Diversity of HEV genotype 3 based on full-length sequences

Dr Florence Nicot

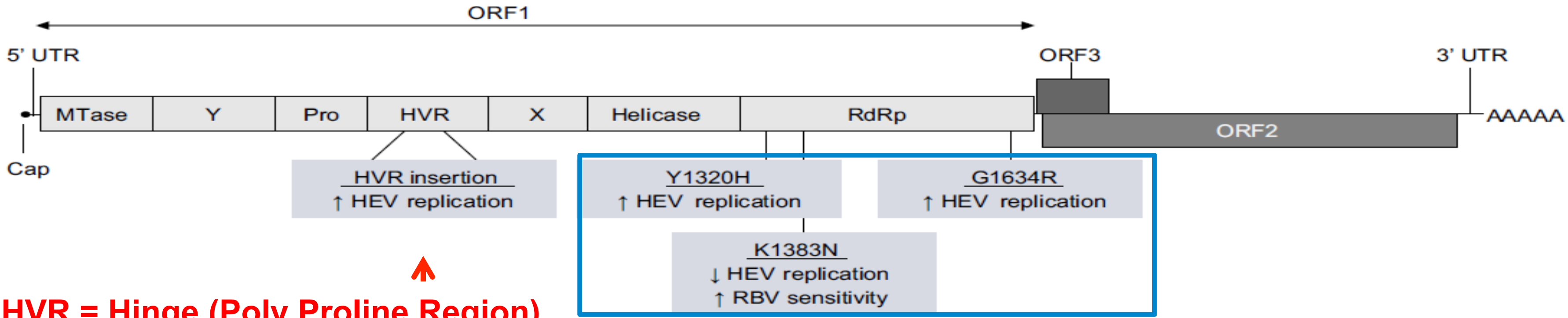
Laboratoire de Virologie, Institut Fédératif de Biologie

Hôpital Purpan, CHU Toulouse

« Long reads: Dream or Reality » 28-11-2017

Hepatitis E virus (HEV)

- *Hepeviridae* family, *Hepevirus* genus
- Single-stranded, positive-sense RNA genome



HVR = Hinge (Poly Proline Region)

Mutations in RdRp involved in ribavirin virological failure

Adapted from Debing, J Hepatol, 2016

HEV transmission routes

❖ Water and interhuman for genotypes 1 et 2

➔ Young people

Severe forms:

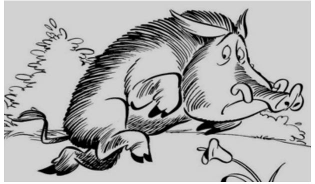
- Pregnant women
- Chronic liver disease

❖ Zoonotic for genotypes 3 et 4

➔ « Old » people

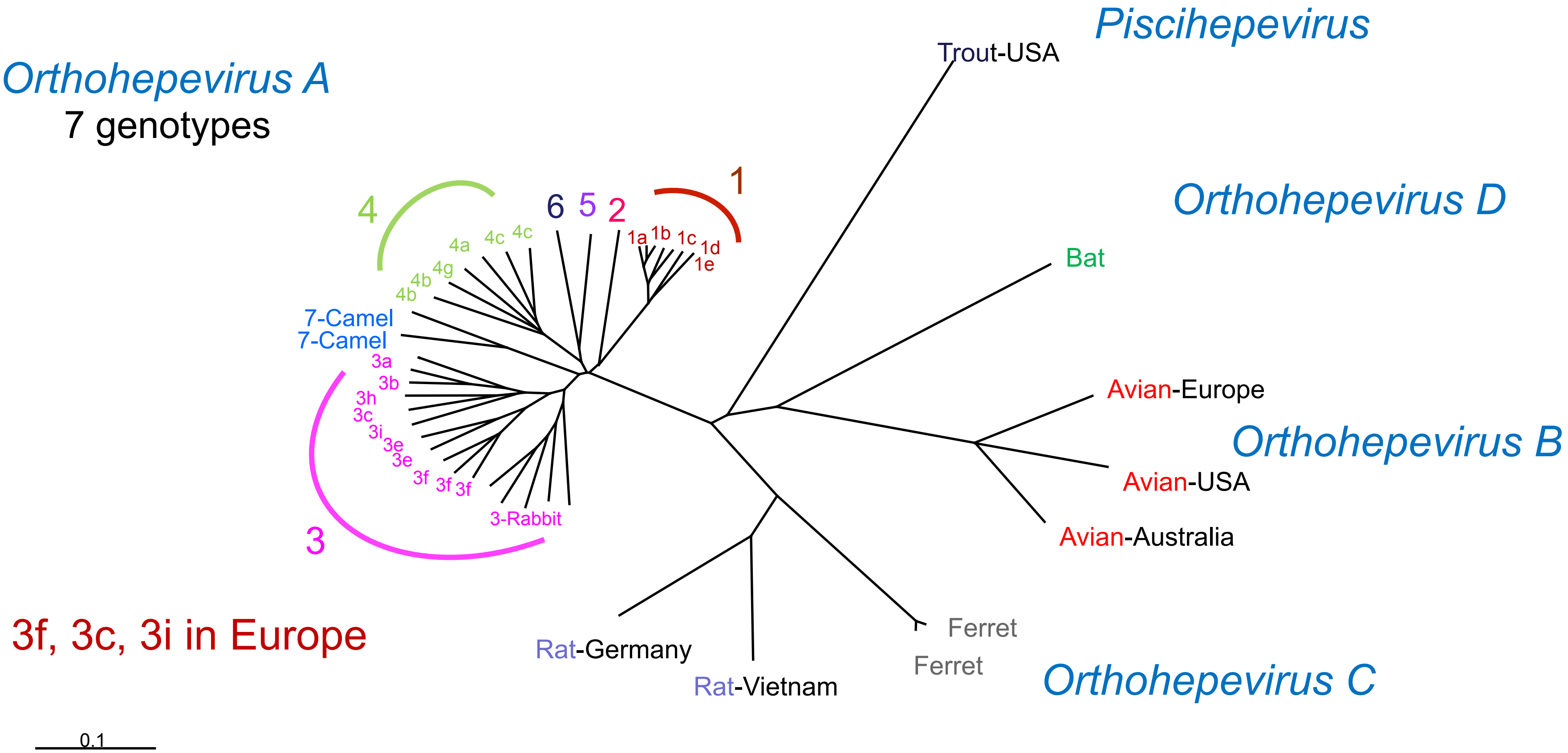
Severe forms :

- Chronic liver disease
- Immunosuppressed : 2/3 of chronic infection, treatment with ribavirin



HEV-3 +++ in industrialized countries

HEV diversity



Orthohepevirus A
7 genotypes

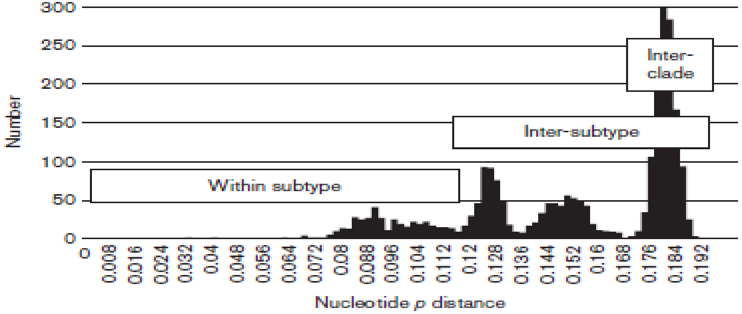
3f, 3c, 3i in Europe

324 full length genomes, only 44 HEV-3 from European origin

HEV-3 strain classification

- Classification:
    - Clades: abchijk, efg, rabbit
    - Subtypes: 3a, 3b, 3c …

- Difficulties to discriminate HEV subtypes from HEV clades based on distance analysis



Smith, JGV, 2015

- Set of reference sequences (full length HEV sequences) recently proposed to harmonize classification

Smith, JGV, 2016

Interests of HEV classification

- **Epidemiologic**: characterisation of strains circulating in an area (genotype, clade, subtype)

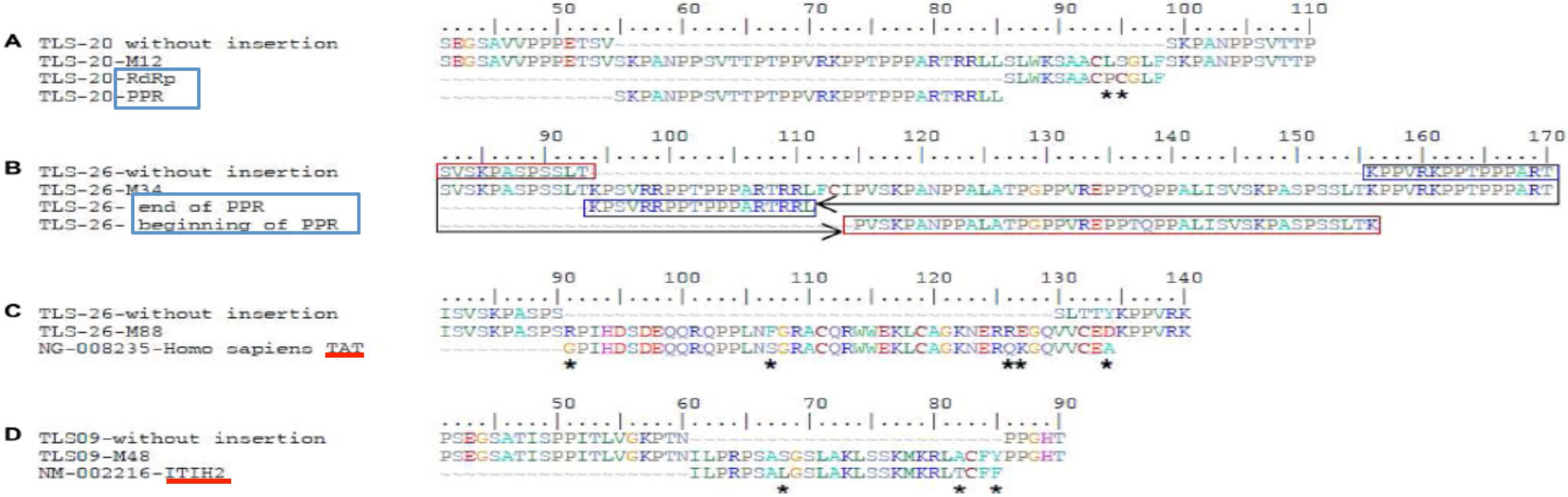
- **Tracing the origin of infection :**
  - Transfusional investigation
  - Comparing strains circulating in a group of persons

- **Regions of the HEV genome usually sequenced**:
  - ORF2: 350 bp +++
  - ORF1: RdRp

Recombinant HEV strains

27 solid organ transplant patients chronically HEV-infected

11% of recombinant events



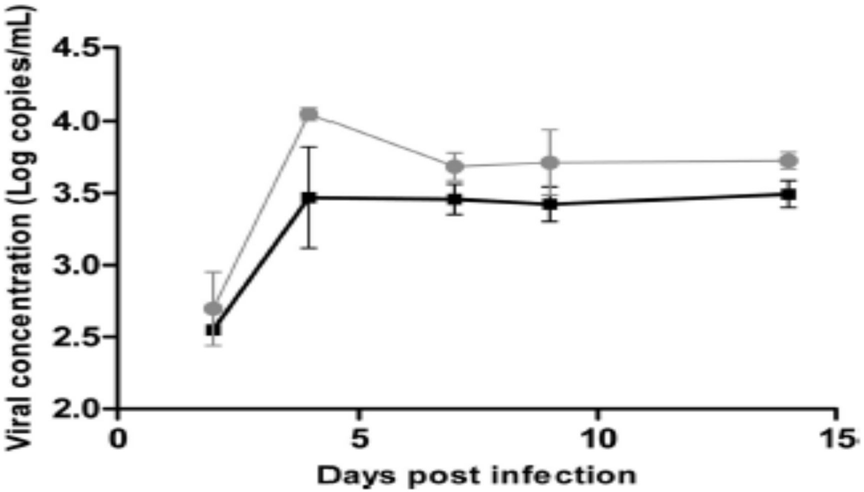
Lhomme, J Virol, 2014

Only 4 strains with an insert of human fragment described.

Sukla, PNAS, 2011; Nguyen, JGV, 2012; Lhomme, J Virol, 2014

Replicative advantage of strains with insertion

Culture cells on HepG2/C3A



Strain with insertion (ITIH2)
Strain without insertion

Lhomme, J Virol, 2014

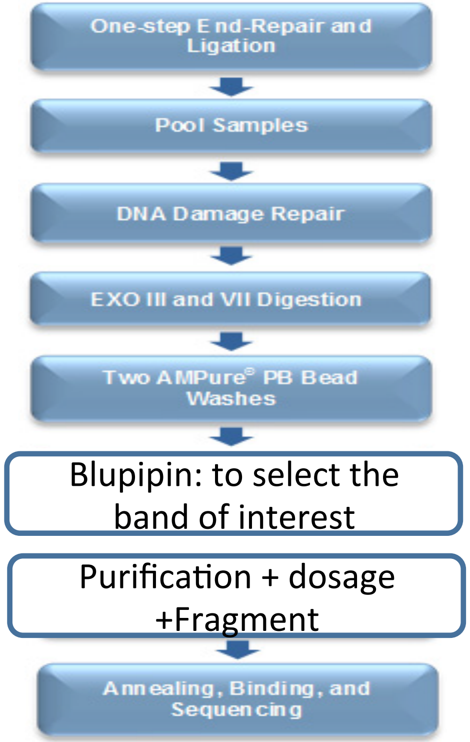
AIMS

- **Increase the database of full-length HEV genome**
- **Identify and describe insertions in the HEV genome (frequency, nature, localisation, features)**

Library preparation

Amplification of 2 overlapping fragments (2x4,500 bp)
Pool of 16 samples by SMRT cells

Virology	Amplifications ORF1+ORF2 (2x4500 bp)	N=32
	Purification 0,5X ORF1 et ORF2	N=32
Get-Plage	Dosage Qubit + Fragment ORF1 et ORF2	N=32
	Equimolar pool ORF1+ORF2	N=16
	Qubit Dosage+ Fragment of each patient	N=16
	Purification	



One-step End-Repair and Ligation N=16

Pool Samples N=1

DNA Damage Repair

EXO III and VII Digestion

Two AMPure® PB Bead Washes

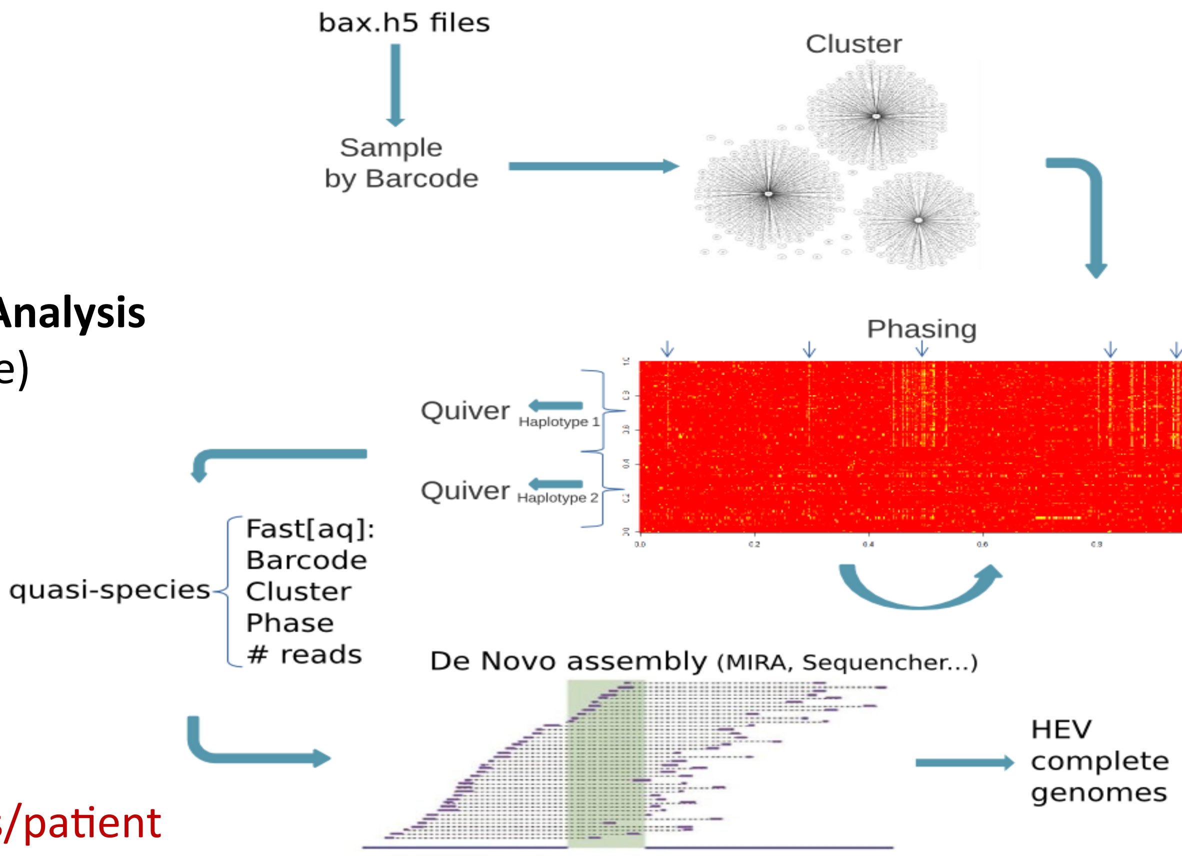
Blupipin: to select the band of interest

Purification + dosage +Fragment

Annealing, Binding, and Sequencing

Bioinformatic analysis

Long Amplicon Analysis
(SMRT Pipe)



bax.h5 files

Sample
by Barcode

Cluster

Phasing

Quiver

Quiver

quasi-species

Fast[aq]:
Barcode
Cluster
Phase
reads

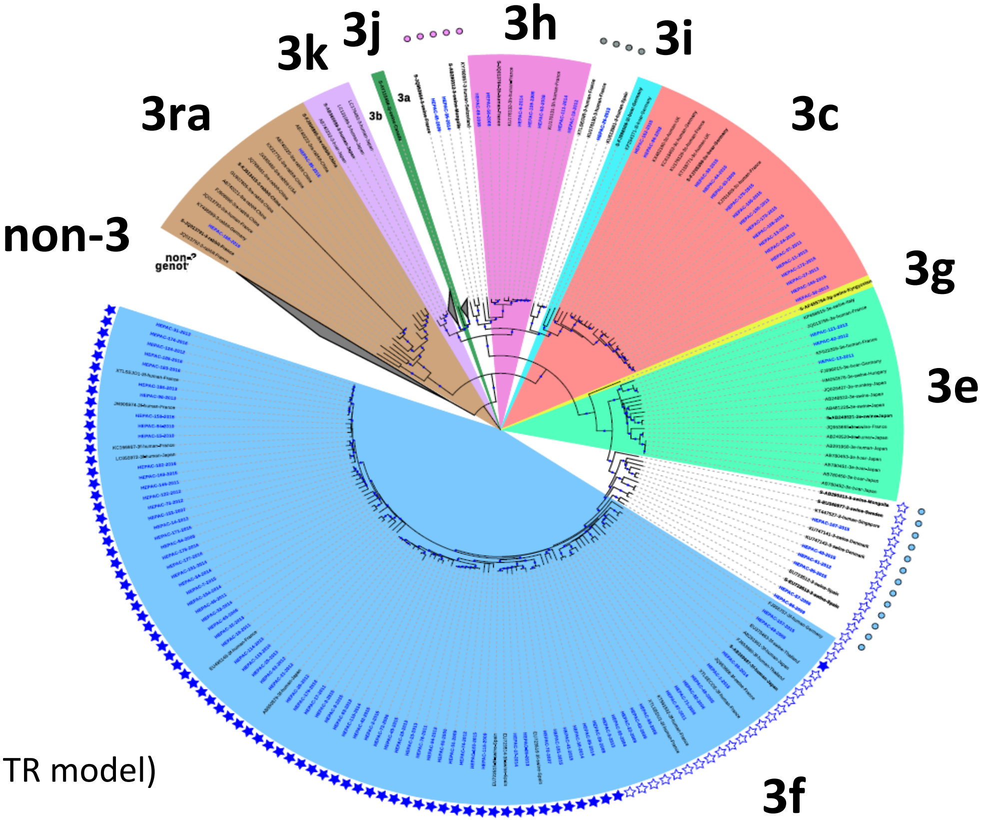
De Novo assembly (MIRA, Sequencher...)

HEV
complete
genomes

1-10 consensus/patient

HEV-3 complete genome sequences (n=250)

115 new full length HEV-3 genomes (74% efg, 24% abjchik, 2% 3rabbit)

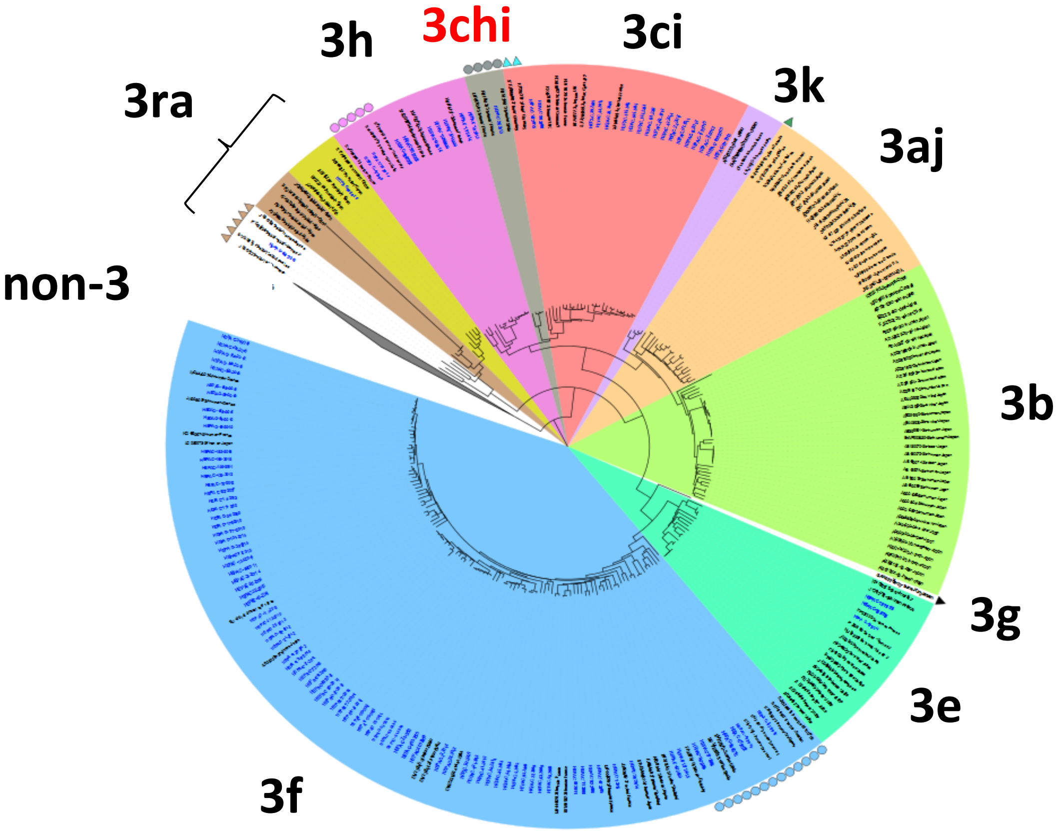


● ● ● 22 unassigned subtypes

ML tree (GTR model)

Nicot, submitted

Automated clustering of HEV subtypes



Automated partition of ML tree

Prosperi, Nat Com, 2011

This method minimize the number of unassigned subtypes (only 4 remained unassigned)

Nicot, submitted

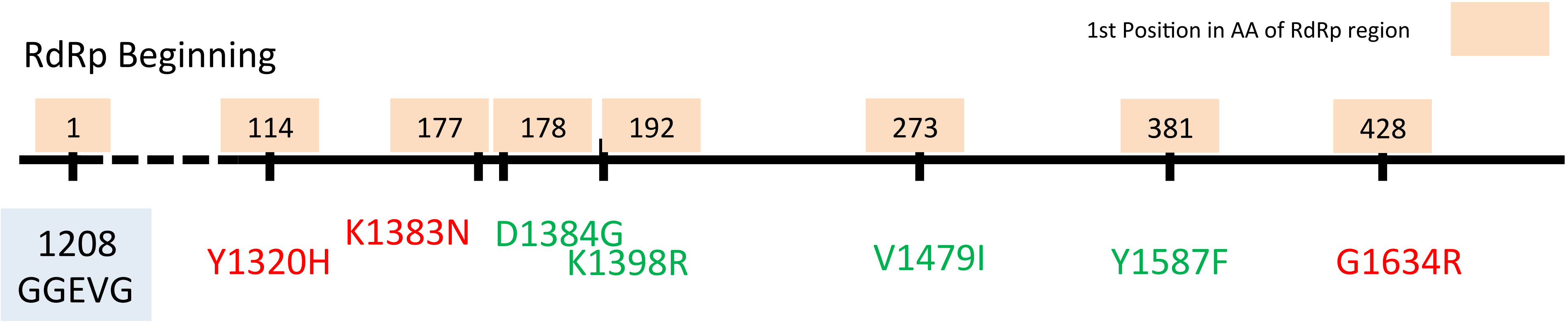
Characterization of Hinge region

Subtype	N	Hinge length Nucleotide
3e	3	228
3f	18	228 (short)
	1	288 (intermediate)
	63	315 (long)
3c	18	225
3h	9	225
3 chi	1	225
3 rabbit	2	189

The length of Hinge region is similar for strains classified in the same clade.

Nicot, submitted

Mutations in RdRp associated with RBV resistance



RdRp Beginning

1st Position in AA of RdRp region

1
114
177
178
192
273
381
428

1208
GGEVG

Y1320H
K1383N
D1384G
K1398R
V1479I
Y1587F
G1634R

Debing et al, 2016; Todt et al, 2016

The frequency of strains harboring these mutations before treatment according to HEV-3 subtypes has not been studied on a large data set.

Frequency of mutation in RdRp

Subtype		3chi-UN2	3aj	3b	3ci	3e	3f	3g	3h	3k	3ra	G3 total
(Number)		(4)	(21)	(37)	(24)	(19)	(107)	(1)	(17)	(4)	(16)	(250)
HEV RdRp mutations	Y1320H	0	0	0	0	0	3 (3%)	0	0	0	0	3 (1%)
	K1383N	0	0	0	0	0	5 (5%)	0	0	0	0	5 (2%)
	D1384G	0	0	0	0	0	0	0	0	0	0	0
	K1398R	0	0	0	0	0	0	0	0	0	0	0
	V1479I	4 (100%)	3 (14%)	7 (19%)	6 (25%)	19 (100%)	107 (100%)	1 (100%)	5 (29%)	4 (100%)	16 (100%)	172 (69%)
	Y1597F	0	0	0	0	0	1 (1%)	0	0	0	0	1 (0.4%)
	G1634R	0	1 (5%)	0	0	18 (95%)	15 (14%)	0	0	3 (75%)	0	37 (15%)
	G1634K	0	0	0	0	1 (5%)	1 (1%)	0	0	0	9 (56%)	11 (4%)

Subtype-specific polymorphisms in the HEV-3 polymerase have been identified.

Nicot, submitted

Strains with insertion

VHP6 strain was used as a control: insert of fragment of human genome was previously observed by shot-gun 454 GS Junior pyrosequencing.

8/115 (7%) strains with a recombination event:

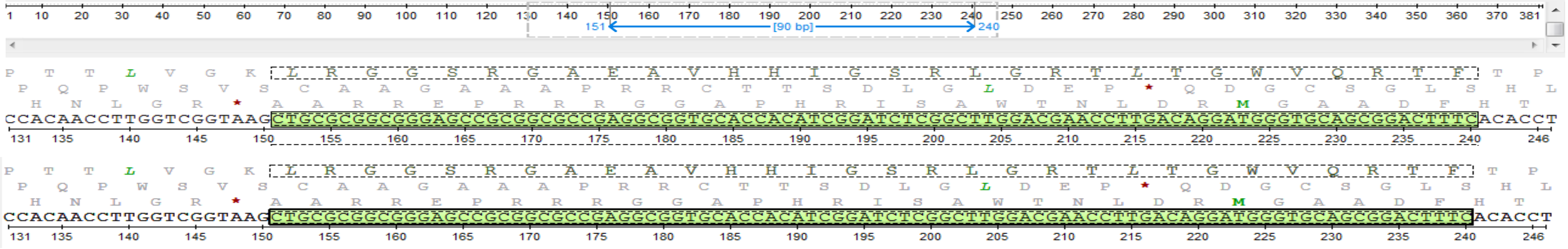
- 6 with insertions of human fragment, 2 with duplications of HEV genome
- located in Hinge region
- identified in HEV-3

Insertions were confirmed by sanger sequencing.

Nicot, manuscript in preparation

Control strain VHP6

VHP6-GATM (L-arginine glycine amidinotransferase)



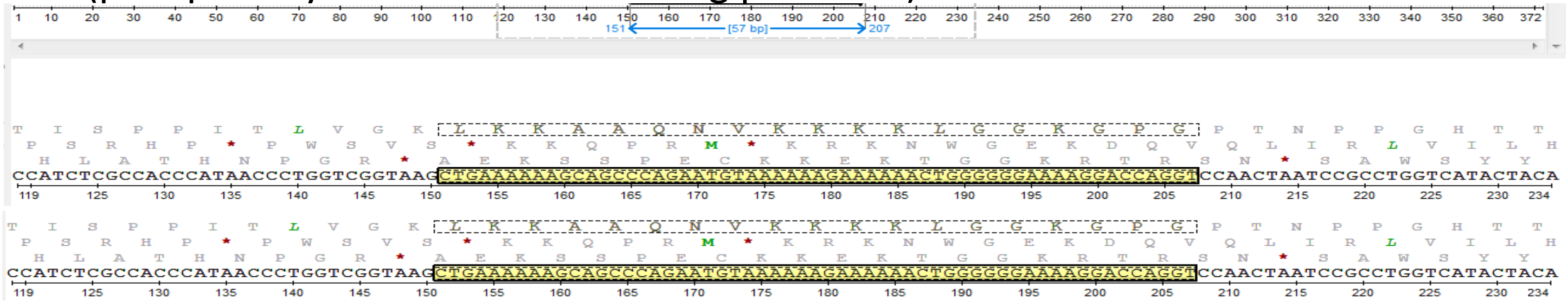
PacBio

```
P  T  T  L  V  G  K  L  R  G  G  S  R  G  A  E  A  V  H  H  I  G  S  R  L  G  R  T  L  T  G  W  V  Q  R  T  F  T  P
 P  Q  P  W  S  V  S  C  A  A  G  A  A  A  P  R  R  C  T  T  S  D  L  G  L  D  E  P  *  Q  D  G  C  S  G  L  S  H  L
  H  N  L  G  R  *  A  A  R  R  E  P  R  R  R  G  G  A  P  H  R  I  S  A  W  T  N  L  D  R  M  G  A  A  D  F  H  T
CCACAACCTTGGTCGGTAAGCTGCGCGGCGGGAGCCGCGGCGCCGAGGCGGTGCACCACATCGGATCTCGGCTTGGACGAACCTTGACAGGATGGGTGCAGCGGACTTTCACACCT
131 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 246
```

Shot-gun

```
P  T  T  L  V  G  K  L  R  G  G  S  R  G  A  E  A  V  H  H  I  G  S  R  L  G  R  T  L  T  G  W  V  Q  R  T  F  T  P
 P  Q  P  W  S  V  S  C  A  A  G  A  A  A  P  R  R  C  T  T  S  D  L  G  L  D  E  P  *  Q  D  G  C  S  G  L  S  H  L
  H  N  L  G  R  *  A  A  R  R  E  P  R  R  R  G  G  A  P  H  R  I  S  A  W  T  N  L  D  R  M  G  A  A  D  F  H  T
CCACAACCTTGGTCGGTAAGCTGCGCGGCGGGAGCCGCGGCGCCGAGGCGGTGCACCACATCGGATCTCGGCTTGGACGAACCTTGACAGGATGGGTGCAGCGGACTTTCACACCT
131 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 246
```

VHP6-PEBP1 (phosphatidylethanolamine binding protein 1)



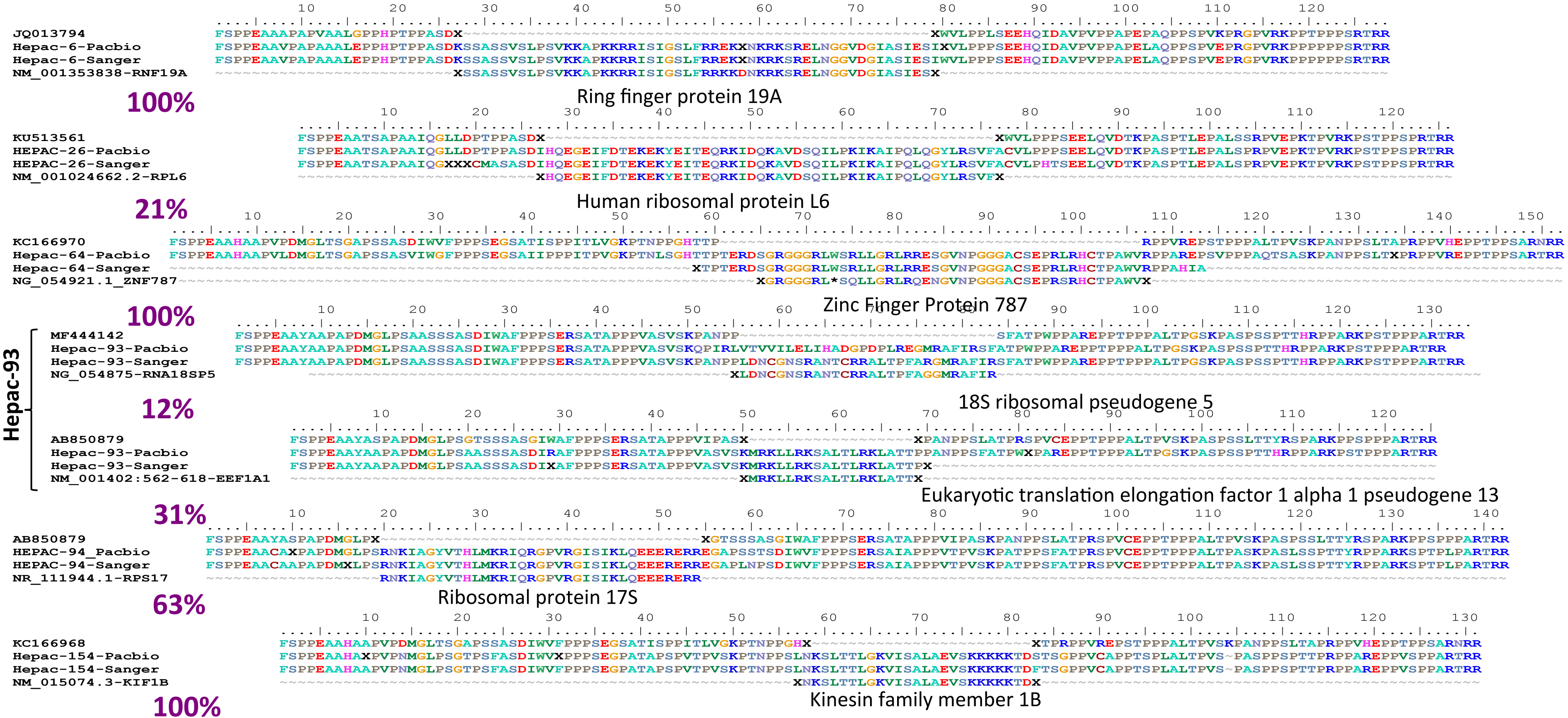
PacBio

```
T  I  S  P  P  I  T  L  V  G  K  L  K  K  A  A  Q  N  V  K  K  K  K  L  G  G  K  G  P  G  P  T  N  P  P  G  H  T  T
 P  S  R  H  P  *  P  W  S  V  S  *  K  K  Q  P  R  M  *  K  R  K  N  W  G  E  K  D  Q  V  Q  L  I  R  L  V  I  L  H
  H  L  A  T  H  N  P  G  R  *  A  E  K  S  S  P  E  C  K  K  E  K  T  G  G  K  R  T  R  S  N  *  S  A  W  S  Y  Y
CCATCTCGCCACCCATAACCCTGGTCGGTAAGCTGAAAAAAGCAGCCCAGAATGTAAAAAAGAAAAAACTGGGGGGAAAAGGACCAGGTCCAACTAATCCGCCTGGTCATACTACA
119 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 234
```

Shot-gun

```
T  I  S  P  P  I  T  L  V  G  K  L  K  K  A  A  Q  N  V  K  K  K  K  L  G  G  K  G  P  G  P  T  N  P  P  G  H  T  T
 P  S  R  H  P  *  P  W  S  V  S  *  K  K  Q  P  R  M  *  K  R  K  N  W  G  E  K  D  Q  V  Q  L  I  R  L  V  I  L  H
  H  L  A  T  H  N  P  G  R  *  A  E  K  S  S  P  E  C  K  K  E  K  T  G  G  K  R  T  R  S  N  *  S  A  W  S  Y  Y
CCATCTCGCCACCCATAACCCTGGTCGGTAAGCTGAAAAAAGCAGCCCAGAATGTAAAAAAGAAAAAACTGGGGGGAAAAGGACCAGGTCCAACTAATCCGCCTGGTCATACTACA
119 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 234
```

Insertions of fragment of human gene



```
                                  10        20        30        40        50        60        70        80        90       100       110       120
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|...
JQ013794                  FSPPEAAAPAPVAALGPPHPTPPASDX~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~XWVLPPLSEEHQIDAVPVPPAPEPAQPPSPVKPRGPVRKPPTPPPSRTRR
Hepac-6-Pacbio            FSPPEAAVPAPAAALEPPHPTPPASDKSSASSVSLPSVKKAPKKRRISIGSLFRREKXNKRKSRELNGGVDGIASIESIXVLPPPSEEHQIDAVPVPPAPELAQPPSPVEPRGPVRKPPPPPPSRTRR
Hepac-6-Sanger            FSPPEAAVPAPAAALEPPHPTPPASDKSSASSVSLPSVKKAPKKRRISIGSLFRREKXNKRKSRELNGGVDGIASIESIWVLPPPSEEHQIDAVPVPPAPELAQPPSPVEPRGPVRKPPPPPPSRTRR
NM_001353838-RNF19A       ~~~~~~~~~~~~~~~~~~~~~~~~~~XSSASSVSLPSVKKAPKKRRISIGSLFRRKKDNKRKSRELNGGVDGIASIESX~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

100%

Ring finger protein 19A

```
                                  10        20        30        40        50        60        70        80        90       100       110       120
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
KU513561                  FSPPEAATSAPAAIQGLLDPTPPASDX~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~XWVLPPPSEELQVDTKPASPTLEPALSSRPVEPKTPVRKPSTPPSPRTRR
HEPAC-26-Pacbio           FSPPEAATSAPAAIQGLLDPTPPASDIHQEGEIFDTEKEKYEITEQRKIDQKAVDSQILPKIKAIPQLQGYLRSVFACVLPPPSEELQVDTKPASPTLEPALSPRPVEPKTPVRKPSTPPSPRTRR
HEPAC-26-Sanger           FSPPEAATSAPAAIQGXXXCMASASDIHQEGEIFDTEKEKYEITEQRKIDQKAVDSQILPKIKAIPQLQGYLRSVFACVLPHTSEELQVDTKPASPTLEPALSPRPVEPKTPVRKPSTPPSPRTRR
NM_001024662.2-RPL6       ~~~~~~~~~~~~~~~~~~~~~~~~~~XHQEGEIFDTEKEKYEITEQRKIDQKAVDSQILPKIKAIPQLQGYLRSVFX~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

21%

Human ribosomal protein L6

```
                                  10        20        30        40        50        60        70        80        90       100       110       120       130       140       150
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|..
KC166970                  FSPPEAAHAAPVPDMGLTSGAPSSASDIWVFPPPSEGSATISPPITLVGKPTNPPGHTTP~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~RPPVREPSTPPPALTPVSKPANPPSLTAPRPPVHEPPTPPSARNRR
Hepac-64-Pacbio           FSPPEAAHAAPVLDMGLTSGAPSSASVIWGFPPPSEGSAIIPPPITPVGKPTNLSGHTTPTERDSGRGGGRLWSRLLGRLRRESGVNPGGGACSEPRLRHCTPAWVRPPAREPSVPPPAQTSASKPANPPSLTXPRPPVREPPTPPSARTRR
Hepac-64-Sanger           ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~XTPTERDSGRGGGRLWSRLLGRLRRESGVNPGGGACSEPRLRHCTPAWVRPPAHIA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
NG_054921.1_ZNF787        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~XGRGGGRL*SQLLGRLRQENGVNPGGGACSEPRSRHCTPAWVX~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

100%

Zinc Finger Protein 787

Hepac-93

```
                                  10        20        30        40        50        60        70        80        90       100       110       120       130
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....
MF444142                  FSPPEAAYAAPAPDMGLPSAASSSASDIWAFPPPSERSATAPPPVASVSKPANPP~~~~~~~~~~~~~~~~~~~~~~~~~~~~SFATPWPPAREPPTPPPALTPGSKPASPSSPTTHRPPARKPSTPPPARTRR
Hepac-93-Pacbio           FSPPEAAYAAPAPDMGLPSAASSSASDIWAFPPPSERSATAPPPVASVSKQPIRLVTVVILELIHADGPDPLREGMRAFIRSFATPWPPAREPPTPPPALTPGSKPASPSSPTTHRPPARKPSTPPPARTRR
Hepac-93-Sanger           FSPPEAAYAAPAPDMGLPSAASSSASDIWAFPPPSERSATAPPPVASVSKPANPPLDNCGNSRANTCRRALTPFARGMRAFIRSFATPWPPAREPPTPPPALTPGSKPASPSSPTTHRPPARKPSTPPPARTRR
NG_054875-RNA18SP5                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~XLDNCGNSRANTCRRALTPFAGGMRAFIR~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

12%

18S ribosomal pseudogene 5

```
                                  10        20        30        40        50        60        70        80        90       100       110       120
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AB850879                  FSPPEAAYASPAPDMGLPSGTSSSASGIWAFPPPSERSATAPPPVIPASX~~~~~~~~~~~~~~~~~~XPANPPSLATPRSPVCEPPTPPPALTPVSKPASPSSLTTYRSPARKPPSPPPARTRR
Hepac-93-Pacbio           FSPPEAAYAAPAPDMGLPSAASSSASDIRAFPPPSERSATAPPPVASVSKMRKLLRKSALTLRKLATTPPANPPSFATPWXPAREPPTPPPALTPGSKPASPSSPTTHRPPARKPSTPPPARTRR
Hepac-93-Sanger           FSPPEAAYAAPAPDMGLPSAASSSASDIXAFPPPSERSATAPPPVASVSKMRKLLRKSALTLRKLATTPX~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
NM_001402:562-618-EEF1A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~XMRKLLRKSALTLRKLATTX~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

Eukaryotic translation elongation factor 1 alpha 1 pseudogene 13

31%

```
                                  10        20        30        40        50        60        70        80        90       100       110       120       130       140
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|..
AB850879                  FSPPEAAYASPAPDMGLPX~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~XGTSSSASGIWAFPPPSERSATAPPPVIPASKPANPPSLATPRSPVCEPPTPPPALTPVSKPASPSSLTTYRSPARKPPSPPPARTRR
HEPAC-94_Pacbio           FSPPEAACAXPAPDMGLPSRNKIAGYVTHLMKRIQRGPVRGISIKLQEEERERREGAPSSTSDIWVFPPPSERSAIAPPPVTPVSKPATPPSFATPRSPVCEPPTPPPALTPASKPASLSSPTTYRPPARKSPTPLPARTRR
HEPAC-94-Sanger           FSPPEAACAAPAPDMXLPSRNKIAGYVTHLMKRIQRGPVRGISIKLQEEERERREGAPLNPSDIWVFPPPSERSAIAPPPVTPVSKPATPPSFATPRSPVCEPPTPPPALTPASKPASLSSPTTYRPPARKSPTPLPARTRR
NR_111944.1-RPS17         ~~~~~~~~~~~~~~~~~~~RNKIAGYVTHLMKRIQRGPVRGISIKLQEEERERR~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

Ribosomal protein 17S

63%

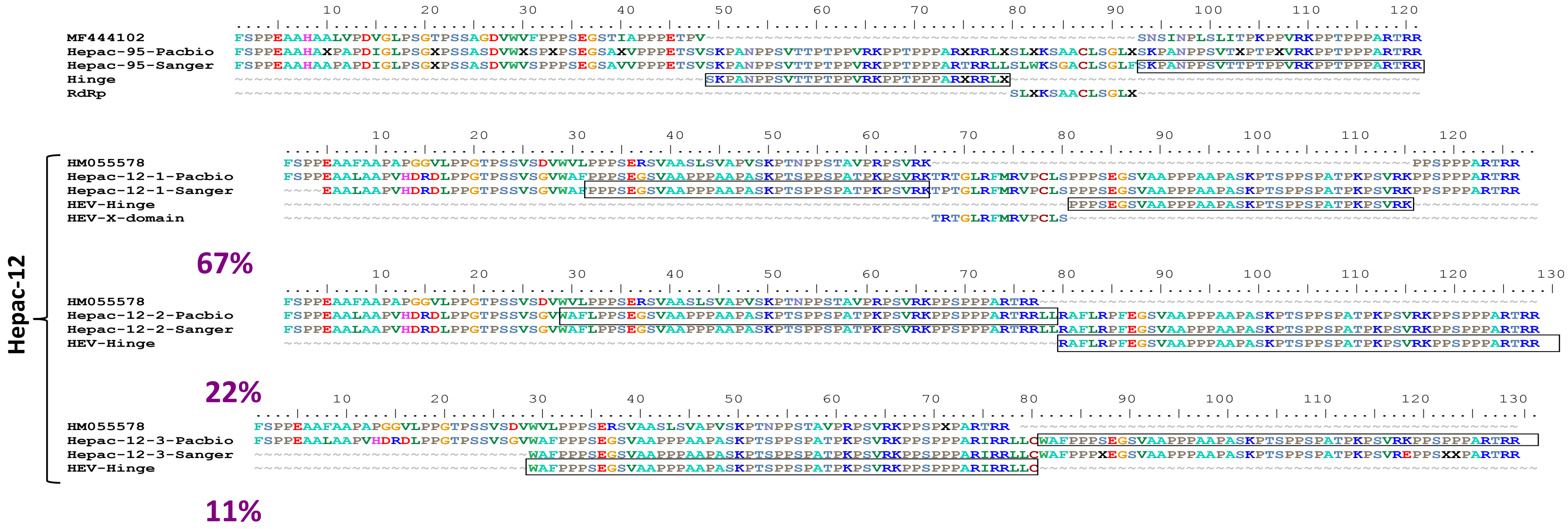
```
                                  10        20        30        40        50        60        70        80        90       100       110       120       130
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
KC166968                  FSPPEAAHAAPVPDMGLTSGAPSSASDIWVFPPPSEGSATISPPITLVGKPTNPPGHX~~~~~~~~~~~~~~~~~~~~~~~~XTPRPPVREPSTPPPALTPVSKPANPPSLTAPRPPVHEPPTPPSARNRR
Hepac-154-Pacbio          FSPPEAAHAXPVPNMGLPSGTPSFASDIWVXPPPSEGPATAPSPVTPVSKPTNPPSLNKSLTTLGKVISALAEVSKKKKKTDSTSGPPVCAPPTSPLALTPVS~PASPPSPTTPRPPAREPPVSPPARTRR
Hepac-154-Sanger          FSPPEAAHAAPVPNMGLPSGTPSFASDIWVFPPPSEGPATAPSPVTPVSKPTNPPSLNKSLTTLGKVISALAEVSKKKKKTDFTSGPPVCAPPTSPLALTPVS~PASPPSPTTPRPPAREPPVSPPARTRR
NM_015074.3-KIF1B         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~XNKSLTTLGKVISALAEVSKKKKKTDX~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

Kinesin family member 1B

100%

Nicot, manuscript in preparation

Duplications of HEV genome



```
                              10        20        30        40        50        60        70        80        90       100       110       120
                      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
MF444102              FSPPEAAHAALVPDVGLPSGTPSSAGDVWVFPPPSEGSTIAPPPETPV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~SNSINPLSLITPKPPVRKPPTPPPARTRR
Hepac-95-Pacbio       FSPPEAAHAXPAPDIGLPSGXPSSASDVWXSPXPSEGSAXVPPPETSVSKPANPPSVTTPTPPVRKPPTPPPARXRRLXSLXKSAACLSGLXSKPANPPSVTXPTPXVRKPPTPPPARTRR
Hepac-95-Sanger       FSPPEAAHAAPAPDIGLPSGXPSSASDVWVSPPPSEGSAVVPPPETSVSKPANPPSVTTPTPPVRKPPTPPPARTRRLLSLWKSGACLSGLFSKPANPPSVTTPTPPVRKPPTPPPARTRR
Hinge                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~SKPANPPSVTTPTPPVRKPPTPPPARXRRLX~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
RdRp                  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~SLXKSAACLSGLX~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

Hepac-12

```
                              10        20        30        40        50        60        70        80        90       100       110       120
                      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|...
HM055578              FSPPEAAFAAPAPGGVLPPGTPSSVSDVWVLPPPSERSVAASLSVAPVSKPTNPPSTAVPRPSVRK~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~PPSPPPARTRR
Hepac-12-1-Pacbio     FSPPEAALAAPVHDRDLPPGTPSSVSGVWAFPPPSEGSVAAPPPAAPASKPTSPPSPATPKPSVRKTRTGLRFMRVPCLSPPPSEGSVAAPPPAAPASKPTSPPSPATPKPSVRKPPSPPPARTRR
Hepac-12-1-Sanger     ~~~~EAALAAPVHDRDLPPGTPSSVSGVWAFPPPSEGSVAAPPPAAPASKPTSPPSPATPKPSVRKTPTGLRFMRVPCLSPPPSEGSVAAPPPAAPASKPTSPPSPATPKPSVRKPPSPPPARTRR
HEV-Hinge             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~PPPSEGSVAAPPPAAPASKPTSPPSPATPKPSVRK~~~~~~~~~~~~~
HEV-X-domain          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~TRTGLRFMRVPCLS~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

67%

```
                              10        20        30        40        50        60        70        80        90       100       110       120       130
                      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HM055578              FSPPEAAFAAPAPGGVLPPGTPSSVSDVWVLPPPSERSVAASLSVAPVSKPTNPPSTAVPRPSVRKPPSPPPARTRR~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Hepac-12-2-Pacbio     FSPPEAALAAPVHDRDLPPGTPSSVSGVWAFLPPSEGSVAAPPPAAPASKPTSPPSPATPKPSVRKPPSPPPARTRRLLRAFLRPFEGSVAAPPPAAPASKPTSPPSPATPKPSVRKPPSPPPARTRR
Hepac-12-2-Sanger     FSPPEAALAAPVHDRDLPPGTPSSVSGVWAFLPPSEGSVAAPPPAAPASKPTSPPSPATPKPSVRKPPSPPPARTRRLLRAFLRPFEGSVAAPPPAAPASKPTSPPSPATPKPSVRKPPSPPPARTRR
HEV-Hinge             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~RAFLRPFEGSVAAPPPAAPASKPTSPPSPATPKPSVRKPPSPPPARTRR
```

22%

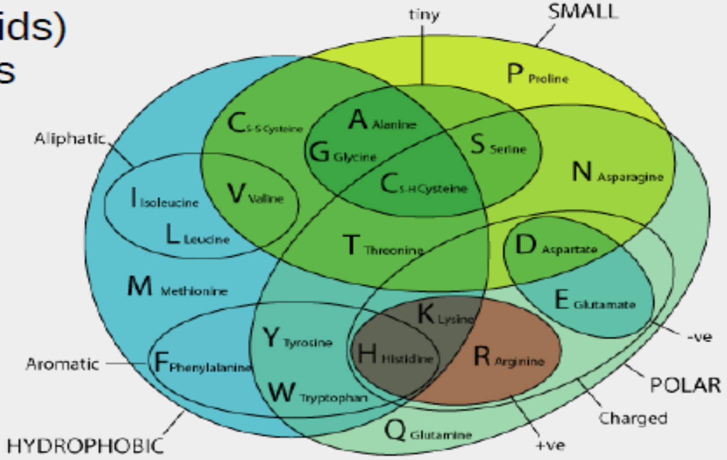
```
                              10        20        30        40        50        60        70        80        90       100       110       120       130
                      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
HM055578              FSPPEAAFAAPAPGGVLPPGTPSSVSDVWVLPPPSERSVAASLSVAPVSKPTNPPSTAVPRPSVRKPPSPXPARTRR ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Hepac-12-3-Pacbio     FSPPEAALAAPVHDRDLPPGTPSSVSGVWAFPPPSEGSVAAPPPAAPASKPTSPPSPATPKPSVRKPPSPPPARIRRLLCWAFPPPSEGSVAAPPPAAPASKPTSPPSPATPKPSVRKPPSPPPARTRR
Hepac-12-3-Sanger     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~WAFPPPSEGSVAAPPPAAPASKPTSPPSPATPKPSVRKPPSPPPARIRRLLCWAFPPPXEGSVAAPPPAAPASKPTSPPSPATPKPSVREPPSXXPARTRR
HEV-Hinge             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~WAFPPPSEGSVAAPPPAAPASKPTSPPSPATPKPSVRKPPSPPPARIRRLLC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

11%

Nicot, manuscript in preparation

Characterization of insertions

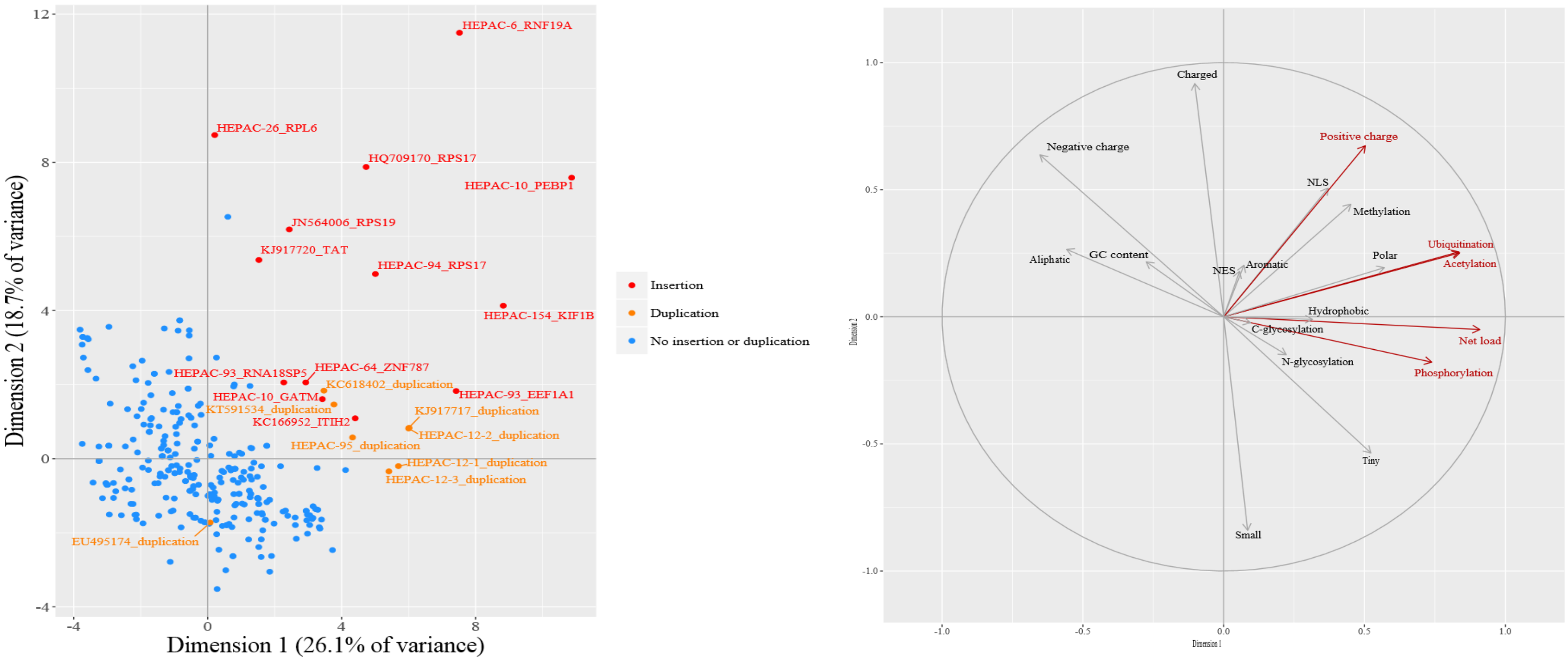
- AA proportions (20 amino-acids)
- AA physico-chemical features



- Net load
- Ubiquitination sites BDM-PUB
- Acetylation sites PAIL
- Methylation sites BPB-PPMS
- Phosphorylation sites NetPhos
- N-Glysosilation sites NetNGlyc
- C-Glycosilation sites NetCglyc
- Nuclear Export Signals (NES) Wregex
- Nuclear Localisation Signals (NLS) SeqNLS

36 features

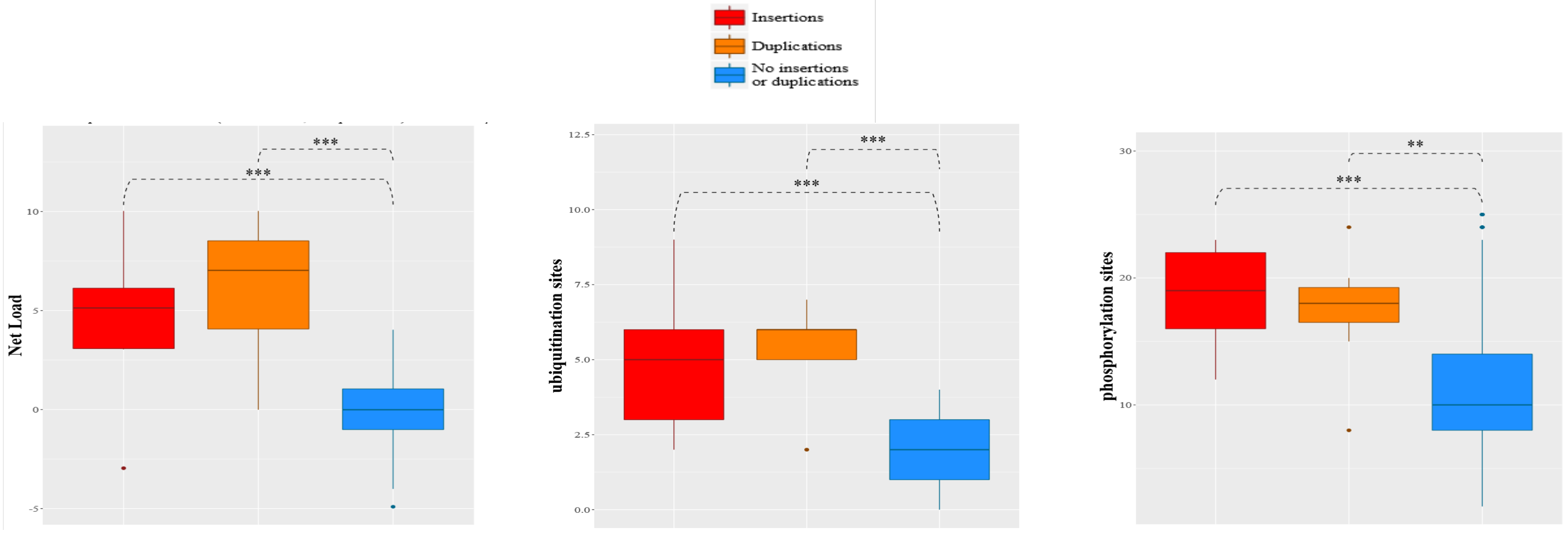
Principal Composant Analysis



Insertions or duplications confer to strains specific characteristics.

Nicot, manuscript in preparation

Contribution of insertions in hinge



Net load

Ubiquitination sites
Acetylation sites

Phosphorylation sites

Insertion: more POS charged AA
Duplication: less NEG charged AA

Nicot, manuscript in preparation

Discussion

- **Dataset** of **HEV-3 complete genome sequences increased** → useful for comparing strains circulating in humans and the animal reservoir

- **Characteristics of HEV-3 subtypes** regarding the **Hinge length** and **polymerase polymorphisms** identified

- Inserted sequences provided **new ubiquitination, acetylation, phosphorylation sites** → regulation of transcription and translation

- Further studies are needed to understand :
  - the biological and clinical relevance of HEV polymerase mutations
  - the role of HEV insertions in virus adaptation

Acknowledgement

Laboratoire de virologie, CNR Hépatite E,
Institut Fédératif de Biologie, CHU Toulouse

N. Jeanne
C. Lefebvre, R. Carcenac, M. Dubois
S. Lhomme, F. Abravanel, J. Izopet

Plateforme Get-Plage, INRA

Plateforme Bioinformatique, INRA