The Perils of Pathogen Discovery: Origin of a Novel Parvovirus-Like Hybrid Genome Traced to Nucleic Acid Extraction Spin Columns

Samia N. Naccache, Alexander L. Greninger, Deanna Lee, Lark L. Coffey, Tung Phan, Annie Rein-Weston, Andrew Aronsohn, John Hackett, Eric L. Delwart, and Charles Y. Chiu

Journal of Virology 2013

The identification and whole-genome assembly of a highly divergent single-stranded DNA (ssDNA) virus situated at the interface between *Circoviridae* and *Parvoviridae* by deep sequencing.
The virus, provisionally named "**p**arvovirus-like" **h**ybrid **v**irus (PHV), was detected in samples from patients with chronic seronegative (non-A-E) hepatitis and diarrhea of unknown etiology.

Parvovirus B19 in humans and porcine circovirus type 2 (PCV2) in pigs have been previously linked to hepatitis.

A study by Xu, *et al.*, recently described the discovery of a hybrid DNA virus in serum samples from Chinese patients with seronegative hepatitis, named NIH-CQV, with a nearly identical sequence to PHV.

chronic seronegative (non-A-E) **hepatitis**

UCSF
169 patsiendi seerum, NA eraldati 3 eri ajal
22, **64(22)**, 105-st seerumist.
0.22 um filter, nukleaas.
QIAamp Viral RNA Mini Kit.
64 patsienti, 16 indekseeritud pooli.
Illumina TrueSeq raamatukogu
Illumina HiSeq
Puhastatakse inimese ja bakterite vastu.
Parvoviiruse sarnased readid (leitud blastX-ga) PRICE assambleri seemikuks (seed) + Geneious. 85% over 25 nt to merge.
9 tsükklit, viirus koos
Kontrollitud PCR-i ja Sangeriga

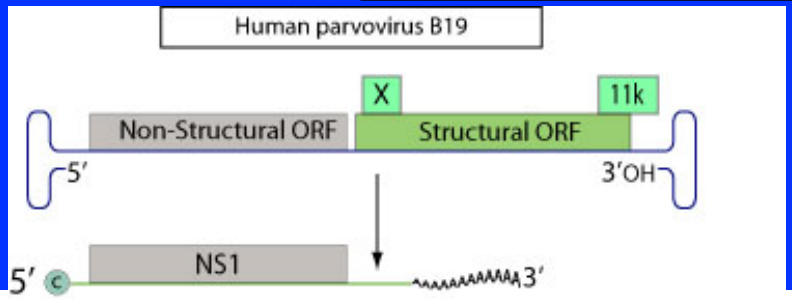
diarrhea of unknown etiology

Blood Systems Research Institute, San Francisco
diarrheal stool samples from Nigeria (75) and Tunisi (50)
0.45 um filter, nukleaas
QIAamp Viral RNA Mini Kit
Raamatukogu ScriptSeq V2 RNA-Seq (Epicentre), 13 sample pools 5-10 each
Illumina MiSeq
Parvoviiruse sarnased readid (leitud blastX-ga)
PRICE assambleri seemikuks (seed) + Geneious.
16 tsüklit viirus koos

Parvovirus like **H**ybrid **V**irus -1 PHV-1
3638 nt

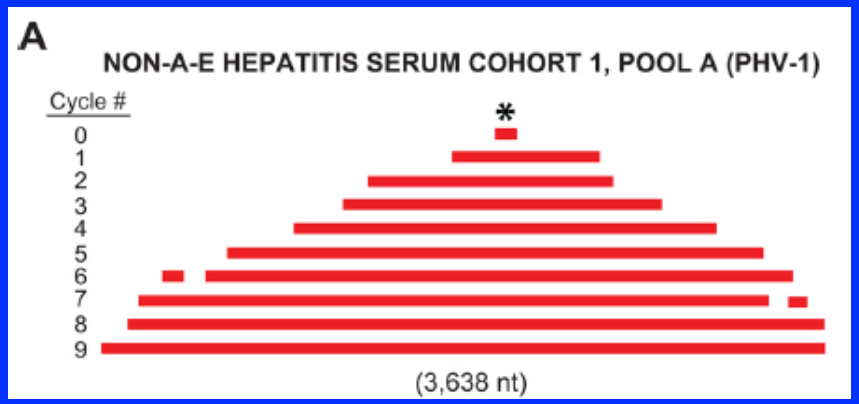
Parvovirus like **H**ybrid **V**irus -2 PHV-2
3625 nt

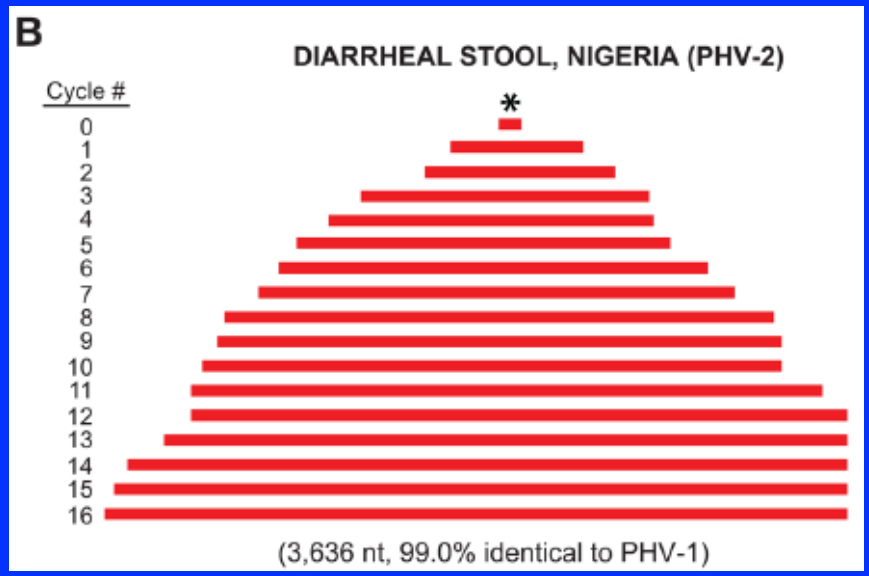
Viirused PHV-1 ja PHV-2
Joondus 3625 nt, identsus 3614 nt (99,7%), 1 gap



NIH-CQV 3780 nt
~20 nt insertsioon +otsad 60 nt pikemad

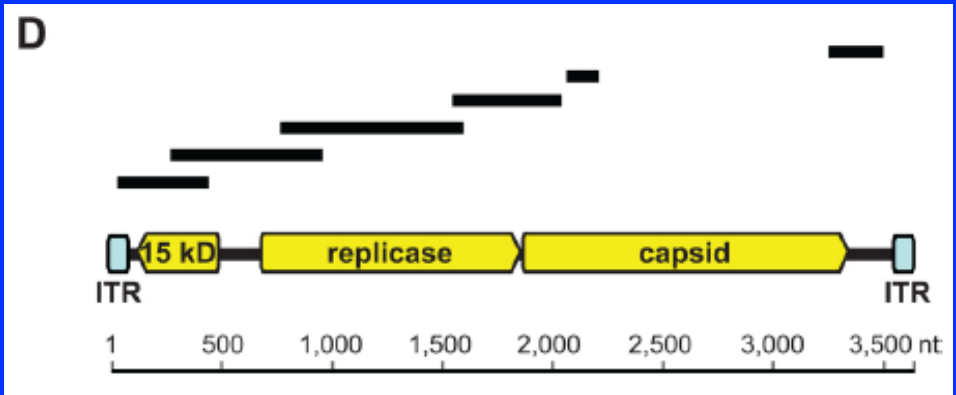
Viiruse genoomi assambleerimine





De novo assembly of PHV. The PRICE assembler was used to assemble the partial or full genomes of PHV-1 (A), PHV-2 (B), corresponding to a sample pool of non-A-E hepatitis sera and a sample pool of diarrheal stool respectively. The asterisks denote the initial seeds used for de novo assembly. Intermediate contiguous sequences (contigs) generated during the assembly (red bars) are mapped to their corresponding location on the PHV genome.

PHV iseloomustus



The genome organization of PHV, showing the open reading frames (ORFs) corresponding to the putative replicase and capsid proteins and hypothetical 15 kD protein of unknown function and two 148-nt inverted terminal repeat (ITR) sequences. Regions of the genome that were confirmed by Sanger sequencing are represented by black lines

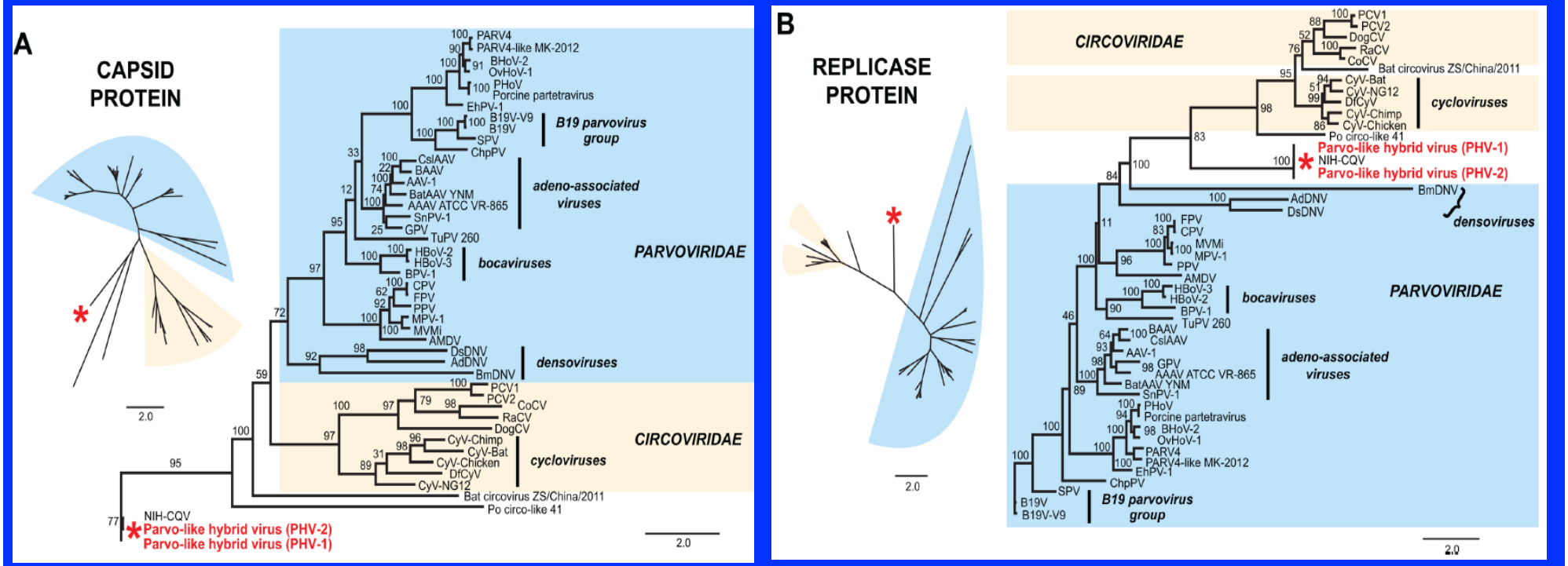
Replikaas sarnane circoviirustele, kapsiid sarnane parvoviirustele

Circoviiruste genoomiks on ssDNA rõngasmolekul

Parvoviiruste genoom lineaarne, pöördkordusjärjestustega otstes.

PHV-del rõngasmolekuli detekteerida ei õnnestunud.

PHV ja NIH-CQV fülogeneetiline asetsemine



Amino acid phylogenetic trees of PHV-1, PHV-2, and NIH-CQV relative to other ssDNA viruses. (A) capsid protein; (B), replicase protein. Representative parvoviruses, circoviruses, and “circo-like” viruses were included in the phylogenetic analysis. Support levels calculated using a maximum likelihood algorithm are displayed at each branching point. Scale bars indicate the number of amino acid substitutions per site.

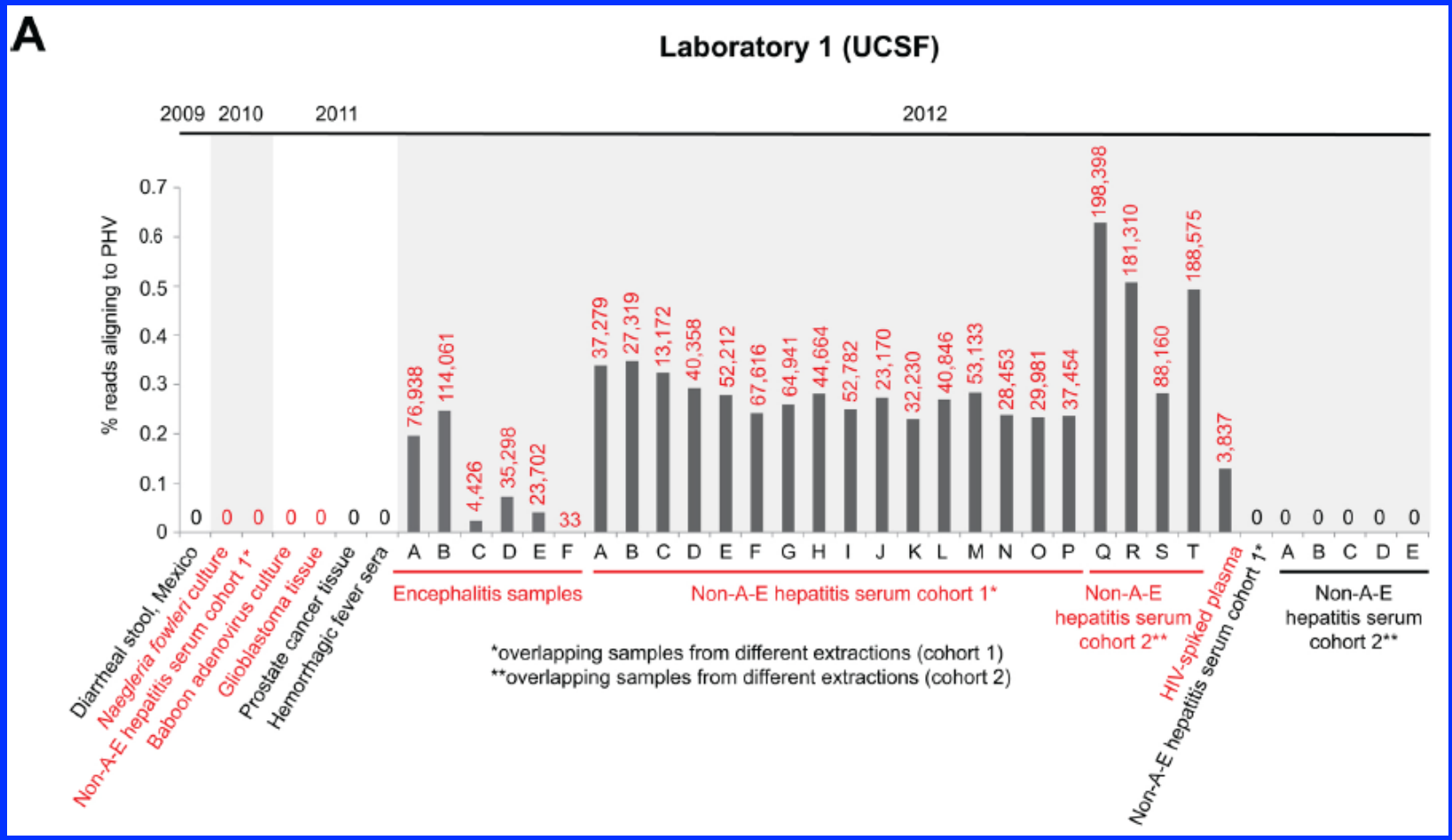
Kuna proovid olid poolitud, siis määrati PHV-d algsetest proovidest PCR-ga.

Esmalt algsest DNA-st mida oli kasutatud poolimiseks.
Viirus leiti kõikidest proovidest.

Seejärel kasutati EZ1 instrumenti (EZ1 Viral Mini Kit v2.0).
Kõik proovid osutusid negatiivseteks PHV suhtes.

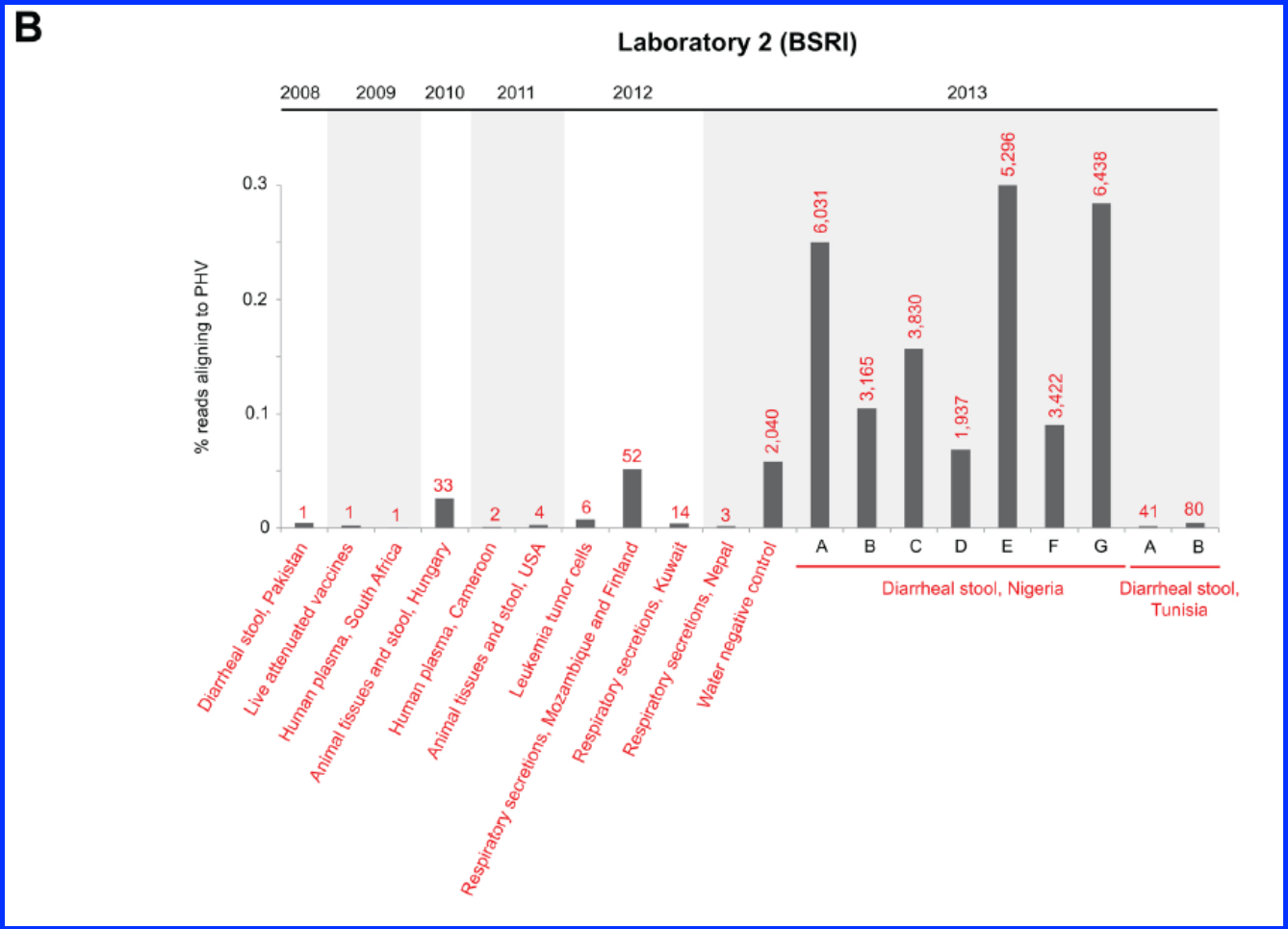
?

Kus PHV-d veel leidub



Screening for PHV in clinical NGS datasets. Sequences corresponding to PHV were identified by BLASTn alignments at an E-score cutoff of 10-30 to sequence reads in NGS datasets processed from 2008 to 2013 in laboratory 1 (A) and laboratory 2 (B). Shown plotted is the percentage of PHV reads in each dataset, with the number of PHV reads indicated by the data labels. Clinical sample sets extracted using **Qiagen spin columns** are highlighted using **red text**; those extracted **using other methods** are shown in **black**.

Kus PHV-d veel leidub



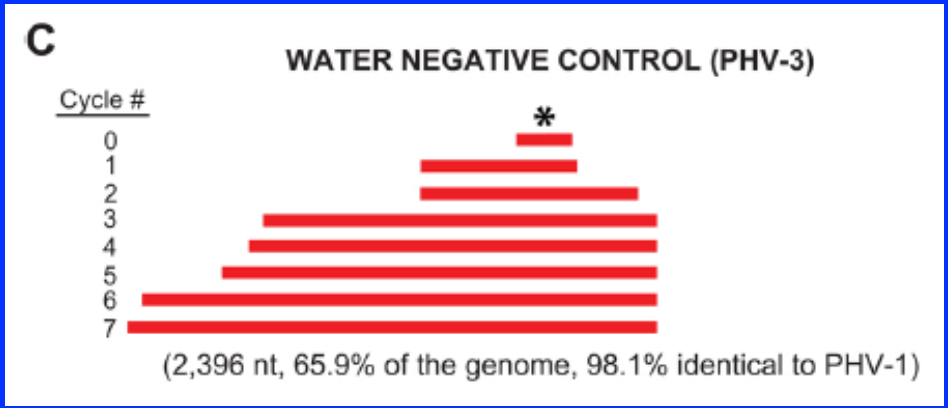
Screening for PHV in clinical NGS datasets. Sequences corresponding to PHV were identified by BLASTn alignments at an E-score cutoff of 10-30 to sequence reads in NGS datasets processed from 2008 to 2013 in laboratory 1 (A) and laboratory 2 (B). Shown plotted is the percentage of PHV reads in each dataset, with the number of PHV reads indicated by the data labels. Clinical sample sets extracted using **Qiagen spin columns** are highlighted using **red text**; those extracted **using other methods** are shown in **black**.

Viiruse olemasolu näib sõltuvat Qiagen-i kindlast kolonnist, eriti alates 2012. aastast.

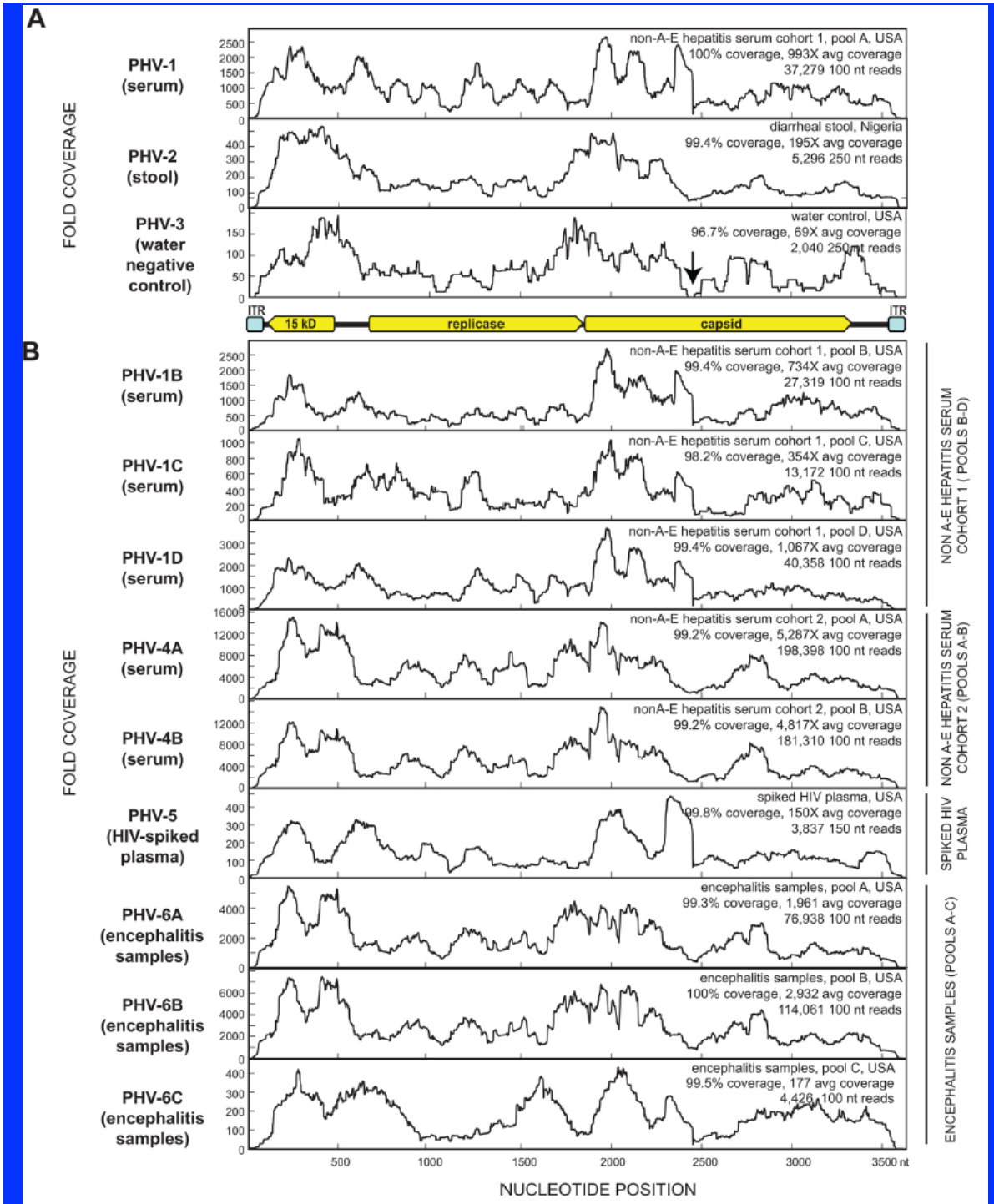
Qiagen-i kolonni ekstraheeriti veega kolmest firmast, kõik andsid PCR-l positiivse tulemuse.

Lähtudes puhtast veest eraldati kolonniga (õigemini küll kolonnist) DNA ja sekveneeriti.

De novo assambleerimisel saadi contig mis kattis 66% PHV genoomist.



Kolonnist eraldatud DNA-st leiti viirus mis oli PHV-ga 97% identne



Mapping of reads from clinical NGS datasets to the PHV genome. The coverage (y-axis) achieved at each nucleotide position along the genome (x-axis) is plotted.

(A) Coverage maps for the independently de novo assembled PHV-1, PHV-2, and PHV-3 genomes. The arrow depicts a break in coverage that precluded whole-genome de novo assembly of PHV-3.

(B) PHV genomes assembled by mapping of NGS reads from various clinical sample datasets to the PHV-1 genome. The inset shows the name of the clinical dataset, as well as the percent (%) and average (avg) genomic coverage achieved.

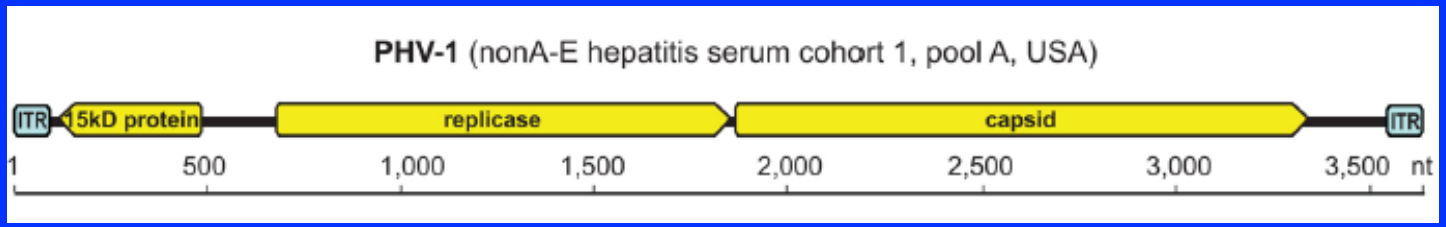
Abbreviations: nt, nucleotide

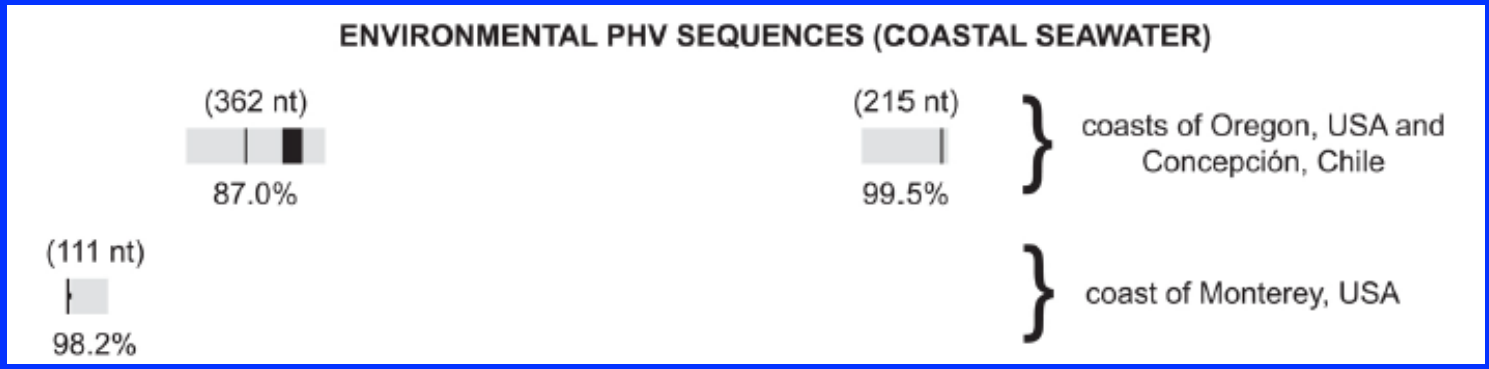
Hiired tulevad mustades sokkidest ja viirus tuleb kolonnist

Table 1. PCR screening of commonly used viral nucleic acid extraction kits for parvo-like hybrid virus (PHV-1). Abbreviations: nt, nucleotide; NCR, non-coding region; C, column elution; F, full extraction; +, positive by PCR; -, negative by PCR; NA, not applicable.

		Replicase (247 nt) 763-1010nt		**Bridge (490 nt)** 1554-2044nt		**Capsid (120 nt)** 1922-2044nt		**Capsid + NCR (160 nt)** 3288-3448nt	
Kit	**Spin Column**	**C**	**F**	**C**	**F**	**C**	**F**	**C**	**F**
RNeasy MinElute Cleanup	RNeasy MinElute	+	+	-	+	+	+	+	+
RNEasy Mini	RNeasy Mini	+	+	+	+	+	+	+	+
QIAamp UltraSens Virus	QIAamp Mini	+	+	-	+	+	+	+	+
QIAamp Viral RNA Mini	QIAamp Mini	-	+	-	-	+	+	+	+
QIAamp DSP Virus	QIAamp MinElute	-	+	-	-	-	+	-	+
PureLink Viral RNA/DNA Mini	Viral	-	-	-	-	-	-	-	-
TRIzol LS	NA	-	-	-	-	-	-	-	-
EZ1 Viral Mini v2.0	NA	-	-	-	-	-	-	-	-
Water, Nuclease-Free (Qiagen, Fisher Scientific, and Epicentre)	NA	-	-	-	-	-	-	-	-

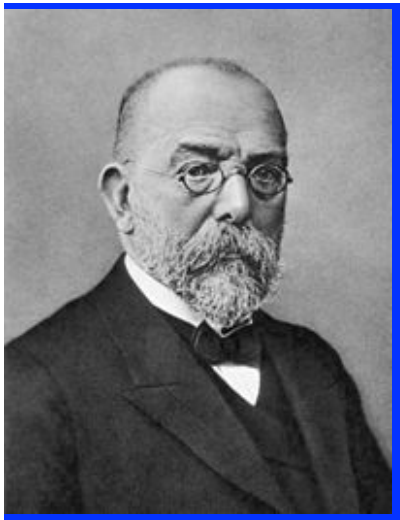
Sarnased järjestusi leidub keskkonnaproovide metagenoomides





BlastN E-score cutoff 10^{-30}

Column-based NA purification is a solid-phase extraction method that binds NA by adsorption to silica, and the silica used in many commercial spin columns is derived from the cell walls of diatoms. If Qiagen's NA extraction kits and "silica gel membrane technology" involve the use of diatoms, it is plausible that PHV is a virus of diatoms and had inadvertently contaminated the spin columns during manufacture.



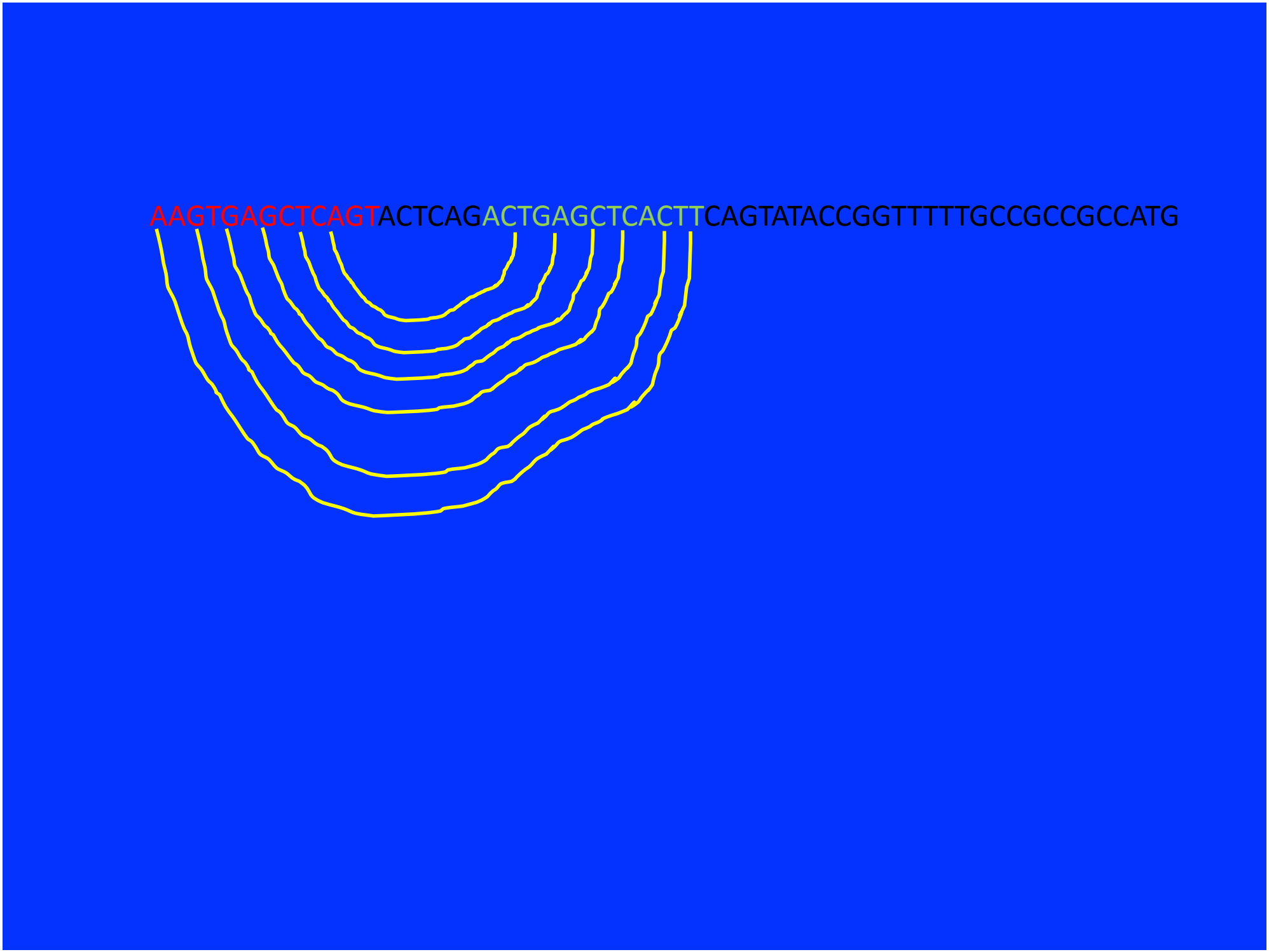
Heinrich Hermann Robert Koch

Koch's postulates are the following:

1. The microorganism must be found in abundance in all organisms suffering from the disease, but should not be found in healthy organisms.
2. The microorganism must be isolated from a diseased organism and grown in pure culture.
3. The cultured microorganism should cause disease when introduced into a healthy organism.
4. The microorganism must be reisolated from the inoculated, diseased experimental host and identified as being identical to the original specific causative agent.

Suur tänu kuulamast!

Ja ärge kaotage valvsust.


AAGTGAGCTCAGTACTCAGACTGAGCTCACTTCAGTATACCGGTTTTTGCCGCCGCCATG

The **retrovirus XMRV** is not associated with chronic fatigue syndrome or prostate cancer and, in fact, originated as a mouse cell line-derived laboratory contaminant

PLoS Pathogens
Journal of Virology
PNAS
Science
Virology
PNAS

Figure 1. Timeline of milestone studies aiming at the characterisation of XMRV as a human pathogen from 2006 to 2012 of interest to blood transfusion (further studies investigating XMRV within a particular disease entity are excluded).
Transfus Med. 2013 The rise and fall of XMRV. Kakisi OK, Robinson MJ, Tettmar KI, Tedder RS.