Detection of new pathogens through surveillance of bank samples

Eric Delwart

Blood Systems Research Institute.

- **Identification of new viruses using only nucleic acid technologies is becoming easier.**

- **Viral replication in tissue culture is helpful but not essential.**

- **Serological reagents are not required.**

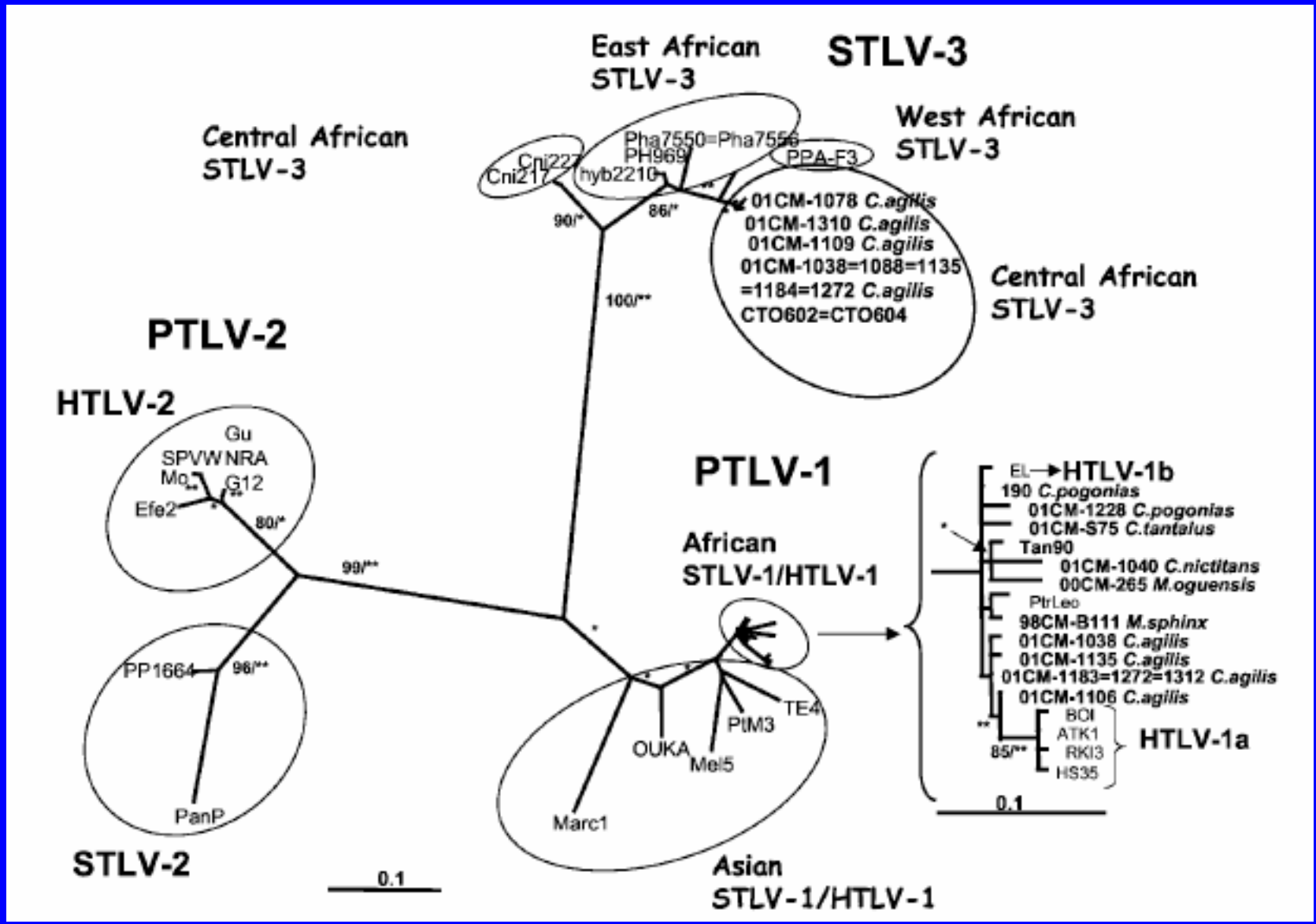
Sources of material for virus discovery

- **Samples from patients with unexplained symptoms of possible infection origins.**

- **Epidemiologically linked disease clusters.**

Possible to target investigations to viruses anticipated to exist or emerge.

Use non-specific, systematic, shotgun-sequencing approach.



Where is HTLV III ?

•New virus may be part of highly prevalent viral flora with no identified clinical consequences (e.g. TTV and HGV/GBV-C)

•Caveat: otherwise "harmless" infection may be pathologic in rare or immuno-suppressed subjects.

Two broad types of nucleic acid viral discovery methods:

1. Methods dependent on nucleic acid level similarities to known viruses

2. Methods independent of detectable nucleic acid similarities to known viruses

1. Methods dependent on nucleic acid level similarities to known viruses

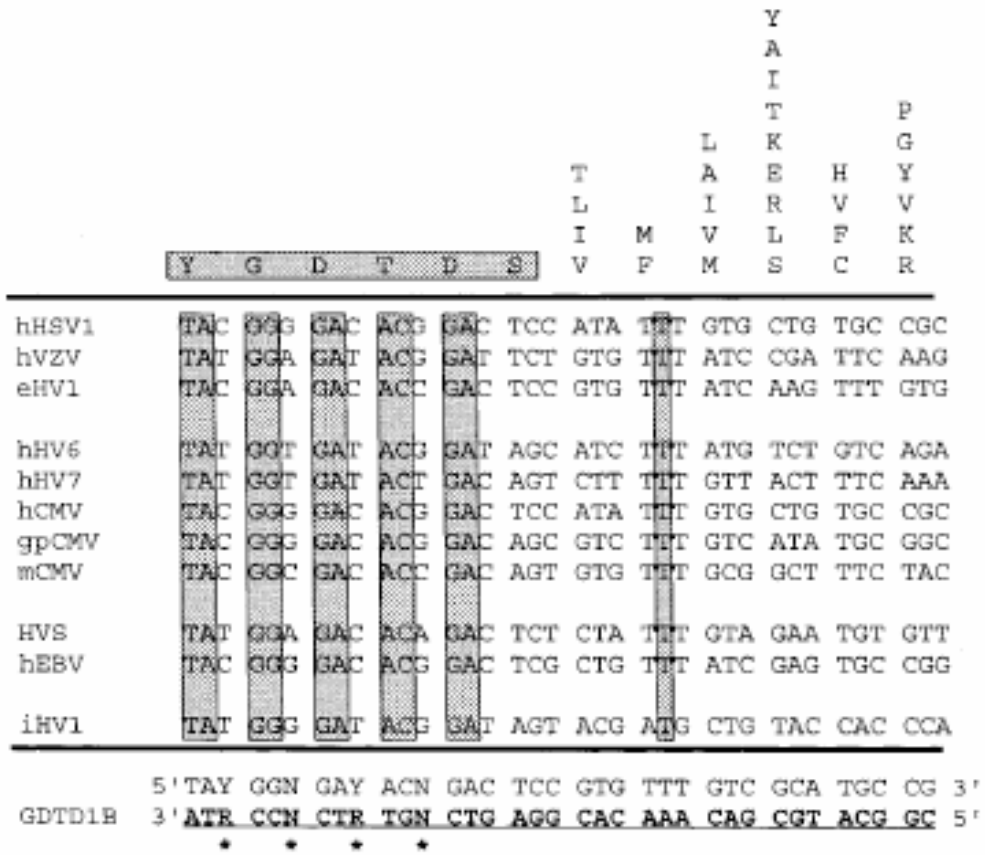
Reliable for known viruses or closely related species

Can target specific viral groups (consensus PCR) or all families (virochip microarray)

Limitations: Fails to identify highly divergent viruses whose nucleic acid will not anneal to oligonucleotide probe/primer

Consensus PCR

RNA de ruses



High diversity of unknown picorna-like viruses in the sea

Alexander I. Culley[1], Andrew S. Lang[2*] & Curtis A. Suttle[1,2,3]

[1]Department of Botany, [2]Department of Earth and Ocean Sciences, [3]Department of Microbiology and Immunology University of British Columbia, 1461-6270 University Boulevard, Vancouver, British Columbia, V6T 1Z4, Canada

A Herpesvirus of Rhesus Monkeys Related to the Human Kaposi's Sarcoma-Associated Herpesvirus

RONALD C. DESROSIERS,* VITO G. SASSEVILLE, SUSAN C. CZAJAK, XIAOMING ZHANG, KEITH G. MANSFIELD, AMITINDER KAUR, R. PAUL JOHNSON, ANDREW A. LACKNER, AND JAE U. JUNG

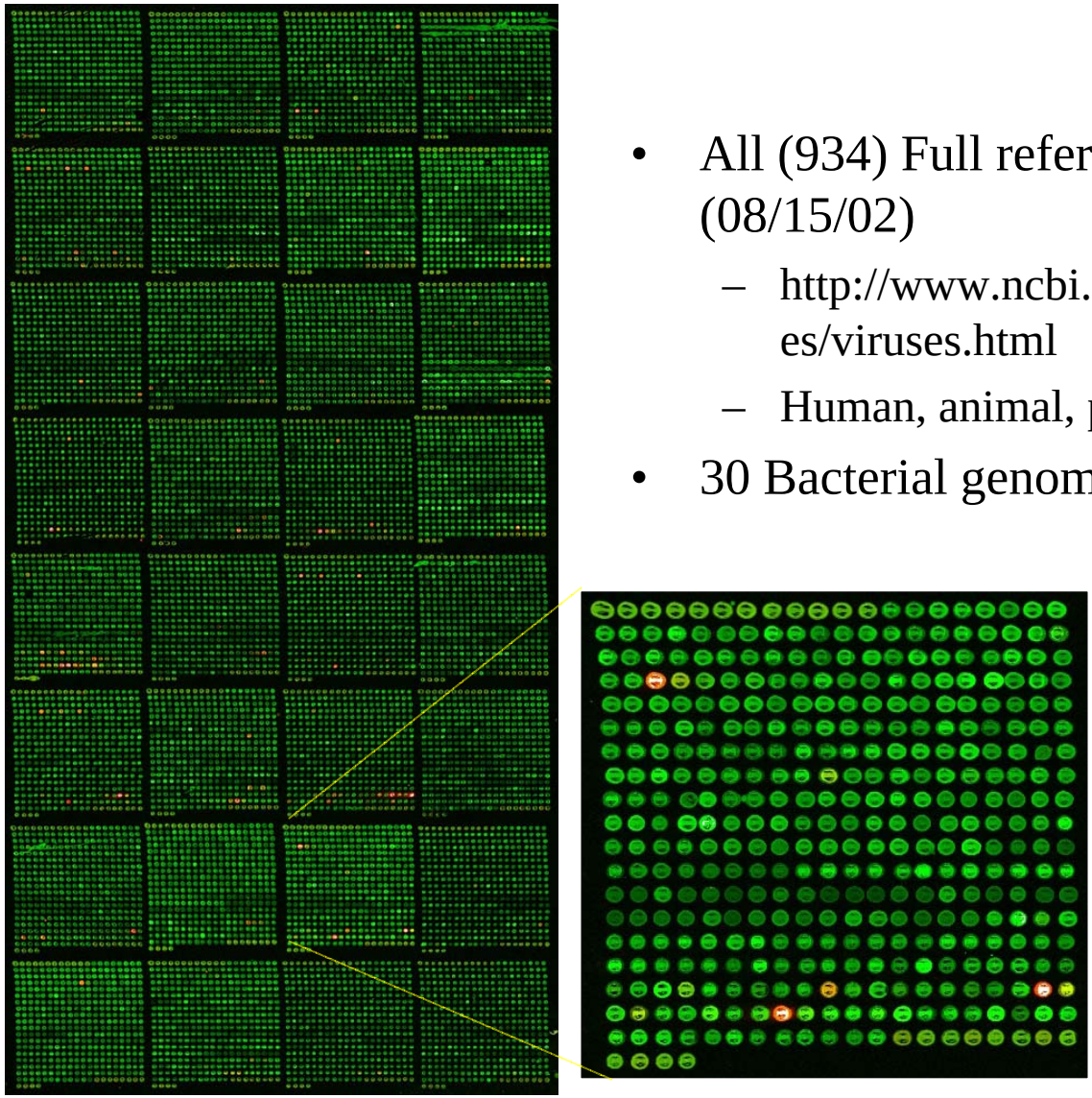
New England Regional Primate Research Center, Harvard Medical School, Southborough, Massachusetts 01772-9102

Detection and Analysis of Diverse Herpesviral Species by Consensus Primer PCR

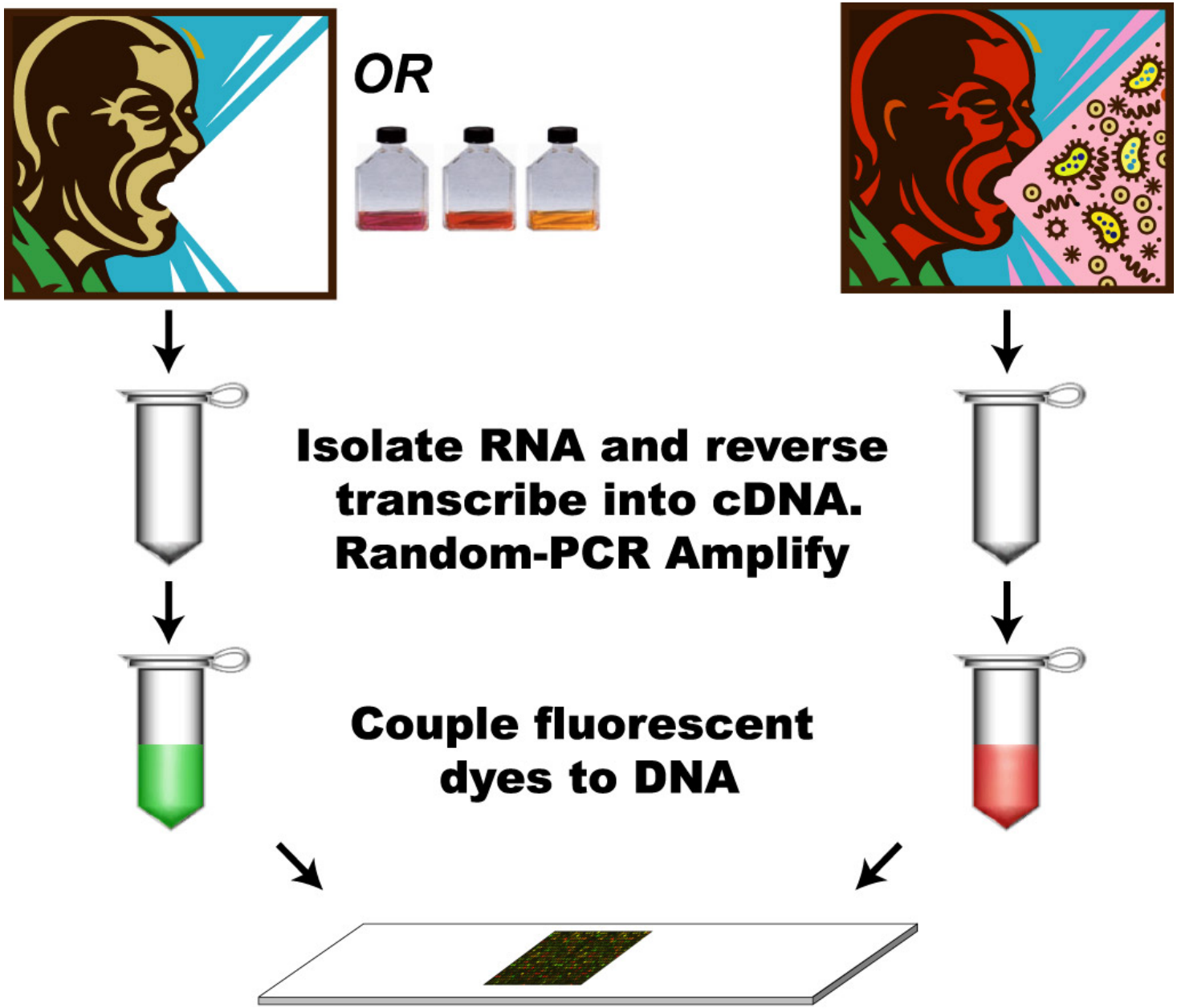
DONALD R. VANDEVANTER,* PAUL WARRENER, LINDSELY BENNETT,† EMILY R. SCHULTZ,‡ SILVIJA COULTER, RICHARD L. GARBER, AND TIMOTHY M. ROSE‡

PathoGenesis Corporation, Seattle, Washington 98119

Virochip 2.0: ~11000 70mers



- All (934) Full reference genome sequences (08/15/02)
  - http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/viruses.html
  - Human, animal, plant, bacteriophage
- 30 Bacterial genomes


OR
Isolate RNA and reverse transcribe into cDNA. Random-PCR Amplify
Couple fluorescent dyes to DNA
Mix fluorescent DNA and hybridize to the microarray

Using virochip SARS virus identified as a coronavirus within 48 hours of amplifying virus in cell culture

2. Methods independent of detectable nucleic acid similarities to known viruses

Sequence-independent amplification (or concentration) + subclone + sequence + search for sequence similarity

Advantages:

a. Theoretically detect all viral groups.

b. Possible to detect more highly divergent viruses following translation of nucleic acid sequence into protein amino acid sequence (i.e viral homology is detectable over longer evolutionary period with protein versus nucleic acid sequences).

Possible due to decreasing cost of sequencing and improving sequence similarity search algorithms

10 million viral particles/ml of seawater
detected using fluorescence techniques

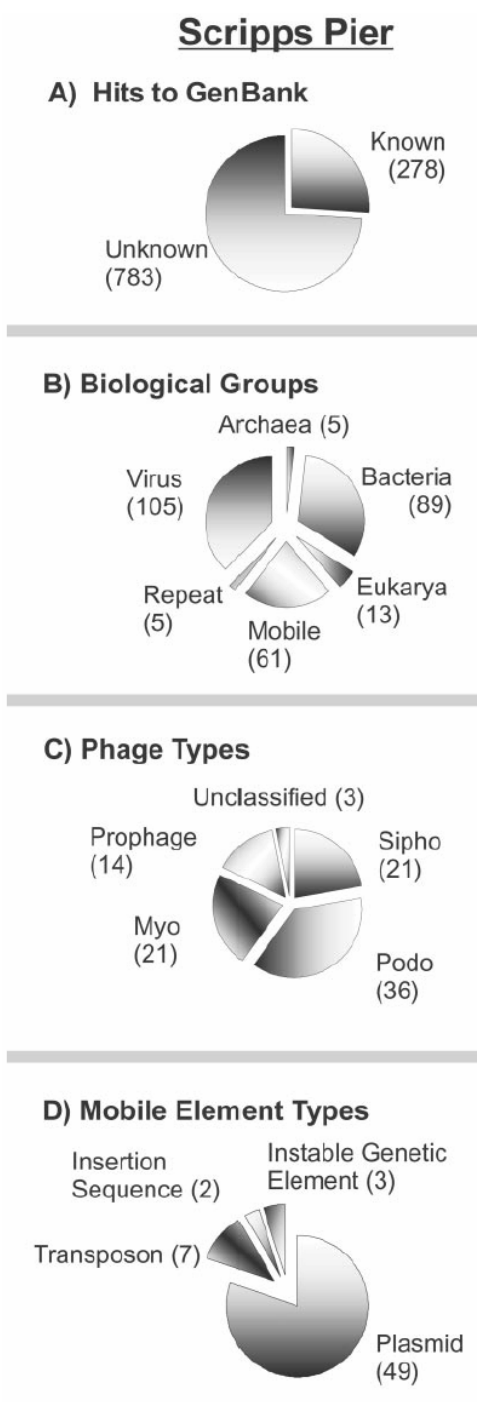
Genomic analysis of uncultured marine viral communities

Mya Breitbart*, Peter Salamon†, Bjarne Andresen†‡, Joseph M. Mahaffy†, Anca M. Segall*, David Mead§, Farooq Azam¶, and Forest Rohwer*‖

*Department of Biology, San Diego State University, San Diego, CA 92182-4614; †Department of Mathematical Sciences, San Diego State University, San Diego, CA 92182-7720; ‡Ørsted Laboratory, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark; §Lucigen, Middleton, WI 53562; and ¶Marine Biology Division, Scripps Institution of Oceanography, La Jolla, CA 92093

filter to remove bacteria and eukaryotes. CsCl gradient banding for viral fraction. Shear dsDNA, ligate linkers, PCR, subclone, sequence 1000 plasmid inserts, Blast against Genbank.

3/4 sequences no similarity hits to Genbank



Estimated 5000 viral species

A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species

Tobias Allander*, Suzanne U. Emerson†, Ronald E. Engle*, Robert H. Purcell*, and Jens Bukh*‡

Sections for *Hepatitis Viruses and †Molecular Hepatitis, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892

Sequence-independent amplification at BSRI:

Six positive controls: HCV RNA positive plasmas

25 plasmas from IDU+/-MSM
with recent high-risk exposure and multiple symptoms
of primary HIV infection but repeatedly HIV RNA negative.

Sequence-independent amplification

Viral nucleic acid protected from DNAse by capsid

Filter plasma/serum or any non-cellular biological sample

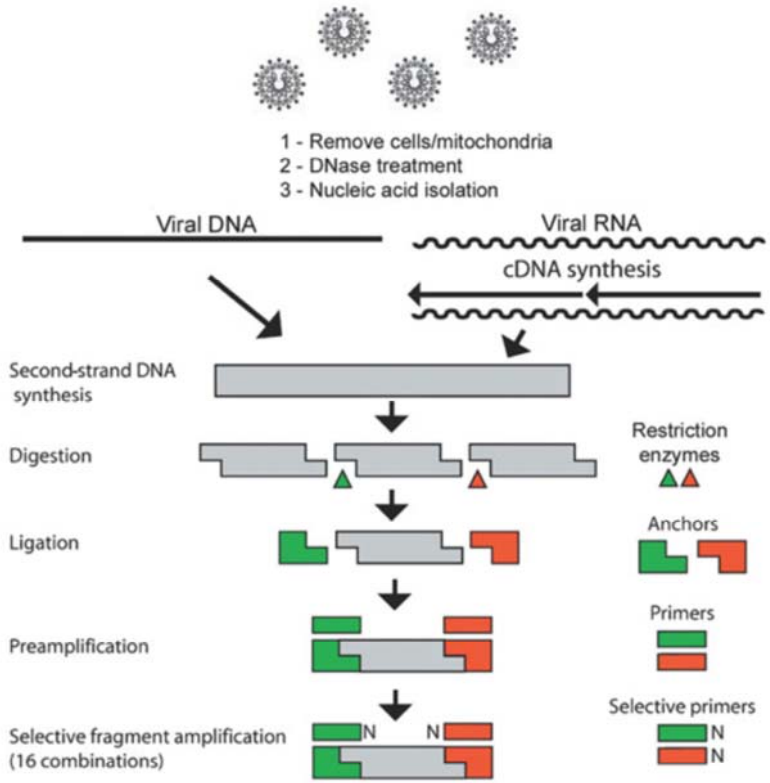
Remove human DNA from plasma by DNAse digestion

Sequence BLAST

Reverse transcribe RNA with random primers RT enzyme alone

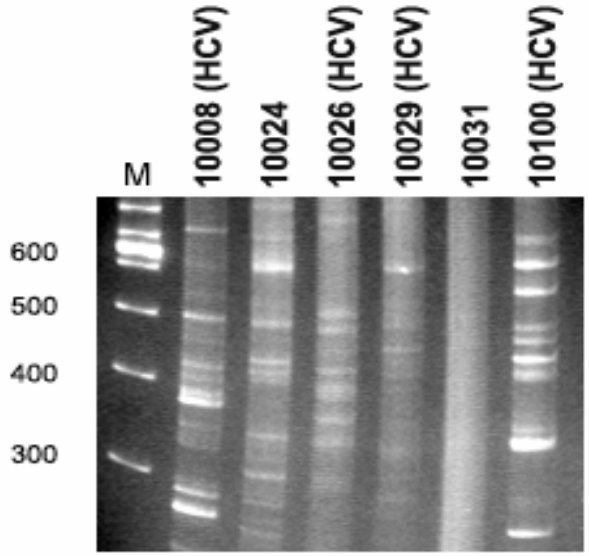
Add linker to restriction overhang.

Amplify using linker as PCR primer.

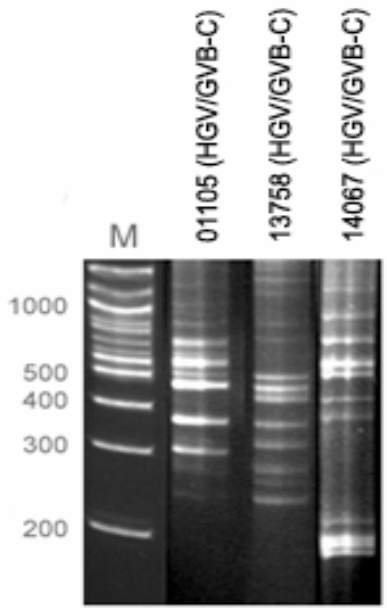


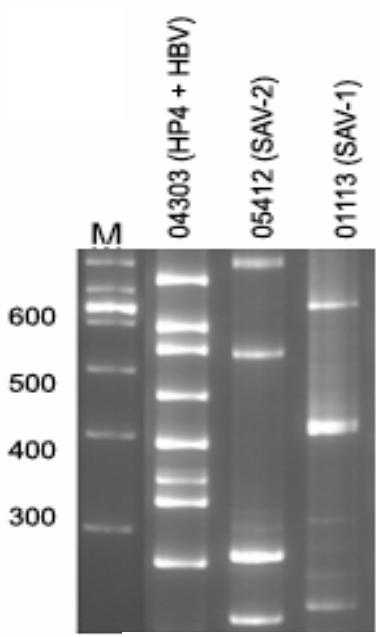
High level of single viral nucleic acid species is expected to yield low complexity amplification products and produce distinct DNA bands by PAGE.

5/6 HCV samples yielded bands



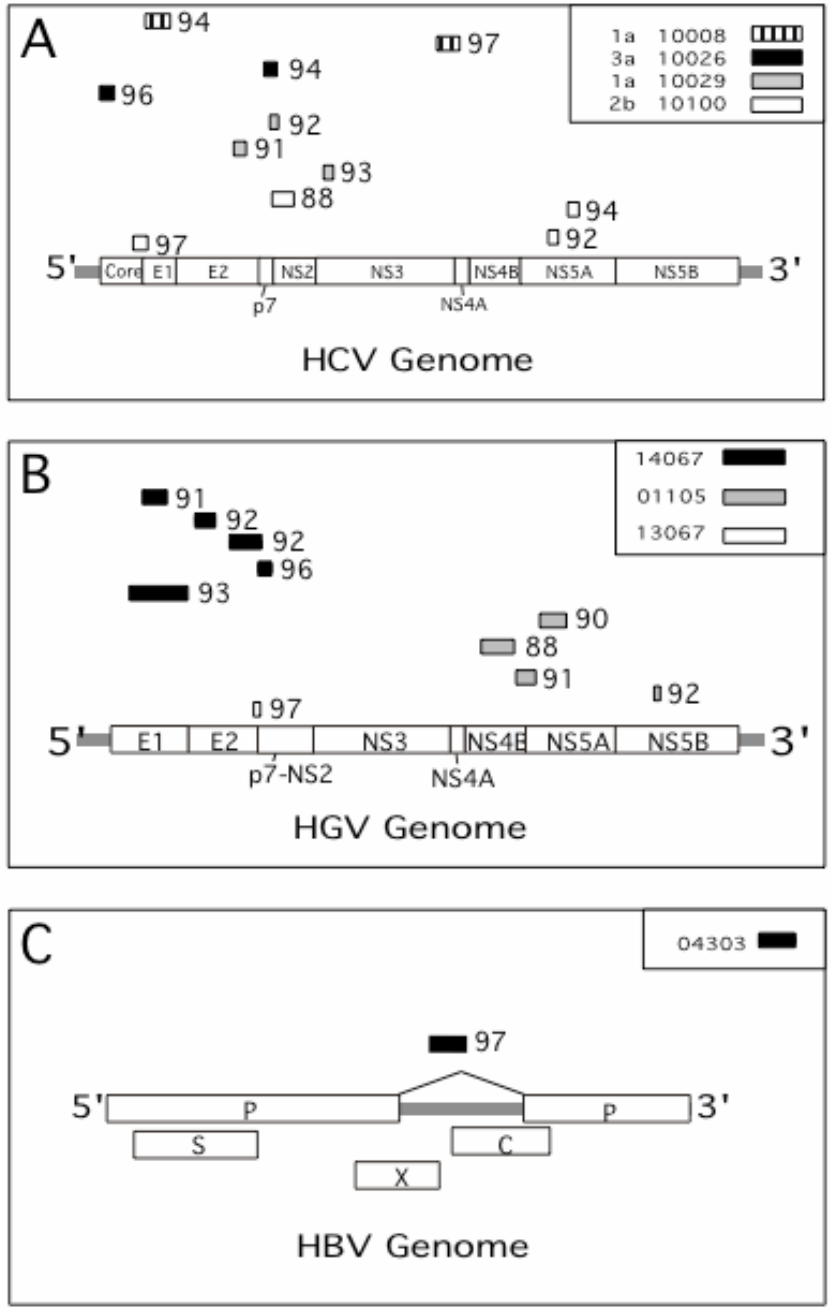
3/25 acute infections plasma RNA yielded bands





3/25 different acute infections plasma DNA yielded bands

After sequencing 10 to 20 plasmid subclones per sample known viruses identified.



GBV-C (previously HGV) is highly prevalent flavivirus.
Early link to transfusion hepatitis not confirmed.

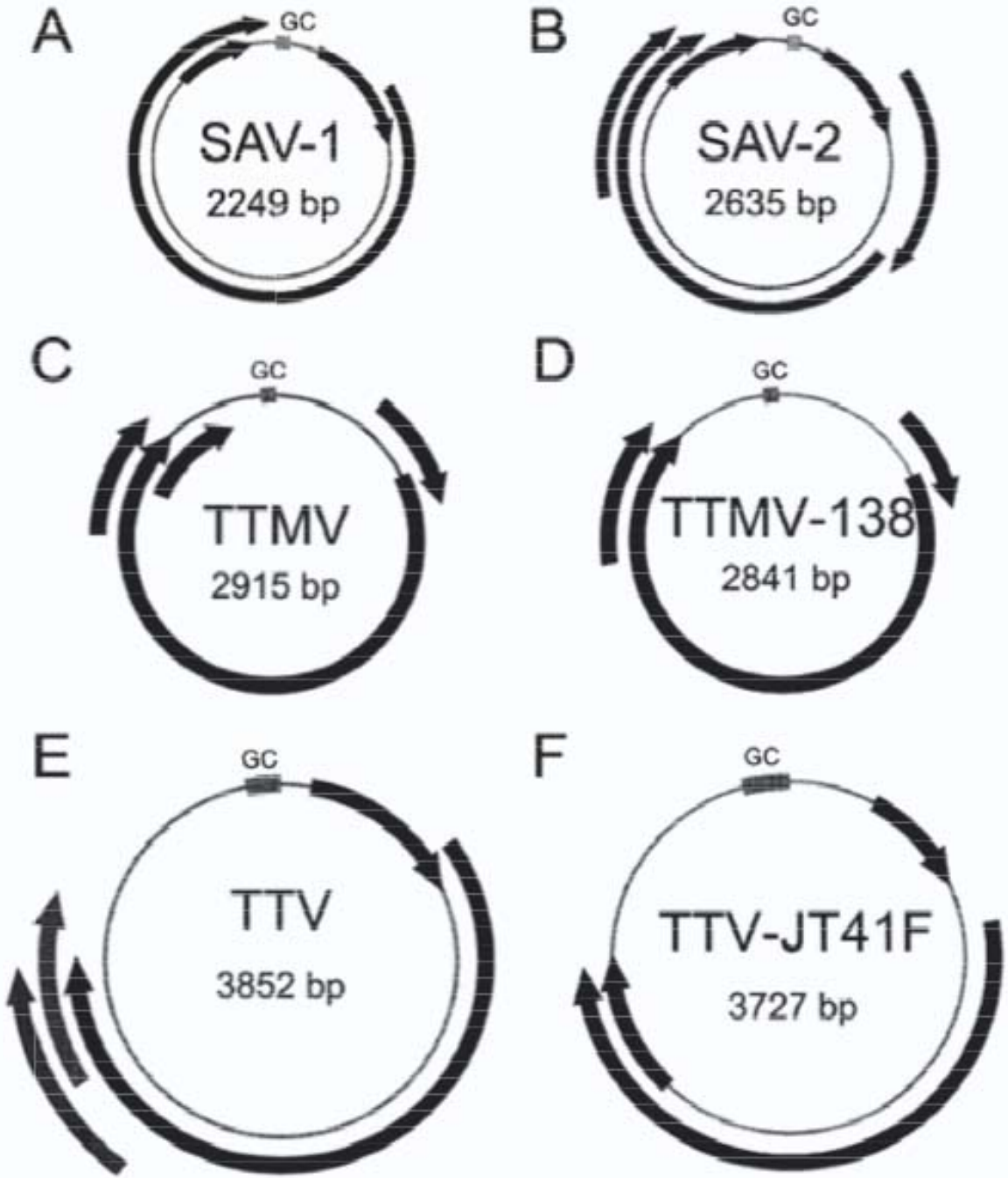
Sequences in two subjects showed detectable DNA level similarities to TT virus.

TTV recently discovered in a blood donor.
Small s.s. circular DNA genome.

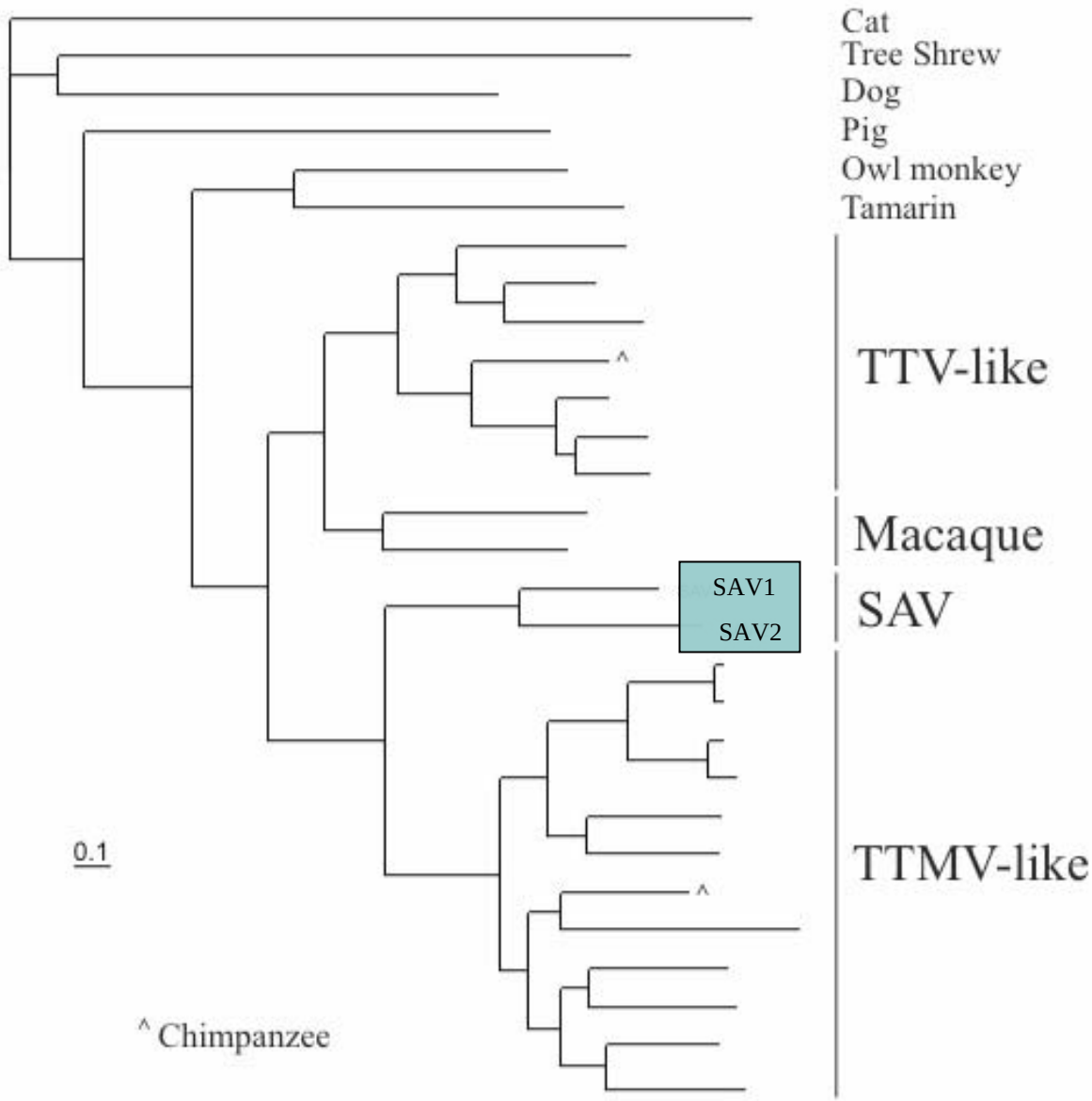
Early link with transfusion hepatitis not confirmed.

Very high human prevalence.

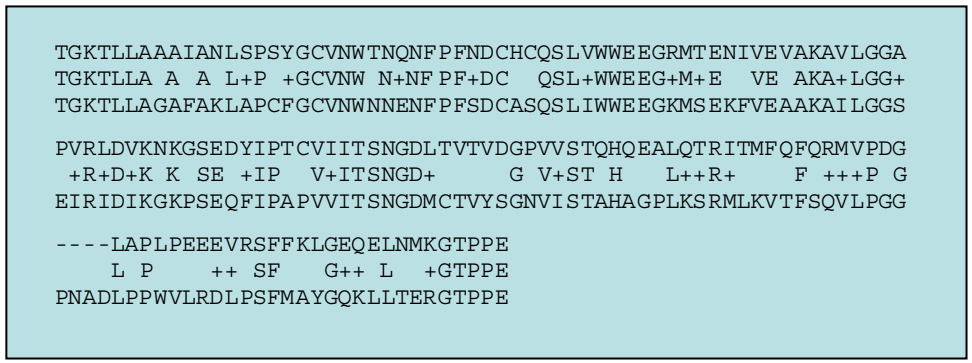
Present in all non-human primate tested.
Possible co-speciation of TTVs and primate species.


A
GC
SAV-1
2249 bp
B
GC
SAV-2
2635 bp
C
GC
TTMV
2915 bp
D
GC
TTMV-138
2841 bp
E
GC
TTV
3852 bp
F
GC
TTV-JT41F
3727 bp

Large ORF phylogenetic analysis of TTV & TTMV circoviruses



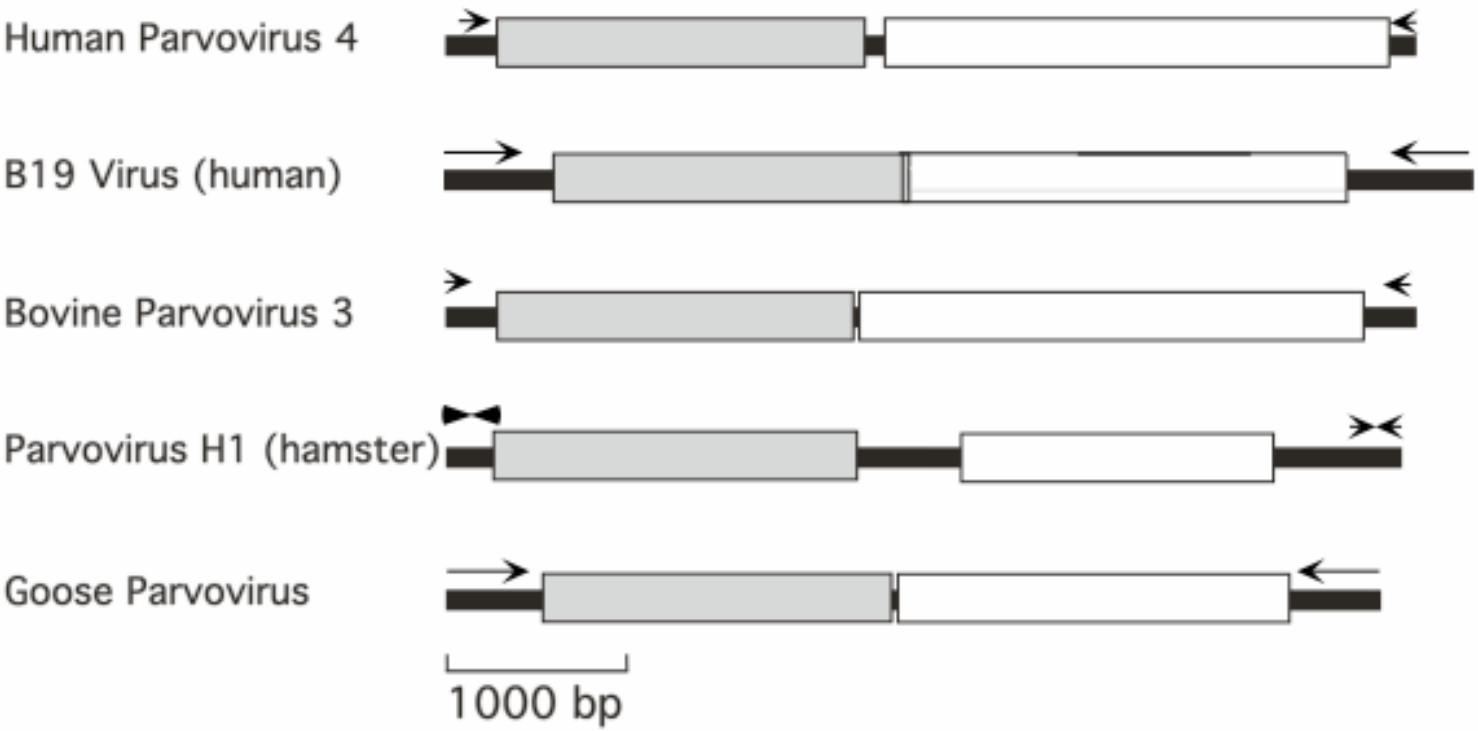
In one patient homology to parvoviruses detected only using amino acid level similarity search (tBLASTx).



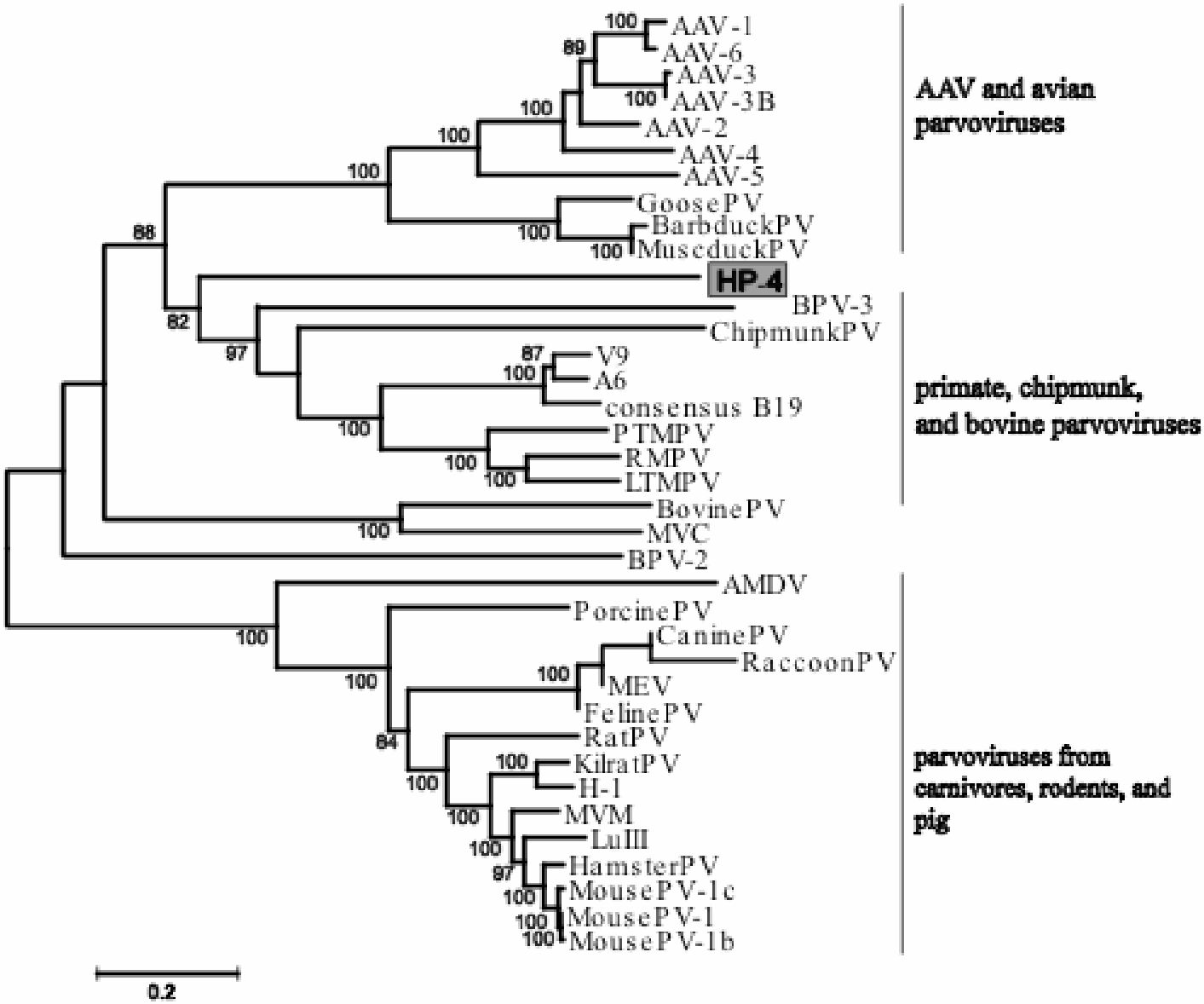
tBLASTx E score =$3e10^{-44}$

Best nucleic acid BLASTn E score to a mouse gene=0.23 (non-significant)

High divergence from previously known parvoviruses would have precluded detection using virochip microarray


Human Parvovirus 4
B19 Virus (human)
Bovine Parvovirus 3
Parvovirus H1 (hamster)
Goose Parvovirus
1000 bp

Full genome new parvovirus phylogenetic analysis



B19 parvovirus is a common innocuous infection in infants.

Adult infection can be pathogenic targeting erythrocyte precursor cells resulting in severe anemia.

B19 resistant to heat inactivation. Plasma pools are screened by PCR to exclude high-level B19 donations.

P CR analysis of plasma pools for new parvovirus DNA

Manufacturer	Numbers positive for sequences
A	8 positive (n=43)*
B	All negative (n=7)
C	All negative (n=4)
D	2 positive (n=3) **
E	All negative (n=14)
F	All negative (n=18)
G	All negative (n=28)
H	All negative (n=12)

Confirmed by amplicon sequencing. New variant ~10% divergent also identified. Similar in genetic diversity to B19/V9/A6 group ?

Focus Diagnostics:

Expressing surface glycoprotein in yeast to develop antibody test and determine seroprevalence.

From 25 patients with acute infection symptoms we found sequencing only ~100 sequences and sequence similarity searches:

1 HBV
3 GBV-C
2 highly divergent TTVs
1 new parvovirus

Worthy of follow-up

Best hit to mouse endogeneous reverse transcriptase

```
YSVIKKNELMPFVVTWMDLEIILSDISQTKTNT
                YH
YS IKKNE M F+ WMDLE IILS+++Q++ N+++
YSAIKKNEFMKFLAKWMDLESIILSEVTQSQR
                NSHN
```

Best hit to human adenovirus F

One third of sequences no sequence similarity to Genbank !!

Highly divergent viruses ?
Human viral/bacterial flora ?

Increased sampling (sequencing) of non-human nucleic acid should:

Yield more conserved regions of highly divergent viruses and permit assembly of longer contigs helping the identification of even more divergent viruses.

Determine if unidentified nucleic acids prevalent in healthy and sick subjects.

Good molecular methods exist to identify viruses both closely and distantly related to known viruses.

Transmissibility through blood product transfusions can be tested using linked donor-recipient samples.

No evidence new human parvovirus group is pathogenic.

Diseases in search of viruses.
Viruses in search of diseases.

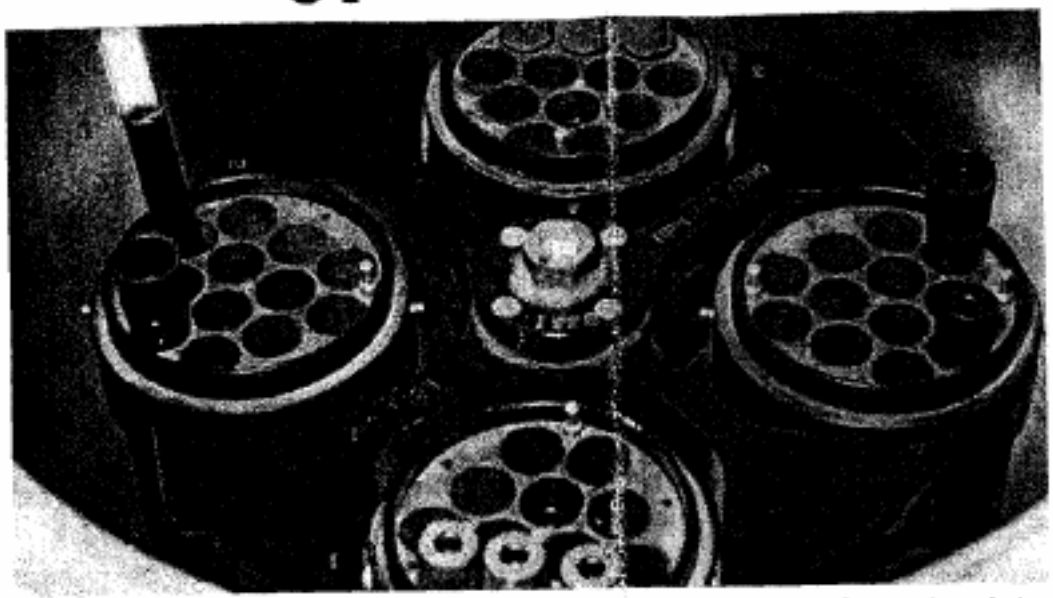
Biologists moot daunting plan to track viruses

Jonathan Knight, San Francisco,
and Alison Abbott, Munich

Three US researchers have proposed an ambitious project to sequence the genetic code of every virus known to infect humans. They plan to use the information in a database to track new and emerging viral infections. But critics say that it is not always possible to tell harmful and benign viruses apart, and that such a system could generate numerous false alarms.

The proposal, to be published in the July issue of *Emerging Infectious Diseases*, would involve collecting viral particles from blood plasma from a variety of sources, including the Red Cross and hospital labs. The genetic material would then be extracted and sequenced, and a database built up.

Methods for using centrifugation to purify viruses from large volumes of plasma were



K. GULDBRANDSEN/SPL

In a spin: centrifuges will be key for separating viral particles from blood samples for genetic analysis.

Future directions:

Develop simpler methods of sequence-independent amplification and subcloning for identifying new RNA and DNA viruses.

Improve viral particle purification methods.

Analyze plasma samples from patients with different symptoms.

Improve distant sequence similarity search algorithms.

Determine prevalence and transfusion transmissibility of new viruses.

Determine pathogenicity new viruses.

- **CDC unexplained death and severe illness, pneumonia and encephalitis projects.**
- **CDC study of persons occupationally exposed to non-human primates (NHP)**
- **NIH: Non viral hepatitis A-E transfusion transmitted hepatitis**
- **California Dept Health Services Unexplained encephalitis**
- **Cell culture supernatants from diagnostic labs showing cytopathic effects from unidentified agent .**

Bioinformatics treatment of sequence data

Seq ID	Total Hits	EMotif	QC Blast	blastn	tblastx	blastx	psiblast	Archaea	Bacteria	Eukaryota	Viroids	Viruses	Humans
Ad5	152								96.7	3.3			
Ad34	152								96.7	3.3			
Ad15	421								99.3	0.2			
Ad32	111								145	4.5			
Ad20	34416							0.4	3.4	96.2			4.6
6	1326							4.1	86.6	8.4			0.2
Ad13	0												
Ad22	2									100			100
Ad29	49									100			2
Ad21	715								1.1	1.8		138.7	
Ad30	34414							0.4	3.4	96.2			4.6
Ad19	286											120.3	
4	1801									100			82.2
1	477								2.5	0.4		105	
Ad10	140								96.4	3.6			
6	1326							4.1	86.6	8.4			0.2
Ad25	32976							0.4	3.6	96			4.8
Ad23	0												

~ half of sequences no-significant homology hit to Genbank

Morris Jones
Eric Delwart
Blood Systems Research Institute
University of California, San Francisco

Acute infection plasma samples

Rick Hecht
San Francisco General Hospital

Bioinformatics
Robert Shafer
Stanford University

Plasma pool analysis
Sally Baylis
National Institute for Biological Standards and Control, UK

Phylogenetic analyses
Peter Simmonds
University of Edinburgh

Vladimir Lukashov
University of Amsterdam

Support from the Blood Systems Foundation