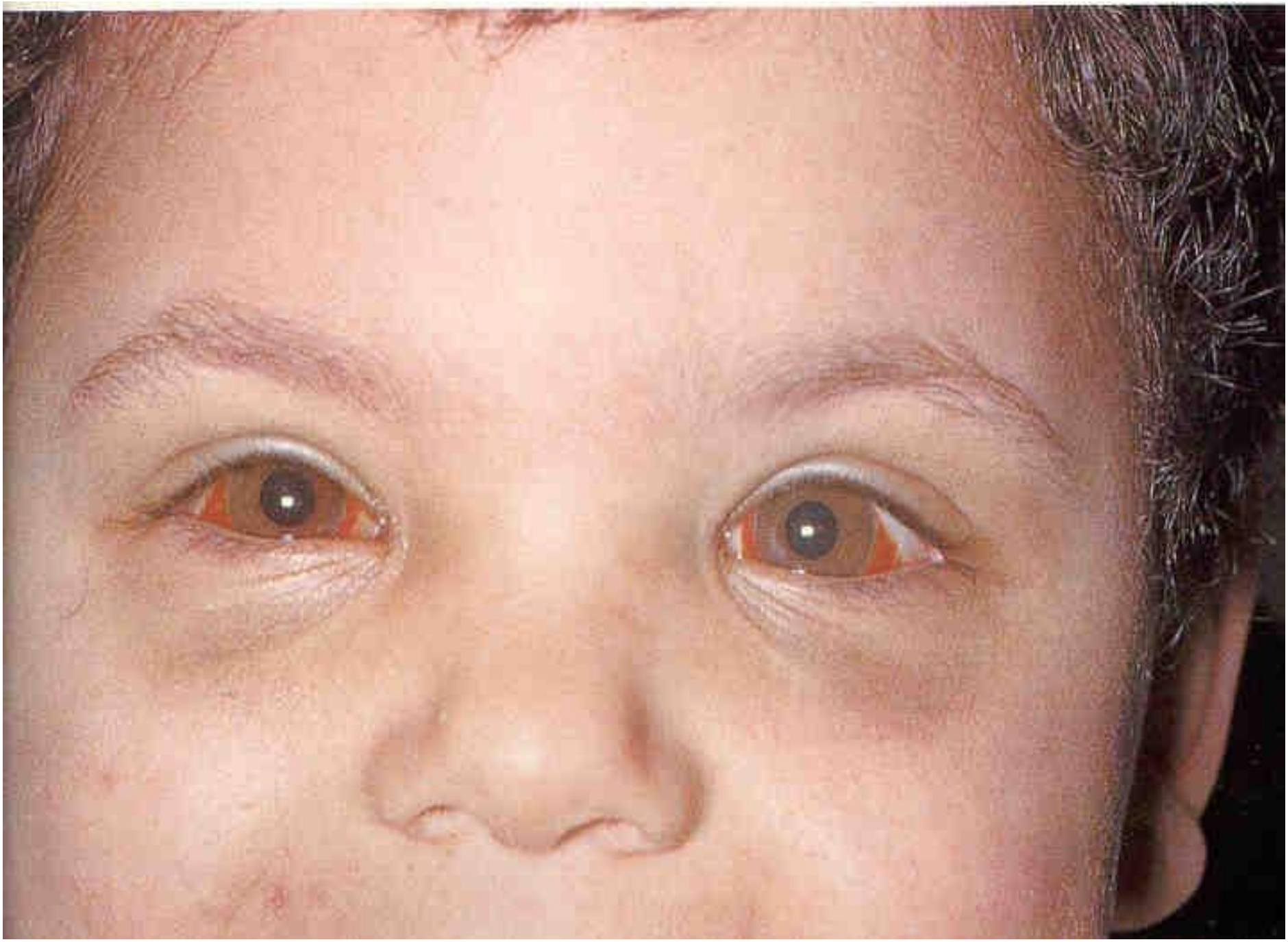


'Real-time' detection

'Taqman' probe

https://www.youtube.com/watch?v=edJEY5g4X8k


Amplification - Flu screen 2_10_02
10^ 1
10^ 0
10^-1
10^-2
ΔRn
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 4
Cycle

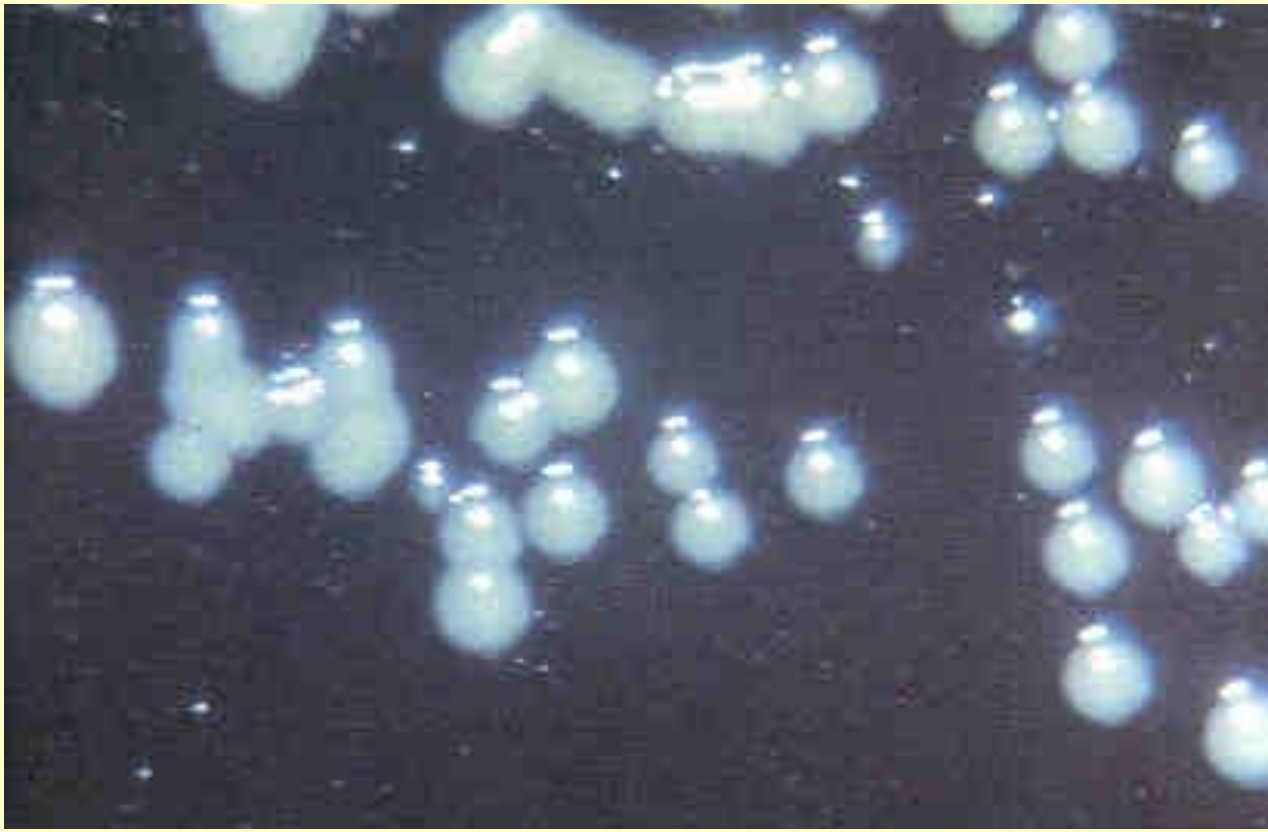


Example 1: Detection of pathogens

- Clinical
- Culture
- Serology

- **PCR**

Bordetella pertussis growing on charcoal medium



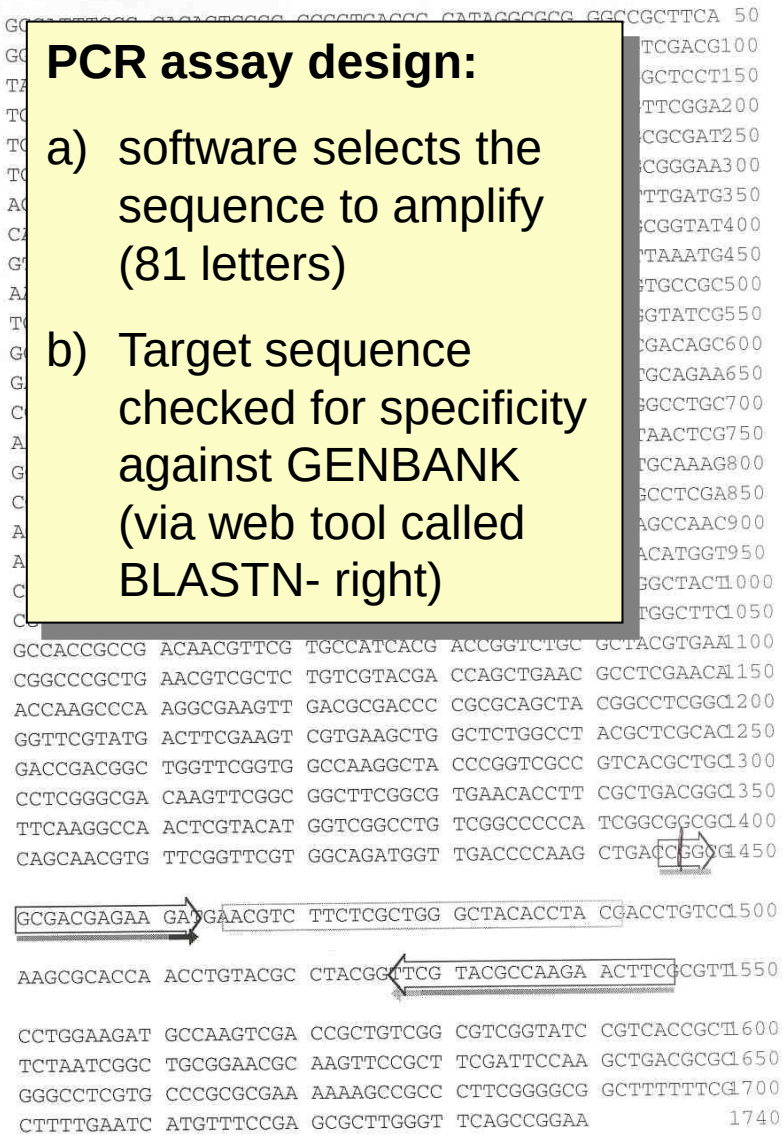
In-house pertussis real-time assay example

- Simple extraction
- Target – Insertion sequence (IS) 481 (*B. pertussis*, *B. holmesii* multiplexed with IS 1001 (*B. parapertussis*)
- Same day result
- Internal amplification control to ensure:
  - human sample present – a conserved human gene sequence is selected
  - Confirms that no inhibition of PCR reaction has occurred

PCR assay design:

a) software selects the sequence to amplify (81 letters)

b) Target sequence checked for specificity against GENBANK (via web tool called BLASTN- right)



BLASTN 2.2.1 [Apr-13-2001]

Reference:
Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

RID: 993103336-9608-4469

Query=
(81 letters) pi amplicon

Database: nt
875,504 sequences; 3,367,259,498 total letters

If you have any problems or questions with the results of this search please refer to the **BLAST FAQs**

Taxonomy reports

Distribution of 6 Blast Hits on the Query Sequence

Mouse-over to show defline and scores. Click to show alignments

Sequences producing significant alignments:	Score (bits)	E Value
gi|48877|emb|X58488.1|PBPORING B.pertussis gene for a porin...	161	7e-38
gi|10581086|gb|AE005072.1|AE005072 Halobacterium sp. NRC-1 ...	38	0.68
gi|11095071|gb|AC084622.1|CBRG44P20 Caenorhabditis briggsae...	38	0.68
gi|9949809|gb|AE004785.1|AE004785 Pseudomonas aeruginosa PA...	36	2.7
gi|211223|gb|M26064.1|CHKATPCA Chicken Ca2+ ATPase, complet...	36	2.7
gi|4520373|dbj|AB024601.1|AB024601 Pseudomonas aeruginosa d...	36	2.7

Alignments

>gi|48877|emb|X58488.1|PBPORING B.pertussis gene for a porin protein
Length = 1740

Score = 161 bits (81), Expect = 7e-38
Identities = 81/81 (100%)
Strand = Plus / Plus

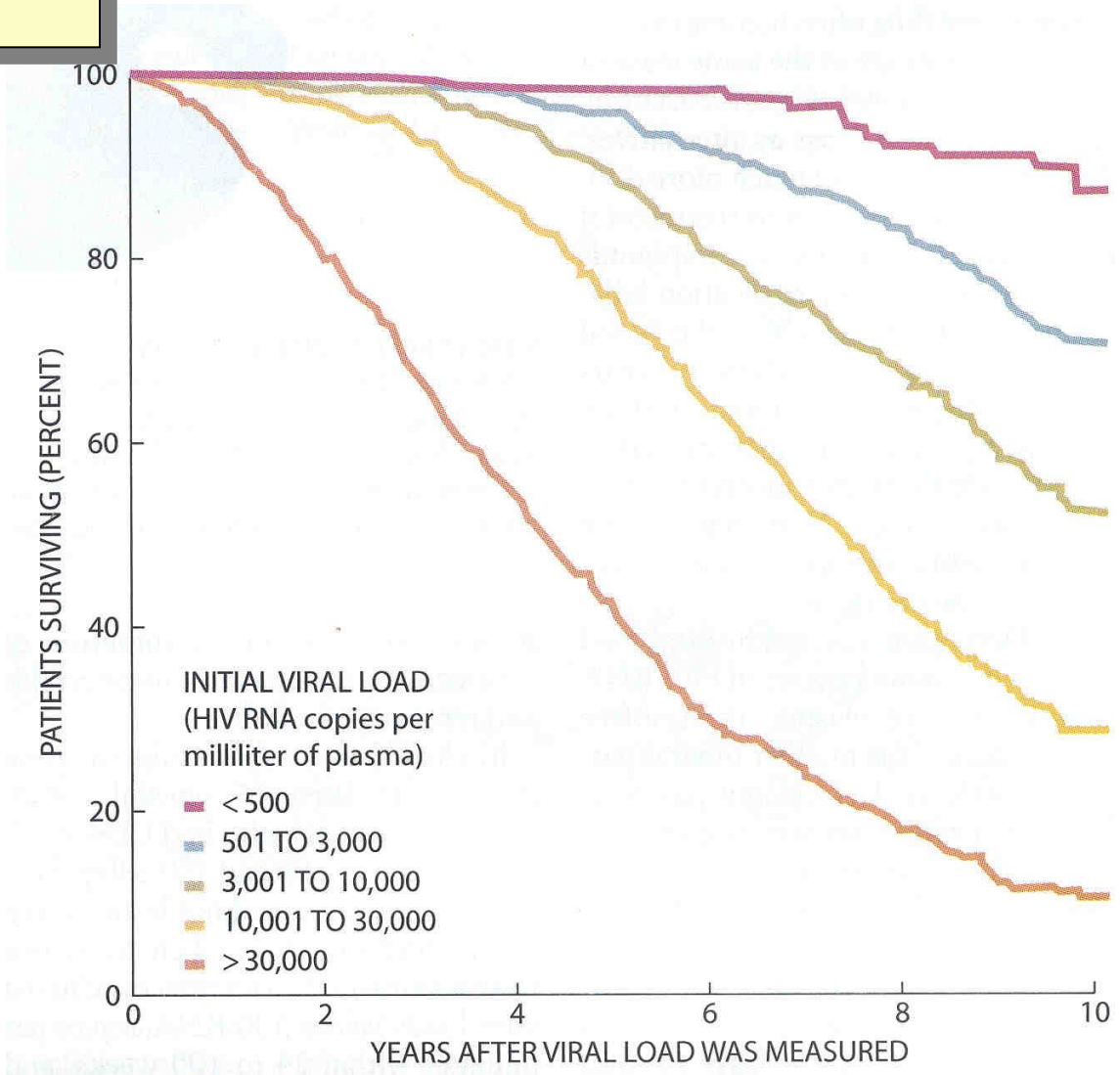
Pertussis assay validation

- Sensitivity: culture assay with dilution series : estimated < 10 bacteria for a positive result
- Specificity: assay tested against many other potential species that might be encountered in resp samples – e.g.
  - *B. pertussis* isolates (15)
  - *B. parapertussis* (1+)
  - *B. bronchiseptica*
  - *B. holmesii*
  - Other bacteria and viruses that may be encountered

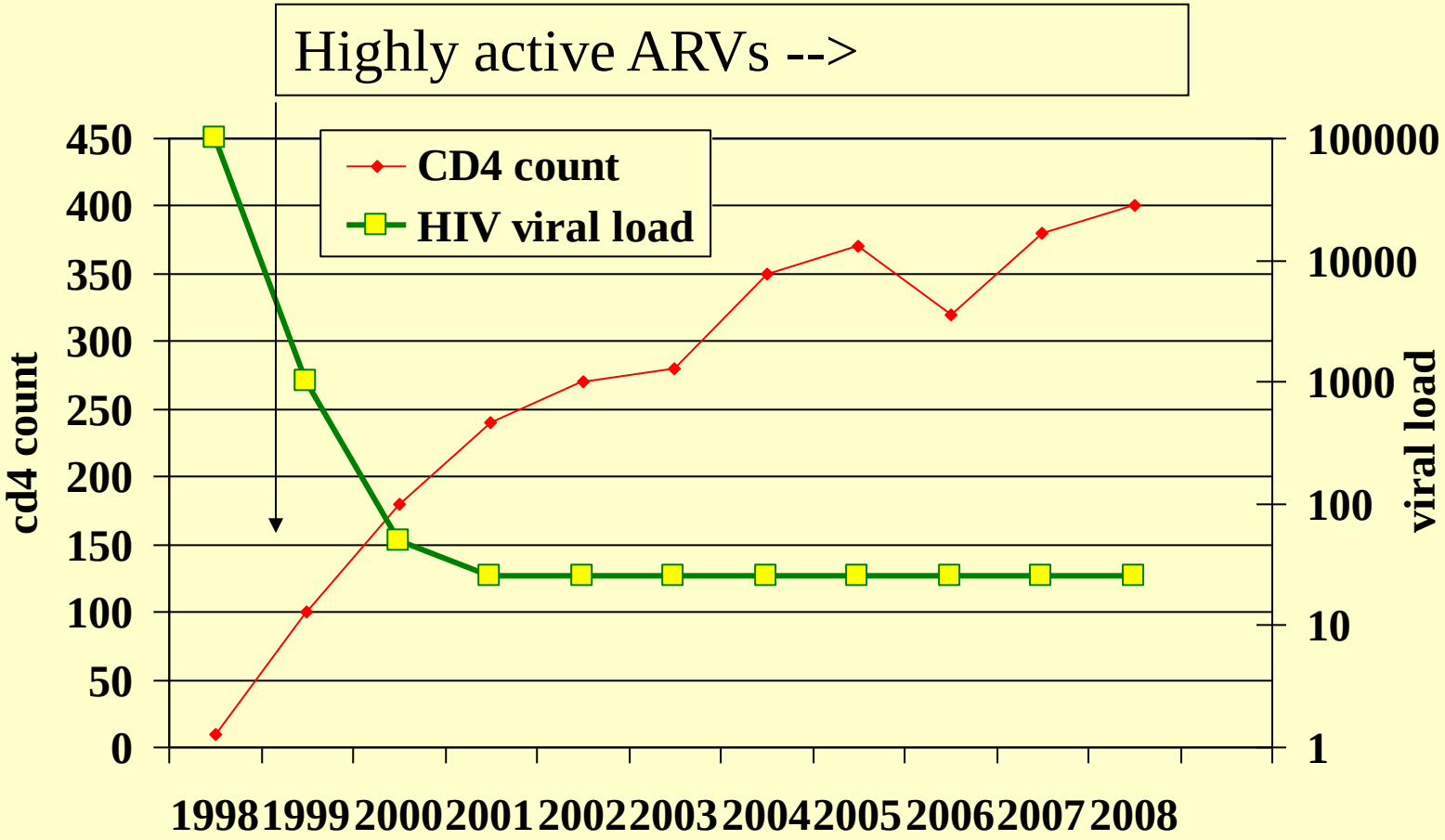
Example 2: Quantification and genotyping

Mellors et al 1998 Scientific American

Survival at 10 years: 1,604 untreated HIV-infected men



HIV management example



HIV resistance genotyping by sequencing

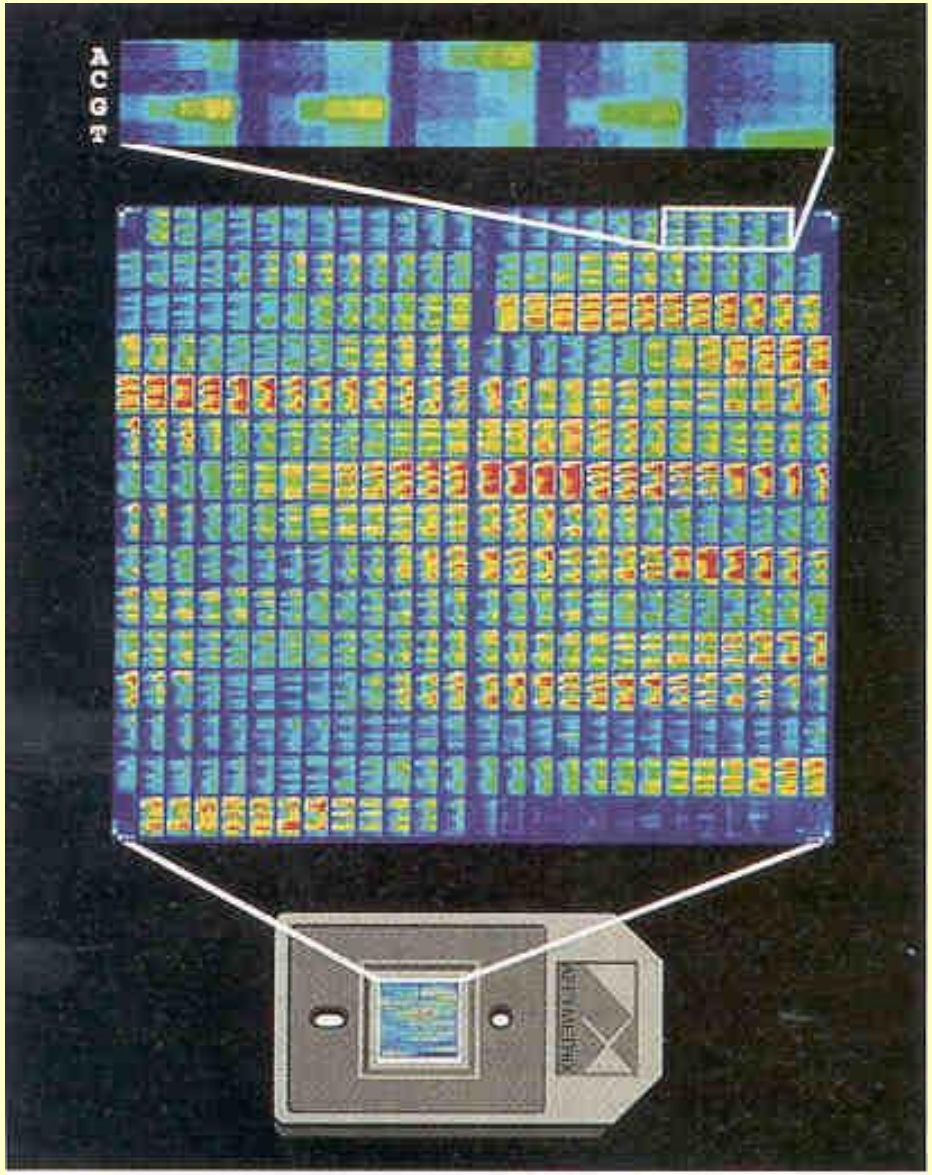
PROTEASE INHIBITORS

DRUG	**Codons implicated in drug resistance** (codons 1-99 sequenced)		
	PRIMARY (major)#	SECONDARY (minor)^	Genotypic Interpretation
Amprenavir	M46I, I50V	none	Resistant
Indinavir	M46I, I50V	none	Resistant
Lopinavir	M46I, I50V	none	Resistant
Nelfinavir	M46I, I50V	none	Resistant
Ritonavir	M46I, I50V	none	Resistant
Saquinavir	M46I, I50V	none	Resistant

REVERSE TRANSCRIPTASE INHIBITORS

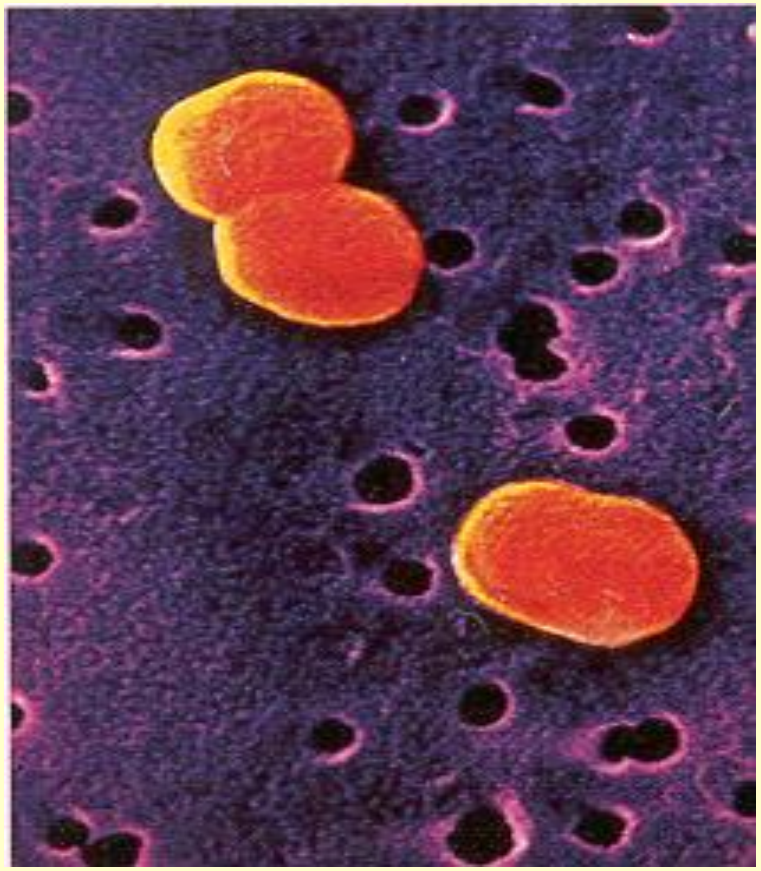
DRUG	**Codons implicated in drug resistance** (codons 1-324 sequenced)		
	PRIMARY (major)	SECONDARY (minor)	Genotypic Interpretation
Abacavir	L74V, Y115F, M184V	T215F	Resistant
Didanosine	L74V	M184V	Resistant
Lamivudine	M184V	none	Resistant
Stavudine	none	T215F	Possibly Resistant
Zalcitabine	L74V	M184V	Resistant
Zidovudine	T215F	none	Resistant
Delavirdine	K103N, Y181C	none	Resistant
Efavirenz	K103N	Y181C	Resistant
Nevirapine	K103N, Y181C	none	Resistant

Multiple probe assays -> 50000 unique probes on a small slide- can detect all relevant drug resistance mutations in one reaction



GENE CHIP FOR HIV
(bottom) *contains thousands of unique DNA probes* (center), *each of which glows* (top) *when a matching sequence is detected.*

Example 3: 16S ribosomal RNA amplification and sequencing

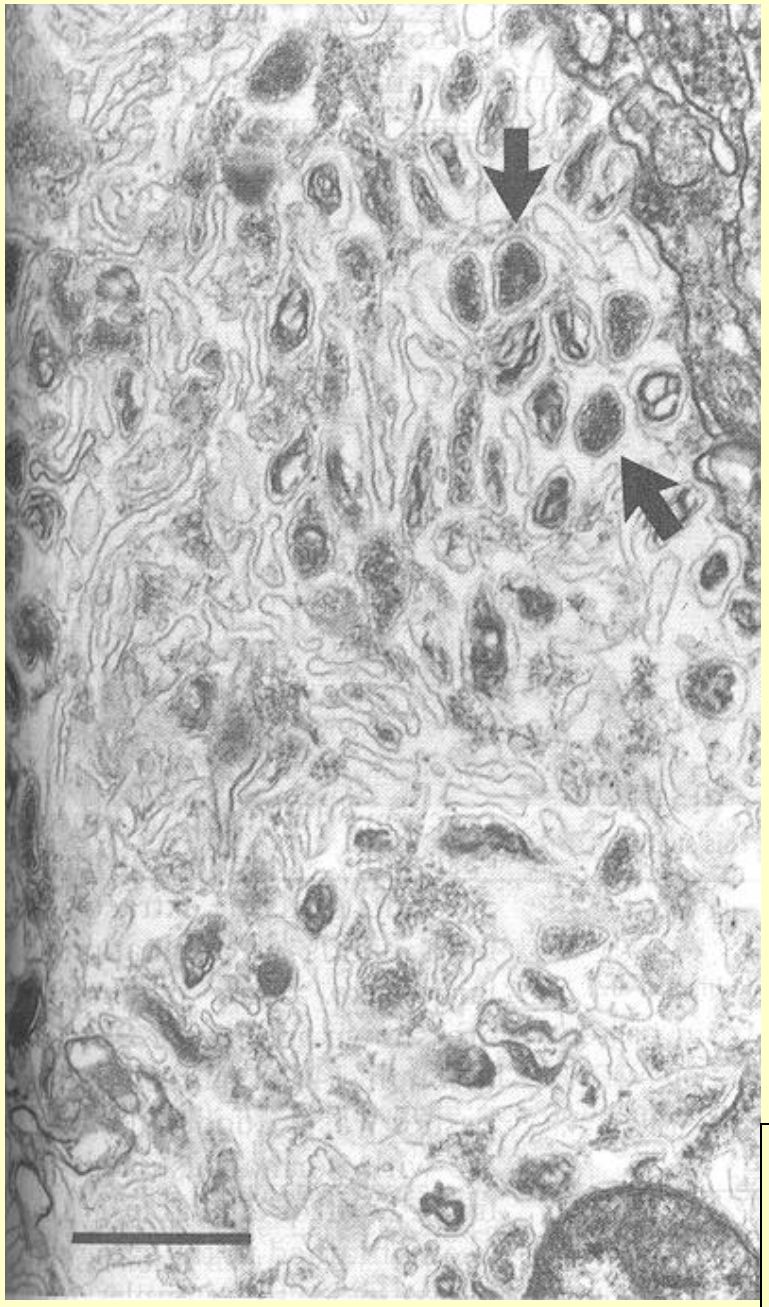


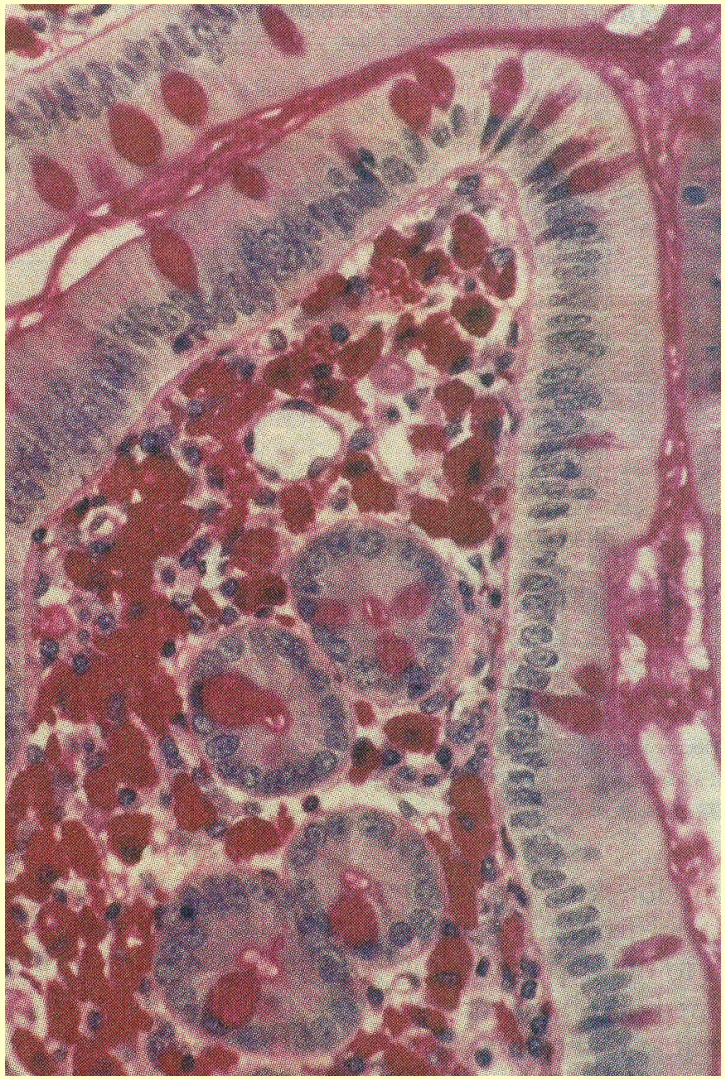
VK, 66 yr female

- Ill for 3 months with high swinging fevers
- steady weight loss over same period
- change in bowel habit to 4-5 loose stools/day from 2/d. No blood.
- No abdominal pain or arthritis. No other symptoms.
- Examination normal

VK (2) Investigations

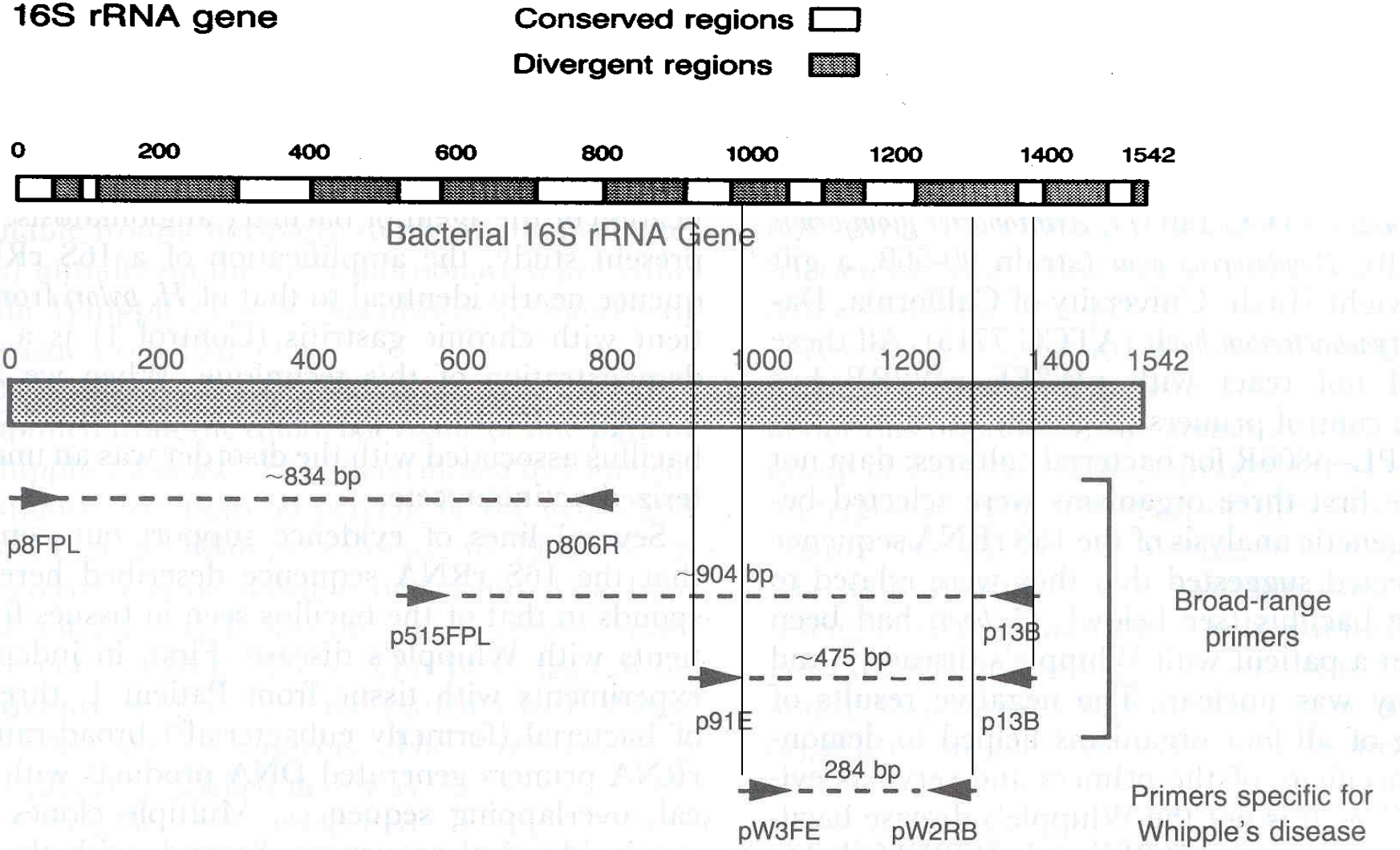
- ESR 100, CRP 137, Hb 101(NN), plt 524
- Gallium and white cell scan normal
- upper and lower endoscopies macroscopically normal …..





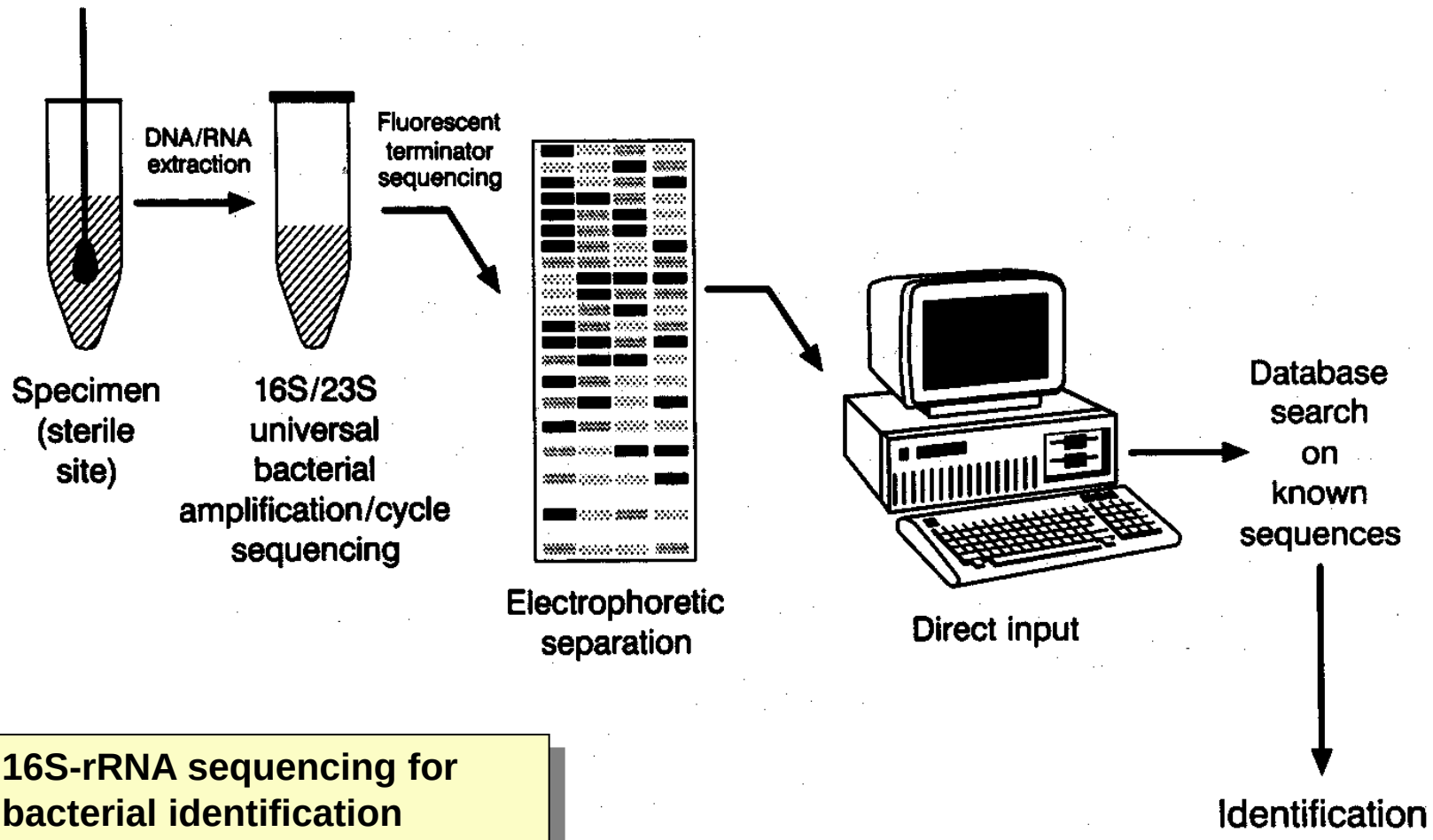
EM x 33,000 and PAS x400

- PAS positive inclusions

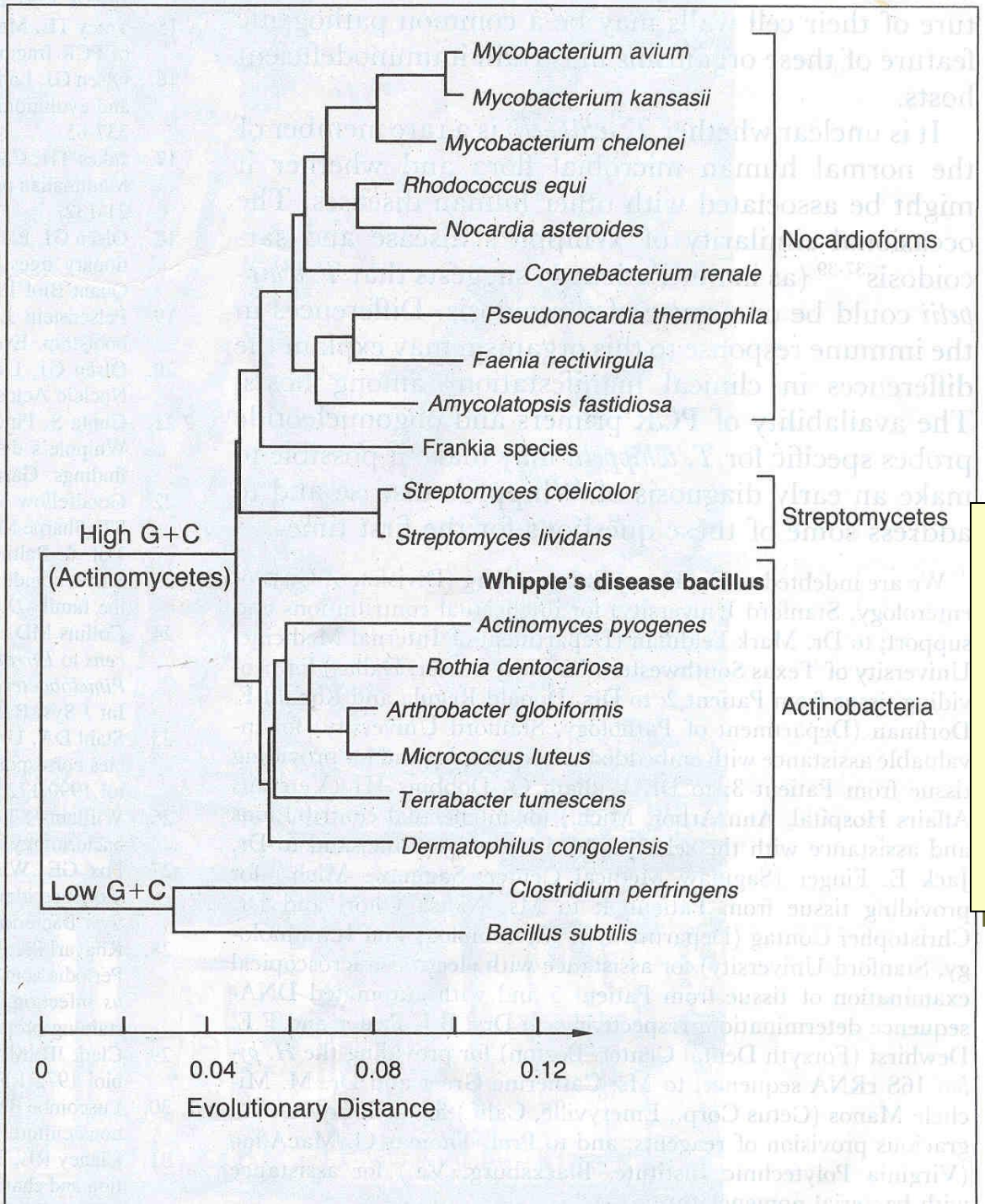


16-S sequencing- PCR primers match conserved regions and flank variable regions that have signature sequences that can identify the bacterial species by comparison with a known database of sequences (GENBANK)

Relman et al. NEJM 1992:327, 293-301.



16S-rRNA sequencing for bacterial identification



***Trophyrema whipellii* – identified originally only by sequencing**

Other applications of 16-s rRNA sequencing

- Identification of unusual bacteria and new species
- Mycobacterial identification
- Fungal identification (18-s rRNA)
- Taxonomic studies
- Rapid detection of multiple bacterial pathogens in blood, tissue or normally sterile fluids – “next generation sequencing” used for this

J Vasc Surg. 2006 Nov;44(5):1055-60.

Bacterial diversity in aortic aneurysms determined by 16S ribosomal RNA gene analysis.

Marques da Silva R, Caugant DA, Eribe ER, Aas JA, Lingaas PS, Geiran O, Tronstad L, Olsen I.

BACKGROUND: Aortic aneurysms are common vascular conditions that cause considerable morbidity and mortality. Understanding of the mechanisms involved in the pathogenesis of the condition remains limited. Recently, infection has been suggested as possible contributor in the development of the disease. The aim of the present study was to examine aortic aneurysms for the presence of bacterial DNA using polymerase chain reaction (PCR) targeting the 16S ribosomal RNA (rRNA) gene, followed by cloning and sequencing. METHODS: Universal eubacterial primers were used to amplify 16S rRNA bacterial genes in 10 specimens from arterial walls of aortic aneurysms. Subsequently, PCR amplicons were cloned into Escherichia coli and sequencing of the cloned inserts was used to determine species identity or closest relatives by comparison with known sequences in GenBank.

RESULTS: Sequences of Stenotrophomonas spp., including S. maltophilia (formerly Pseudomonas homology group V) were detected in six aneurysm samples. Propionibacterium acnes was identified in five samples, and Brevundimonas diminuta (formerly P. diminuta) in four samples. Other species previously assigned to the Pseudomonas genus such as Comamonas testosteroni, Delftia acidovorans, Burkholderia cepacia, Herbaspirillum sp., and Acidovorax sp. were also detected. Some clones fell into other environmental species, including Methylobacterium sp. and Bradyrhizobium elkanii, and others represented bacteria that have not yet been cultivated. DNA sequences from oral bacteria, including Streptococcus sanguinis, Tannerella forsythia, and Leptotrichia buccalis were detected. Sequences from Prevotella melaninogenica and Lactobacillus delbrueckii, which are commonly found in both mouth and gastrointestinal tract, were also detected. Additional species included Dermacoccus spp. and Corynebacterium vitaeruminis.

CONCLUSIONS: A wide variety of bacteria, including oral bacteria, was found to colonize aortic aneurysms and may play a role in their development. Several of these microorganisms have not yet been cultivated. CLINICAL RELEVANCE: Although Chlamydophila pneumoniae has been detected in aneurysmal walls, its exact role in the condition remains inconclusive. Overall, there is scarce information about the role of microorganisms in aneurysmal disease. In the present study, we used molecular genetics to detect a diversity of bacteria in arterial walls of aortic aneurysms. The presence of multiple microorganisms in aneurysmal disease may have implications for chemoprophylaxis and antibiotic treatment if directed only at C.pneumoniae.

illumina

APPLICATIONS | SYSTEMS | CLINICAL | SERVICES | SCIENCE

Systems / **MiSeq Benchtop Sequencer**

Overview

NEW. MORE READS. LONGER READ LENGTHS.

Cluster Generation and Sequencing

MISEQ REAGENT KIT V2

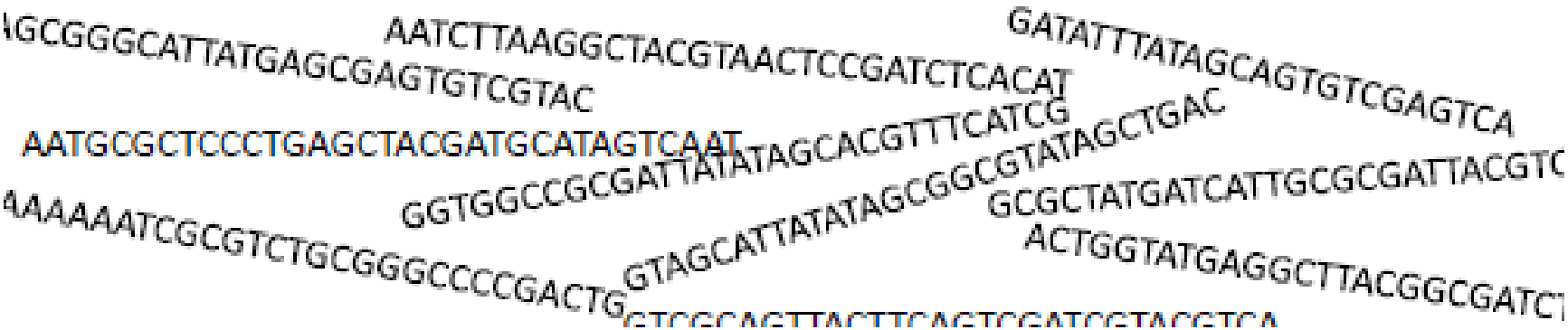
READ LENGTH	TOTAL TIME*	OUTPUT
1 × 36 bp	~4 hrs	540-610 Mb
2 × 25 bp	~5.5 hrs	750-850 Mb
2 × 150 bp	~24 hrs	4.5-5.1 Gb
2 × 250 bp	~39 hrs	7.5-8.5 Gb

http://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-you-will-learn/what-next-generation-dna-

What is it?

- All next generation sequencing techniques produce millions of short fragments of DNA sequence from the ~4.3Mbp *C. difficile* genome, from which the "whole" genome must be reconstructed
  - different sequencing platforms have different properties, advantages and limitations; all produce this kind of data

Sarah Walker, ECCMID 2013

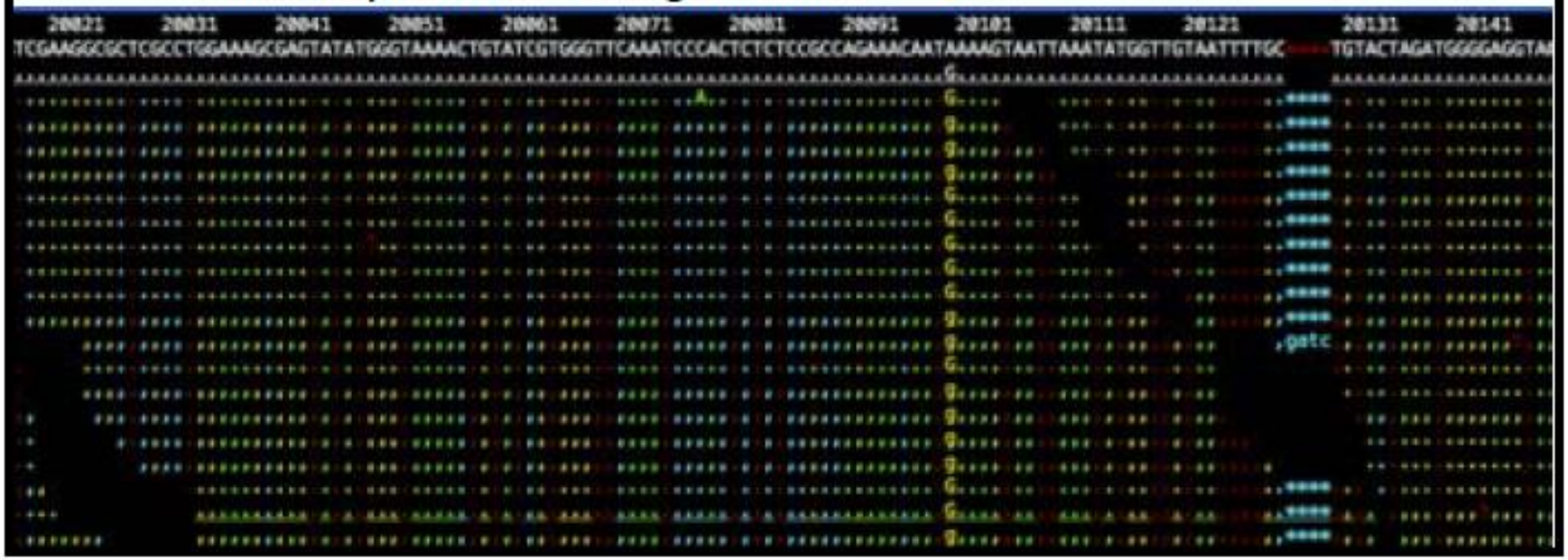


Strategies for whole genome reconstruction

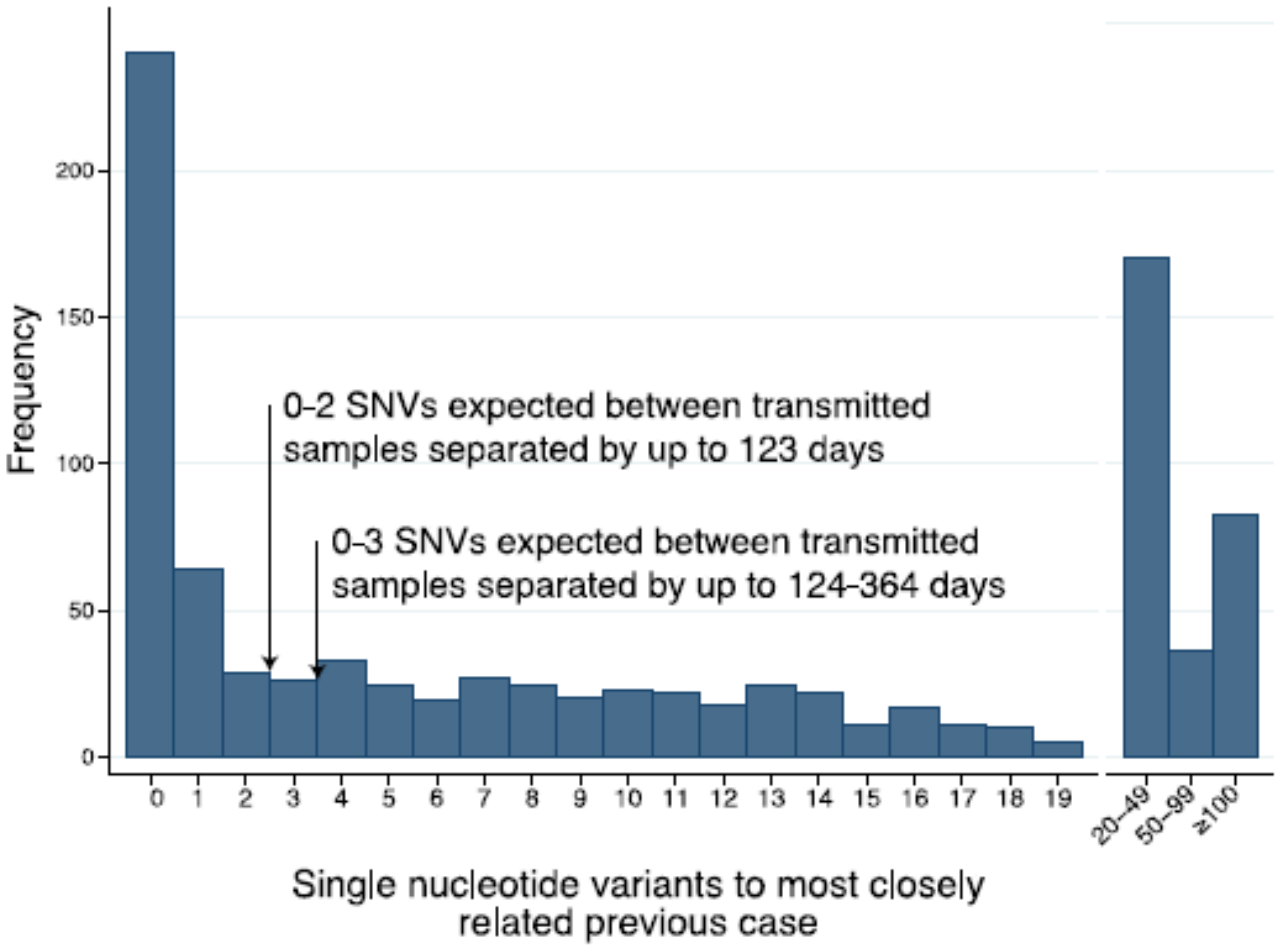
- Reference-based mapping
  - sequence data aligned to a specified reference Sanger sequence (eg CD630, CD196)
  - identify **high quality variants** (single nucleotide variants, insertions, deletions) versus this reference
  - non-repetitive "core genome"
- *De novo* assembly
  - sequence data aligned to itself
  - identify presence/absence of specific mobile genetic elements
  - "non-core/accessory genome"
- Classic tension between sensitivity/specificity

Strategies for whole genome reconstruction

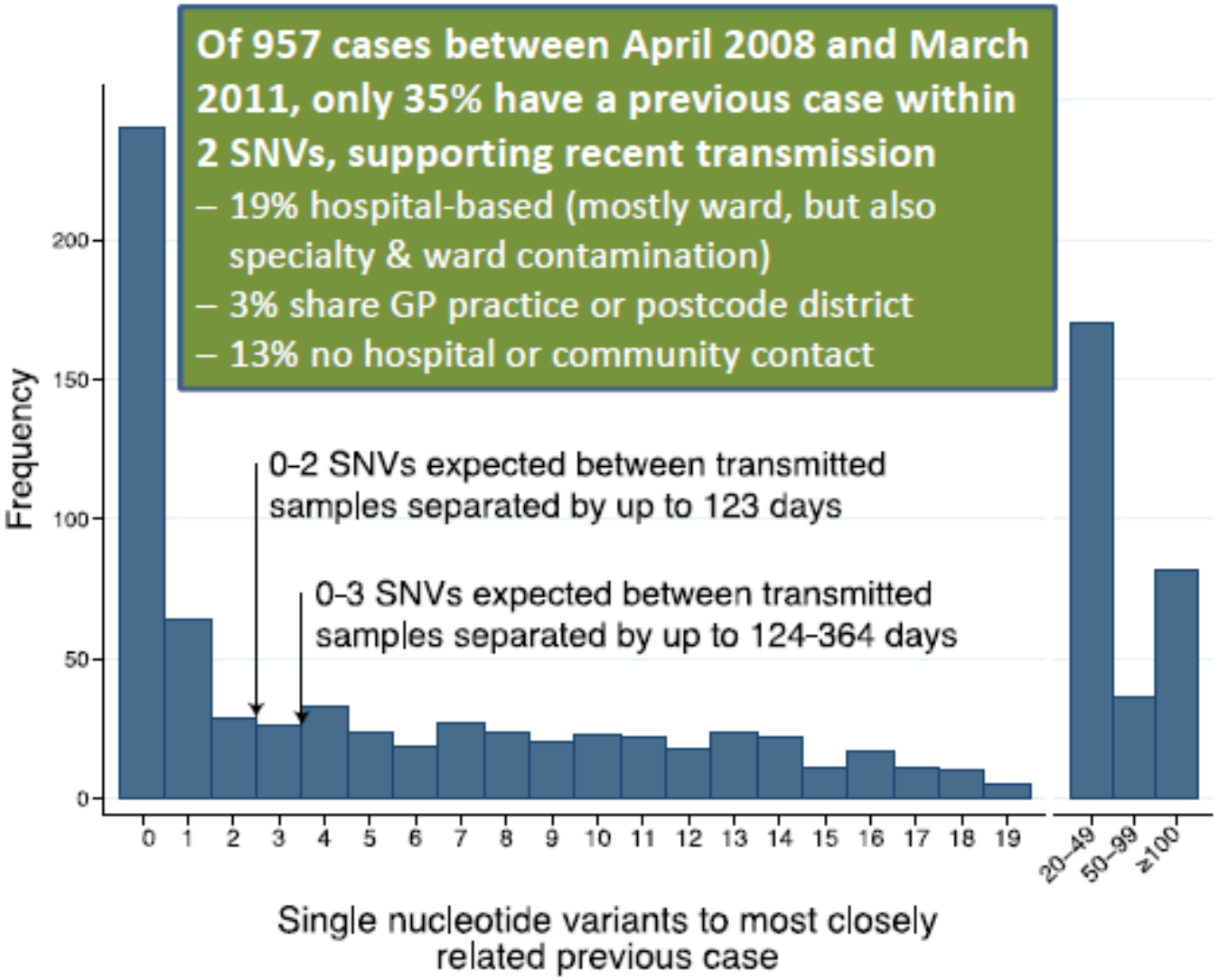
- Reference-based mapping
  - sequence data aligned to a specified reference Sanger sequence (eg CD630, CD196)
  - identify **high quality variants** (single nucleotide variants, insertions, deletions) versus this reference
  - non-repetitive "core genome"



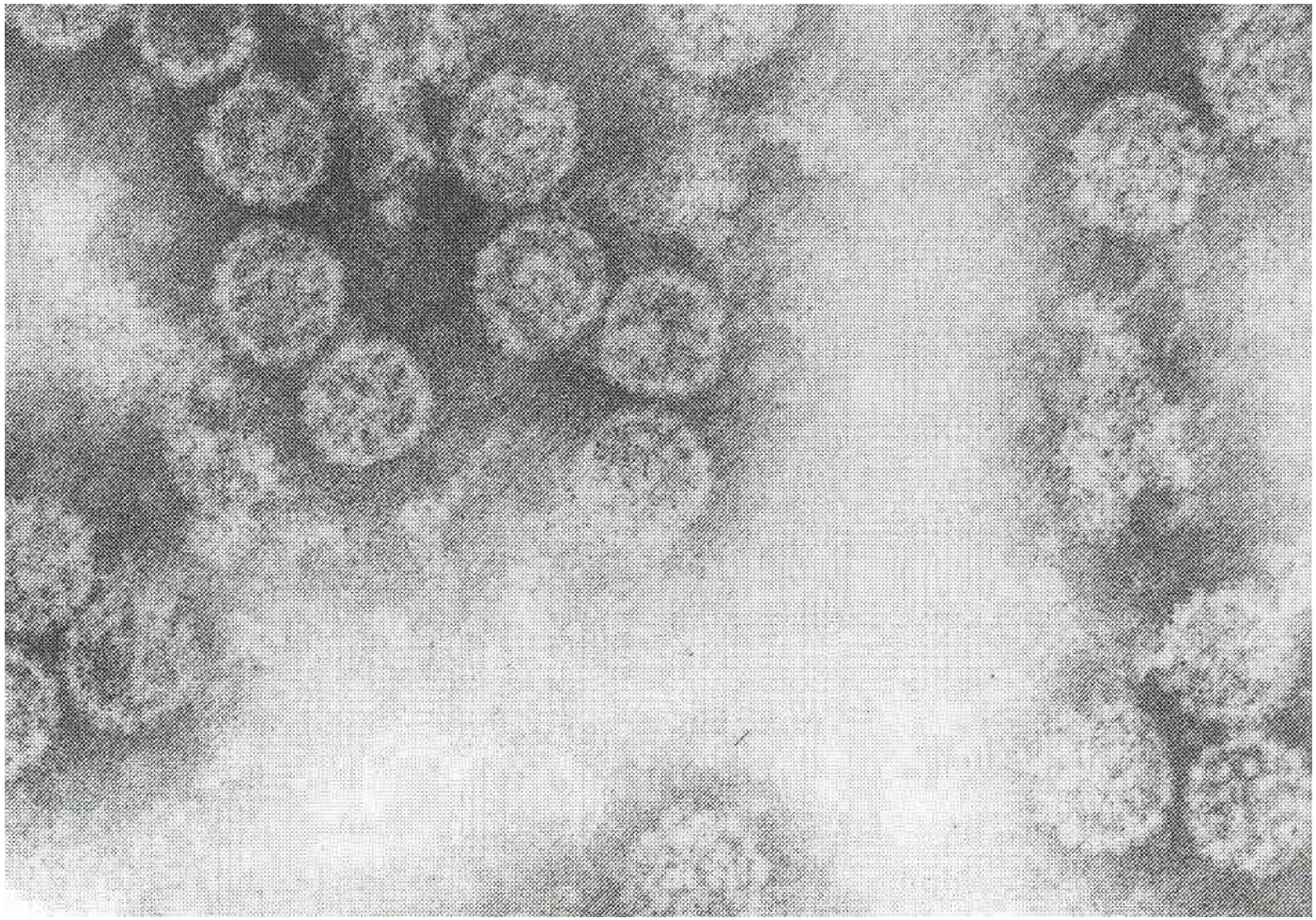
Distribution of SNVs from each case to closest previous case



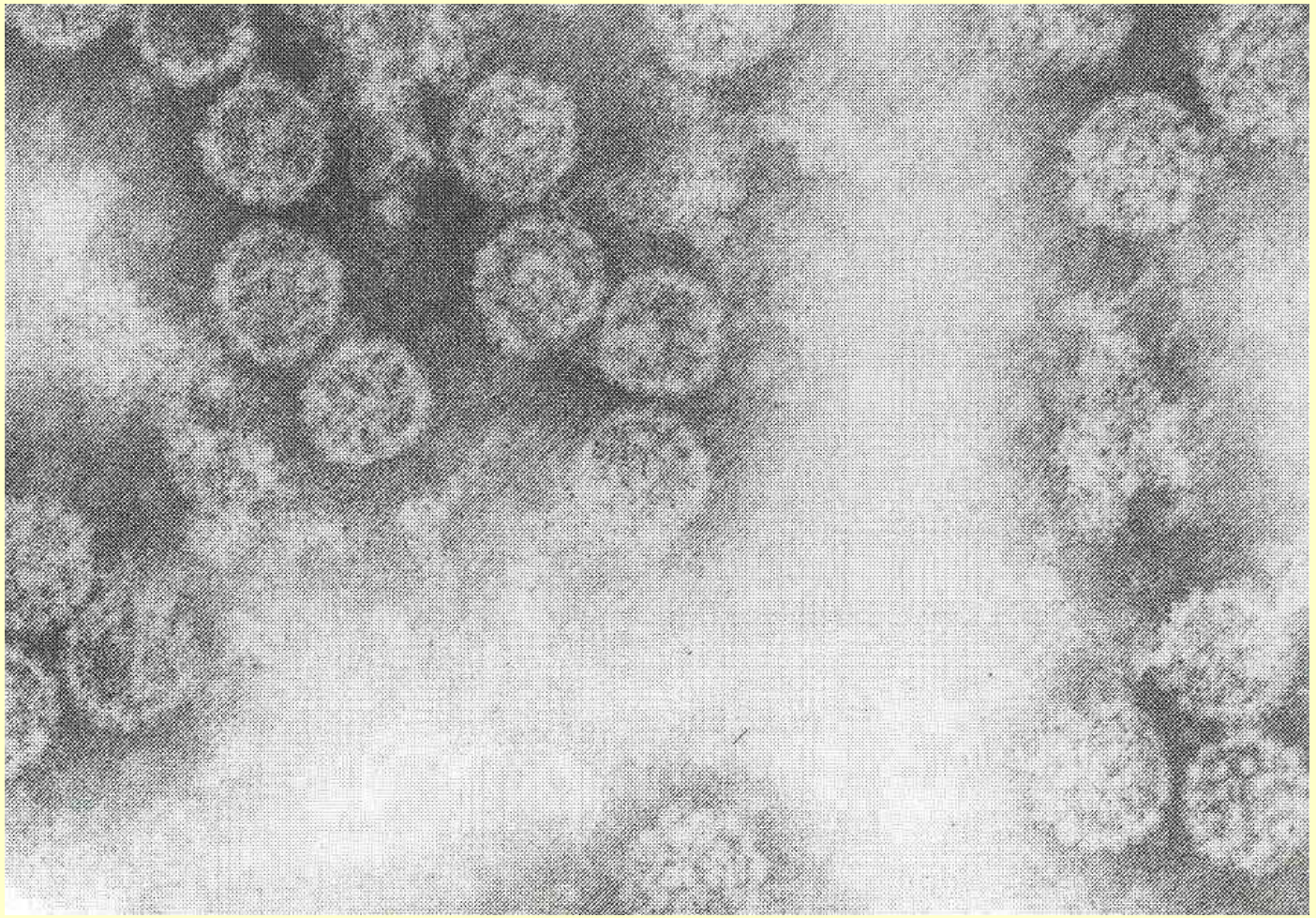
Distribution of SNVs from each case to closest previous case

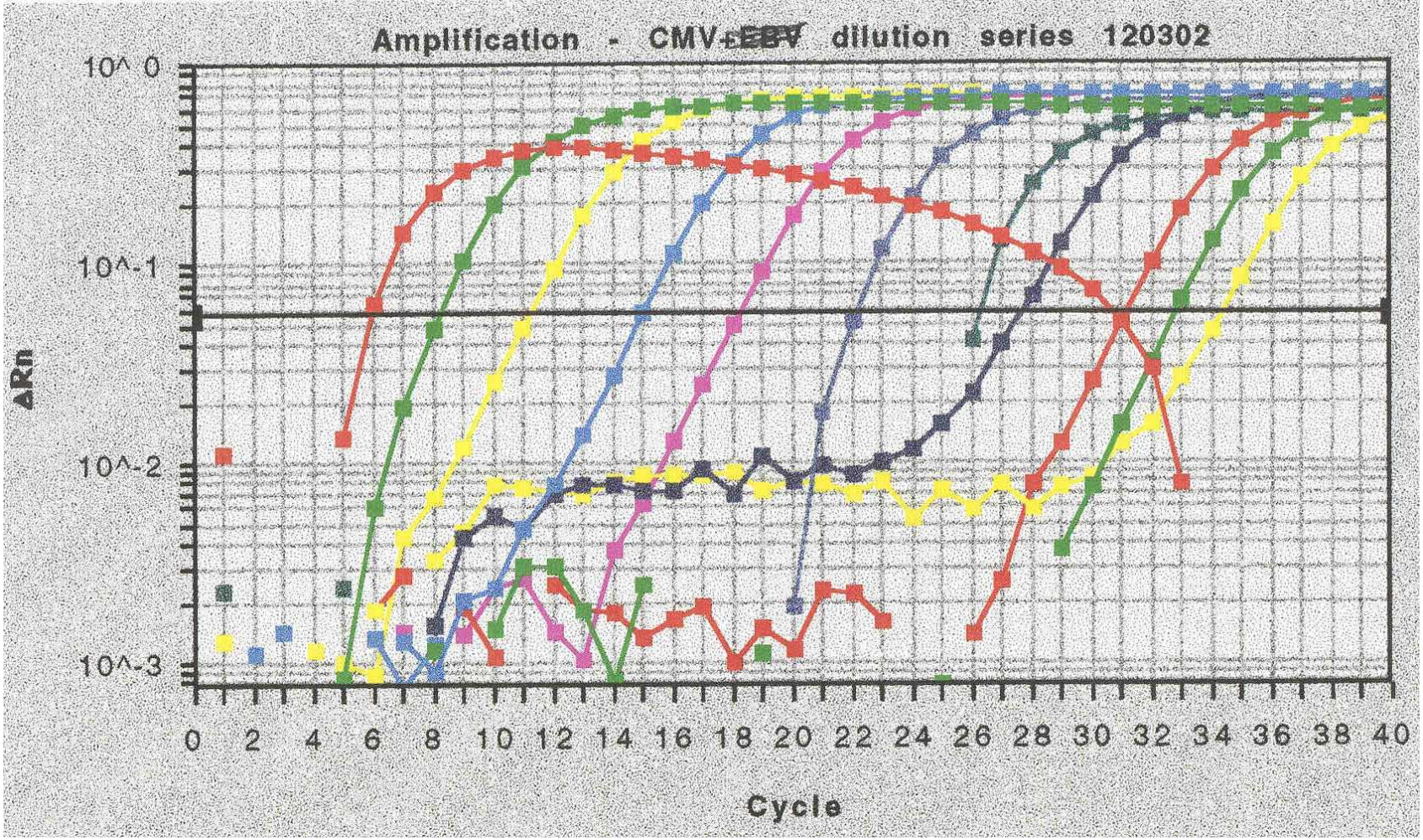


Example 4: Quantification to distinguish latent infection from significant reactivation

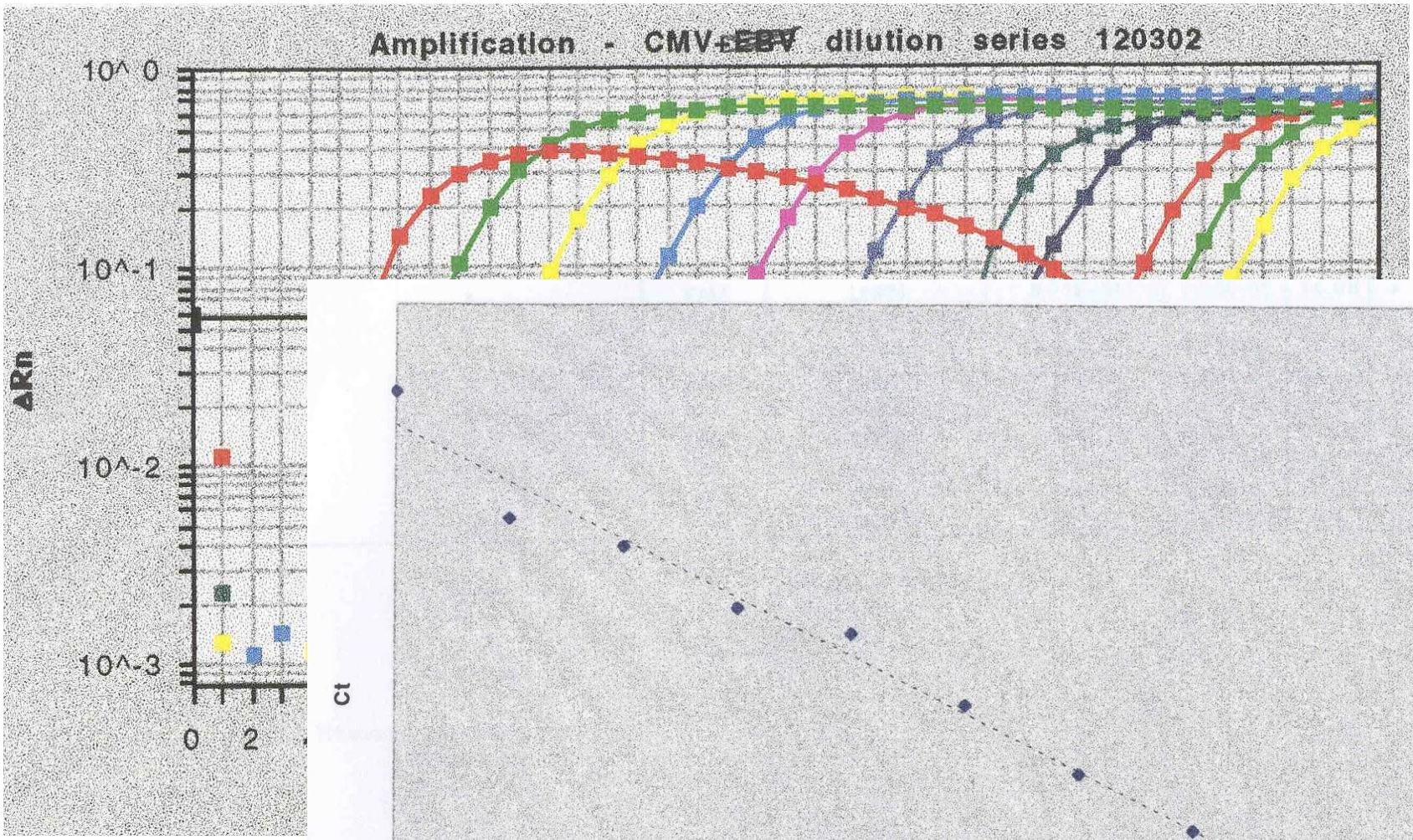


Example 4: Quantification to distinguish latent infection from significant reactivation





1 in 10 dilution series : read threshold cycle number by where the plot crosses the solid horizontal line



Calibration curve created – can be used to correlate quantify with threshold cycle number

Utility of quantification

- Cytomegalovirus
  - early diagnosis of CMV disease (plasma)
- Ebstein-barr virus
  - diagnosis of lymphoproliferative syndromes associated with immunosuppression
- Disease pathogenesis

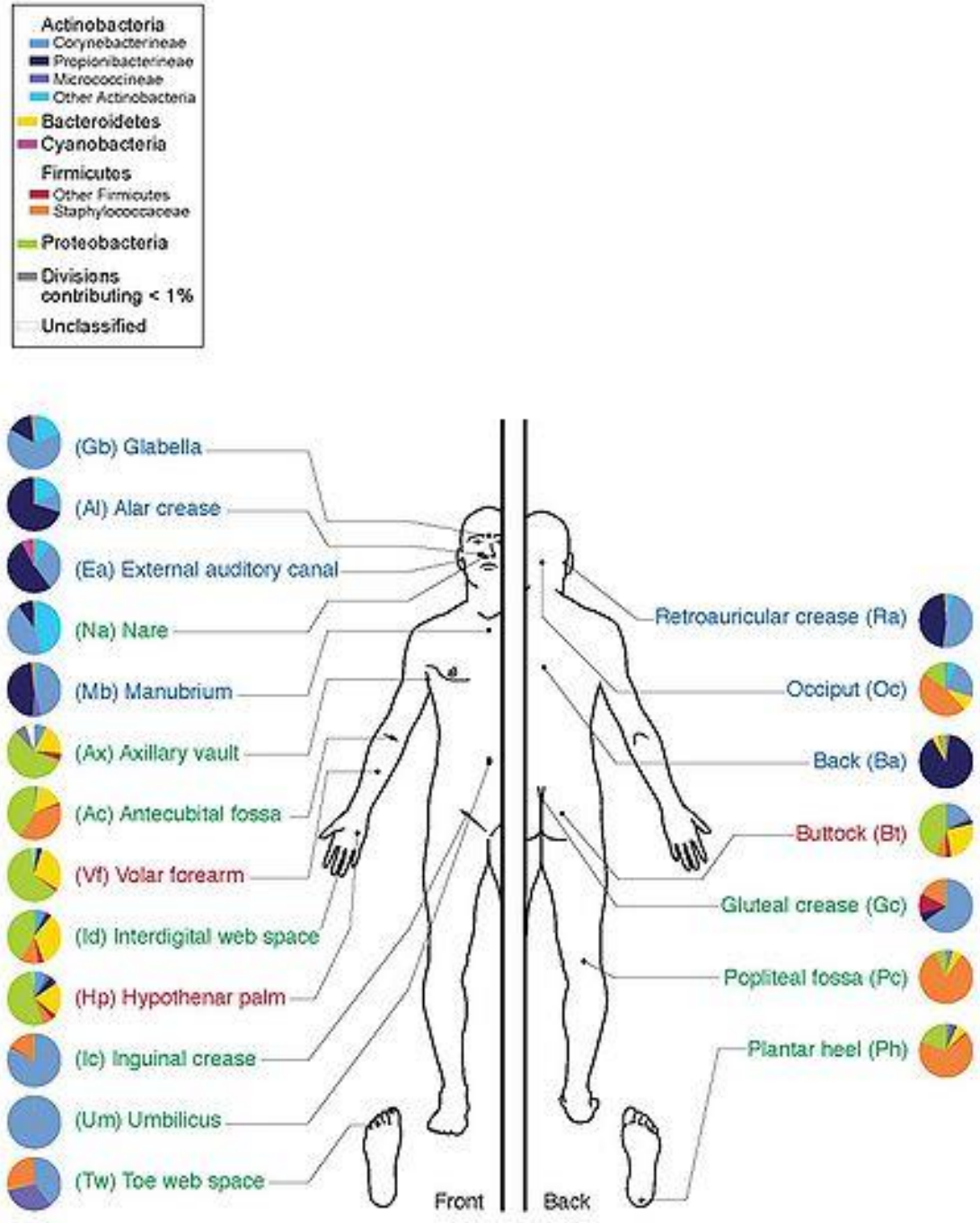
The study of the human "microbiome"

Relman and Falkow (2001) called for a "second human genome project" that "would entail a comprehensive inventory of microbial genes and genomes at the four major sites of microbial colonization in the human body: mouth, gut, vagina, and skin."

This approach, coupled with a "study of host genome-wide expression analysis," would yield major "insights into the role of the endogenous flora in health and disease."

Microbiome (2)

- Total microbial cells found in association with humans may exceed the total number of cells making up the human body by a factor of ten-to-one. The total number of genes associated with the human microbiome could exceed the total number of human genes by a factor of 100-to-one.
- 20% to 60% of the human-associated microbiome, depending on body site, is uncultivable. Rapid gene sequencing technologies enable description of the total spectrum of genomes.
- The NIH Human Microbiome Project
  - Studies of microbiomes in inflam bowel disease, obesity, skin with different diseases etc.
  - Cross species comparisons of gut microbiomes
  - Studies of ‘viromes’ in children



Unique distribution patterns for key bacterial genera

Two approaches

Phenotypic approach

- biological amplification
- slow
- definitive identification sometimes impossible
- not feasible for many pathogens

Molecular approaches

- nucleic acid amplification and sequencing
- rapid
- sensitive
- specific

Molecular diagnosis: the flip side

False positives

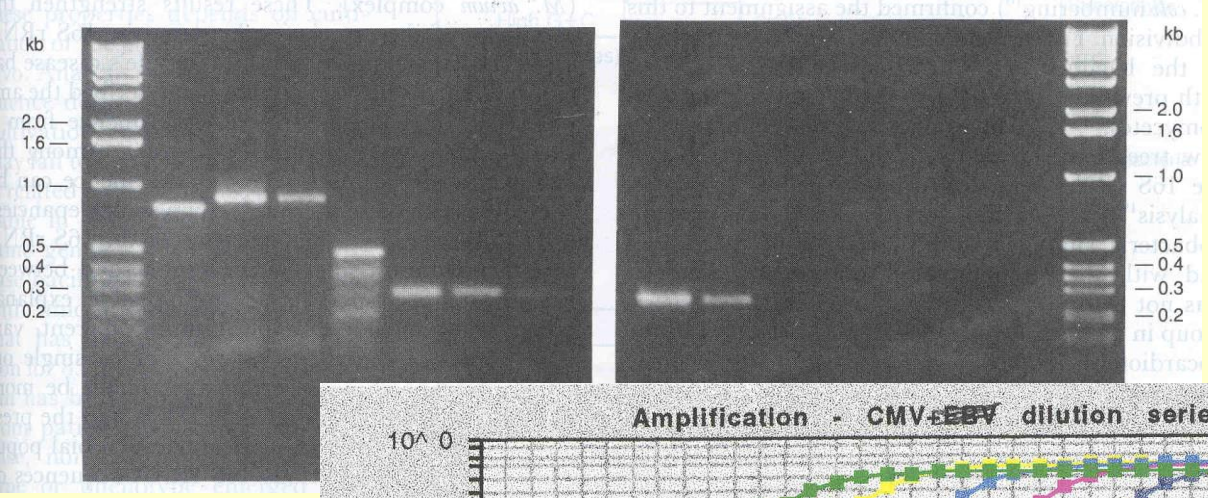
- Contamination
- Cross-reaction
- Poor assay design

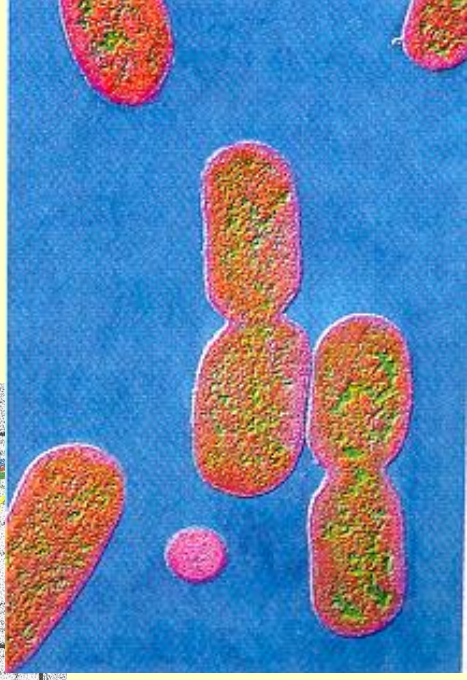
Assay sensitivity issue

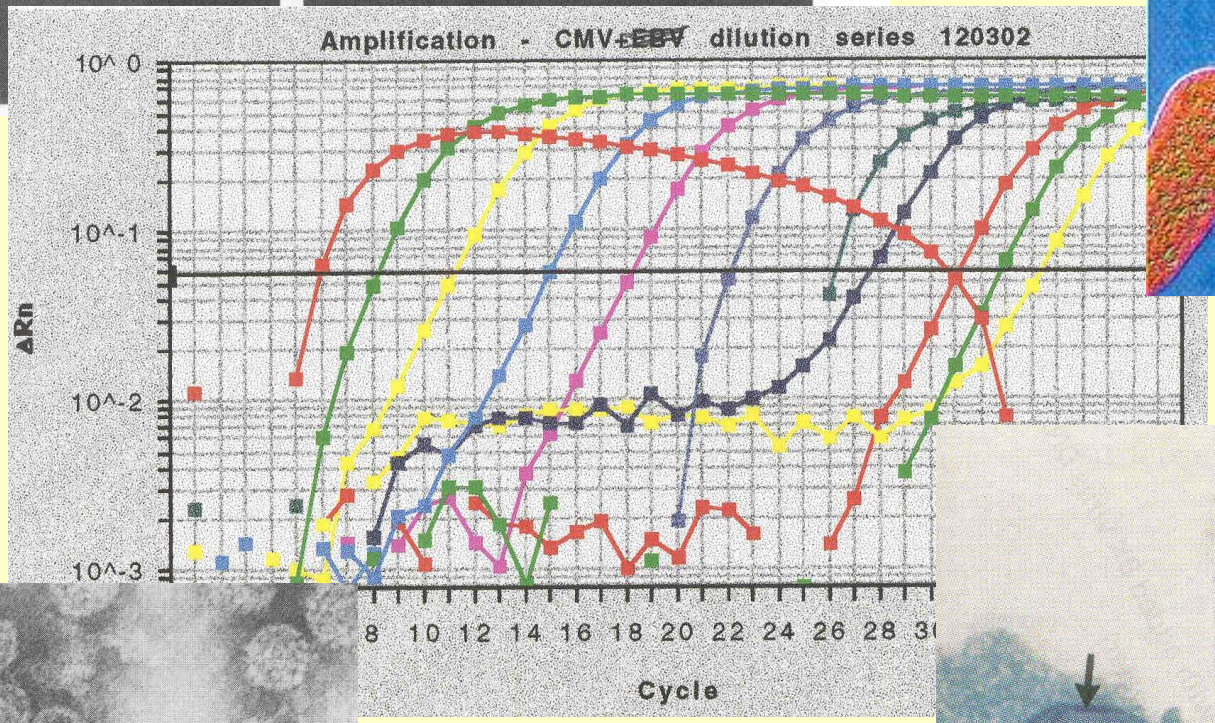
- Detection of NA from dead micro-organisms
- Chronic viral infections

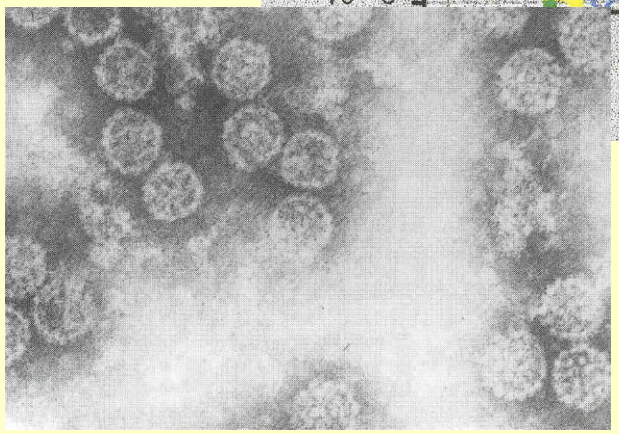
False negatives

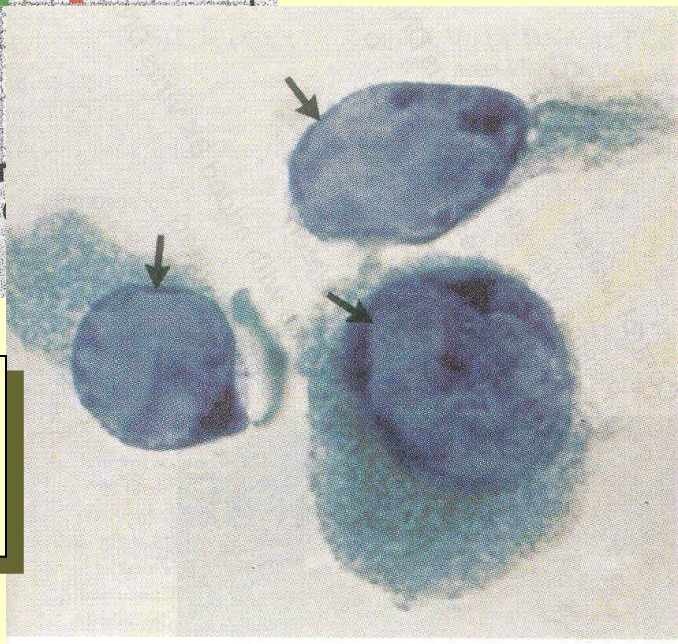
- NAA inhibitors
- sample error during extraction
- wrong organism sought


kb
2.0
1.6
1.0
0.5
0.4
0.3
0.2
kb
2.0
1.6
1.0
0.5
0.4
0.3
0.2




Amplification - CMV+EBV dilution series 120302
10^ 0
10^-1
10^-2
10^-3
ΔRn
8 10 12 14 16 18 20 22 24 26 28 3
Cycle





Thank you!