Approches de métagénomique

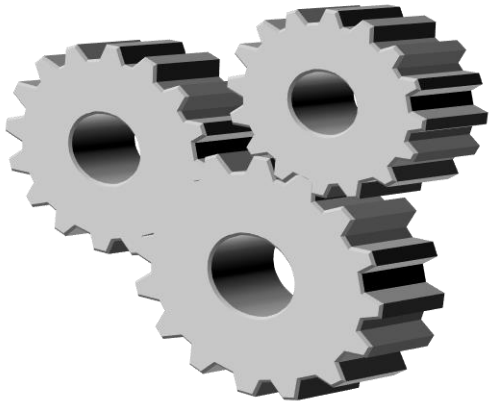
Contexte

Stratégies mises en place dans la validation des dons de sang

optimisation de la sécurité des produits sanguins

Connaissance toujours mise à jour des différents agents infectieux infectant l'homme et transmissibles par le sang

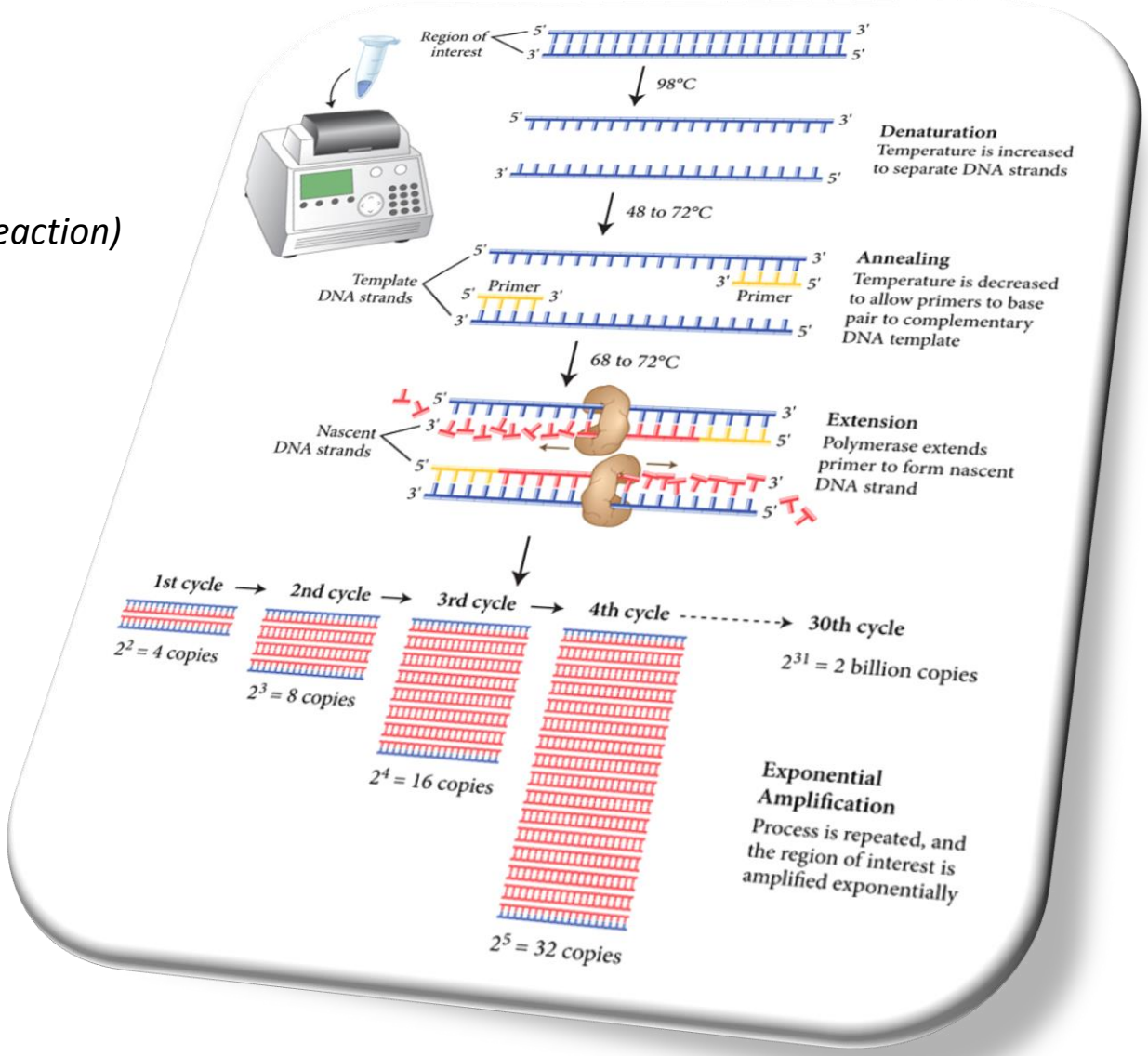
indissociable des avancées technologiques…



Recherche de nouveaux virus: quelles approches moléculaires…?

Approches séquences-spécifiques:

PCR *(polymerase chain reaction)*



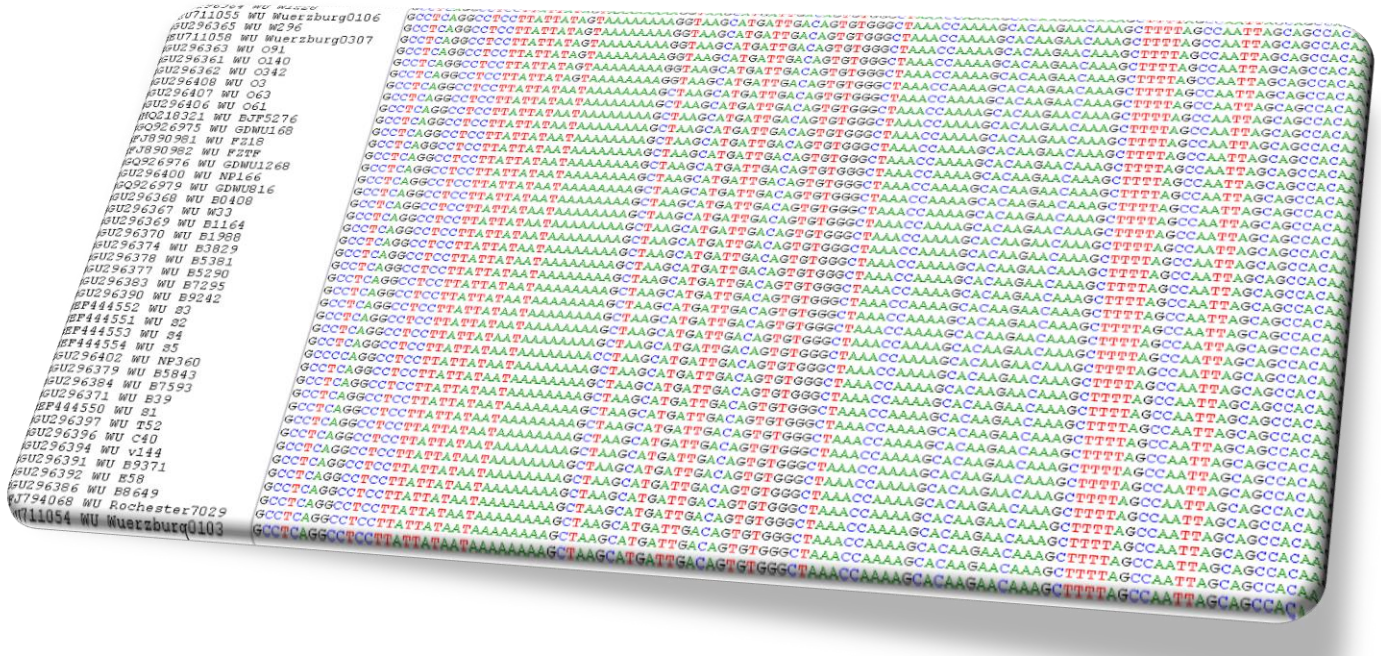
Utilisation de PCR consensus:

recherche de nouveaux variants d'une famille virale par ex.

 # données bibliographiques

```
>gi|363585120|gb|JN687847.1| HIV-1 isolate BP00060_nef from USA
envelope glycoprotein (env) and rev protein (rev) genes, partial cds;
and nef protein (nef) gene, complete cds
GGGTGGGAGATCCTCAAATATTGGTGGAACCTTCTACAATATTGGAGTCAGGAACTAAGGAATAGTGCTG
TTAGCTTGCTTAACGCCACAGCCATAGCAGTAGCTGAGGGGACAGATAGGGTTATAGAAGCAACACAAAG
AATTTTTAGAGCTTTTCTCCATATACCTAGAAGAATAAGACAGGGCTTTGAAAGGTGTTTGCTATAAGAT
GGGTGGCAAGTGGTCAAAAAGAAGTRTGGTTGGATGGCCTGCTGTAAGGGAACGGATGAGAAGAGCTGAG
CCAGCAGCAGAAGCAGCAGCAGACGGGGTGGGAGCAGTATCTCGAGACCTGGCAAGTCATGGAGCAATCA
CAAGTAGCAATACAGCAAGTAATAATGCTGATTGTGCCTGGCTGGAAGCACAAGAGGATGAGGAAGTGGG
TTTTCCAGTCAAACCCCAGGTACCTTTAAGACCAATGACTTACAAGGGAGCTGTAGATCTTAGCCACTTT
TTAAAAGAAAAGGGGGGACTGGATGGGTTAGTTTACTCCCAGCAAAGGCAAGATATCCTTGATCTGTGGG
TCTACCACACACAAGGCTACTTCCCTGATTGGCAGAATTACACACCAGGGCCAGGGGTCAGATATCCACT
GACCTTTGGATGGTGCTACAAGCTAGTACCAGTAGAGCCAGAGGAGGTAGAAAAGGCCAACGAAGGAGAG
AGAAACAACTTGTTACACCCTATGAGCCTGCATGGGATGGATGACCCAGAGAAAGAAGTGTTACAGTGGA
AGTTTGACAGCCGCCTAGCATTTCATCACATGGCCCGAGAGCTGCATCCGGAGTACTACAAGAACTGCTG
ACACCGAGCTTTCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGRAGGCGTGGCCTGGGCGGGACTGGG
```

 # alignements de séquences

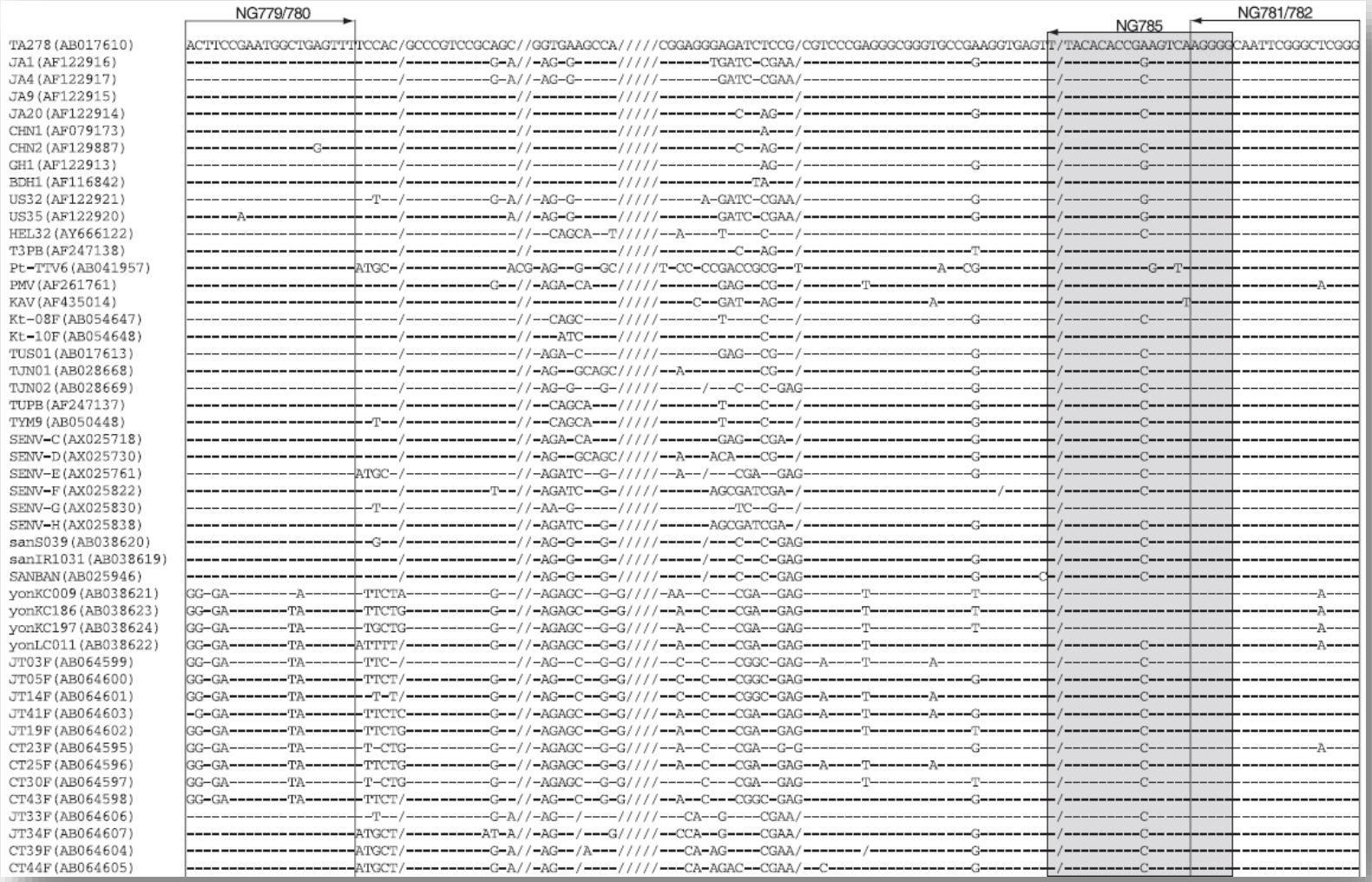


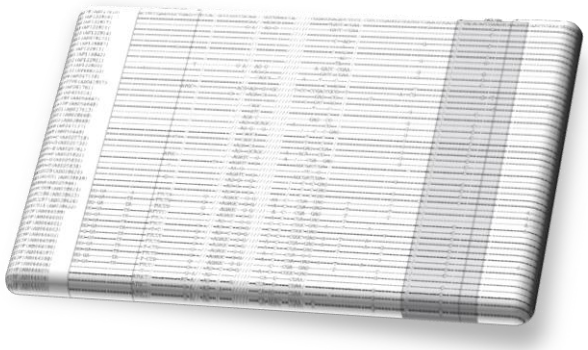
système de détection

sens 5'TTAGGATTGCTCTCGTAGGA3'

TTAGGATTGCTCTCGTAGGATCGATTGCTAGATGATTCATGGATG.....CTATGAGATCGCACCATAGCTAGATTAGGTTATTTA

3'CGATCTAATCCAATAAAT5' *rev*





Utilisation de <u>systèmes PCR consensus</u>…

Informations sur une courte portion de séquence

Long travail à la suite

Risque de faux-négatifs (variants très divergents)

Genre / Famille spécifique

Approches méthodologiques séquences-indépendantes

informations potentielles sur l'ensemble du génome

faisabilité ARN / ADN

genre / famille indépendant

découverte d'agents viraux

Approches méthodologiques séquences-indépendantes:

Méthodologie PEER (Primer extension enrichment reaction)

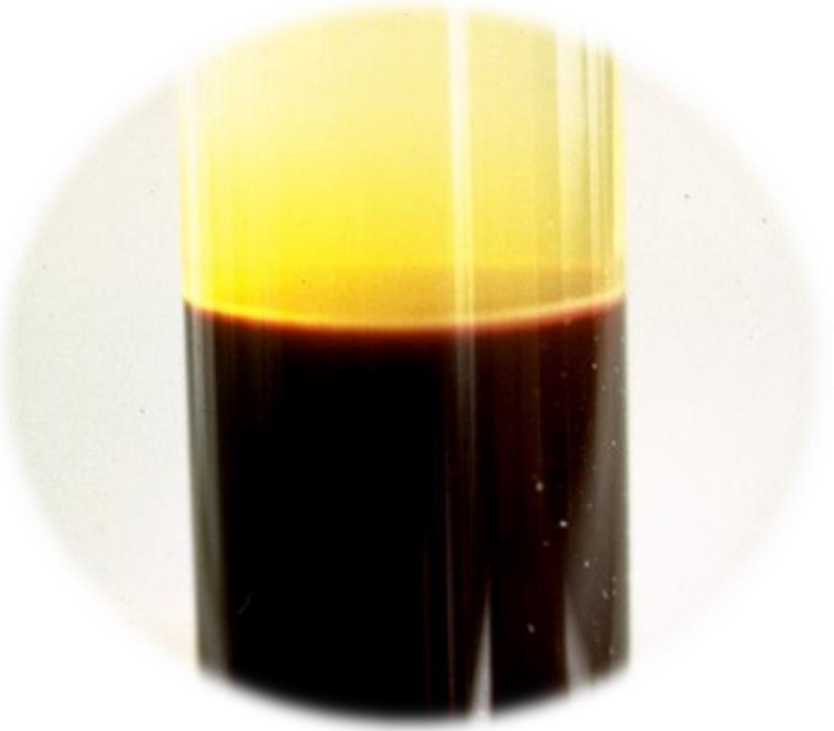
Soustraction génique

Méthodologie SISPA (Sequence-independent single primer amplification)

Restriction-ligation

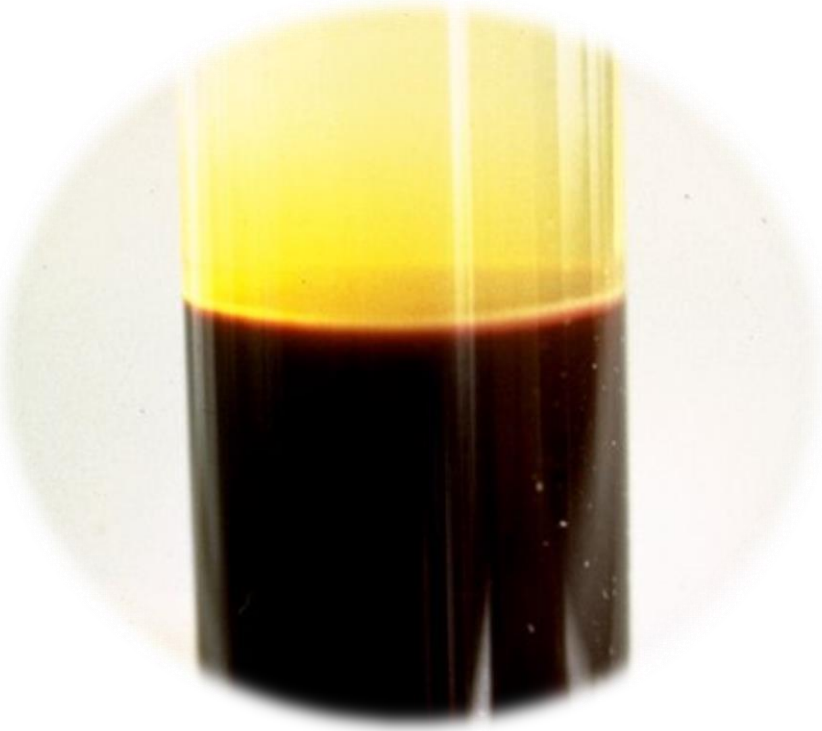
Allander T. et al. Proc Natl Acad Sci 2001; Ganova-Raeva L. et al. Nucl Acids Res 2006

PEER

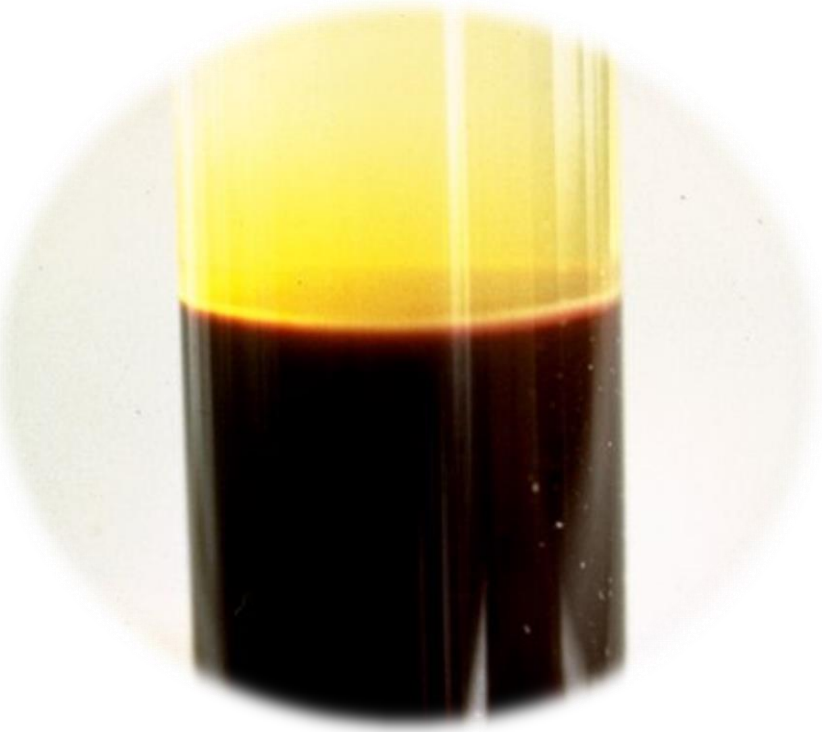


Temps t

PEER

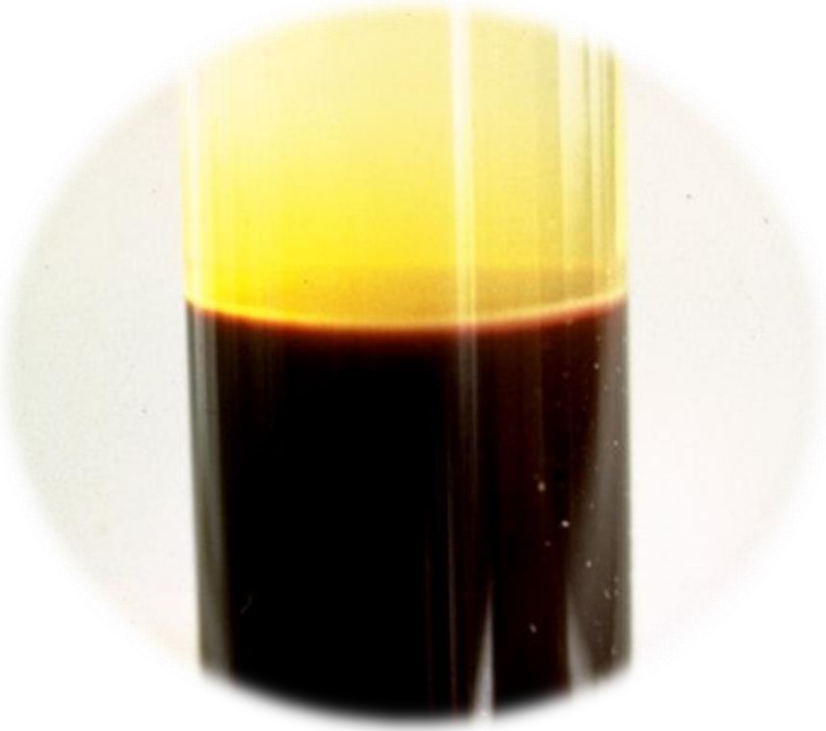


Temps t

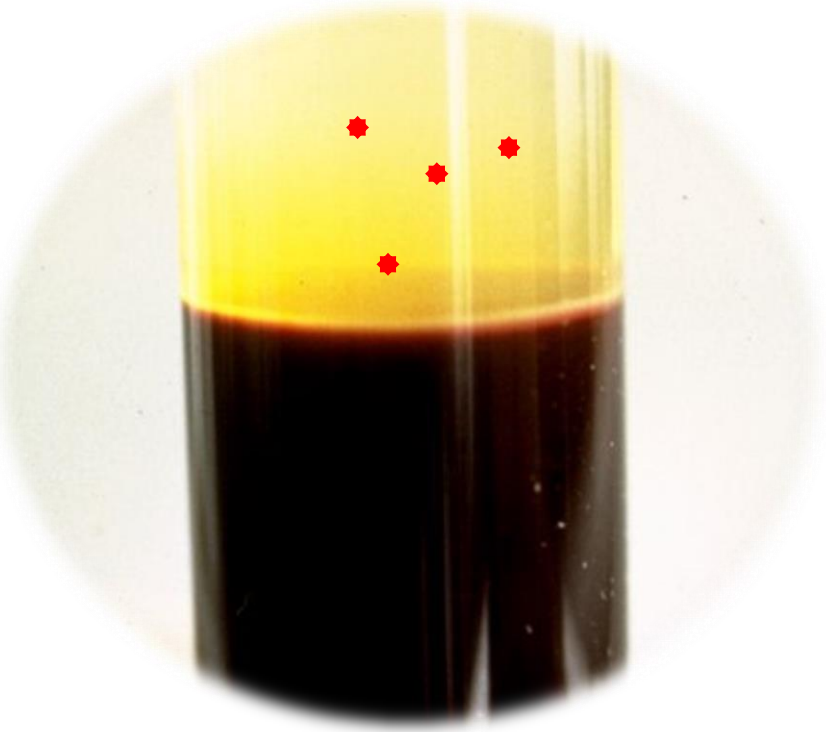


Temps t + x

PEER



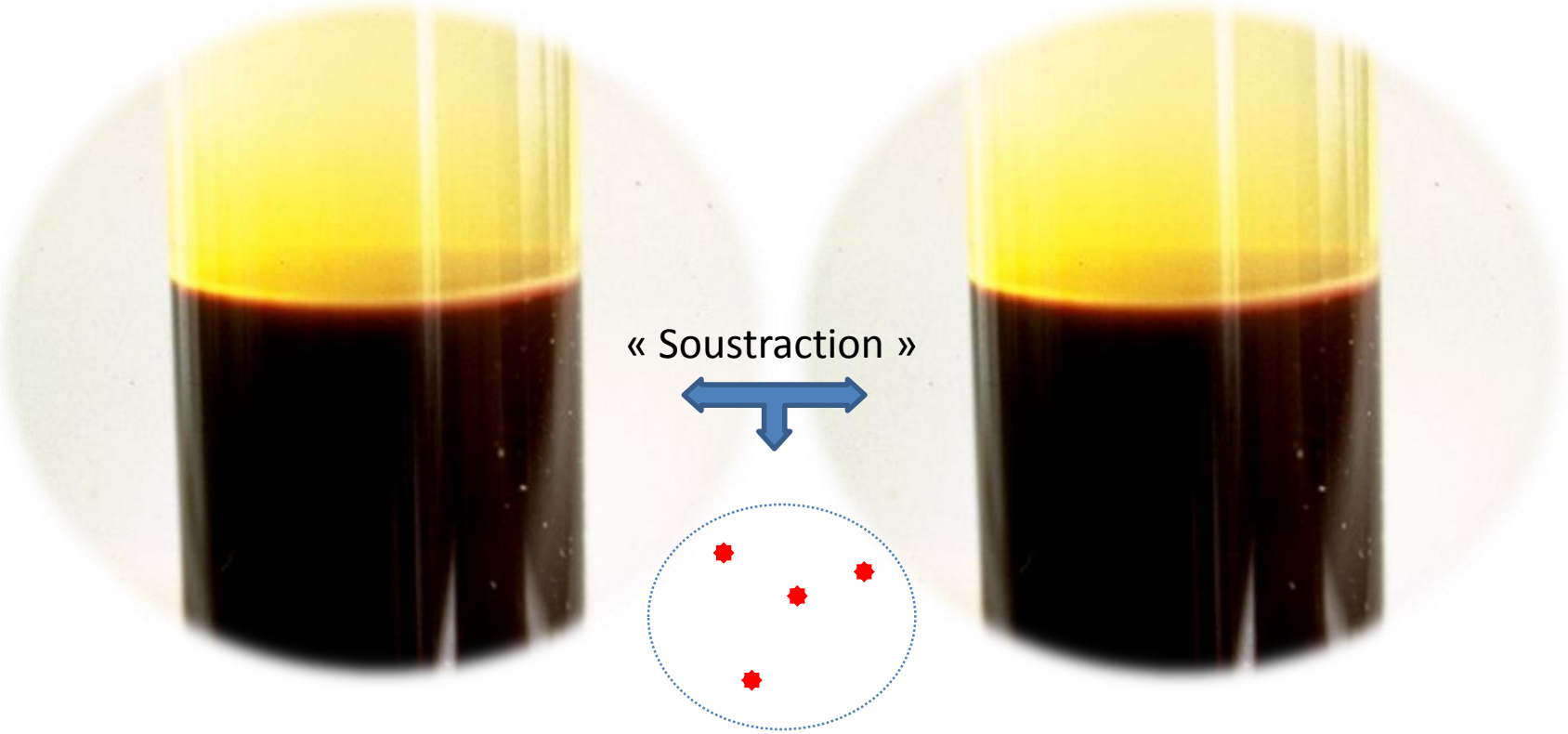
Temps t



Temps t + x

Evènement infectieux

PEER


« Soustraction »
Temps t
Temps t + x
Evènement infectieux
Processing…

PEER

Extraction des acides nucléiques

(tester : échantillon infecté / driver : échantillon non infecté)

RT-PCR avec incorporation d'amorces spécifiques

tester 1 (1 amorce commune aux 2 extrémités)

tester 2 (2 amorces distinctes aux 2 extrémités)

driver (1 amorce commune biotinylée aux 2 extrémités)

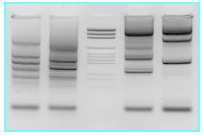
Digestion avec enzymes de restriction à site GC

Fixation d'adaptateurs / Digestion enzymatique

Hybridation amorces-tester 1 / driver

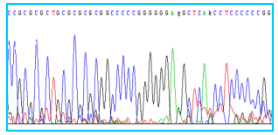
Purification des amorces-tester 1 non hybridées

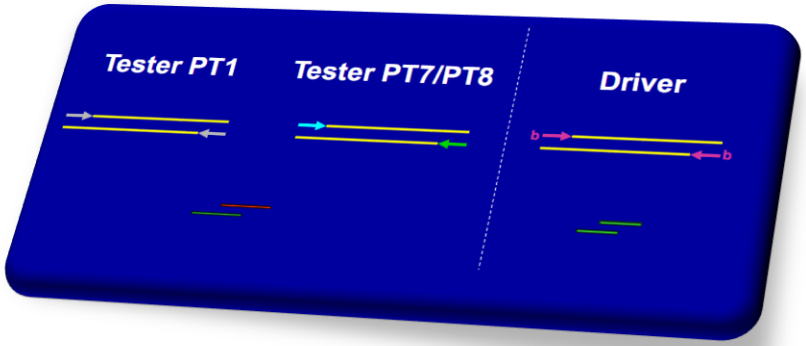
Hybridation amorces-tester 1 / tester 2



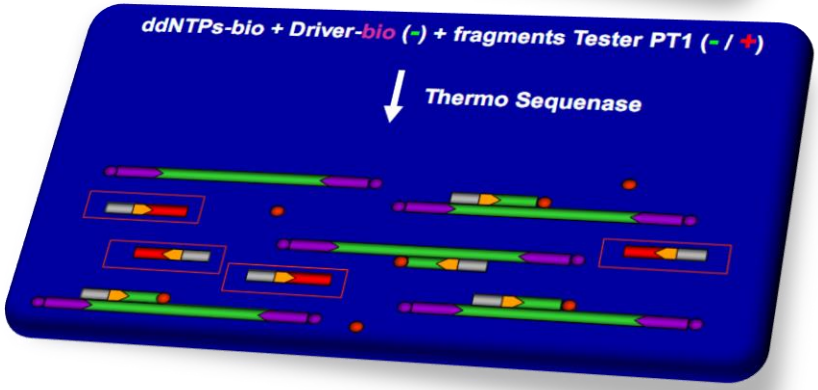
Elongation / PCR

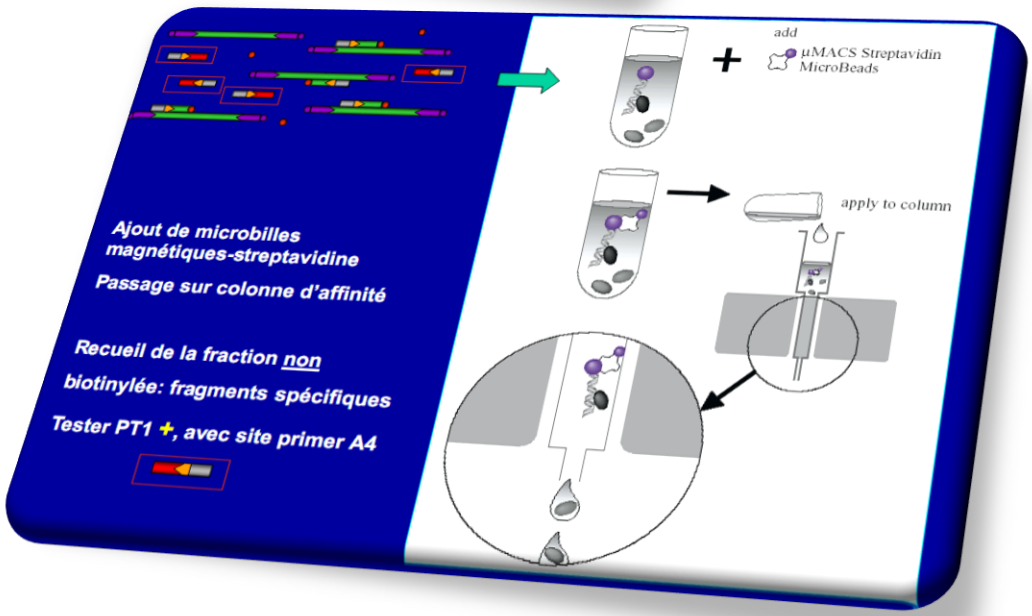
Clonages / Séquençages

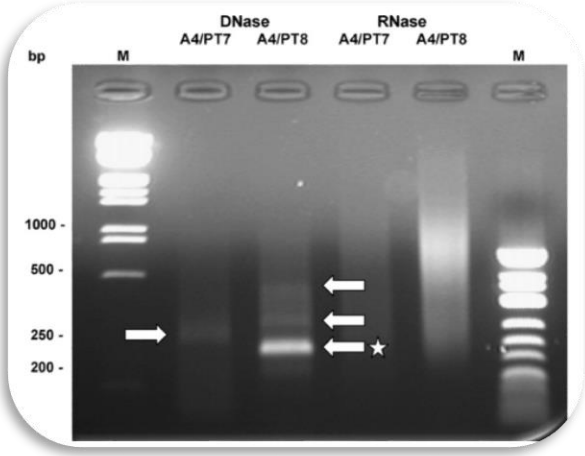



Tester PT1
Tester PT7/PT8
Driver
b
b


Digestion par BpmI:
GTAAACACAGAAGGAGGGCTGGAGnnnnnnnnnnnnnnnnnnnnnnnnn . . .
ATTTGTGTCTTCCTCCCGACCTCnnnnnnnnnnnnnnnnnnnnnnnnn . . .
site BpmI
5′ AdaptCTGGAGnnnnnnnnnnnnnnnnnnn... 3′
3′ AdaptGACCTCnnnnnnnnnnnnnnnnnnn... 5′
BpmI: coupure différée
« Amorces »
Tester PT1 (- / +)
avec extrémités
connues (A4)
16
5′ AdaptCTGGAGnnnnnnnnnnnnnnnn 3′
3′ AdaptGACCTCnnnnnnnnnnnnnnn 5′
14


ddNTPs-bio + Driver-bio (-) + fragments Tester PT1 (- / +)
Thermo Sequenase


Ajout de microbilles
magnétiques-streptavidine
Passage sur colonne d'affinité
Recueil de la fraction non
biotinylée: fragments spécifiques
Tester PT1 +, avec site primer A4
add
µMACS Streptavidin
MicroBeads
apply to column


DNase
RNase
A4/PT7
A4/PT8
A4/PT7
A4/PT8
bp
M
M
1000 -
500 -
250 -
200 -

Nucleic Acids Research, 2006, Vol. 34, No. 11 **e76**
doi:10.1093/nar/gkl391

Primer Extension Enrichment Reaction (PEER): a new subtraction method for identification of genetic differences between biological specimens

Lilia Ganova-Raeva*, Xinjian Zhang, Fengli Cao, Howard Fields and Yury Khudyakov

Centers for Disease Control and Prevention, National Center for Infectious Diseases, Division of Viral Hepatitis/Laboratory Branch, Atlanta, GA 30329, USA

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Journal of Virological Methods 139 (2007) 106–110

Journal of Virological Methods

www.elsevier.com/locate/jviromet

Short communication

Effective detection of highly divergent viral genomes in infected cell lines using a new subtraction strategy (primer extension enrichment reaction—PEER)

Philippe Biagini [a,*], Xavier de Lamballerie [b], Philippe de Micco [a,b]

[a] *Unité des Virus Emergents EA3292, Etablissement Français du Sang Alpes-Méditerranée, 27 Bd. Jean Moulin, 13285 Marseille, France*
[b] *Unité des Virus Emergents EA3292, Faculté de Médecine, Marseille, France*

Received 4 May 2006; received in revised form 12 September 2006; accepted 19 September 2006
Available online 25 October 2006

Abstract

The development of new molecular approaches for identification or discovery of known and unknown pathogens is essential. The purpose of this study was to evaluate the utility of a new method designated primer extension enrichment reaction (PEER) in detecting viral genomes from highly divergent viral families (ssDNA, dsDNA, ssRNA, dsRNA) in infected cell lines. Findings indicate that the PEER methodology is a powerful tool since it permitted successful *ex vivo* detection of infectious agents from the *Circoviridae, Herpesviridae, Paramyxoviridae, Picornaviridae* and *Reoviridae* families. The great potential of PEER was further confirmed in tests using serial dilutions of West Nile virus (*Flaviviridae*) and in a "blind" test in which a member of the *Poxviridae* family was successfully identified. The technique was shown to be useful for identifying causative viral pathogens, following inoculation of susceptible cell lines with clinical biological samples.

Keywords: Subtraction technique; Nucleic acids; PEER; Virus; Cell line

PEER

lan), resuspended in 50 μl of nuclease-free water, and processed immediately.

- *Step#2: half of each nucleic acid solution was treated using either 40 units of RNase-free DNase I (Roche) or 50 ng/μl of DNase-free RNase (Roche) at 37 °C for 60 min. Enzyme denaturation was then carried out for 3 min at 94 °C.*
- *Step #3: one driver and two testers were processed using 10 pmol of different primers:*

(i) *driver* PDSMART (5′-AAGCAGTGGTAACAACGCAGAGTACGCGGG-3′)
PDON7 (5′-AAGCAGTGGTAACAACGCAGAGTAIIIIIII-3′)

(ii) *tester I* PT1G (5′-ACACTCGAGGAGGTCTGGAGGG-3′)
PT1N7 (5′-ACACTCGAGGAGGTCTGGAGIIIIIII-3′)

(iii) *tester II* PT8G (5′-AAGCAGAGGCAGCATTGGAGGG-3′)
PT7N7 (5′-GAGCTGTGGTGAGTTGGTTGGAAIIIIIII-3′)

Five microliter of nuclease-treated nucleic acids were added to 2 μl of each primer, incubated at 70 °C during 3 min and placed on ice. Eleven microliter of the following solution was subsequently added: 1 μl powerscript reverse transcriptase (Clontech, Ozyme, Saint-Quentin Yvelines), 4 μl 5× first-strand buffer (Clontech), dNTPs 0.5 mM each (Roche), DTT 5 mM (Roche), 40 units RNase inhibitor (Invitrogen, Cergy Pontoise). The mix was incubated at 42 °C for 60 min followed by 3 min at 94 °C.

- *Step #4: for amplification reactions, 100 pmol of the following primers were used:*

(i) *driver* PDObio (5′biotin-AAGCAGTGGTAACAACGCAGAGT-3′)
(ii) *tester I* PT1 (5′-AACACTCGAGGAGGTCTGGAG-3′)
(iii) *tester II* PT7 (5′-AGCTGTGGTGAGTTGGTTGG-3′) PT8 (5′-AGCAGAGGCAGCATTGGAGG-3′)

Ten microliter of the products obtained in step #3 were used in a 100 μl reaction containing 2 μl of 50× advantage 2 polymerase mix (Clontech), 10 μl 10× polymerase buffer (Clontech), dNTPs 0.2 mM each (Clontech), and corresponding primers. Amplifications were carried out using 30 cycles of denaturation at 94 °C for 30 s followed by annealing and extension at 68 °C for 2 min. PCR products were subsequently purified using the QIAquick PCR purification kit (Qiagen, Courtaboeuf) and resuspended in 75 μl of nuclease-free water.

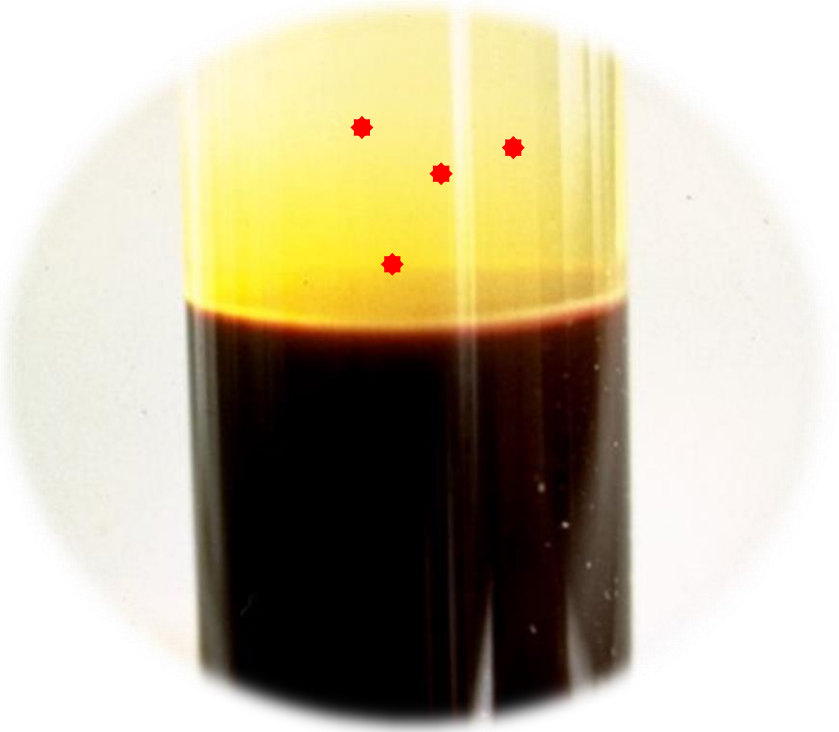
- *Step #5: 70 μl of tester I obtained in step #4 were incubated overnight at 37 °C in a 100 μl reaction mix containing 10× TaqI buffer (New England Biolabs, Ozyme, Saint-Quentin Yvelines) and the following endonucleases: TaqI, HpaII, HinPII, AciI, (New England Biolabs) and MaeII (Roche).*

Digested products were then heat-inactivated for 3 min at 94 °C, further purified using the MinElute Reaction Cleanup kit (Qiagen), and resuspended in 55 μl EB buffer (Tris–HCl 10 mM, pH 8.5).

- *Step #6: 50 μl of digested products were treated using 2 units of Klenow Enzyme (Roche) in a 60 μl reaction containing dCTP (0.17 mM) and 6 μl 10× polymerase buffer. The mix was incubated at 37 °C for 90 min, then heat-inactivated and purified using the MinElute kit. Klenow-treated products were resuspended in 50 μl EB buffer.*
- *Step #7: an equimolar proportion (200 pmol each) of primers A4FS (5′-AAGTAAACACAGAAGGAGGGCTGGAG-3′) and A4RSP (5′phosphate-GCTCCAGCCCTCCTTCTGTGTTTAC-3′) were mixed, heated at 94 °C for 5 min and cooled to room temperature. Five microliter of the resulting adapter (Ad4SP) were combined with 45 μl of products from step #6, 4 units of T4 DNA Ligase (Roche), and 6 μl 10× ligation buffer in a 60 μl reaction mix, and incubated at room temperature for 3 h. The mix was purified using the MinElute kit and resuspended in 55 μl EB buffer.*
- *Step #8: 52.4 μl of ligation products were incubated at 37 °C for 2 h in a 60 μl mix containing 1× BSA, 6 μl 1× buffer #3 and 2 units of BpmI (New England Biolabs), then purified (MinElute kit) and resuspended in 50 μl EB buffer.*
- *Step #9: 23 μl of digested products were mixed with 10 μl of the driver obtained in step #4 in a 50 μl reaction containing 32 units of Thermo Sequenase DNA Polymerase (Amersham Biosciences, Saclay), biotin-11-ddNTPs (0.1 mM each) (PerkinElmer Life Sciences, Boston, MA), 1× TS reaction buffer, and 1× TS dilution buffer (Amersham Biosciences). The blocking reaction consisted of 55 cycles of denaturation at 94 °C for 20 s, annealing at 45 °C for 20 s, and extension at 72 °C for 20 s.*
- *Step #10: products from step #9 were mixed with 50 μl of streptavidin-coated magnetic beads (μMACS Streptavidin kit, Miltenyi Biotec, Paris), incubated at 60 °C, and loaded into a μMACS column previously equilibrated with 200 μl nucleic acid equilibration buffer (μMACS Streptavidin kit, Miltenyi Biotec). The flow-through was collected and 100 μl EB buffer were added, immediately, to the column and further collected in the same tube.*
- *Step #11: 38.5 μl of column-purified products were used in a 50 μl reaction mix containing 5 μl of tester II obtained at step #4, 2.5 units of Takara Ex taq polymerase (Takara Shuzo, BioWhittaker, Emerainville), dNTPs (0.2 mM each) (Roche), and 1× polymerase buffer (Takara Shuzo). Amplification consisted of: (i) 15 cycles of denaturation at 94 °C for 20 s, annealing at 45 °C for 30 s and extension at 68 °C for 90 s, followed by (ii) 15 cycles of denaturation at 94 °C for 20 s, annealing at 50 °C for 30 s and extension at 68 °C for 90 s.*
- *Step #12: 5 μl of the products obtained at step #11 were used as templates in 100 μl PCR reactions using primers A4 (5′-ACACAGAAGGAGGGCTGGAG-3′) and PT7 (5′-AGCTGTGGTGAGTTGGTTGG-3′), or A4 and PT8 (5′-AGCAGAGGCAGCATTGGAGG-3′) (100 pmol each), 5 units of Takara Ex taq polymerase (Takara Shuzo), dNTPs (0.2 mM each) (Roche) and 1× polymerase buffer (Takar-*

J Virol Methods 2007

SISPA



Temps t

Evènement infectieux

SISPA

x1

Pré-traitement de l'échantillon à la nucléase

Extraction des acides nucléiques

(soit ARN, soit ADN)

Transcription inverse-Polymérase

(synthèse ADN double brin)

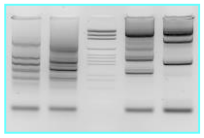
Incubation avec enzyme de restriction

Csp6 / TaqI

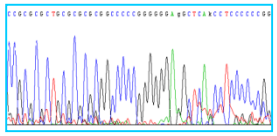
Ligation des produits obtenus

à un adaptateur **-Csp6 / -TaqI**

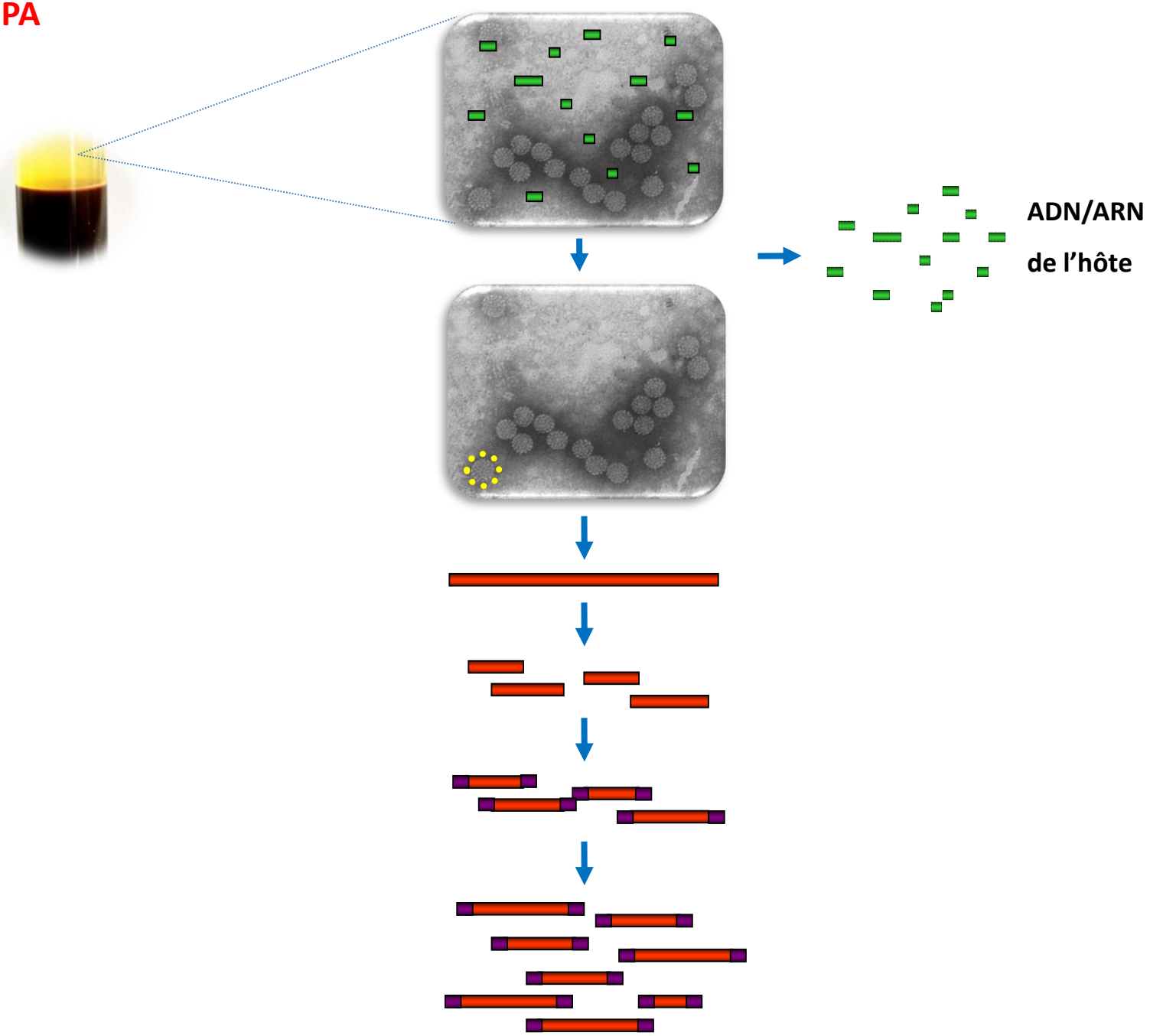
PCR



Purifications / Clonages / Séquençages



SISPA



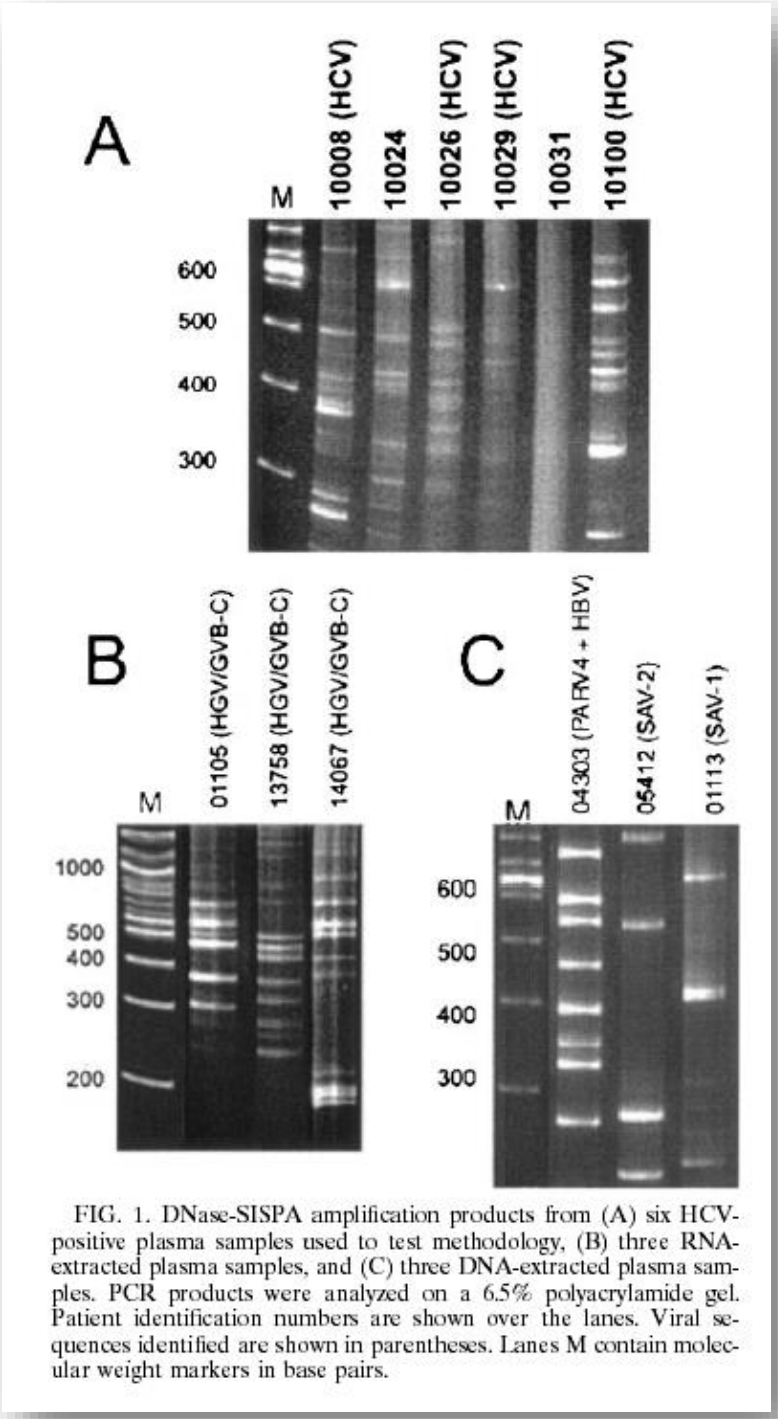


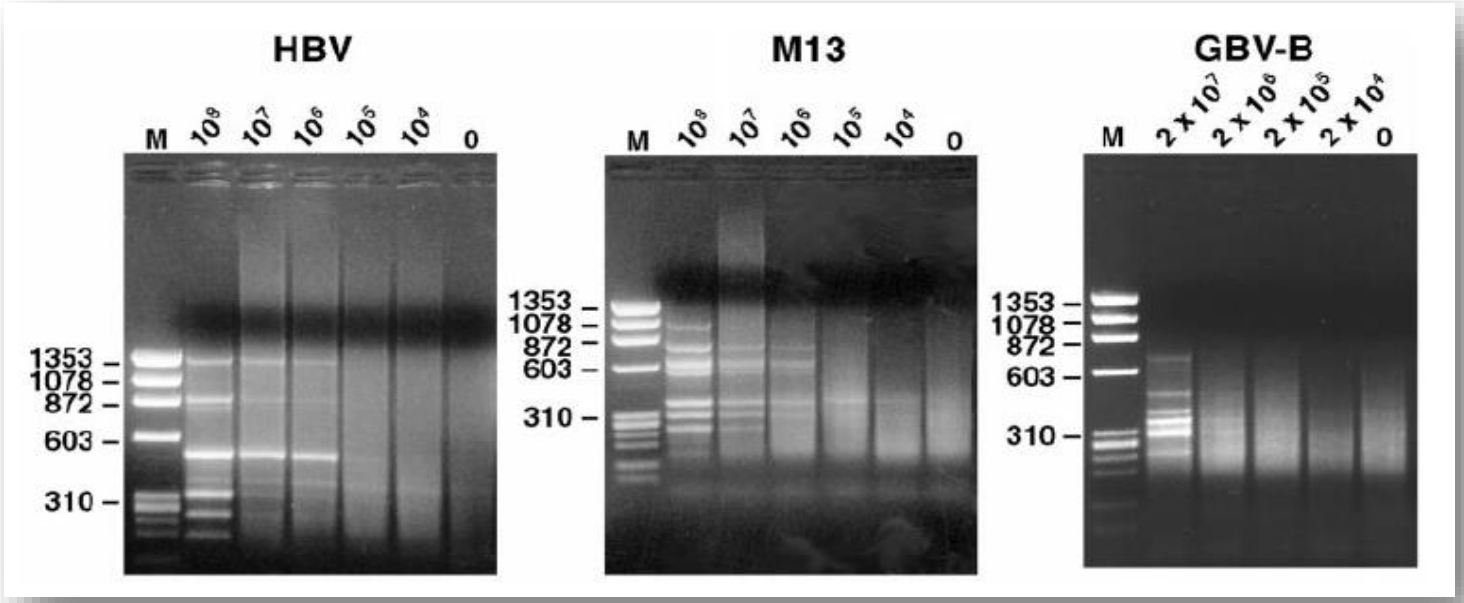
FIG. 1. DNase-SISPA amplification products from (A) six HCV-positive plasma samples used to test methodology, (B) three RNA-extracted plasma samples, and (C) three DNA-extracted plasma samples. PCR products were analyzed on a 6.5% polyacrylamide gel. Patient identification numbers are shown over the lanes. Viral sequences identified are shown in parentheses. Lanes M contain molecular weight markers in base pairs.

Découverte:

Parvovirus 4

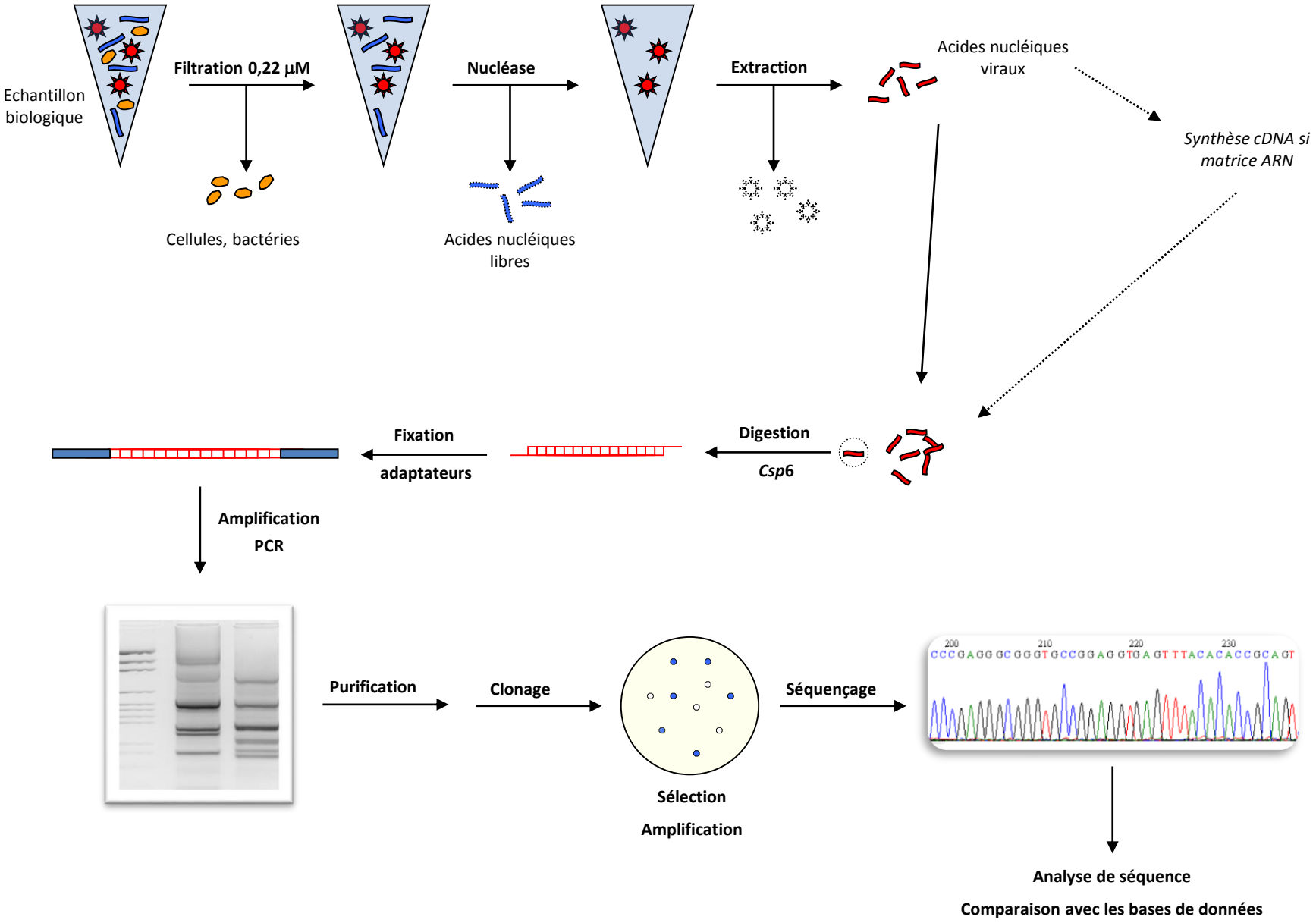
Small anellovirus

…Sensibilité…

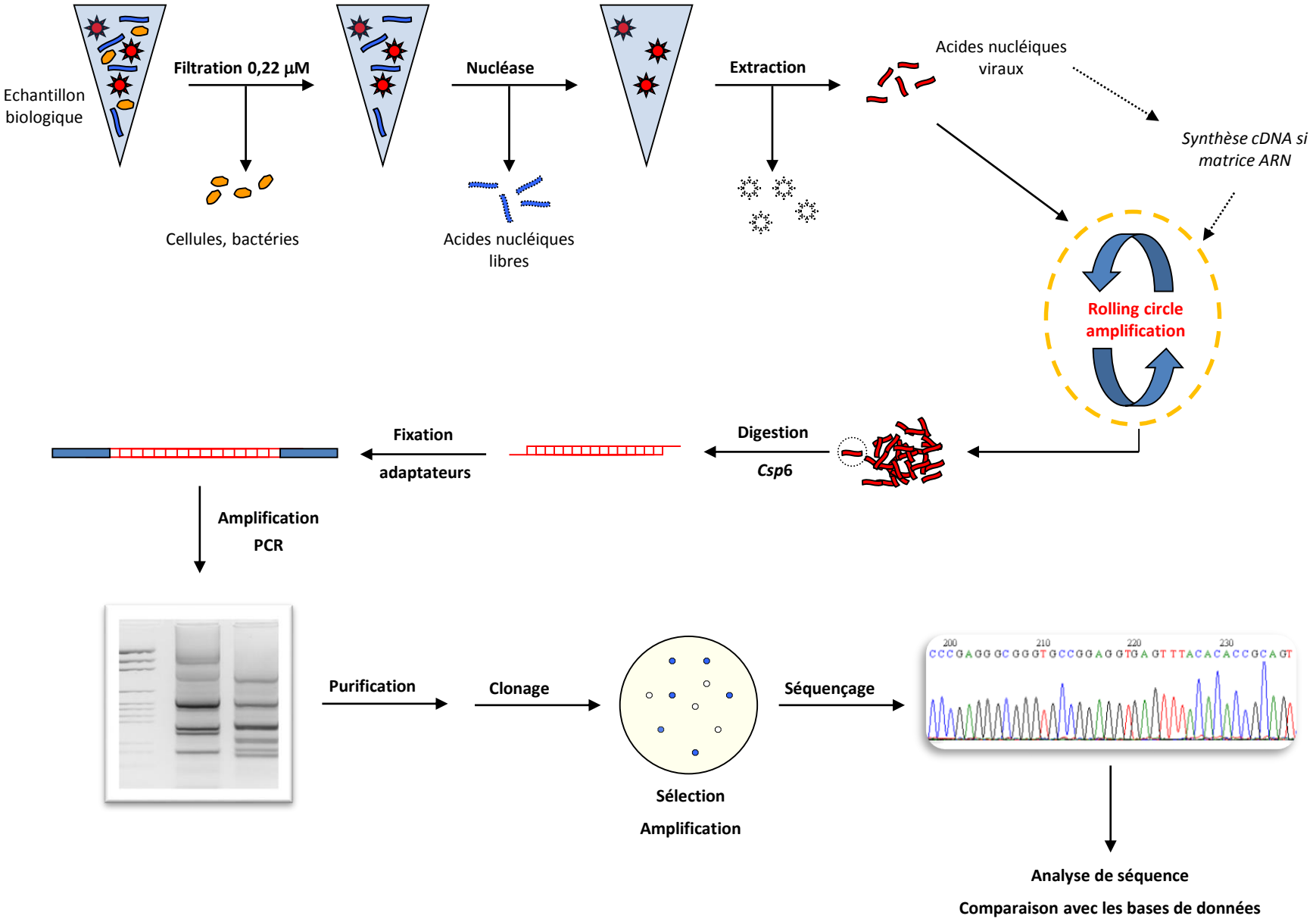


… <u>Faible</u> pouvoir de détection en dessous de 10^6 copies/ml…

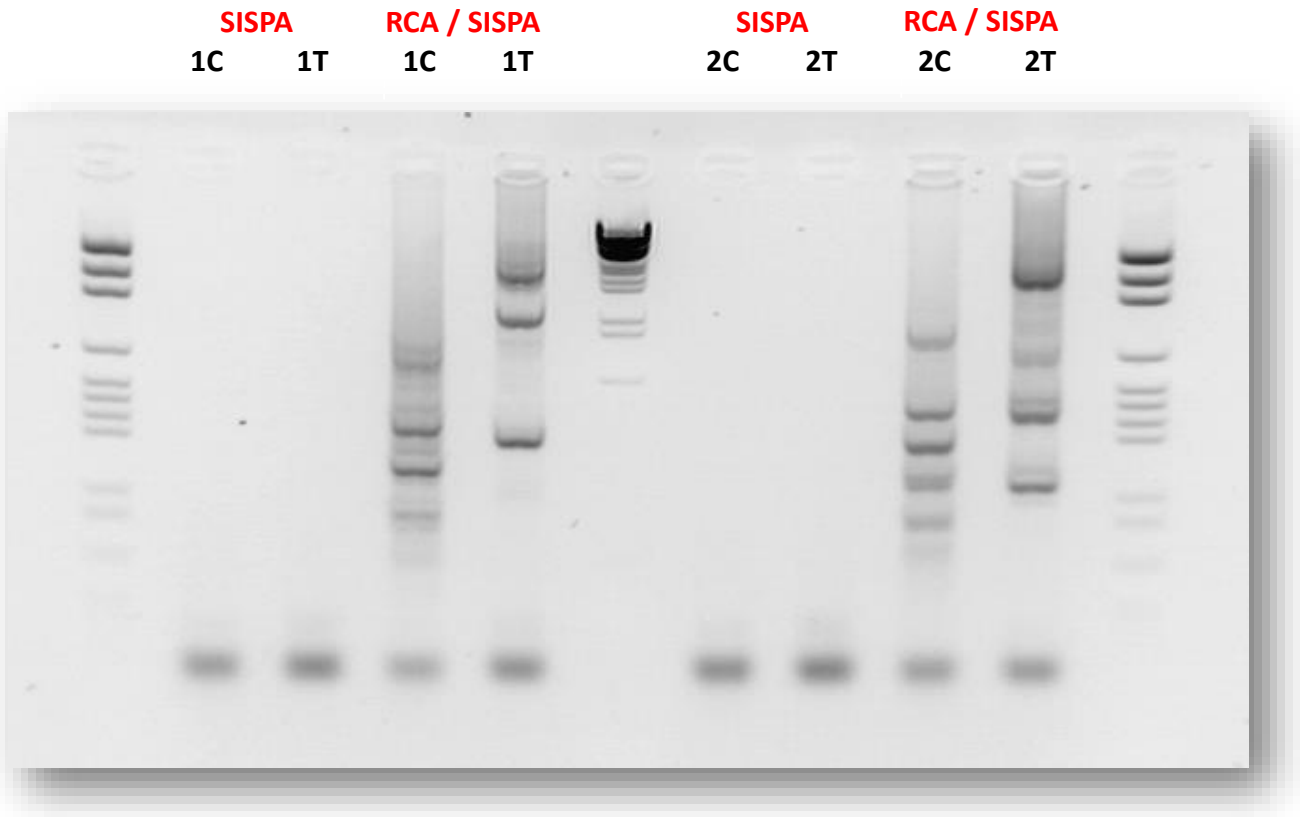
SISPA


Echantillon biologique
Filtration 0,22 µM
Cellules, bactéries
Nucléase
Acides nucléiques libres
Extraction
Acides nucléiques viraux
Synthèse cDNA si matrice ARN
Digestion
Csp6
Fixation
adaptateurs
Amplification
PCR
Purification
Clonage
Sélection
Amplification
Séquençage
200
210
220
230
CCCGAGGGCGGGTGCCGGAGGTGAGTTTACACACCGCAGT
Analyse de séquence
Comparaison avec les bases de données

RCA-SISPA


Echantillon biologique
Filtration 0,22 µM
Cellules, bactéries
Nucléase
Acides nucléiques libres
Extraction
Acides nucléiques viraux
Synthèse cDNA si matrice ARN
Rolling circle amplification
Digestion
Csp6
Fixation
adaptateurs
Amplification
PCR
Purification
Clonage
Sélection
Amplification
Séquençage
200
210
220
230
CCCGAGGGCGGGTGCCGGAGGTGAGTTTACACACCGCAGT
Analyse de séquence
Comparaison avec les bases de données

Echantillons plasma 1 et 2…



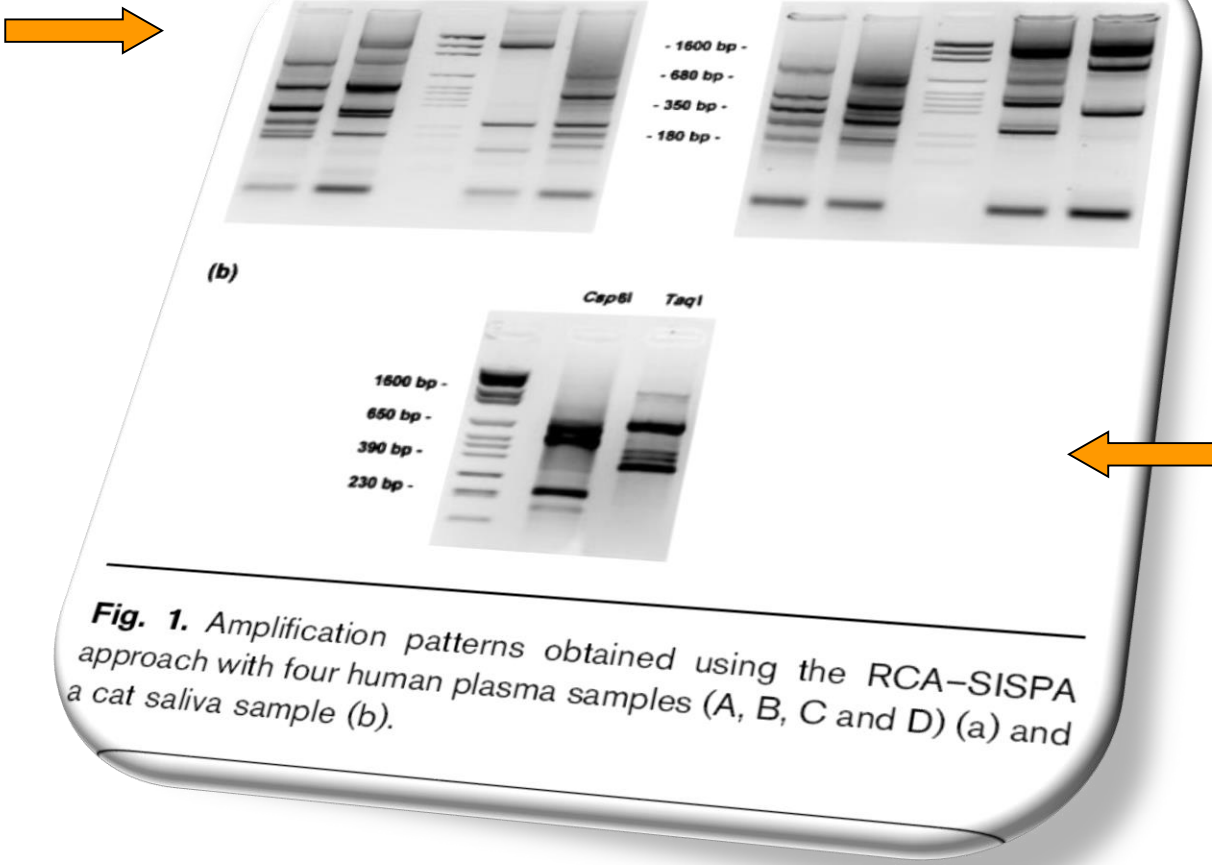
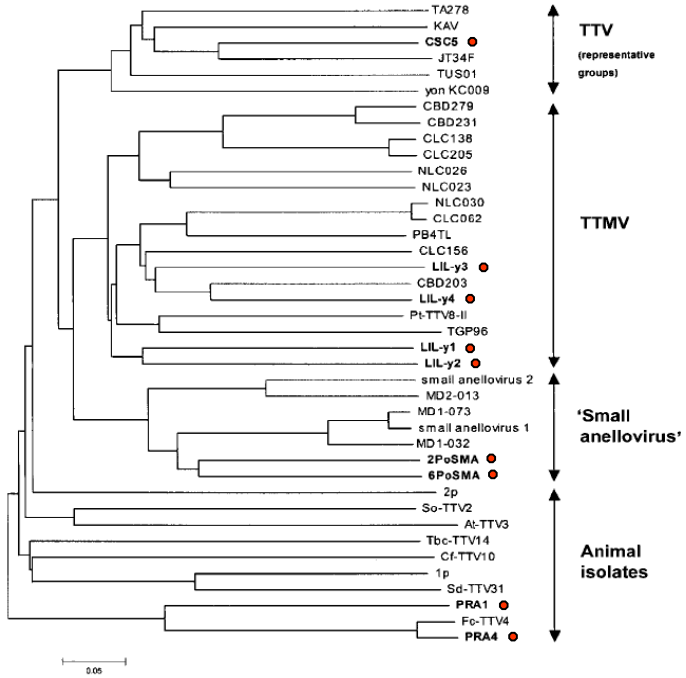


Fig. 1. Amplification patterns obtained using the RCA–SISPA approach with four human plasma samples (A, B, C and D) (a) and a cat saliva sample (b).

J Gen Virol 2007

Fig. 2. Predicted genomic structure of representative isolates characterized in this study: (a) LIL-y1 and LIL-y2, TTMV species; (b) 2PoSMA and 6PoSMA, 'small anellovirus' species; (c) PRA1, animal species. The closed arrows represent major ORFs (above 50 aa). (d) Predicted secondary structure formed by the GC-rich region of isolates LIL-y1 and LIL-y2.

➡ **Identification de nouveaux variants hautement divergents (famille *Anelloviridae*)**



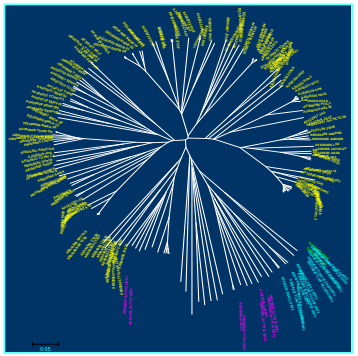


Fig. 3. Neighbour-joining phylogenetic tree constructed using the entire nucleotide sequence of ORF1 (due to the large number of complete sequences already described for species TTV, only highly divergent representative sequences were included).

J Gen Virol 2007

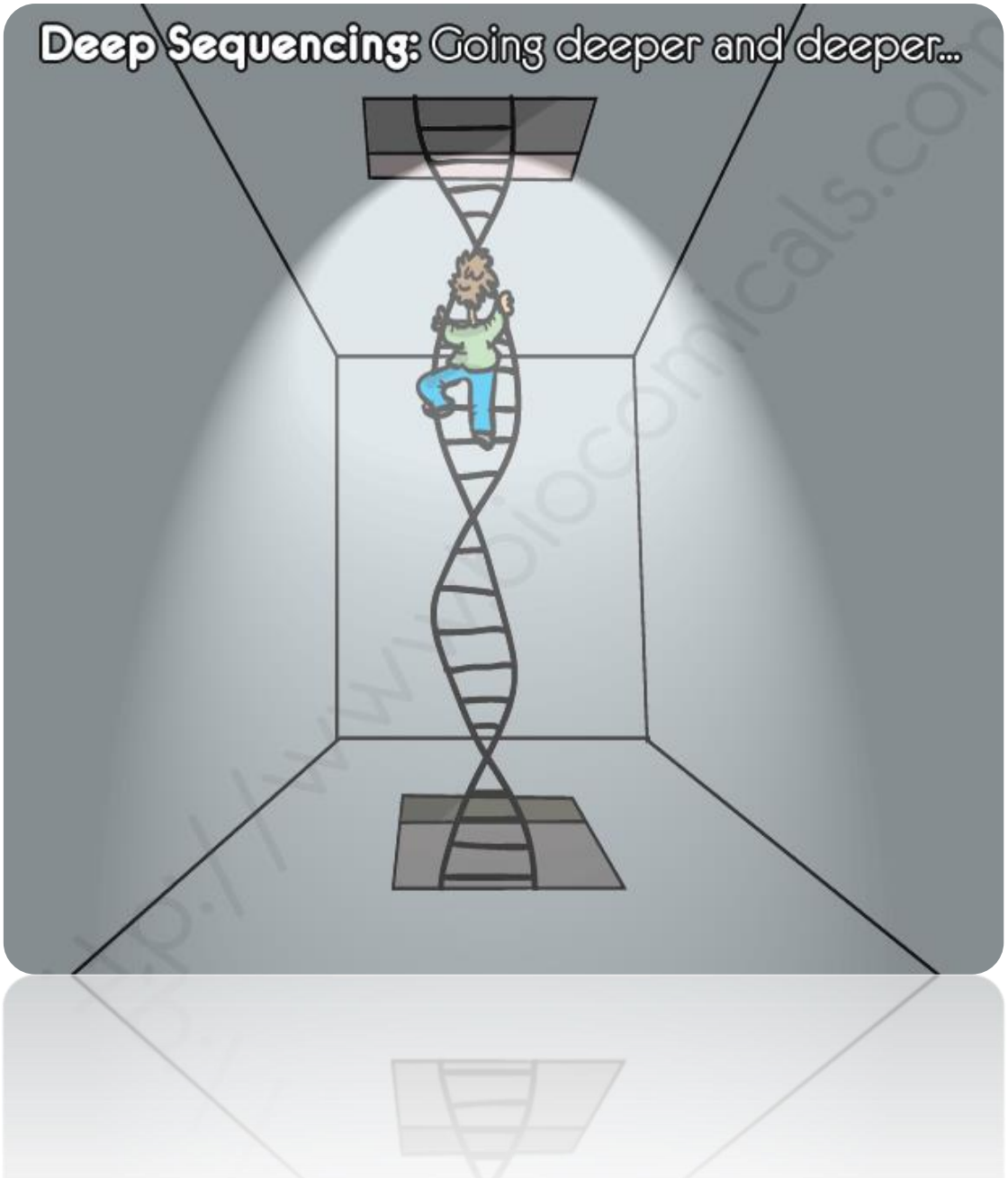
… Analyse des séquences…

clones: < 100

Métagénomique

NGS

séquences: > 100.000



+

représentées

Gradient de densité en espèces virales

-

représentées

+

représentées

Gradient de densité en espèces virales

-

représentées

SISPA

+

représentées

Gradient de densité en espèces virales

-

représentées

RCA-SISPA

+

représentées

Gradient de densité en espèces virales

-

représentées

NGS

+

représentées

Gradient de densité en espèces virales

-

représentées

NGS

+ [ ]

Exploring the viral world through metagenomics

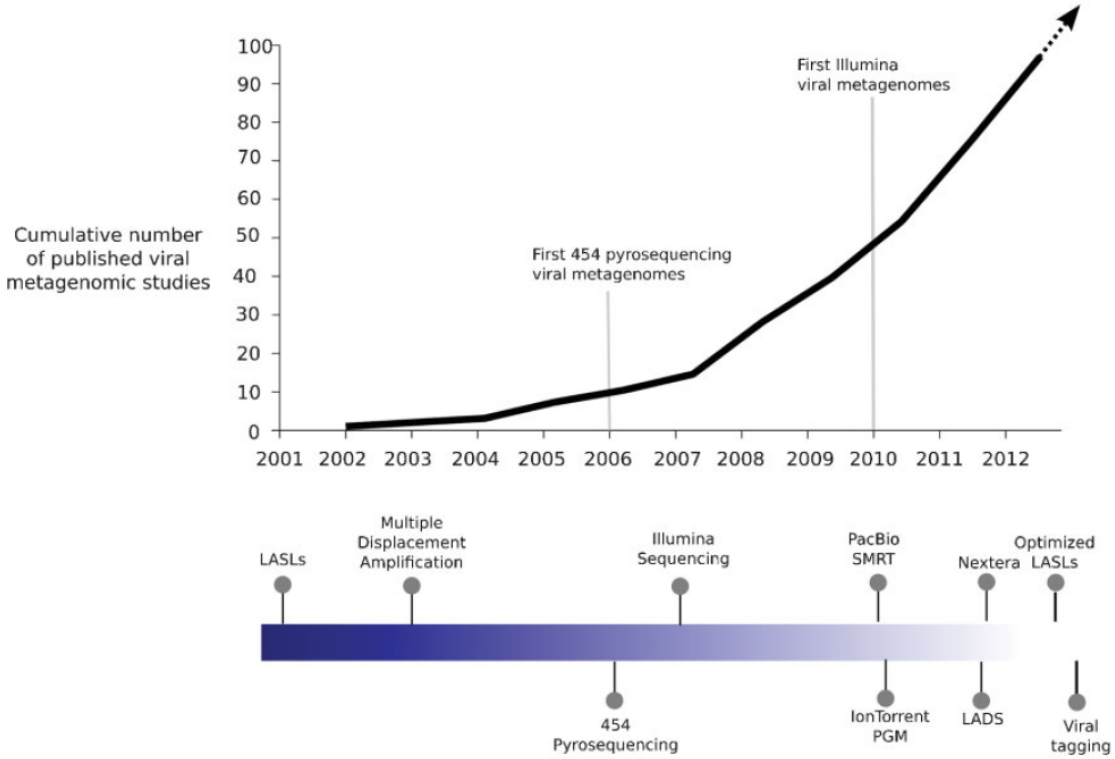


Figure 1. Advances in viral metagenomics from 2002 to 2012. The cumulative number of published studies is shown on the figure, beginning with the publication of the first viral metagenomes in 2002. Technological advancements are highlighted on the bar below the figure, presented according to the year of their publication and/or widespread release.

DNA viral families identified in environmental samples and organisms through viral metagenomic appr

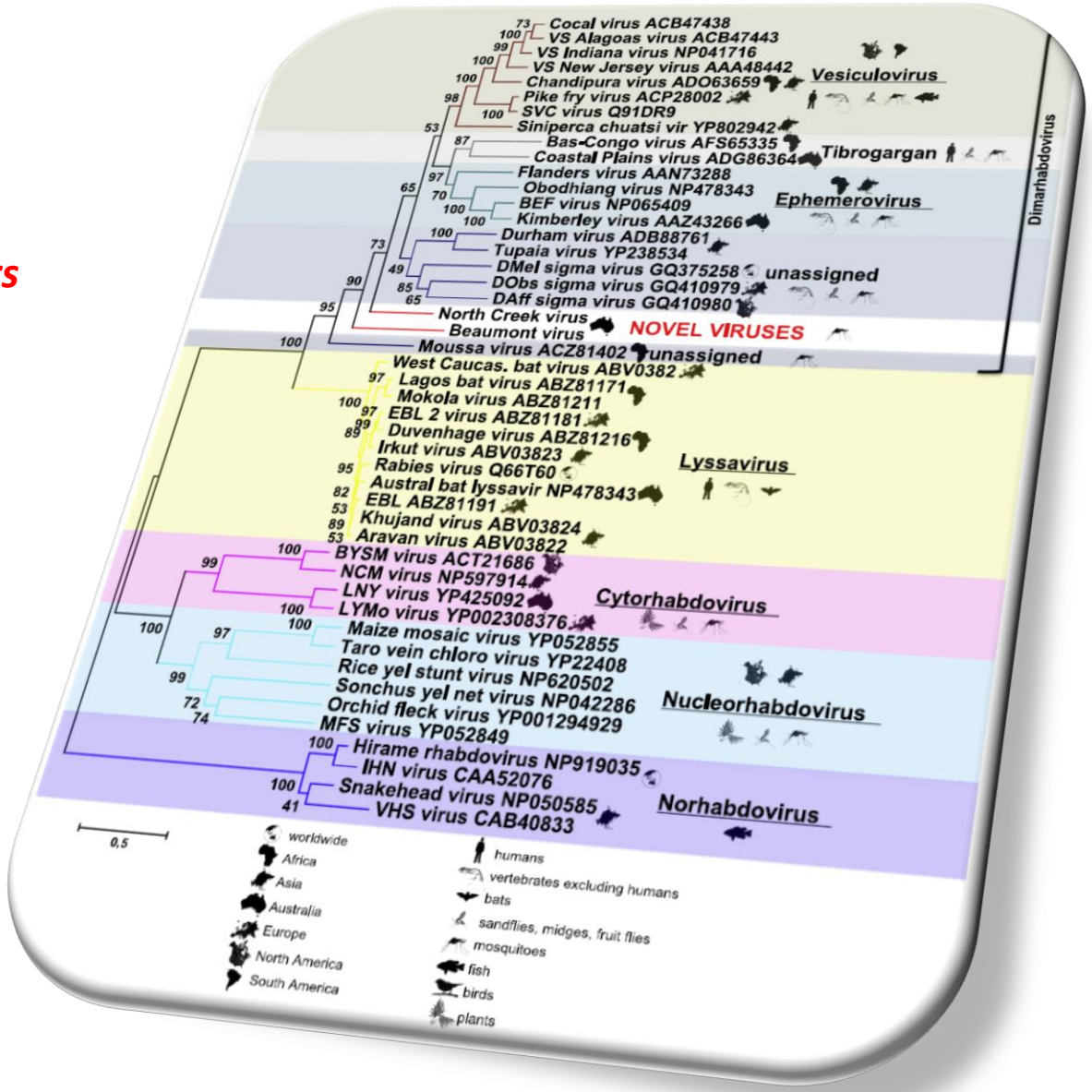
Viral family	Environment or organism	Known host
Single-stranded DNA		
Anelloviridae	Harbor seals, Humans, Mosquitoes, Sea lions	Mammals
Gyrovirus genus	Sea turtles	Birds
Circoviridae	Antarctic lake, Man-made lake, Reclaimed water, British Columbia coastal waters, Chesapeake Bay, Sargasso Sea, Rice paddy soil, Bats, Chimpanzees, Dragonflies, Mosquitoes	Mammals/Birds
Microviridae	Antarctic lake, Reclaimed water, British Columbia and California coastal waters, Gulf of Mexico, Sargasso Sea, Tampa Bay, Rice paddy soil, Bats, Freshwater and marine microbialites	Bacteria
Inoviridae	Antarctic lake, Reclaimed and potable water, Arctic Ocean, Marine sediment, Fermented food	Bacteria
Nanoviridae	Antarctic lake, Reclaimed water, Strait of Georgia, Tampa Bay, Rice paddy soil, Chimpanzees, Corals, Mosquitoes	Plants
Geminiviridae	Antarctic lake, Reclaimed water, Rice paddy soil, Chimpanzees, Corals, Mosquitoes	Plants
Parvoviridae	Bats, Corals, Mosquitoes, Turkeys, Fermented food	Insects/Mammals
Double-stranded DNA		
Siphoviridae	All DNA metagenomes	Bacteria/Archaea
Podoviridae	All DNA metagenomes	Bacteria
Myoviridae	All DNA metagenomes	Bacteria/Archaea
Rudiviridae	Antarctic lake, Hot springs, Fermented food	Archaea
Fuselloviridae	Hot springs, Fermented food	Archaea
Herpesviridae	Antarctic lake, Reclaimed water, Hot springs, Arctic Ocean, Gulf of Mexico, Corals, Mosquitoes, Fermented food	Mammals
Mimiviridae	Antarctic lake, Reclaimed water, Arctic Ocean, Hydrothermal vent, Gulf of Mexico, Corals, Fermented food	Algae/Protists
Phycodnaviridae	Antarctic lake, Reclaimed water, Arctic Ocean, Chesapeake Bay, Hydrothermal vent, Corals, Fermented food	Algae
Iridoviridae	Antarctic lake, Reclaimed water, British Columbia coastal waters, Hydrothermal vent, Corals, Mosquitoes, Fermented food	Fish/Amphibians/Insects
Poxviridae	Antarctic lake, Reclaimed water, Hydrothermal vent, Bats, Corals, Horses, Mosquitoes	Mammals/Insects
Nimaviridae	Antarctic lake, Arctic Ocean, Corals	Crustaceans
Adenoviridae	Bats, Corals, Fermented food	Mammals
Asfarviridae	Antarctic lake, Arctic Ocean, Corals	Mammals
Polydnaviridae	Antarctic lake, Corals	Insects
Baculoviridae	Antarctic lake, Arctic Ocean, Corals	Insects/Crustaceans
Papillomaviridae	Arctic Ocean, Corals, Mosquitoes	Mammals
Lipothrixviridae	Antarctic lake	Archaea
Plasmaviridae	Antarctic lake	Bacteria
Ascoviridae	Antarctic lake, Fermented food	Insects
Tectiviridae	Rice paddy soil, Fermented food	Bacteria
Caulimoviridae	Corals, Fermented food	Plants

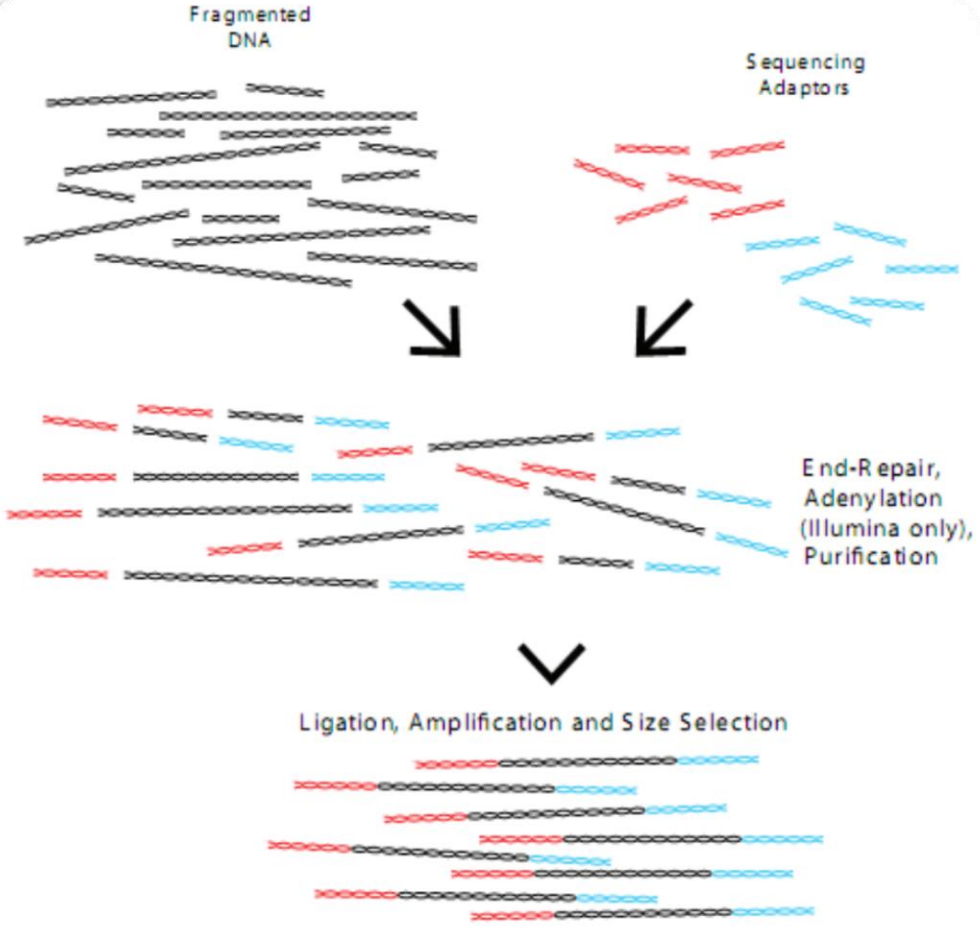
RNA viral families identified in environmental samples and organisms through viral metagenomic approaches[a]

	Viral family	Environment or organism	Known host
Positive (+) and negative (−) sense single-stranded RNA			
+	Retroviridae	Arctic Ocean, Corals, Fermented food	Birds/Mammals
	Comovirinae	Man-made lake, Reclaimed water, English Bay	Plants
	Dicistroviridae	Man-made lake, Reclaimed water, English Bay, Strait of Georgia, Bats	Insects
	Marnaviridae	Man-made lake, English Bay	Protists
	Picornaviridae	Man-made lake, Reclaimed water, English Bay, Bats, Ringed seal, Turkeys	Birds/Mammals
	Sequiviridae	Man-made lake, Reclaimed water	Plants
	Iflavirus	Man-made lake, Bats	Insects
	Tombusviridae	Man-made lake, Reclaimed water, Strait of Georgia, Humans	Plants
	Umbravirus	Man-made lake, Strait of Georgia	Plants
	Nodaviridae	Man-made lake, Bats	Fish/Insects
	Sobemovirus	Man-made lake, Bats, Humans	Plants
	Hepeviridae	Man-made lake	Birds/Mammals
	Flaviviridae	Man-made lake	Mammals
	Virgaviridae	Man-made lake, Reclaimed water, Humans	Plants
	Flexiviridae	Man-made lake	Plants
	Potyviridae	Reclaimed water	Plants
	Nora virus	Man-made lake, Reclaimed water	Insects
	Leviviridae	Man-made lake, Turkeys	Bacteria
	Tymoviridae	Man-made lake, Bats, Humans, Turkeys	Plants
	Tetraviridae	Man-made lake, Bats	Insects
	Astroviridae	Bats, Turkeys	Birds/Mammals
	Luteoviridae	Man-made lake, Bats	Plants
	Secoviridae	Bats	Plants
	Coronaviridae	Arctic Ocean, Bats	Birds/Mammals
	Bromoviridae	Man-made lake, Humans	Plants
−	Orthomyxoviridae	Man-made lake	Birds/Mammals
Double-stranded RNA			
	Picobirnavirus	Man-made lake, Humans, Turkeys	Mammals
	Reoviridae	Man-made lake, Reclaimed water, English Bay	Birds/Mammals/Insects/ Plants/Fungi
	Partiviridae	Man-made lake, Bats	Plants/Protists/Fungi

Enhanced arbovirus surveillance with deep sequencing: Identification of novel rhabdoviruses and bunyaviruses in Australian mosquitoes

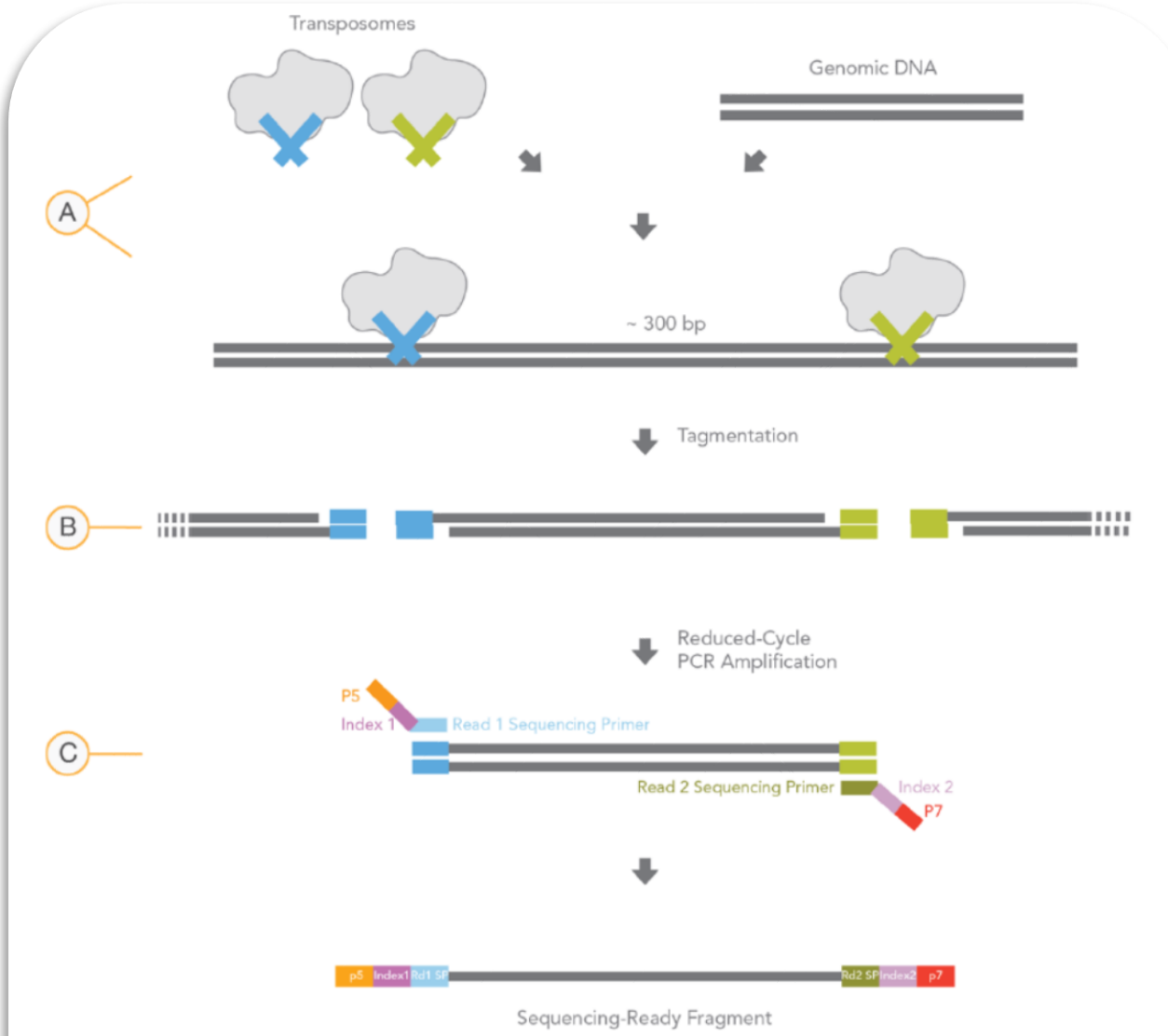
Vecteurs





Library Preparation.

DNA fragment libraries are prepared by sonication shearing (e.g., Covaris S2) or physical disruption (HydroShear). Final fragment size: Illumina, 200-500 bp; SOLiD, 60-110 bp; and Roche 454, 300-800 bp. Sheared DNA is end-repaired, adenylated (Illumina), purified, and ligated to unique sequencing adaptors. (The resultant library is amplified using the sequencing adaptors in the case of Illumina.) Size-selection of the template/adaptor ligation product follows.

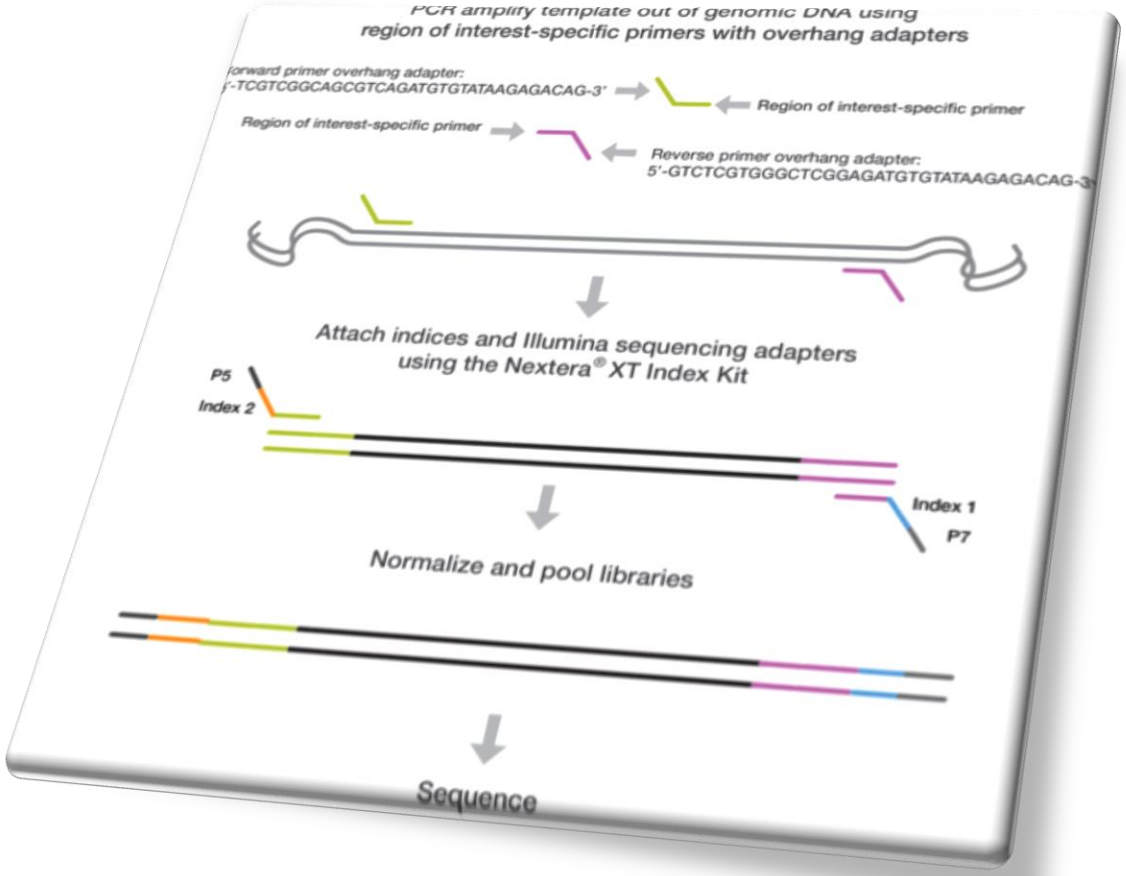


A Nextera XT transposome with adapters is combined with template DNA
B Tagmentation to fragment and add adapters
C Limited cycle PCR to add sequencing primer sequences and indicies

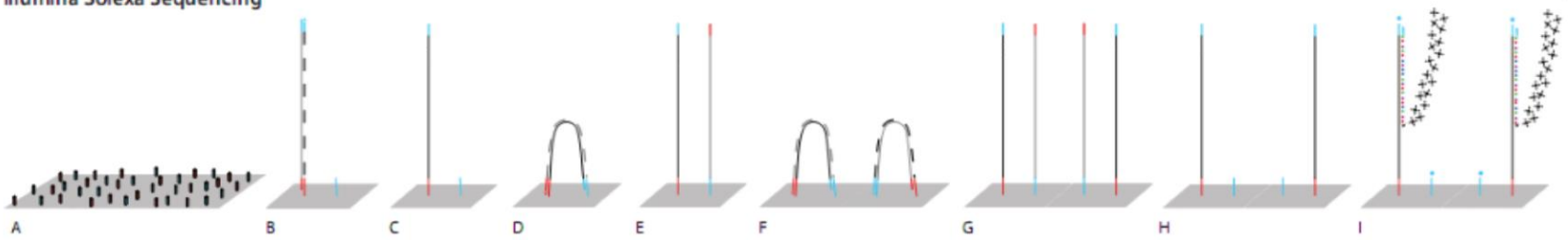
Détection / génotypage bactéries:

PCR 16S

séquençage massif



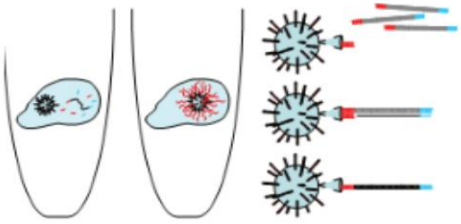
Illumina Solexa Sequencing



Cluster Generation. Single-stranded templates hybridize randomly to a lawn of solid-phase oligonucleotides (A). A new strand is extended by polymerization from the anchored oligonucleotide (B), with subsequent removal of the template by denaturation (C). The free 3'end of the de novo synthesized product hybridizes to its surface-bound complement, forming a bridge template (D) for further polymerization. Denaturation (E), followed by additional rounds of bridge amplification and denaturation (F, G), and, finally, cleavage and removal of the complement (H) results in the formation of anchored clusters of identical sequence.

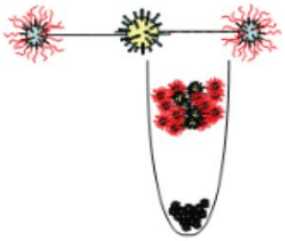
Sequencing. Four labeled reversible terminators, primers, and DNA polymerase are added, initiating the first sequencing cycle. Upon laser excitation, each cluster emits a detectable fluorescent signal associated with one of four bases. Termination is reversed and each cycle repeats this process (H).

ABI SOLiD Sequencing



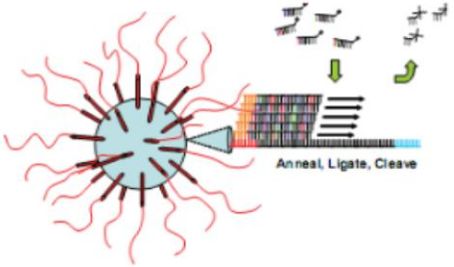
Emulsion PCR. Clonal bead populations are generated in water-in-oil microreactors containing a single capture bead, template, PCR components, and primers. A template anneals to a capture bead primer.

Polymerase extension from the bead-bound primer is followed by template dissociation.



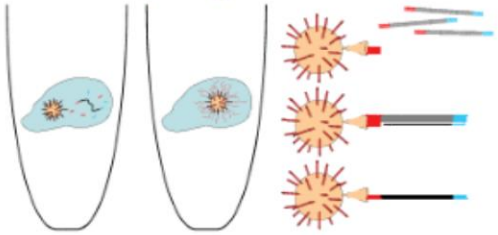
Enrichment of Templated Beads. Polystyrene bead-bounnd adapter-spedific oligonucleotides bind and collect templated beads. Templated beads are then enriched in a glycerol gradient.

3'-end Modification and Bead Deposition. Templated beads are 3'modified to facilitate a covalent linkage to a glass slide. Beads are deposited randomly on the slide.

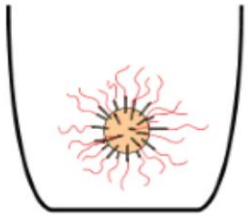


Sequencing. Four fluorescently labeled di-base probes compete to ligate to the sequencing primer, driven by complementarity to the 1st and 2nd base of the template. After multiple cycles of ligation, detection and cleavage, the extension product is removed. Five ligation cycles are performed with the primer for each successive cycle reset to the -1 position.

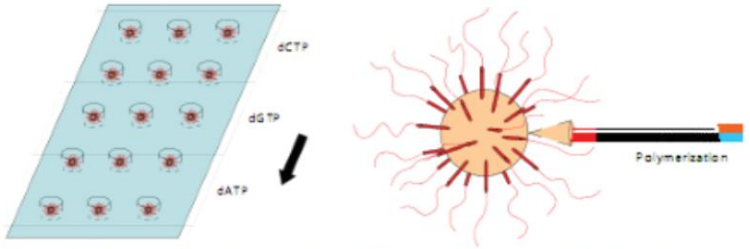
Roche 454 Sequencing



Emulsion PCR. Clonal bead populations are generated in water-in-oil microreactors containing a single capture bead, template, PCR components, and primers. A template anneals to a capture bead primer. Polymerase extension from the bead-bound primer is followed by template dissociation.



Enrichment of Clones and Bead Deposition. Clonally amplified fragments are enriched manually or using an automated system. Fragments are loaded one bead per flow cell well.



Sequencing. Nucleotides flowed sequentially in a fixed order across the through the flow cell device during a sequencing run. Individual beads each carrying identical copies of a unique single-stranded DNA molecule are sequenced in parallel. When a nucleotide complementary to the template strand flows into a well, the polymerase extends the existing DNA strand by adding the nucelotide, generating a detectable light signal.

454

	GS Jr.	GS FLX+
Amplification Method	Emulsion PCR on beads	
Chemistry	Synthesis (pyrosequencing)	
Read length (bp)	400	700
Yield/run (Gb)	0.05	0.9
Primary Error	Indel	
Error rate	~1%	
Run time (hours)	10	20
Virus-related Publications	187	
Advantage(s)	Long reads, maturity	
Disadvantage(s)	Homopolymer misreads, high cost/Mb	

Illumina

	MiSeq	HiSeq
Amplification Method	Bridge PCR in situ	
Chemistry	Synthesis (reversible termination)	
Read length (bp)	250	125
Yield/run (Gb)	8	1,000
Primary Error	Substitution	
Error rate	~0.1%	
Run Time (hours)	39	276
Virus-related Publications	129	
Advantage(s)	Easy work flow, maturity	
Disadvantage(s)	Shortest reads, long run	

Ion Torrent

	PGM	Proton
Amplification Method	Emulsion PCR on beads	
Chemistry	Synthesis (H^+ detection)	
Read length (bp)	400	200
Yield/run (Gb)	2	10
Primary Error	Indel	
Error rate	~1%	
Run time (hours)	7	4
Virus-related Publications	13	
Advantage(s)	Low cost, fast run	
Disadvantage(s)	Homopolymer misreads	

PacBio

	RS II
Amplification Method	No PCR
Chemistry	Single-molecule real-time sequencing
Read length (bp)	8,500
Yield/run (Gb)	0.15
Primary Error	Indel
Error rate	~13%
Run time (hours)	2
Virus-related Publications	6
Advantage(s)	Longest reads
Disadvantage(s)	High error rate, expensive

		Run time	Read length (bp)	# reads per run	Output per run
Roche	GS FLX Titanium XL+	23 h	700	1 million	700 Mb
	GS Junior System	10 h	400	0.1 million	40 Mb
LifeTechnologies	Ion torrent	4 h	200–400	4 million	1.5–2 Gb
	Proton	4 h	125	60–80 million	8–10 Gb
	Abi/solid	10 days	75 + 35	2.7 billion	300 Gb
Illumina/solexa	HiSeq2000/2500	12 days	2 × 100	3 billion	600 Gb
	MiSeq	65 h	2 × 300	25 million	15 Gb
Pacific biosciences	RSII	2 days	50% of reads > 10 kb	0.8 million	5 Gb
Helicos	Heliscope	10 days	~30	500 million	15 Gb

Tagmentation of Genomic DNA

Hands-On: 7 min/8 samples
Total: 17 min/8 samples

Reagents
ATM
TD
NT

Output
96-well Plate

PCR Amplification

Hands-On: 7 min/8 samples
Cycle Time: 38 min

Reagents
NPM
Index 1
Index 2

Output
Amplification PCR Plate

PCR Clean-Up

Hands-On: 15 min/8 samples
Total: 30 min/8 samples

Reagents
Ampure XP Beads
Fresh 80% EtOH
RSB

Output
Post-PCR TCY Plate

Library Normalization

Hands-On: 30 min/96 samples
Processing: 80 min/96 samples

Reagents
LNA1
LNB1
LNW1
LNS1
NaOH

Output
SGP Plate

Library Pooling for MiSeq Sequencing

Hands-On: 5 min

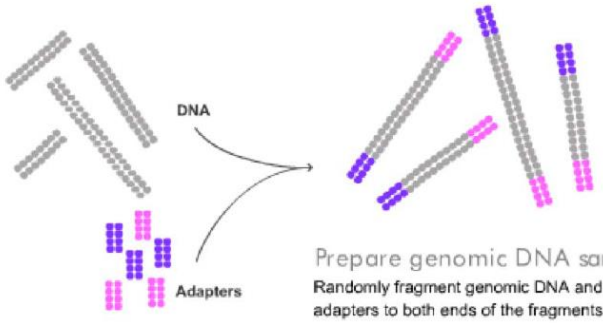
Reagents
HT1

Output
PAL & DAL tubes

Cold Storage Option

Fill in the lab tracking form as you perform the assay

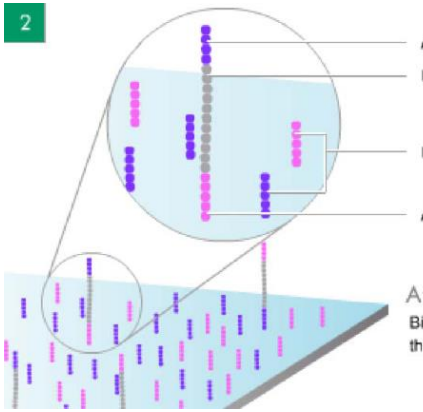
1



Prepare genomic DNA sample

Randomly fragment genomic DNA and ligate adapters to both ends of the fragments

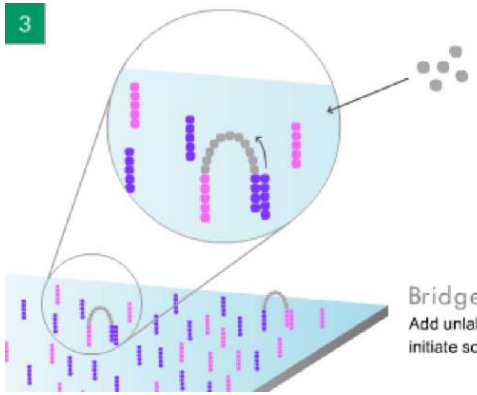
2



Attach DNA to surface

Bind single stranded fragments randomly to the inside surface of the flow cell channels.

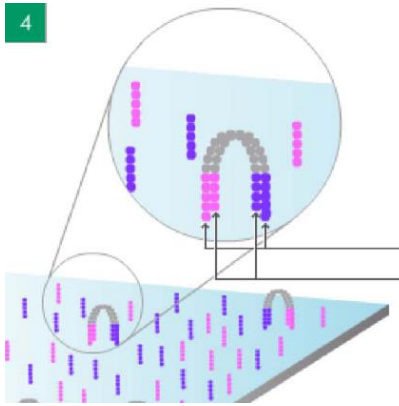
3



Bridge amplification

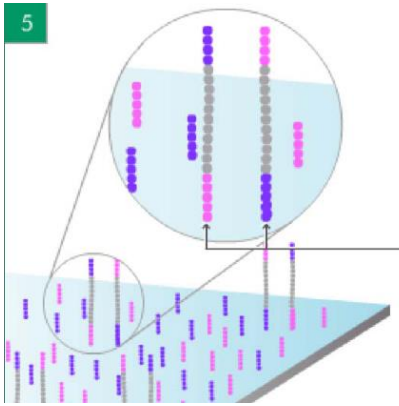
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

4



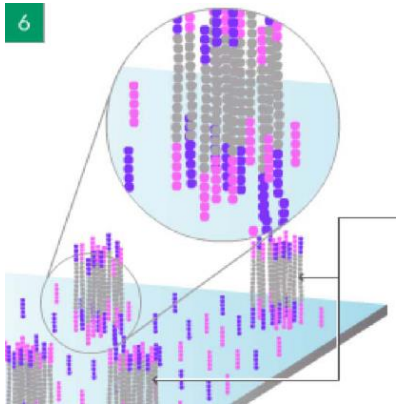
Fragments become double stranded

5



Denature the double stranded molecules

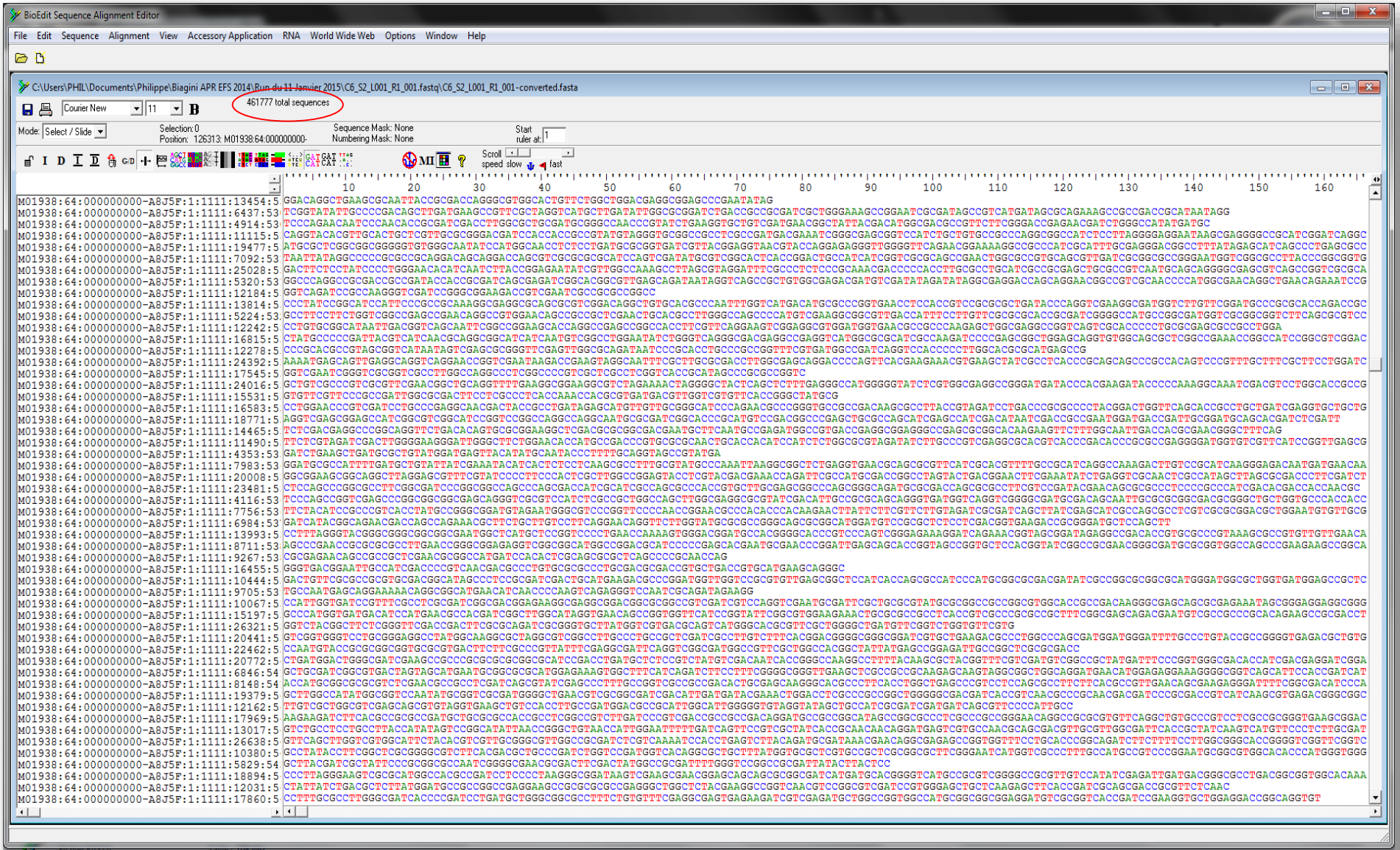
6

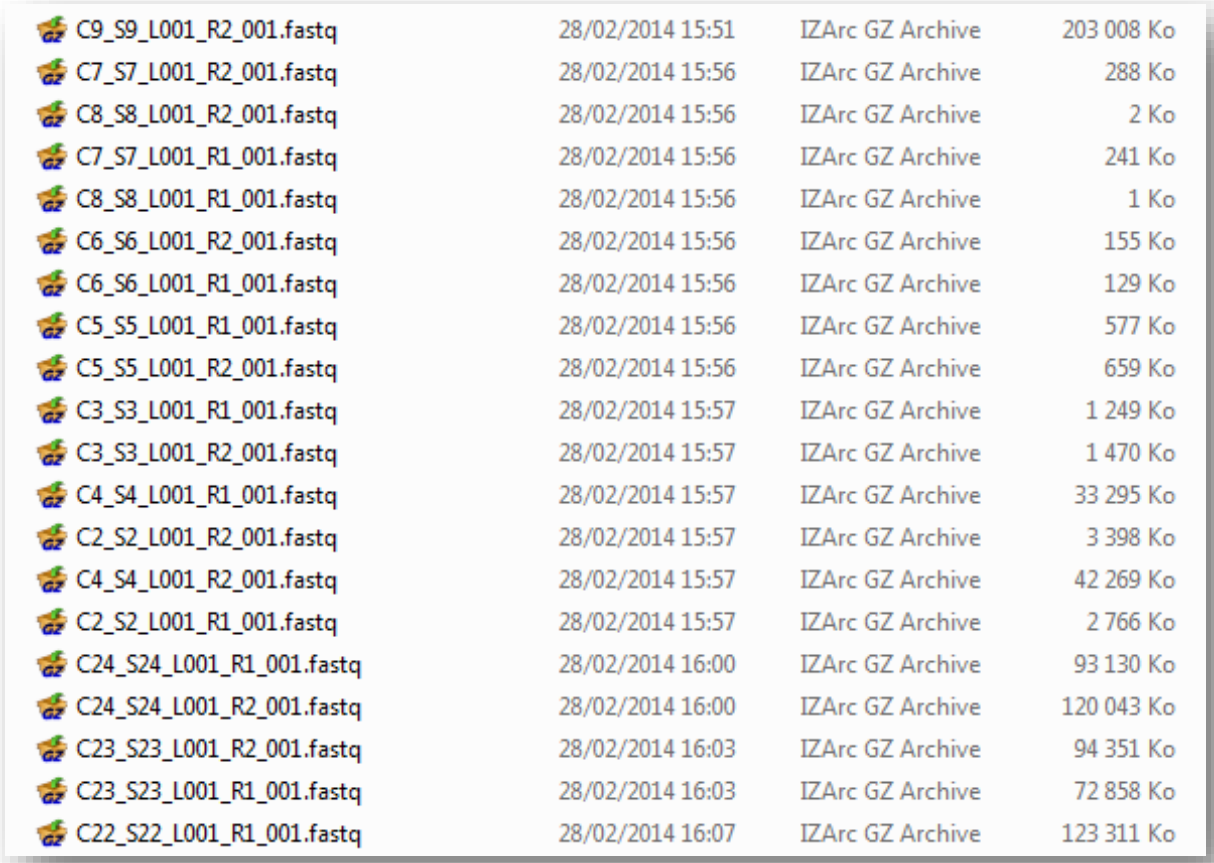


Completion of amplification

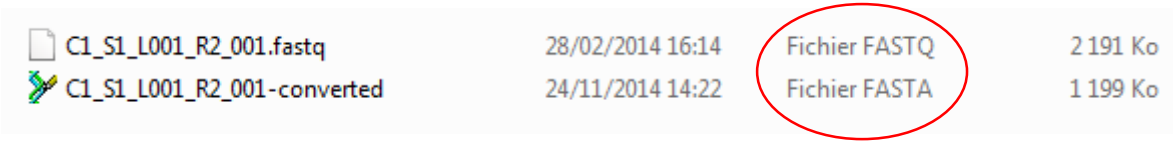
On completion, several million dense clusters of double stranded DNA are generated in each channel of the flow cell.

10^5 – 10^6 séquences / échantillon



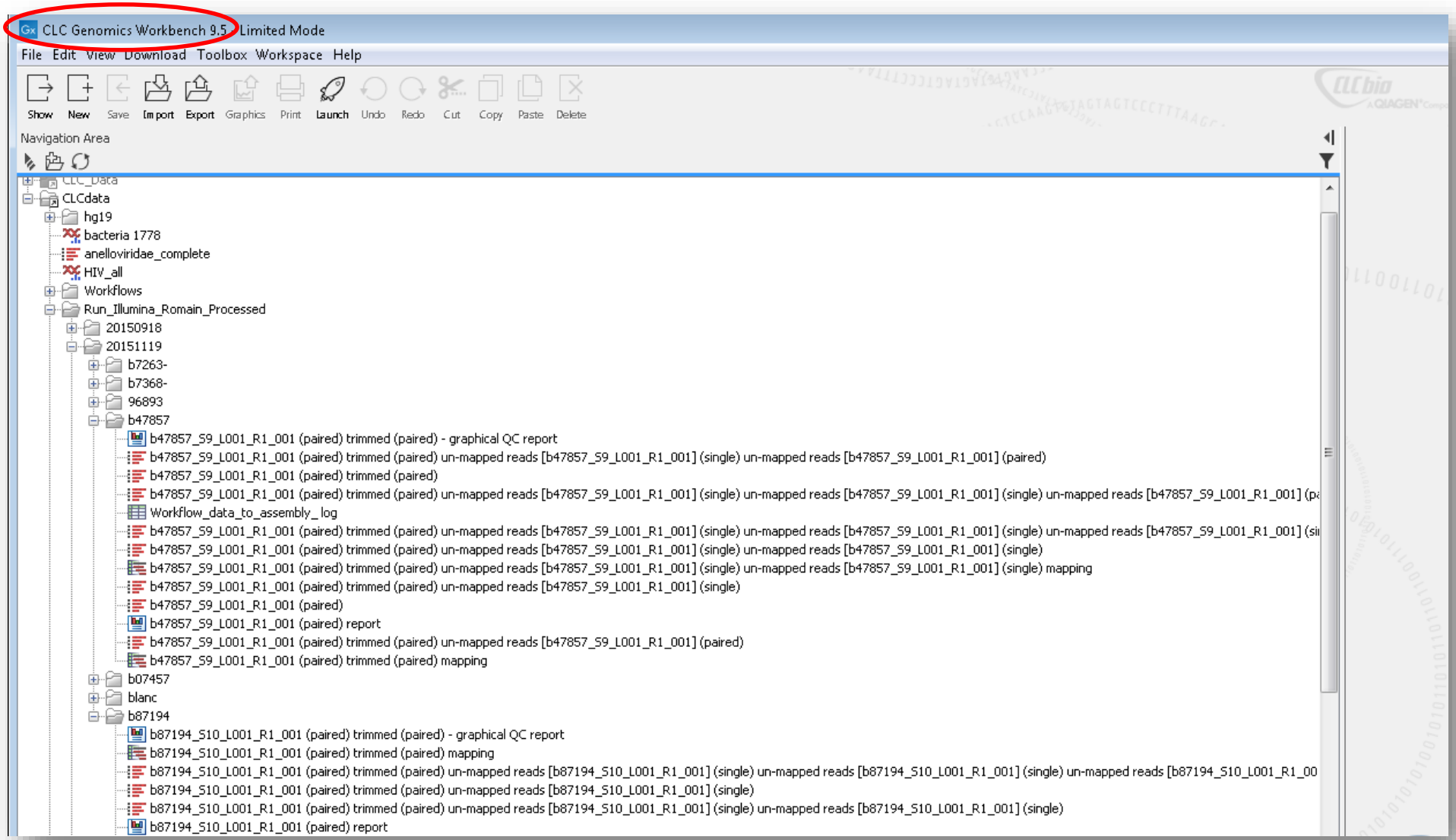


C9_S9_L001_R2_001.fastq	28/02/2014 15:51	IZArc GZ Archive	203 008 Ko
C7_S7_L001_R2_001.fastq	28/02/2014 15:56	IZArc GZ Archive	288 Ko
C8_S8_L001_R2_001.fastq	28/02/2014 15:56	IZArc GZ Archive	2 Ko
C7_S7_L001_R1_001.fastq	28/02/2014 15:56	IZArc GZ Archive	241 Ko
C8_S8_L001_R1_001.fastq	28/02/2014 15:56	IZArc GZ Archive	1 Ko
C6_S6_L001_R2_001.fastq	28/02/2014 15:56	IZArc GZ Archive	155 Ko
C6_S6_L001_R1_001.fastq	28/02/2014 15:56	IZArc GZ Archive	129 Ko
C5_S5_L001_R1_001.fastq	28/02/2014 15:56	IZArc GZ Archive	577 Ko
C5_S5_L001_R2_001.fastq	28/02/2014 15:56	IZArc GZ Archive	659 Ko
C3_S3_L001_R1_001.fastq	28/02/2014 15:57	IZArc GZ Archive	1 249 Ko
C3_S3_L001_R2_001.fastq	28/02/2014 15:57	IZArc GZ Archive	1 470 Ko
C4_S4_L001_R1_001.fastq	28/02/2014 15:57	IZArc GZ Archive	33 295 Ko
C2_S2_L001_R2_001.fastq	28/02/2014 15:57	IZArc GZ Archive	3 398 Ko
C4_S4_L001_R2_001.fastq	28/02/2014 15:57	IZArc GZ Archive	42 269 Ko
C2_S2_L001_R1_001.fastq	28/02/2014 15:57	IZArc GZ Archive	2 766 Ko
C24_S24_L001_R1_001.fastq	28/02/2014 16:00	IZArc GZ Archive	93 130 Ko
C24_S24_L001_R2_001.fastq	28/02/2014 16:00	IZArc GZ Archive	120 043 Ko
C23_S23_L001_R2_001.fastq	28/02/2014 16:03	IZArc GZ Archive	94 351 Ko
C23_S23_L001_R1_001.fastq	28/02/2014 16:03	IZArc GZ Archive	72 858 Ko
C22_S22_L001_R1_001.fastq	28/02/2014 16:07	IZArc GZ Archive	123 311 Ko

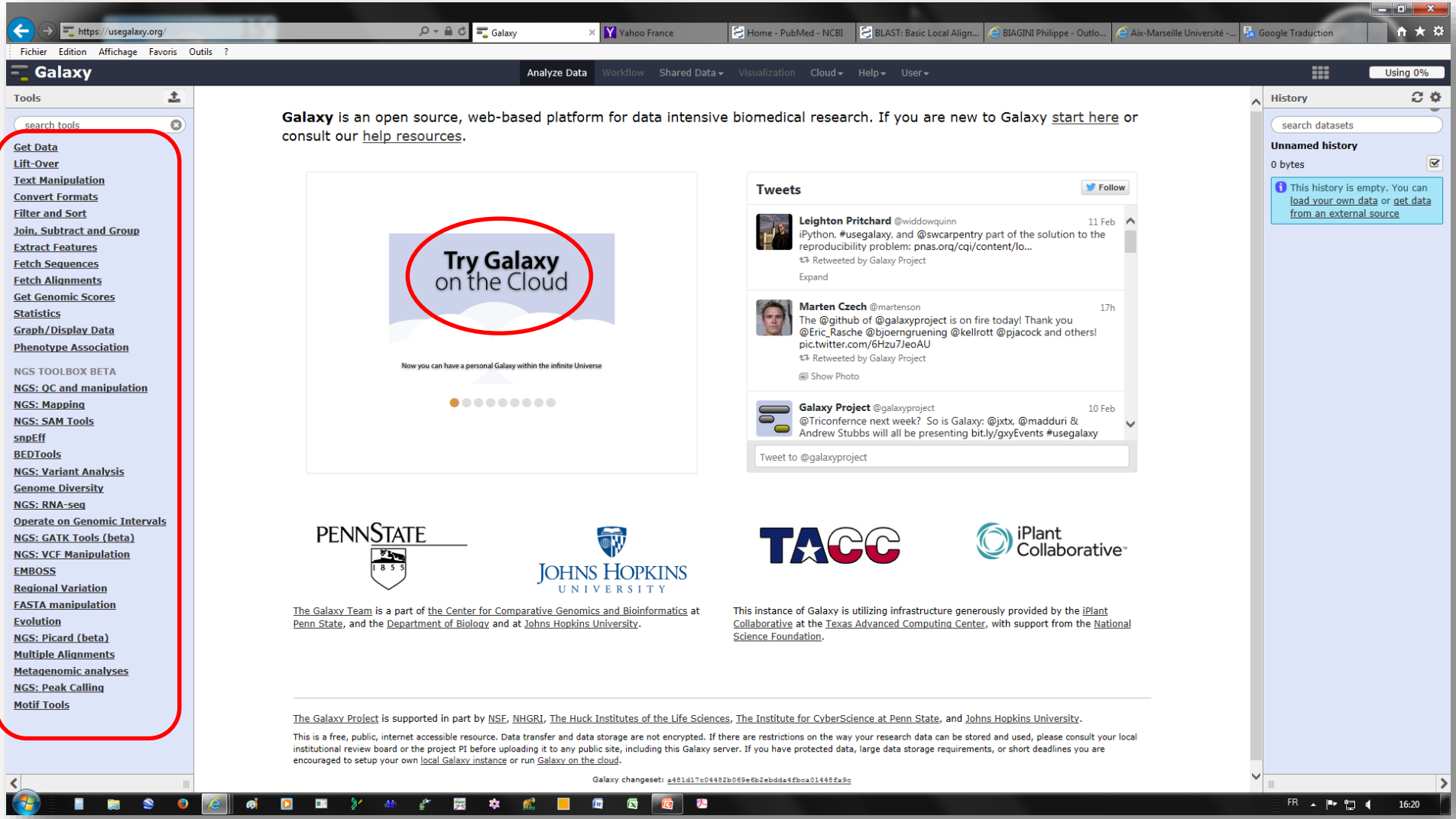


C1_S1_L001_R2_001.fastq	28/02/2014 16:14	Fichier FASTQ	2 191 Ko
C1_S1_L001_R2_001-converted	24/11/2014 14:22	Fichier FASTA	1 199 Ko

Bioinformatique



Bioinformatique



WebMGA

web server for metagenomic analysis

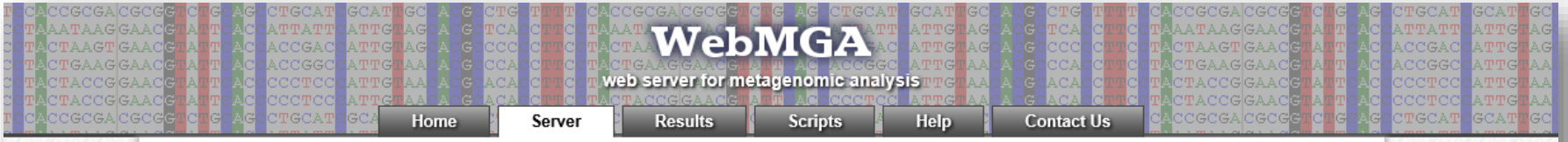
Home | Server | Results | Scripts | Help | Contact Us

- clustering
- rRNA prediction
- tRNA prediction
- orf prediction
- function annotation
- pathway annotation
- sequence info
- quality control
- filtering sequence
- taxonomy binning
- OTU finder
- format conversion

Currently we provide the following single metagenomic analyses:

Category	Name	Description
clustering	cd-hit-est	fast DNA clustering
	cd-hit	protein clustering
	h-cd-hit	hierarchical protein clustering
	cd-hit-454	filtering 454 duplicate reads
rRNA prediction	blastn_rRNA	rRNA prediction by blastn program
	hmm_rRNA	rRNA prediction by hmmer 3.0 program
tRNA prediction	tRNA	tRNA prediction by tRNAscan-SE program
orf prediction	orf_finder	orf prediction by six-reading-frame technique
	metagene	orf prediction by metagene program
	fraggene_scan	orf prediction by fraggene_scan program
function annotation	cog	protein function annotation by COG database
	kog	protein function annotation by KOG database
	prk	protein function annotation by NCBI PRK database
	pfam	protein function annotation by pfam database
	tigrfam	protein function annotation by tigrfam database
pathway annotation	kegg	pathway annoation by KEGG database
sequence info	fna_stat	statistics of DNA sequences including length and GC content
	faa_stat	statistics of protein sequences including length
quality control	qc_filter_fastq	removing reads with low average quality for a read with fastq format
	qc_filter_fasta_qual	removing reads with low average quality for a read with fasta+quality format
	trimm	trimming the low-quality tail of illumina reads
filtering sequence	filter_human	filtering human sequences from reads
taxonomy binning	rdp_binning	taxonomic binning by rdp classifier program
	frhit_binning	taxonomic binning by frhit program
OTU finder	cd-hit-otu	OTU finder by cd-hit-otu program
format conversion	fastq2fasta	converting fastq file to fasta file

Note: every single analysis can also be accessed by using a script.



- clustering
- rRNA prediction
- tRNA prediction
- orf prediction
- function annotation
- pathway annotation
- sequence info
- quality control
- filtering sequence
- taxonomy binning
- OTU finder
- format conversion

Currently we provide the following single metagenomic analyses:

Category	Name	Description
clustering	cd-hit-est	fast DNA clustering
	cd-hit	protein clustering
	h-cd-hit	hierarchical protein clustering
	cd-hit-454	filtering 454 duplicate reads
rRNA prediction	blastn_rRNA	rRNA prediction by blastn program
	hmm_rRNA	rRNA prediction by hmmer 3.0 program
tRNA prediction	tRNA	tRNA prediction by tRNAscan-SE program
orf prediction	orf_finder	orf prediction by six-reading-frame technique
	metagene	orf prediction by metagene program
	fraggene_scan	orf prediction by fraggene_scan program
function annotation	cog	protein function annotation by COG database
	kog	protein function annotation by KOG database
	prk	protein function annotation by NCBI PRK database
	pfam	protein function annotation by pfam database
	tigrfam	protein function annotation by tigrfam database
pathway annotation	kegg	pathway annoation by KEGG database
sequence info	fna_stat	statistics of DNA sequences including length and GC content
	faa_stat	statistics of protein sequences including length
quality control	qc_filter_fastq	removing reads with low average quality for a read with fastq format
	qc_filter_fasta_qual	removing reads with low average quality for a read with fasta+quality format
	trimm	trimming the low-quality tail of illumina reads
filtering sequence	filter_human	filtering human sequences from reads
taxonomy binning	rdp_binning	taxonomic binning by rdp classifier program
	frhit_binning	taxonomic binning by frhit program
OTU finder	cd-hit-otu	OTU finder by cd-hit-otu program
format conversion	fastq2fasta	converting fastq file to fasta file

Note: every single analysis can also be accessed by using a script.

WebMGA

web server for metagenomic analysis

Home | Server | Results | Scripts | Help | Contact Us

- clustering
- rRNA prediction
- tRNA prediction
- orf prediction
- function annotation
- pathway annotation
- sequence info
- quality control
- filtering sequence
- taxonomy binning
- OTU finder
- format conversion

Currently we provide the following single metagenomic analyses:

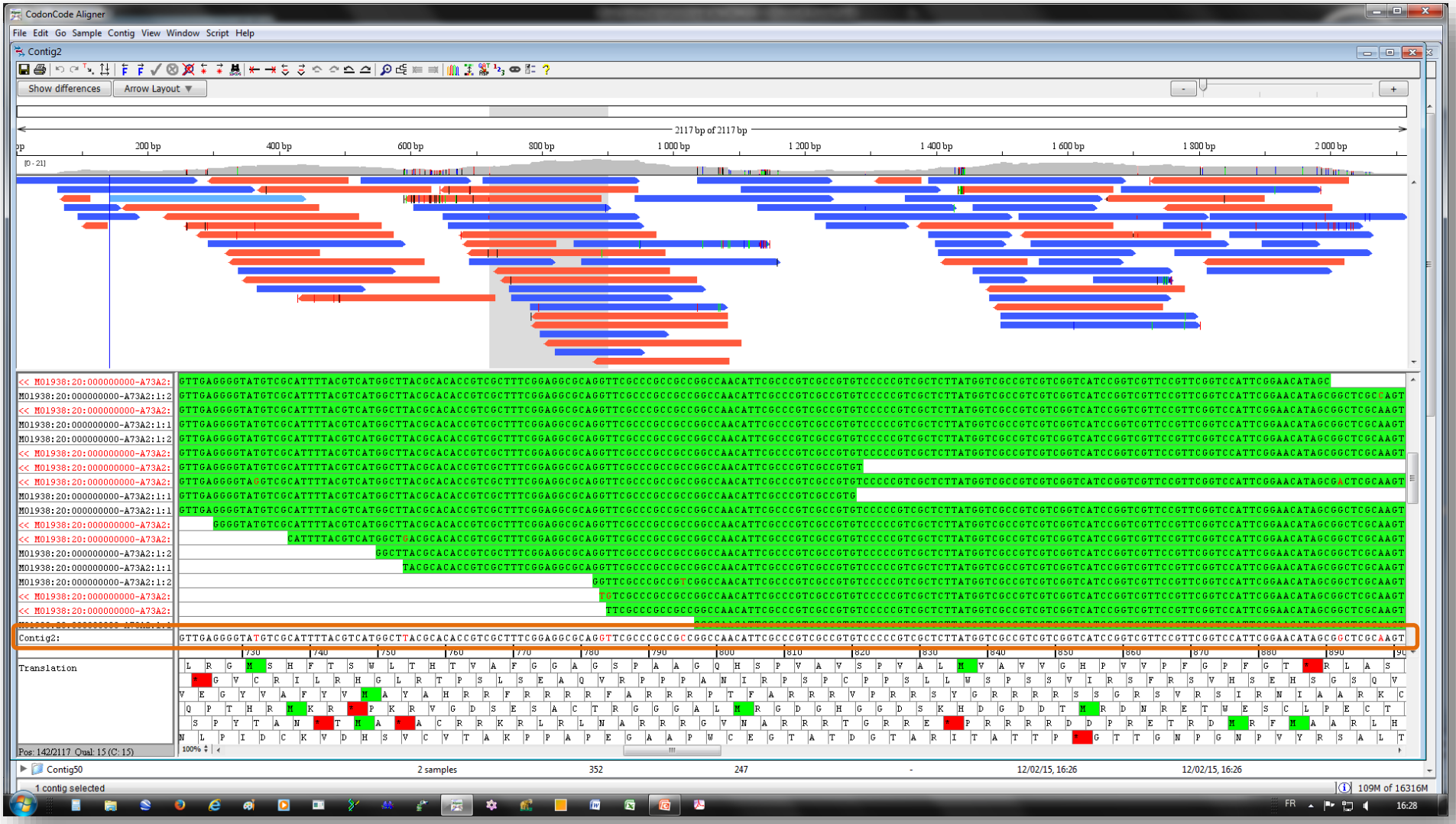
Category	Name	Description
clustering	cd-hit-est	fast DNA clustering
	cd-hit	protein clustering
	h-cd-hit	hierarchical protein clustering
	cd-hit-454	filtering 454 duplicate reads
rRNA prediction	blastn_rRNA	rRNA prediction by blastn program
	hmm_rRNA	rRNA prediction by hmmer 3.0 program
tRNA prediction	tRNA	tRNA prediction by tRNAscan-SE program
orf prediction	orf_finder	orf prediction by six-reading-frame technique
	metagene	orf prediction by metagene program
	fraggene_scan	orf prediction by fraggene_scan program
function annotation	cog	protein function annotation by COG database
	kog	protein function annotation by KOG database
	prk	protein function annotation by NCBI PRK database
	pfam	protein function annotation by pfam database
	tigrfam	protein function annotation by tigrfam database
pathway annotation	kegg	pathway annoation by KEGG database
sequence info	fna_stat	statistics of DNA sequences including length and GC content
	faa_stat	statistics of protein sequences including length
quality control	qc_filter_fastq	removing reads with low average quality for a read with fastq format
	qc_filter_fasta_qual	removing reads with low average quality for a read with fasta+quality format
	trimm	trimming the low-quality tail of illumina reads
filtering sequence	filter_human	filtering human sequences from reads
taxonomy binning	rdp_binning	taxonomic binning by rdp classifier program
	frhit_binning	taxonomic binning by frhit program
OTU finder	cd-hit-otu	OTU finder by cd-hit-otu program
format conversion	fastq2fasta	converting fastq file to fasta file

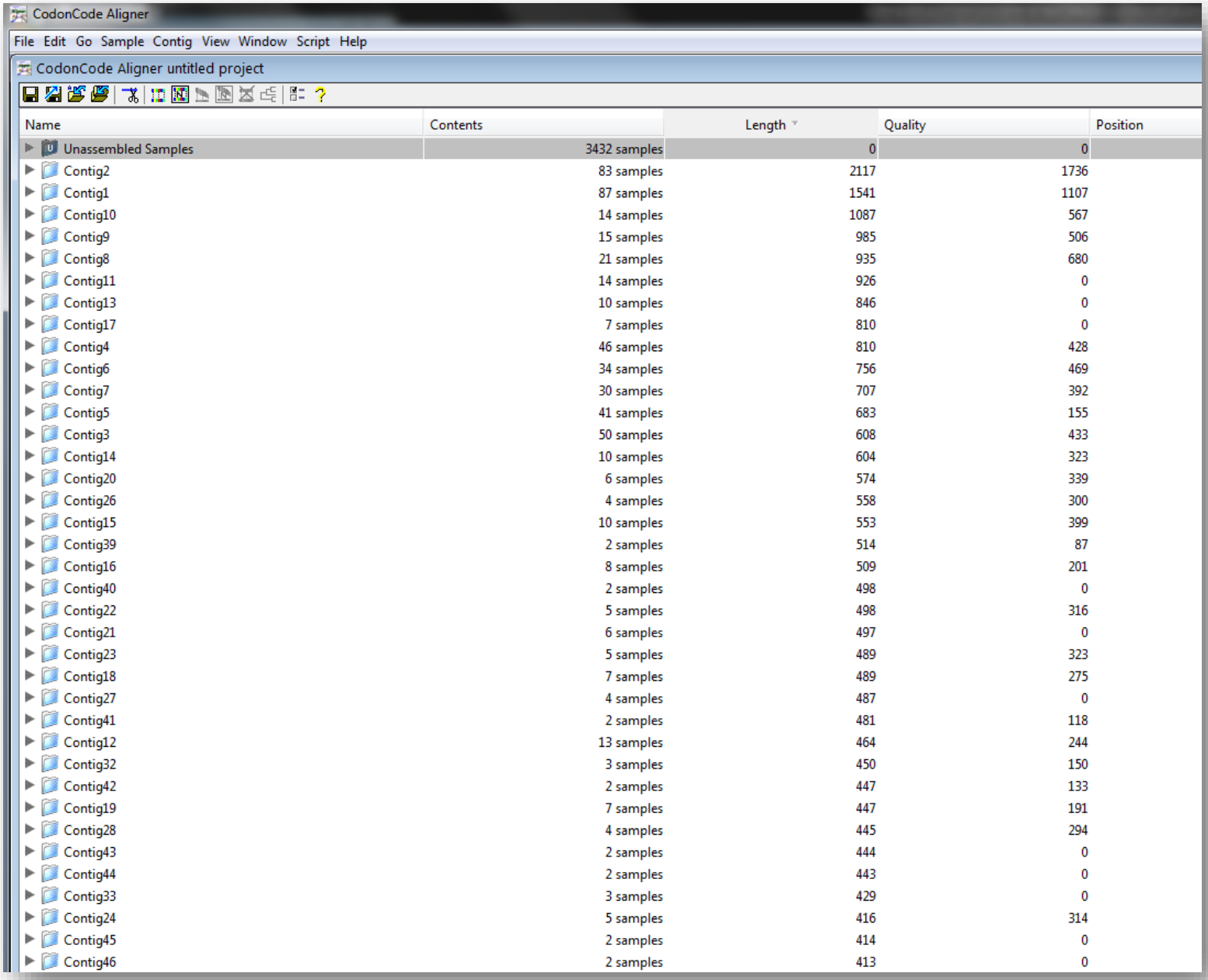
Note: every single analysis can also be accessed by using a script.

Qualité


NextGen Workbench - C1_S1_L001_R1_001.fastq - Mem: 23.23 MB
File View Clipping & filtering Tools
Fastq Editor Settings Info
Sample Explorer
Clipping & filtering Sequence viewer Split/Convert Demultiplex file Overrepresented/Contaminants Clustering & dereplication
File details
File: C1_S1_L001_R1_001.fastq
Size: 2.12 MB
Total reads: 4128. Reads after filtering: 4069
Detected encoding: Sanger
Multiplexing type: Illumina v1.8
Instrument: M01938
Lane: 1 Tile number: 1101 Pair: 1 Multiplex ID: 1
Per sequence quality scores (Y= Number of reads)
Before processing After processing Expand
Sequence length distribution (Y= No of reads)
Per Base Sequence Quality (Y= Quality)
10th-90th percentile Interquartile (25-75%) Median Mean
Per base sequence content (Y= Percent)
A C G T N Up=Before/Down=After
Encoding
Autodetect Solexa Sanger Illumina 1+ Illumina 1.3+ Illumina 1.5+ Illumina 1.8+
Detected encoding: Sanger
Loading C1_S1_L001_R1_001.fastq
Number of reads containing N bases: 0
Overrepresented sequences before filtering: 1
Overrepresented sequences after filtering: 0
Done filtering.
Overrepresented sequences save to C:\Users\PHIL\AppData\Roaming\Nex
Clip adaptor/tag/barcode/MID
Clip low quality ends using:
Clip poly-A/T tails with minimum length of
Cut reads shorter than
Cut reads longer than
Cut reads with average QV under
Cut reads if they contain:
Cut read if
Cut reads with GC under
Fast refresh
Apply
Apply & save...
Seq duplication parameters
Overrepresented sequences threshold
Check duplication level for the first
File size: 2.12 MB
Reads: 4128
Time: 109ms
File: C:\Users\PHIL\Documents\Philippe\Biagini APR EFS 2014\ILLUMINA second run FULL\c1\C1_S1_L001_R1_001.fastq

Assemblages





Name	Contents	Length	Quality	Position
Unassembled Samples	3432 samples	0	0	
Contig2	83 samples	2117	1736	
Contig1	87 samples	1541	1107	
Contig10	14 samples	1087	567	
Contig9	15 samples	985	506	
Contig8	21 samples	935	680	
Contig11	14 samples	926	0	
Contig13	10 samples	846	0	
Contig17	7 samples	810	0	
Contig4	46 samples	810	428	
Contig6	34 samples	756	469	
Contig7	30 samples	707	392	
Contig5	41 samples	683	155	
Contig3	50 samples	608	433	
Contig14	10 samples	604	323	
Contig20	6 samples	574	339	
Contig26	4 samples	558	300	
Contig15	10 samples	553	399	
Contig39	2 samples	514	87	
Contig16	8 samples	509	201	
Contig40	2 samples	498	0	
Contig22	5 samples	498	316	
Contig21	6 samples	497	0	
Contig23	5 samples	489	323	
Contig18	7 samples	489	275	
Contig27	4 samples	487	0	
Contig41	2 samples	481	118	
Contig12	13 samples	464	244	
Contig32	3 samples	450	150	
Contig42	2 samples	447	133	
Contig19	7 samples	447	191	
Contig28	4 samples	445	294	
Contig43	2 samples	444	0	
Contig44	2 samples	443	0	
Contig33	3 samples	429	0	
Contig24	5 samples	416	314	
Contig45	2 samples	414	0	
Contig46	2 samples	413	0	

BLAST®

Basic Local Alignment Search Tool

Home | Recent Results | Saved Strategies | Help

NCBI/ BLAST/ blastn suite

Standard Nucleotide BLAST

blastn | blastp | blastx | tblastn | tblastx

BLASTN programs search nucleotide databases using a nucleotide query.

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

```
CCCTCGAACATGTCCCCGCCGGGGCTGTTGATGTGCACCTCGACGGGGCGCTCGCCGATCGCGCGGA
GCTGCGCCGTAACCCGCTTGGCCGTCACACCGCCGCCGGACCAGAAGTCCTCGCCAATGGCGTCGAA
CATAGTGATGACGTTGTCGTCACTCTGGACTGATCGGATGCCGGCCGCCTCGTTCGACCACTTCTCG
AACACGTCGGGGGAGGTGAACGCGGARATGTTCCGCTTCGCCGGCATCGGGAGCGCGGCCGGGCGCT
CCGCCTTCGCTTTGATCCTGCTCATCACGCCCCTCGGGCGCTTCAGGAGGT
```

Query subrange

From

To

Or, upload file — Parcourir...

Job Title — 3000 sequences (Contig488)

Enter a descriptive title for your BLAST search

☐ **Align two or more sequences**

Choose Search Set

Database ○ Human genomic + transcript ○ Mouse genomic + transcript ◉ Others (nr etc.):

Nucleotide collection (nr/nt)

Organism Optional — Enter organism name or id--completions will be suggested ☐ Exclude +

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude Optional — ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to Optional — ☐ Sequences from type material

Entrez Query Optional — YouTube Create custom database

Enter an Entrez query to limit search

Program Selection

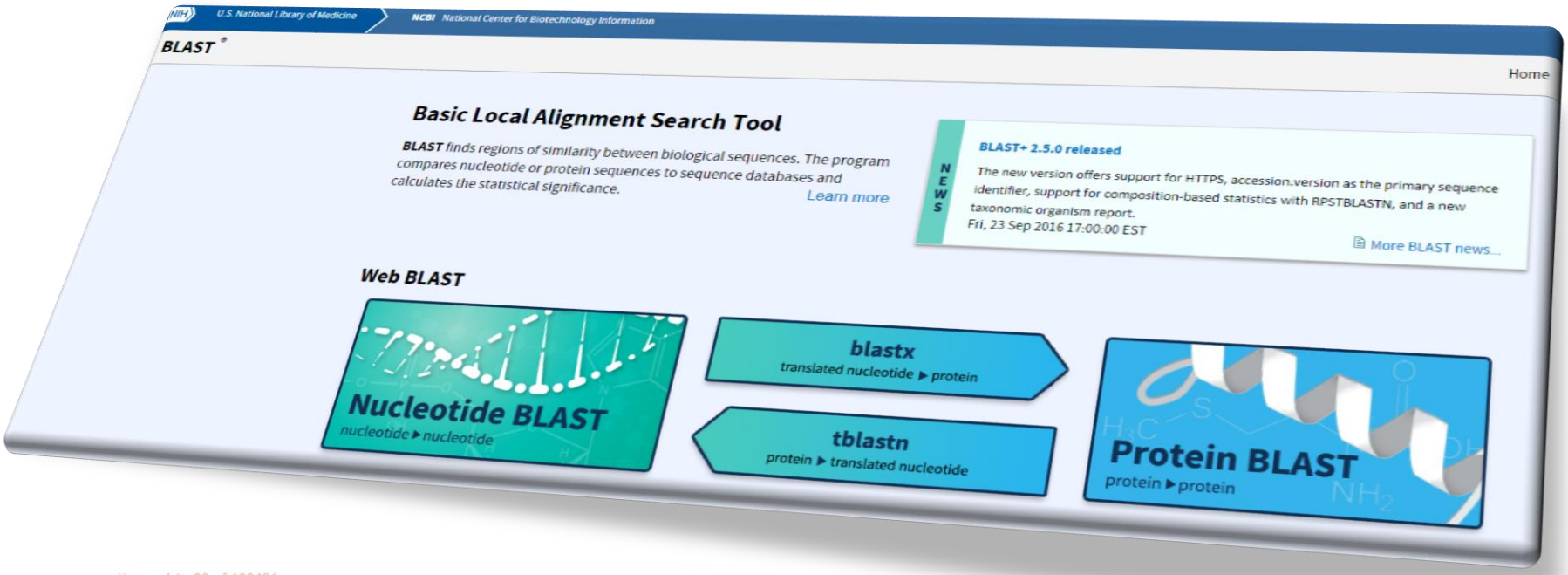
Optimize for

○ Highly similar sequences (megablast)

○ More dissimilar sequences (discontiguous megablast)

◉ Somewhat similar sequences (blastn)

Choose a BLAST algorithm



Items: 1 to 20 of 103431

<< First < Prev Page 1 of 5172 Next > Last >>

Found 103987 nucleotide sequences. Nucleotide (103431) EST (89) GSS (467)

1. Hepatitis B virus isolate pt1_S07003334 reverse transcriptase and surface antigen genes, partial cds
740 bp linear DNA
Accession: KT340641.1 GI: 1050243117
GenBank FASTA Graphics PopSet

2. Hepatitis B virus isolate pt1_S08005564 re... cds
740 bp linear DNA
Accession: KT340642.1 GI: 1050243120
GenBank FASTA Graphics PopSet

3. Hepatitis B virus isolate pt1_S08021045 re... cds
740 bp linear DNA
Accession: KT340643.1 GI: 1050243123
GenBank FASTA Graphics PopSet

4. Hepatitis B virus isolate pt1_S10013154 re... cds
740 bp linear DNA
Accession: KT340644.1 GI: 1050243126
GenBank FASTA Graphics PopSet

5. Hepatitis B virus isolate pt2_E11003948 re... cds
740 bp linear DNA
Accession: KT340645.1 GI: 1050243129
GenBank FASTA Graphics PopSet

6. Hepatitis B virus isolate pt2_H0405094 rev... cds

Items: 1 to 20 of 8211

<< First < Prev

1. Torque teno virus 2 isolate HNSCC_tumor_53 ORF1 gene, partial cds
222 bp linear DNA
Accession: KU720361.1 GI: 1057447234
GenBank FASTA Graphics PopSet

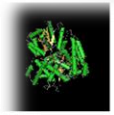
2. Torque teno virus 3 isolate HNSCC_tumor_52 ORF1 gene, partial cds
222 bp linear DNA
Accession: KU720362.1 GI: 1057447236
GenBank FASTA Graphics PopSet

3. Torque teno virus 2 isolate HNSCC_tumor_56 ORF1 gene, partial cds
222 bp linear DNA
Accession: KU720363.1 GI: 1057447238
GenBank FASTA Graphics PopSet

4. Torque teno virus 2 isolate HNSCC_tumor_66 ORF1 gene, partial cds
222 bp linear DNA
Accession: KU720364.1 GI: 1057447240
GenBank FASTA Graphics PopSet

5. Torque teno virus 1 isolate HNSCC_tumor_67 ORF1 gene, partial cds
222 bp linear DNA
Accession: KU720365.1 GI: 1057447242
GenBank FASTA Graphics PopSet

6. Torque teno virus 2 isolate HNSCC_tumor_62 ORF1 gene, partial cds

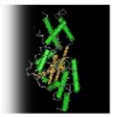


rep: ATP-dependent DNA helicase Rep
Designed to identify **rep** members of the uvrD/**rep** subfamily. [DNA metabolism, DNA replication, recomb...
Accession: TIGR01074 ID: 130146
View in Cn3D Specific Protein Protein

REP: Plasmid rolling circle replication initiator protein **REP** and truncated derivatives [Mobilome: prophages transposons]
Accession: COG5655 ID: 227942
View in Cn3D Specific Protein Protein Superfamily Superfamily Members

rep
RNA replicase, beta subunit
Accession: PHA00028 ID: 222774
View in Cn3D Specific Protein Protein Superfamily Superfamily Members PubMed

Rep_1: Replication protein
Replication proteins (**rep**) are involved in plasmid replication. The **Rep** protein binds to the plasmid...
Accession: cl02412 ID: 295292
View in Cn3D Protein Superfamily Members PubMed



UvrD-helicase: UvrD/**REP** helicase N-terminal domain
The **Rep** family helicases are composed of four structural domains. The **Rep** family function as dimers....
Accession: pfam00580 ID: 278977
View in Cn3D Specific Protein Protein Superfamily Superfamily Members PubMed

Gemini_AL1: Geminivirus **Rep** catalytic domain
The AL1 proteins encodes the replication initiator protein (**Rep**) of geminiviruses, which is a replic...
Accession: pfam00799 ID: 279178
View in Cn3D Specific Protein Protein Superfamily Superfamily Members PubMed

>gb|EF611223.1| Variola virus strain Brazil_131 right variable genomic sequence
Length=10891

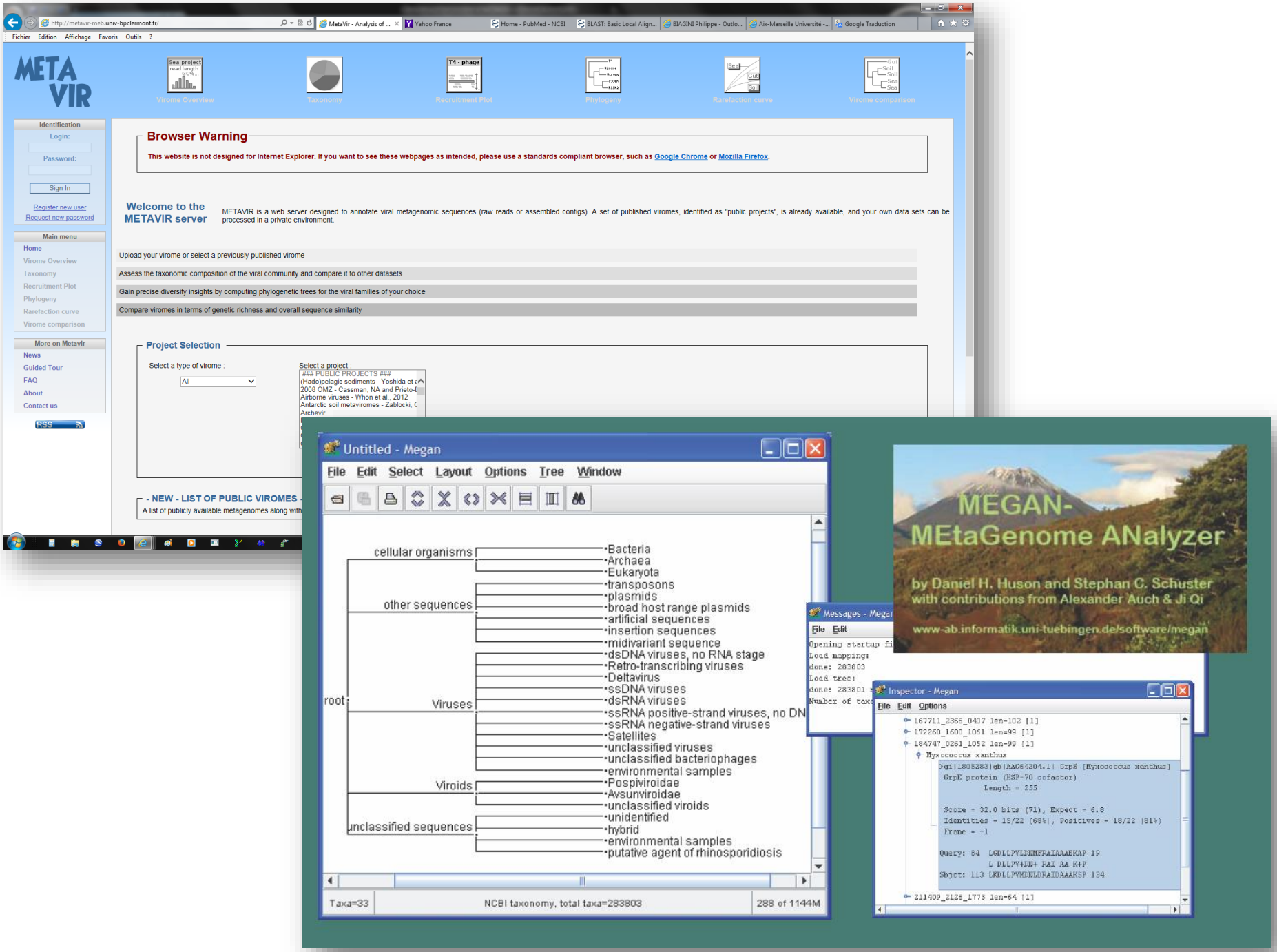
GENE ID: 1486563 B7R | B7R [Variola virus] **(10 or fewer PubMed links)**

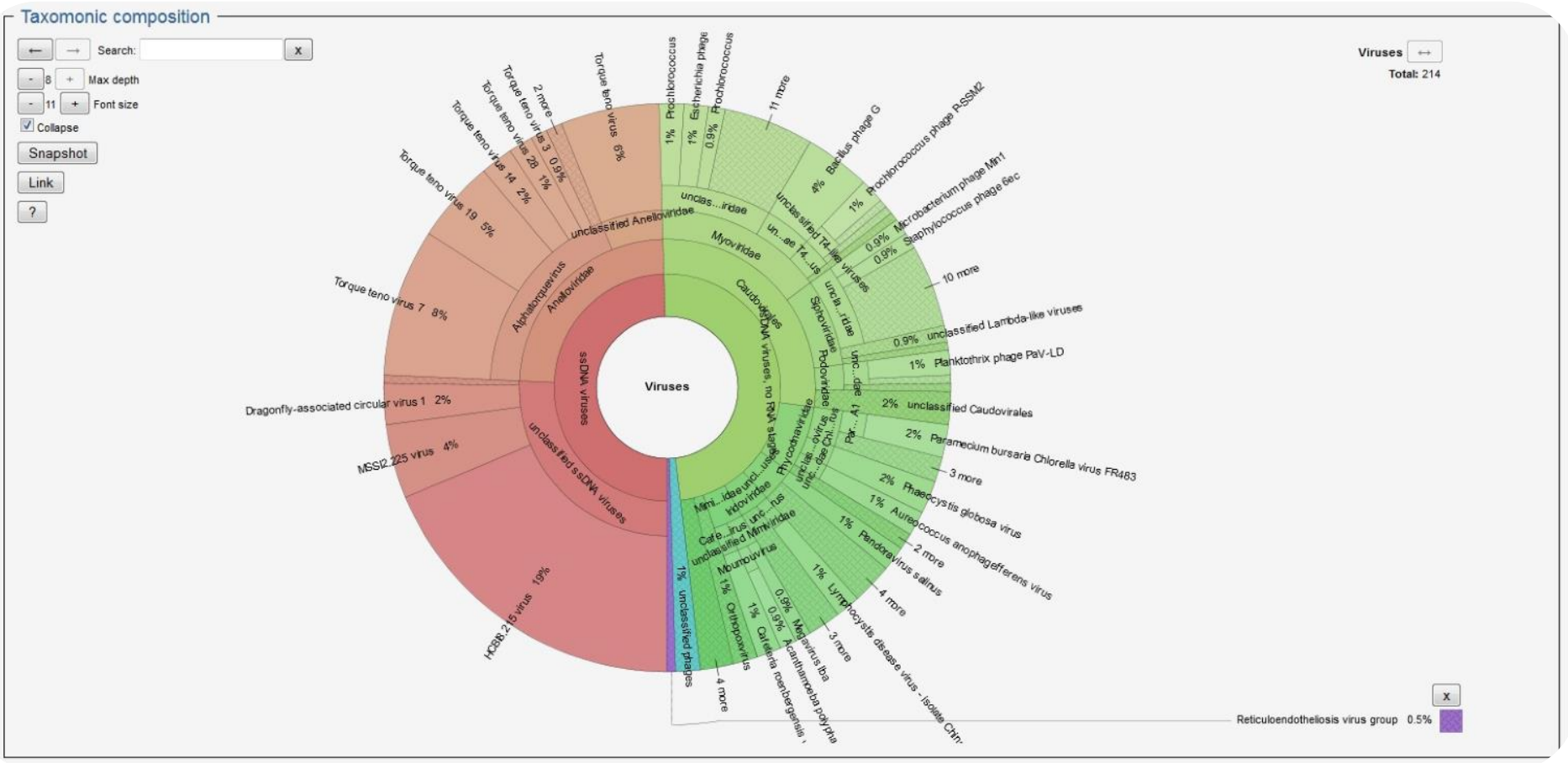
Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

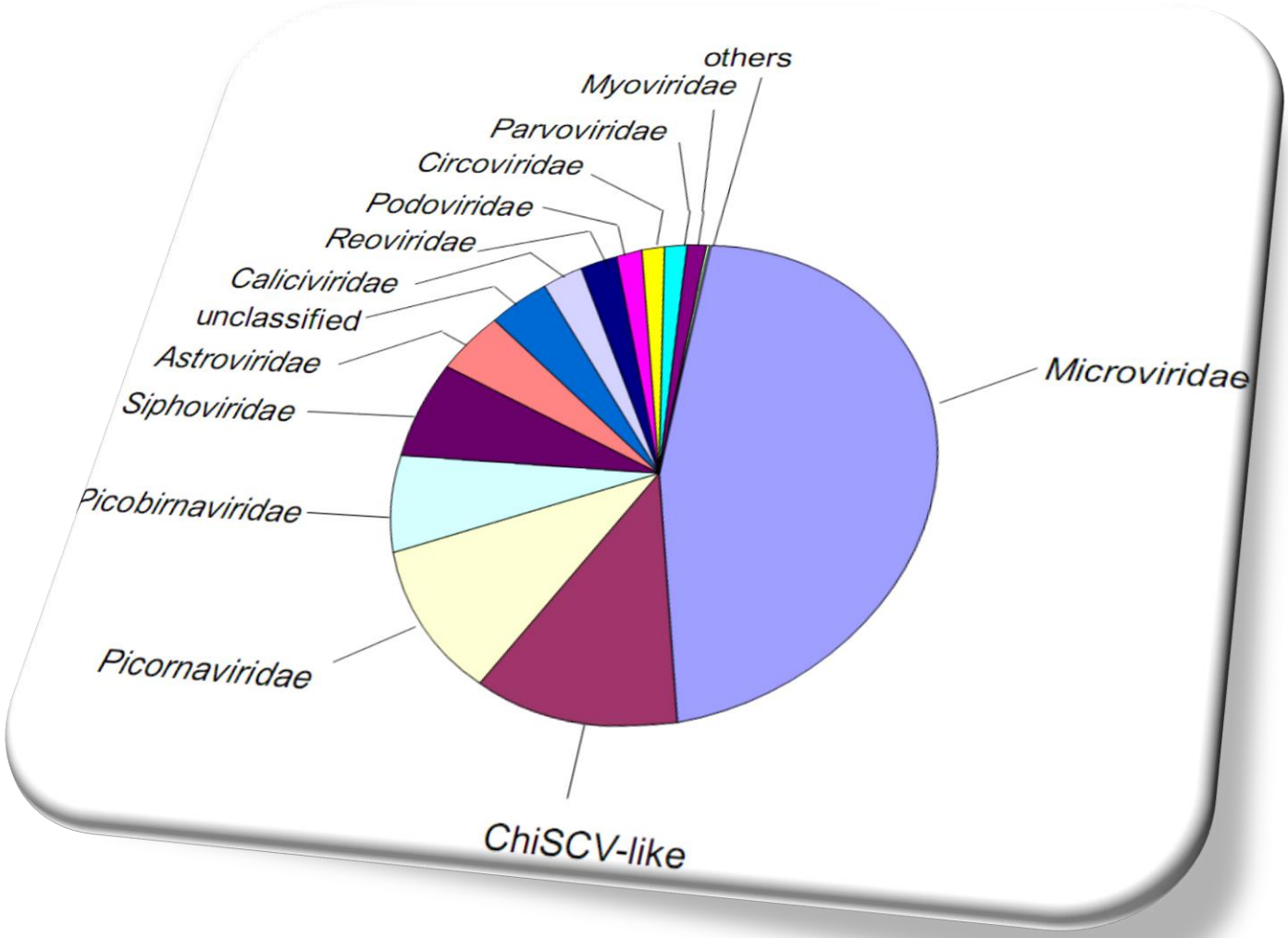
Score = 250 bits (276), Expect = 2e-63
Identities = 138/138 (100%), Gaps = 0/138 (0%)
Strand=Plus/Plus

```
Query  1     CATCTGGAGAATCCACAACAGACAAGACTTCGGGACCAATTACTAATAAAGAAGATCATA  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2203  CATCTGGAGAATCCACAACAGACAAGACTTCGGGACCAATTACTAATAAAGAAGATCATA  2262

Query  61    CAGTCACAGACACTGTCTCATACACTACAGTAAGTACATCATCTGAAATTGTCACTACTA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2263  CAGTCACAGACACTGTCTCATACACTACAGTAAGTACATCATCTGAAATTGTCACTACTA  2322

Query  121   AATCAACCGCCAATGATG  138
             ||||||||||||||||||
Sbjct  2323  AATCAACCGCCAATGATG  2340
```

Score = 64.4 bits (70), Expect = 2e-07
Identities = 71/93 (76%), Gaps = 14/93 (15%)
Strand=Plus/Plus


METAVIR
Virome Overview
Taxonomy
Recruitment Plot
Phylogeny
Rarefaction curve
Virome comparison
Identification
Login:
Password:
Sign In
Register new user
Request new password
Main menu
Home
Virome Overview
Taxonomy
Recruitment Plot
Phylogeny
Rarefaction curve
Virome comparison
More on Metavir
News
Guided Tour
FAQ
About
Contact us
RSS
Browser Warning
This website is not designed for Internet Explorer. If you want to see these webpages as intended, please use a standards compliant browser, such as Google Chrome or Mozilla Firefox.
Welcome to the METAVIR server
METAVIR is a web server designed to annotate viral metagenomic sequences (raw reads or assembled contigs). A set of published viromes, identified as "public projects", is already available, and your own data sets can be processed in a private environment.
Upload your virome or select a previously published virome
Assess the taxonomic composition of the viral community and compare it to other datasets
Gain precise diversity insights by computing phylogenetic trees for the viral families of your choice
Compare viromes in terms of genetic richness and overall sequence similarity
Project Selection
Select a type of virome :
All
Select a project :
PUBLIC PROJECTS
(Hado)pelagic sediments - Yoshida et
2008 OMZ - Cassman, NA and Prieto-
Airborne viruses - Whon et al., 2012
Antarctic soil metavirome - Zablocki,
Archevir
- NEW - LIST OF PUBLIC VIROMES -
A list of publicly available metagenomes along with
Untitled - Megan
File Edit Select Layout Options Tree Window
cellular organisms
Bacteria
Archaea
Eukaryota
other sequences
transposons
plasmids
broad host range plasmids
artificial sequences
insertion sequences
midivariant sequence
root
Viruses
dsDNA viruses, no RNA stage
Retro-transcribing viruses
Deltavirus
ssDNA viruses
dsRNA viruses
ssRNA positive-strand viruses, no DN
ssRNA negative-strand viruses
Satellites
unclassified viruses
unclassified bacteriophages
environmental samples
Viroids
Pospiviroidae
Avsunviroidae
unclassified viroids
unclassified sequences
unidentified
hybrid
environmental samples
putative agent of rhinosporidiosis
Taxa=33
NCBI taxonomy, total taxa=283803
288 of 1144M
Messages - Megan
File Edit
Opening startup fi
Load mapping:
done: 283803
Load tree:
done: 283801
Number of taxo
Inspector - Megan
File Edit Options
167711_2368_0407 len=102 [1]
172260_1600_1061 len=99 [1]
184747_0261_1052 len=99 [1]
Myxococcus xanthus
>gi|1805283|qb|AAC84204.1| GrpE [Myxococcus xanthus]
GrpE protein (HSP-70 cofactor)
Length = 255
Score = 32.0 bits (71), Expect = 6.8
Identities = 15/22 (68%), Positives = 18/22 (81%)
Frame = -1
Query: 84 LGDLLPVLDMMFRAIAAAEKAP 19
L DLLPV+DM+ RAI AA K+P
Sbjct: 113 LKDLLPVMDMLDRAIDAAAKSP 134
211409_2126_1773 len=64 [1]
MEGAN-
MEtaGenome ANalyzer
by Daniel H. Huson and Stephan C. Schuster
with contributions from Alexander Auch & Ji Qi
www-ab.informatik.uni-tuebingen.de/software/megan


Taxomonic composition
Search:
Max depth
Font size
Collapse
Snapshot
Link
?
Viruses
Total: 214
Viruses
ssDNA viruses
dsDNA viruses, no RNA stage
unclassified phages
Anelloviridae
Alphatorquevirus
unclassified Anelloviridae
Myoviridae
Caudovirales
Siphoviridae
Podoviridae
Phycodnaviridae
Mimiviridae
unclassified Mimiviridae
Mimivirus
unclassified ssDNA viruses
Torque teno virus 6%
Torque teno virus 3 0.9%
Torque teno virus 28 1%
Torque teno virus 14 2%
Torque teno virus 19 5%
Torque teno virus 7 8%
Dragonfly-associated circular virus 1 2%
MSSI2_225 virus 4%
HCBB_215 virus 19%
Prochlorococcus 1%
Escherichia phage 1%
Prochlorococcus 0.9%
11 more
Bacillus phage G 4%
Prochlorococcus phage P-SSM2 1%
Microbacterium phage Min1 0.9%
Staphylococcus phage 6ec 0.9%
10 more
unclassified Lambda-like viruses 0.9%
Planktothrix phage PaV-LD 1%
unclassified Caudovirales 2%
Paramecium bursaria Chlorella virus FR483 2%
3 more
Phaeocystis globosa virus 2%
Aureococcus anophagefferens virus 1%
Pandoravirus salinus 1%
2 more
Lymphocystis disease virus - isolate Chin 1%
3 more
Megavirus lba 0.9%
Acanthamoeba polyphaga 0.9%
Cafeteria roenbergensis 1%
Orthopoxvirus 1%
4 more
Reticuloendotheliosis virus group 0.5%


others
Myoviridae
Parvoviridae
Circoviridae
Podoviridae
Reoviridae
Caliciviridae
unclassified
Astroviridae
Siphoviridae
Picobirnaviridae
Picornaviridae
ChiSCV-like
Microviridae

… Informatique… ?!

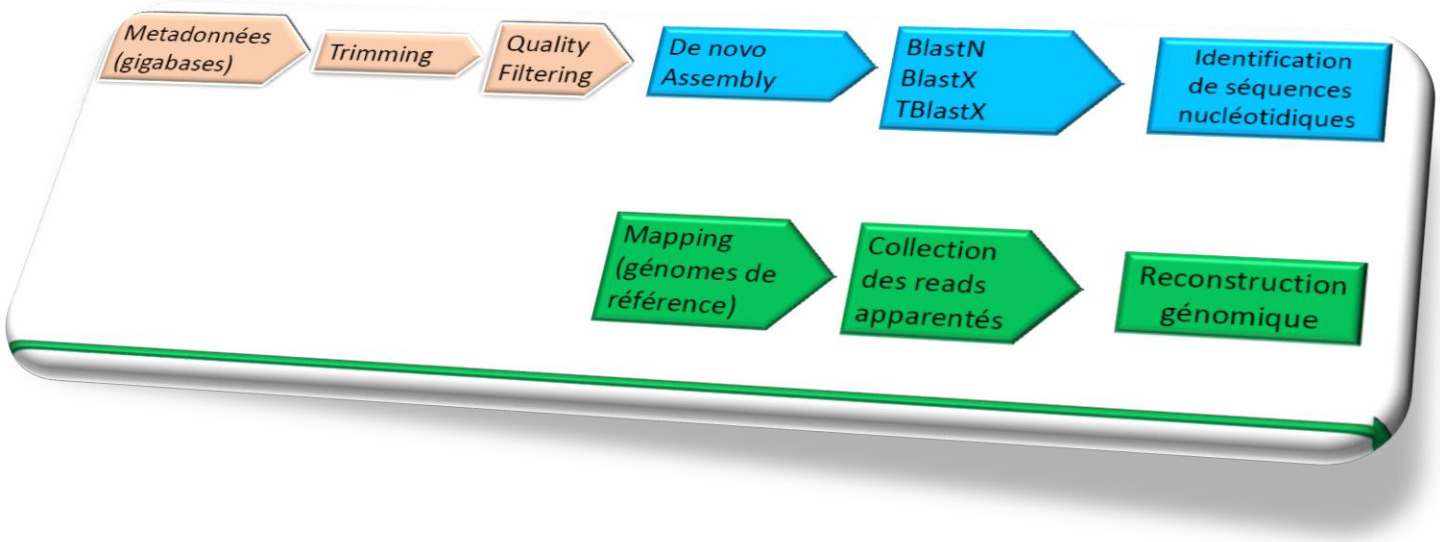
Projet APR EFS « Métagénomique virale ».

Etude moléculaire des co-infections sur prélèvements HBV+, HCV+, HIV+.

Caractérisation moléculaire des séquences virales d'intérêt.

Design systèmes de détection PCR/PCR real-time. Etudes épidémiologiques.

Renseignements sur séquences complètes / génotypes de chaque virus témoin.



- Utilisation de pipelines dédiés pour le traitement des métadonnées (Galaxy)
- Mise en place de stations de calcul dédiées (x cœurs, RAM, stockage,…)
- Implémentation de logiciels (CLC Genomics Workbench, CodonCode Aligner, MEGAN,…)

1ère publication:

Divergent Gemycircularvirus in HIV-Positive Blood, France

Rathviro Uch, Pierre-Edouard Fournier, Catherine Robert, Caroline Blanc-Tailleur, Vital Galicher, Romain Barre, François Jordier, Philippe de Micco, Didier Raoult, Philippe Biagini

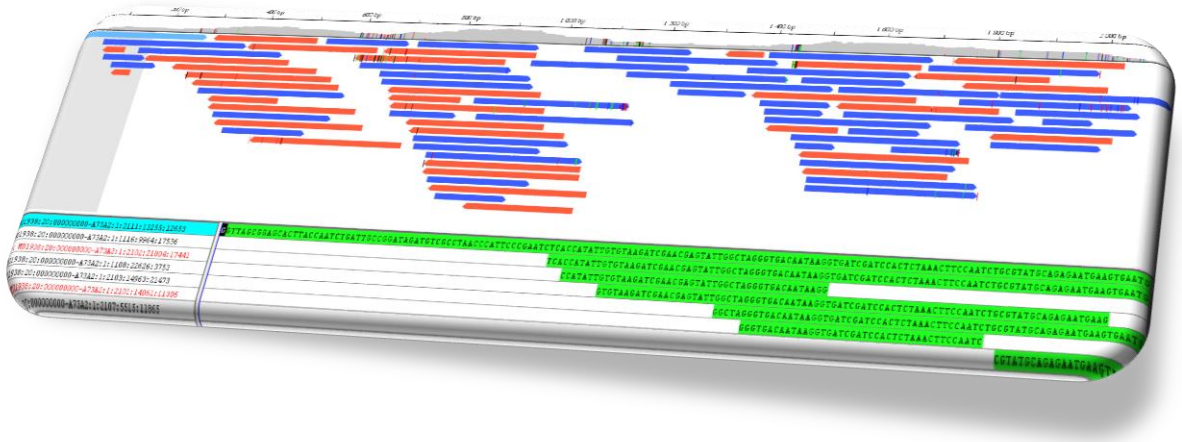
Emerging Infectious Diseases • www.cdc.gov/eid • Vol. 21, No. 11, November 2015

Séquences de type **gemycircularvirus**:

(virus circulaires à ADNsb / histoire naturelle ?)

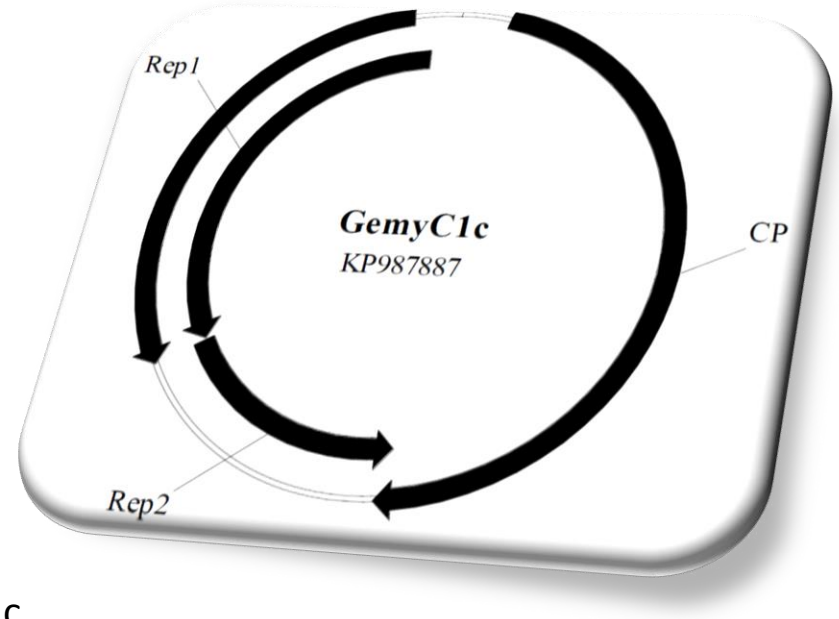
- ~2%

- assemblage de novo



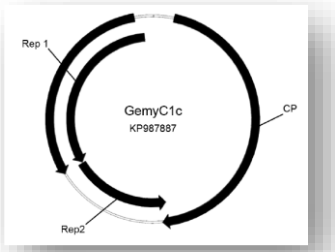
- séquence validée par PCR inversée et clonages

Plasma VIH-1 ARN+
sous-type B
~530 copies/ml



Gemycircularvirus C1c

(2109 nt)



Sclerotinia sclerotiorum SsHADV-1 (YP 003104795)
Sheep Gemycircularvirus 1a (YP 009109740)
Bird Gemycircularvirus 5 (YP 009109728)
Dragonfly CV1 (YP 009021855)
Bird Gemycircularvirus 1b (YP 009109738)
Bird Gemycircularvirus 1c (YP 009109736)
Odonata Gemycircularvirus 1 (AJD07459)
Odonata Gemycircularvirus 2 (AJD07463)
Hypericum japonicum CV (AGS12482)
Dragonfly CV3 (YP 009021851)
Bird Gemycircularvirus 8 (YP 009109722)
Sewage Gemycircularvirus 3 (YP 009115528)
Bird Gemycircularvirus 10 (YP 009109718)
Seal Gemycircularvirus 4 (YP 009109730)
Cassava CV (YP 009021042)
Bird Gemycircularvirus 9 (YP 009109720)
Sewage Gemycircularvirus 8 (YP 009115536)
Sewage Gemycircularvirus 7a (YP 009115535)
Mosquito SDBVL-G (AEF58774)
Dragonfly CV1 (YP 009021859)
Human feces Gemycircularvirus BZ1 (AJD20390)
Human CSF Gemycircularvirus SL1 (YP 009130660)
Mongoose Gemycircularvirus a (YP 009130623)
Badger Gemycircularvirus (YP 009130621)
Human blood HIV+ GemyC1c (KP987887)
Mongoose Gemycircularvirus c (YP 009130627)
Cattle HCBI8.215 (YP 009051830)
Mongoose Gemycircularvirus b (YP 009130625)
Mongoose Gemycircularvirus d (YP 009130629)
Human blood MSSI2.225 (YP 009051836)
Sewage Gemycircularvirus 10b (AIF34857)
Meles meles VS4700006 (AEW87507)

70% div.

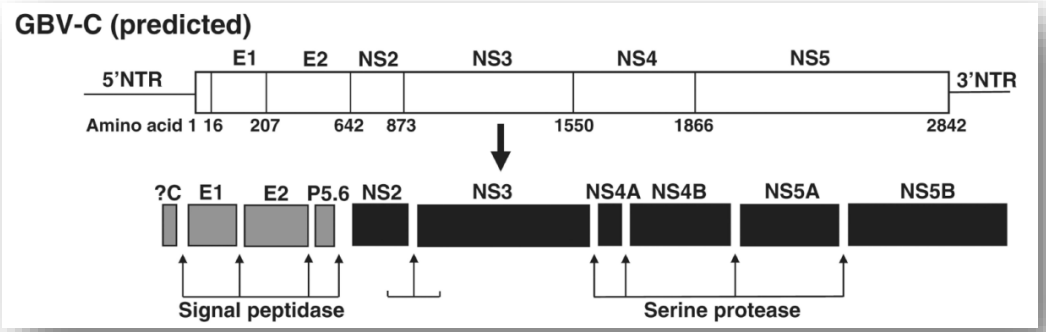
CAP aa

0.05

2ème publication en préparation:

- 16 échantillons VHC+ / VIH+ contenant des séquences du virus GBV-C

(Pegivirus, *Flaviviridae*, ARNsb lin., ~9.300 nt)



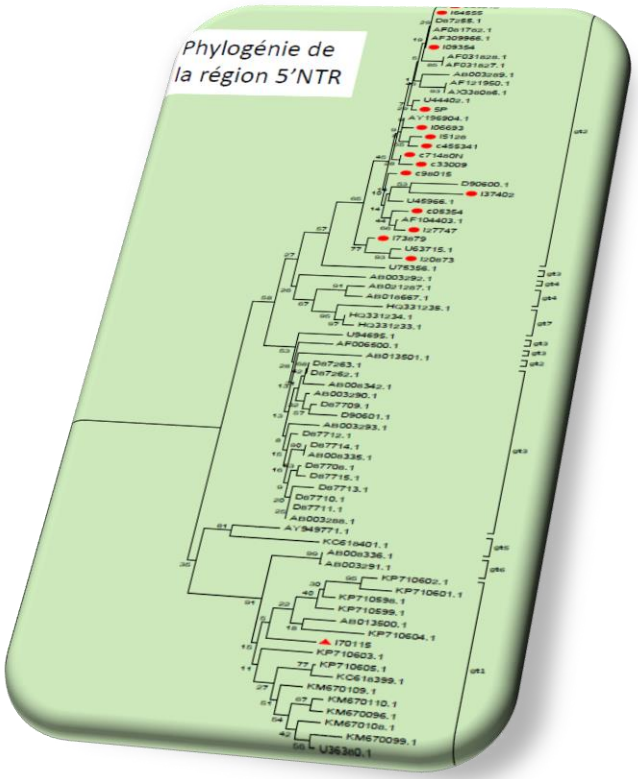
- Co-infections avec VHC, VIH
- Impact ?
- Ralentissement évolution VIH SIDA…

- 6 génotypes connus
- France: 1 séquence complète…

- Assemblages / Reconstruction 16 séquences complètes – partielles

- Analyse phylogénétique / Recombinaisons…

- Meilleure connaissance de la diversité génétique du virus en France



Exploitation des données en cours:

- Analyse des séquences: # variants (papillomavirus, circo-cyclovirus,…)

inconnues…

- Inventaire des espèces virales / échantillon testé

- Si séquences d'intérêt: épidémiologie, fondamental,…

Merci !