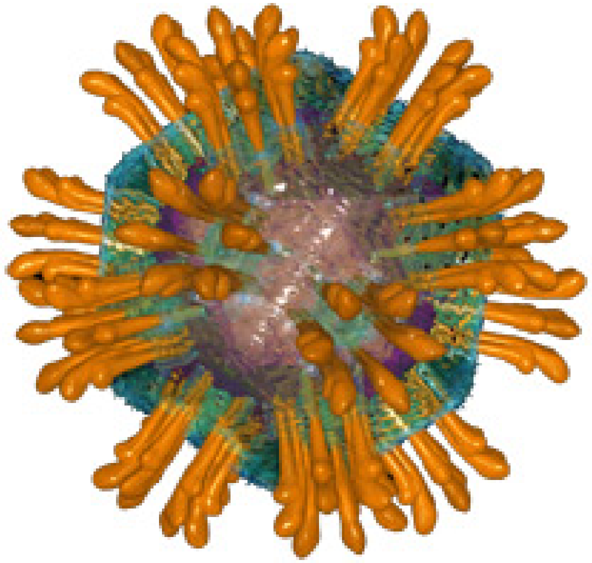
St. Luke's
Medical Center
Quezon City
One of the world's best

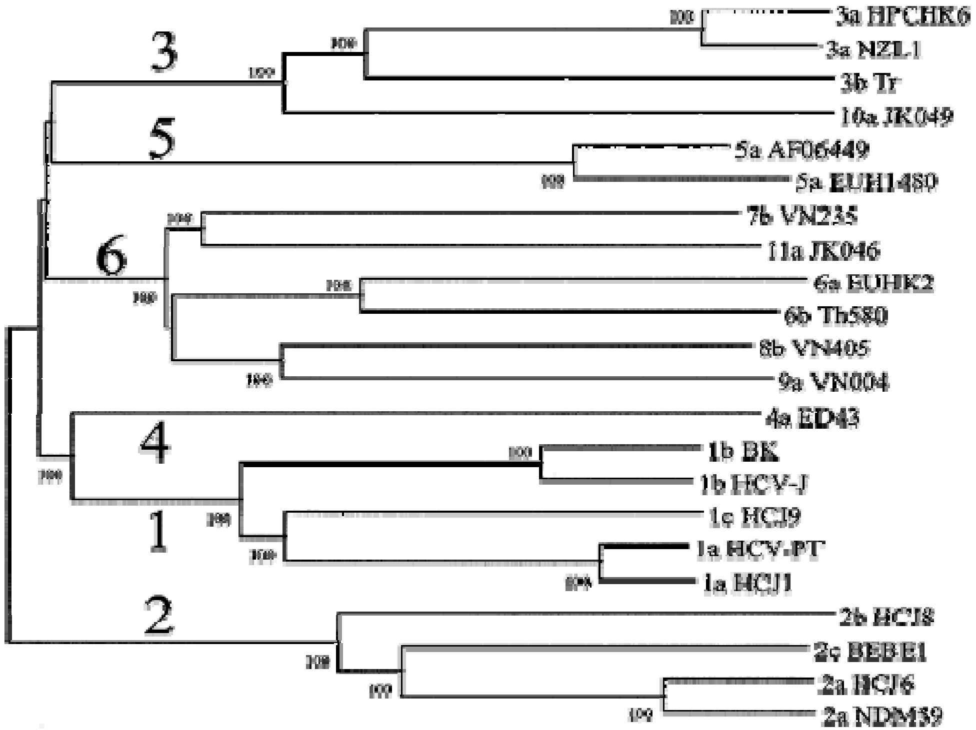
GENOMIC IDENTIFICATION OF HCV SUBTYPES BY SEQUENCE ANALYSIS OF THE NS5A REGION

MO. Baclig,[1] MM. Rivera,[1] RZ. Predicala,[1] MPS. Dimamay,[1] RR. Matias,[1]
FF. Natividad,[1] JG Cervantes,[2] and the Liver Diseases Study Group
[1]Research and Biotechnology Division, St. Luke's Medical Center, Philippines
[2]Center for Liver Diseases, St. Luke's Medical Center, Philippines

HCV

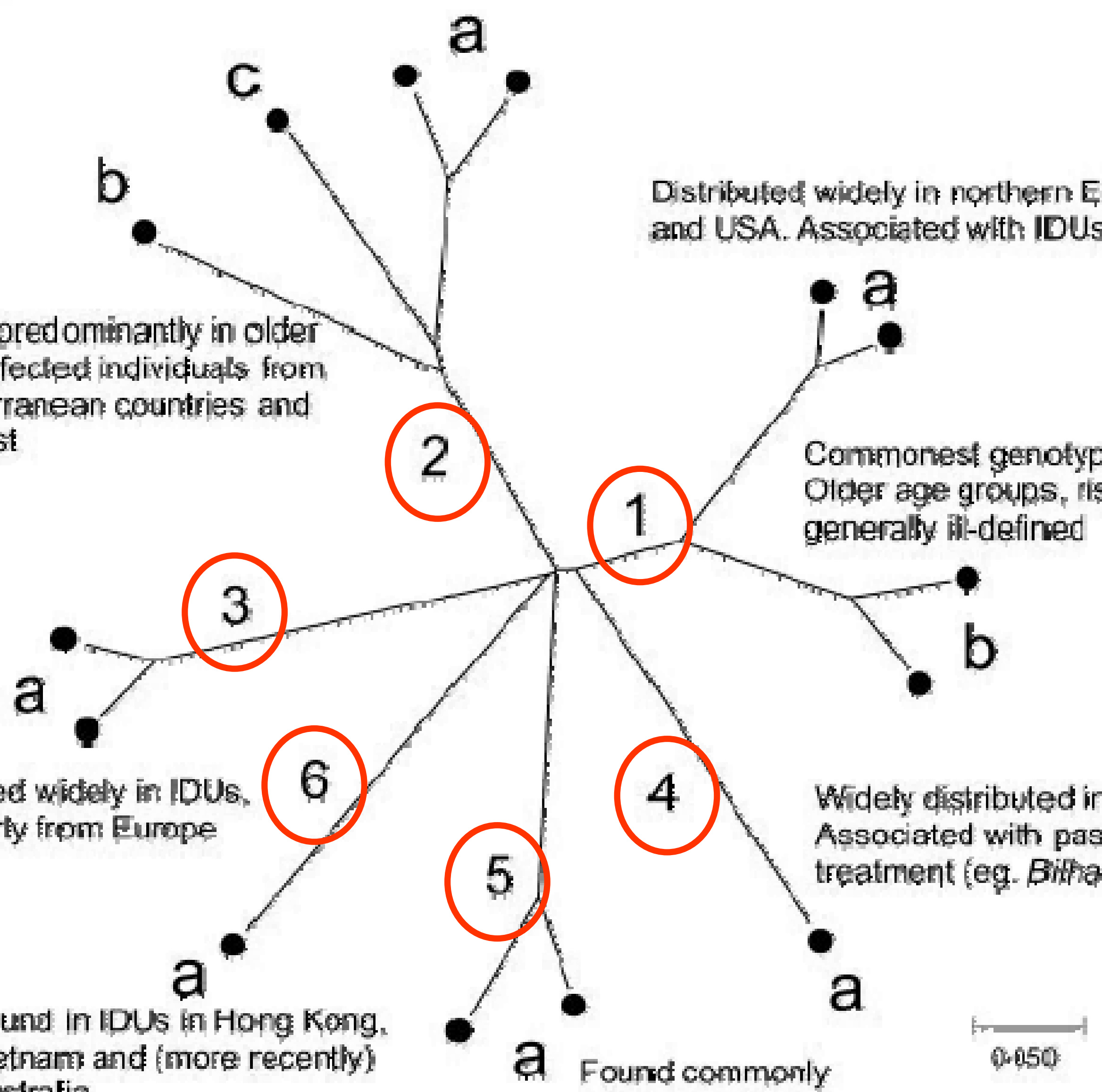


www.chiron.com



Simmonds P *et al.*, 2000

- Single-stranded, positive-sense RNA genome of ~9.6kb
- Six major genotypes and >70 subtypes



Distributed widely in northern Europe and USA. Associated with IDUs

Found predominantly in older HCV-infected individuals from Mediterranean countries and Far East

Commonest genotype worldwide. Older age groups, risk factors generally ill-defined

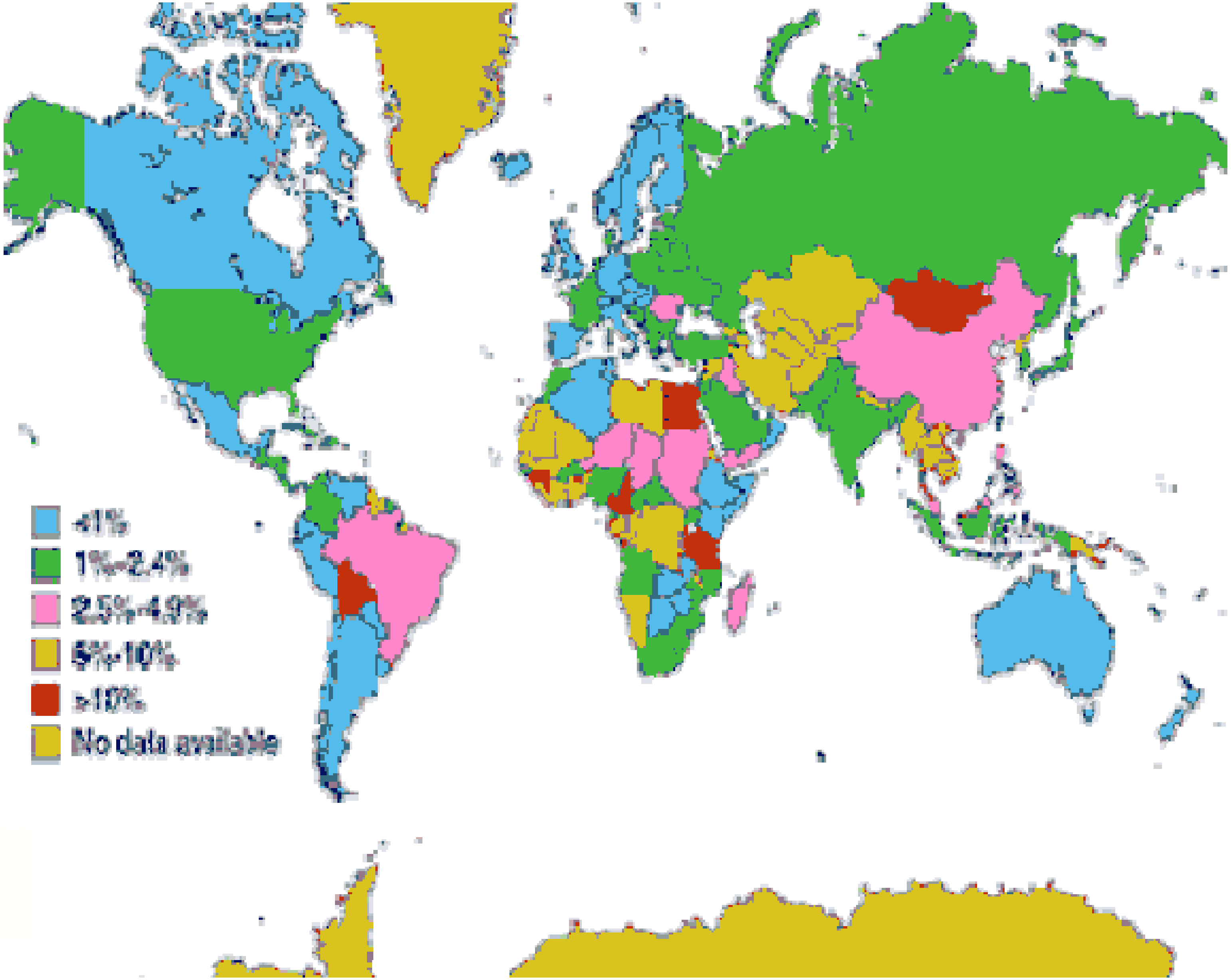
Distributed widely in IDUs, particularly from Europe

Widely distributed in Middle East, Associated with past medical treatment (eg. *Bilharzia* injections)

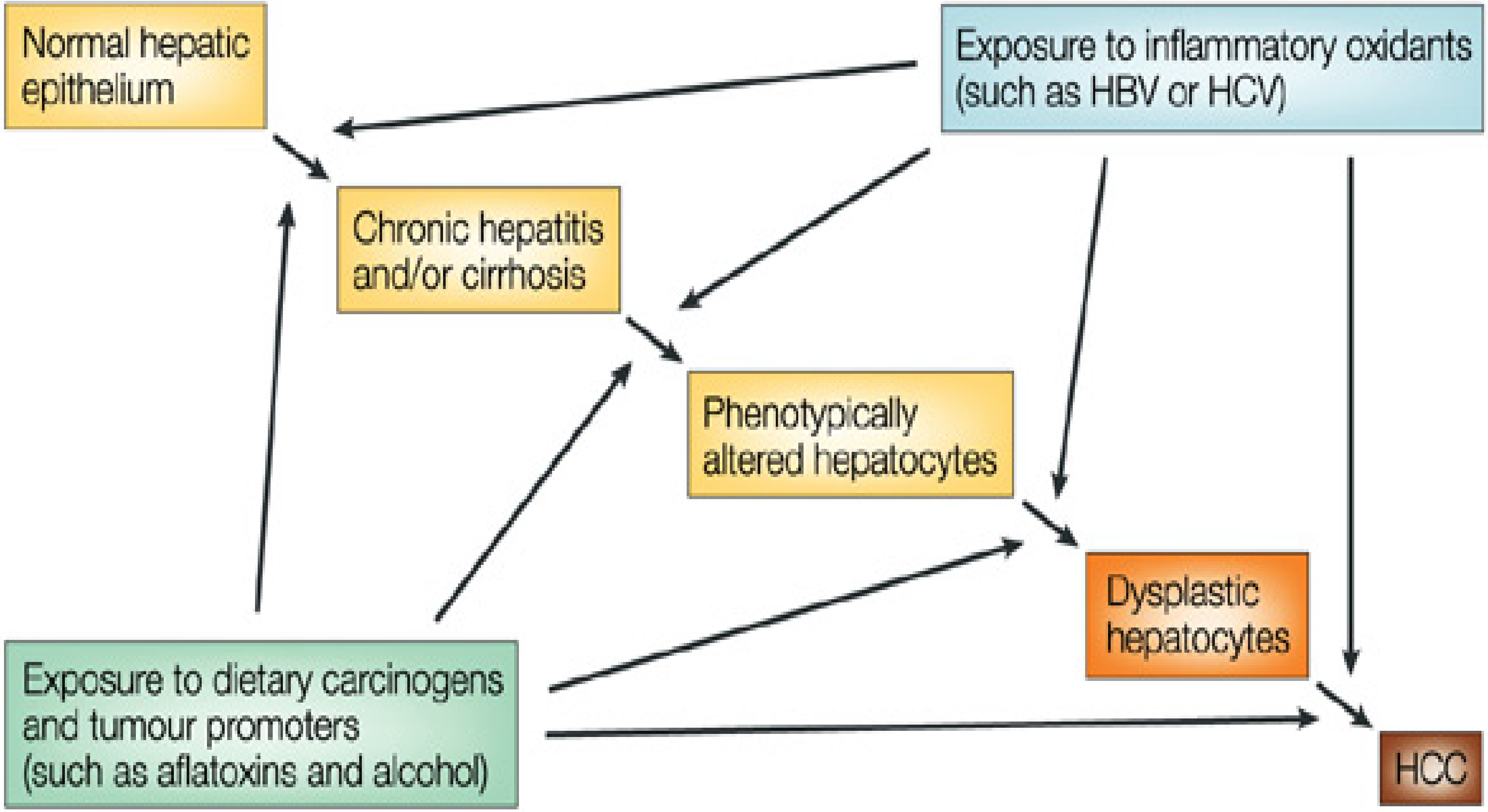
Found in IDUs in Hong Kong, Vietnam and (more recently) Australia

Found commonly only in South Africa

Simmonds P 2004. Genetic Diversity and Evolution of HCV. *J Gen Virol*. 85:3173-88



Cohen J. *Science*. 1999;225-26



Normal hepatic epithelium

Exposure to inflammatory oxidants (such as HBV or HCV)

Chronic hepatitis and/or cirrhosis

Phenotypically altered hepatocytes

Dysplastic hepatocytes

Exposure to dietary carcinogens and tumour promoters (such as aflatoxins and alcohol)

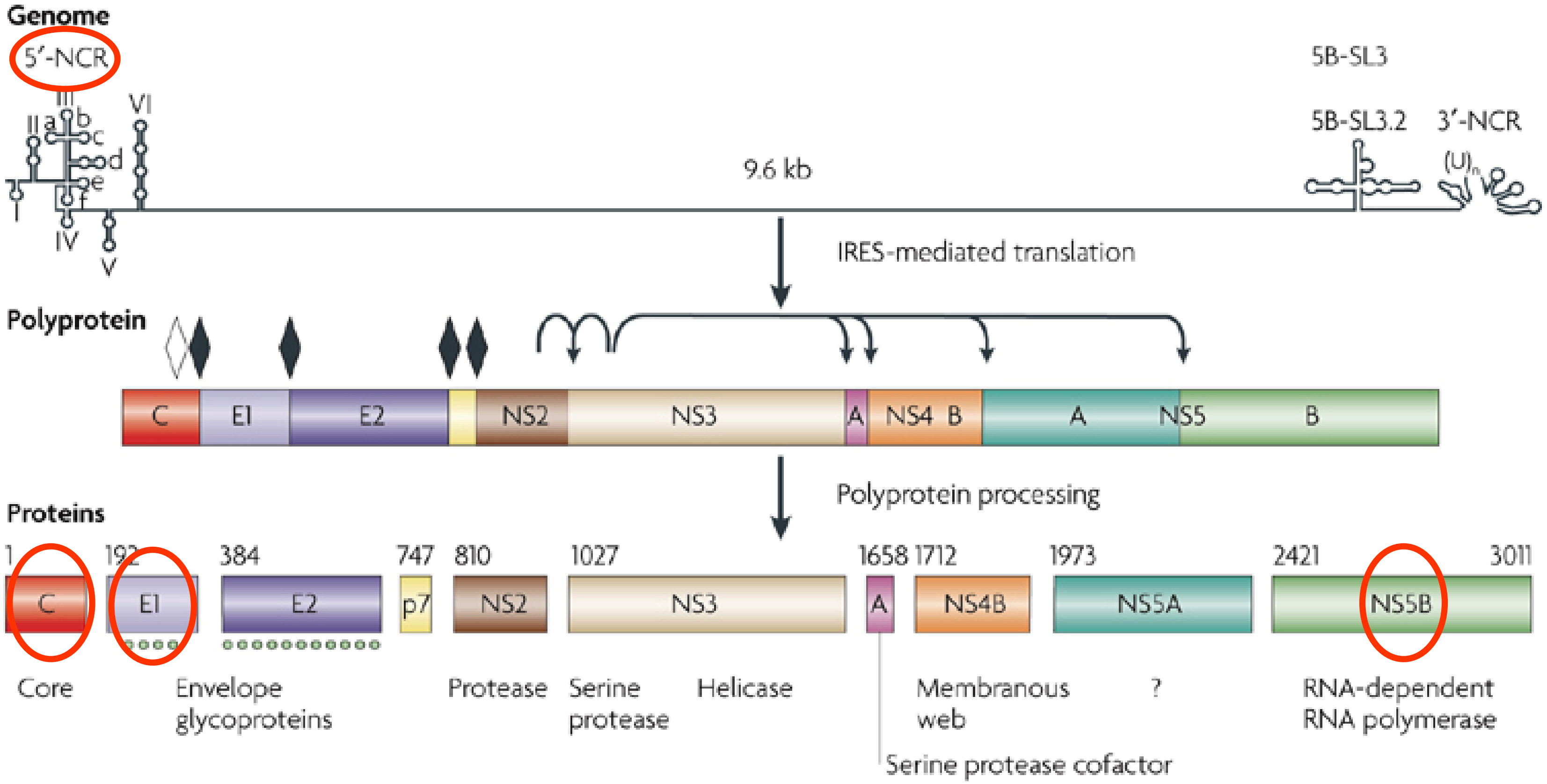
HCC

Nature Reviews | Cancer

Percent Nucleotide Difference

Types	1a	1b	1c	2a	2b	2c	3a	3b	4a	5a	6a
1a	0	19	15	35	34	37	33	34	32	31	36
1b		0	23	36	33	36	33	29	36	30	35
1c			0	32	30	33	35	30	36	39	39
2a				0	18	23	33	33	34	34	32
2b					0	19	36	31	35	33	32
2c						0	36	35	35	34	35
3a							0	21	35	33	36
3b								0	34	32	39
4a									0	34	34
5a										0	32
6a											0

HCV Genome



Moradpour D *et al.*, 2007. Nature Reviews Microbiology 5:453-63

- Genotyping based on 5’NCR is **accurate for most genotypes**

- Methods based on the use of 5’NCR are **unable to distinguish subtypes 1a from 1b in 5-10 %**

Chen Z *et al.*, *J Clin. Microbiol.* 2002

- Some of the **genotype-specific motifs** that were initially identified in the 5’NCR are **no longer found to be conserved**
- The **G residue at position 243**, originally considered to be representative of 1b, is found to occur in a relative proportion of 1a

Murphy D *et al.*, 2007. *J Clin. Microbiol.*

Objectives

- To identify the subtypes of HCV-1 isolates by RFLP of the RT-PCR amplified 5’NCR
- To identify the subtypes of the isolates by nucleotide sequencing of the 5’NCR and NS5A region
- To identify key amino acids in the NS5A region that can be used to differentiate between HCV subtypes

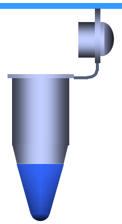
Significance

- Provide researchers with an epidemiologic marker
    - Origin and spread
    - Distribution
    - Routes of transmission
        - Outbreak studies
        - Novel transmission risks
    - Association with risk groups
    - Viral evolution

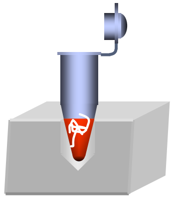
- Data may have major implications in designing optimal strategies for disease management, as well as, treatment benefits

- Characterization is likely to facilitate in the development of vaccine

Methodology



Viral RNA Extraction



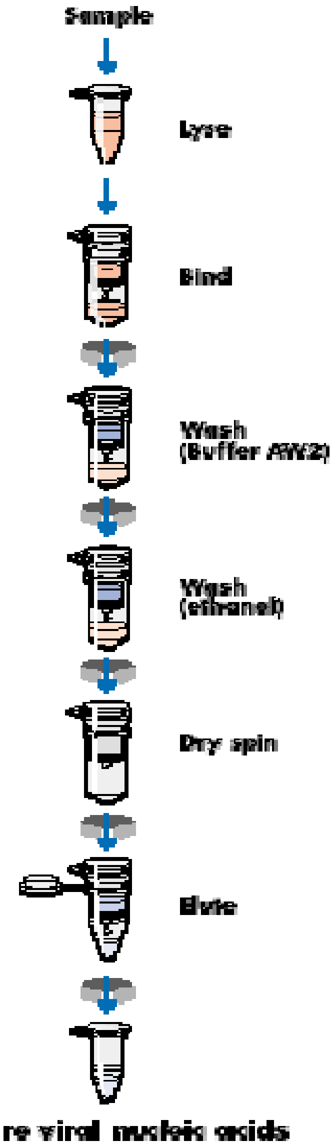
cDNA synthesis

Nested RT-PCR and RFLP

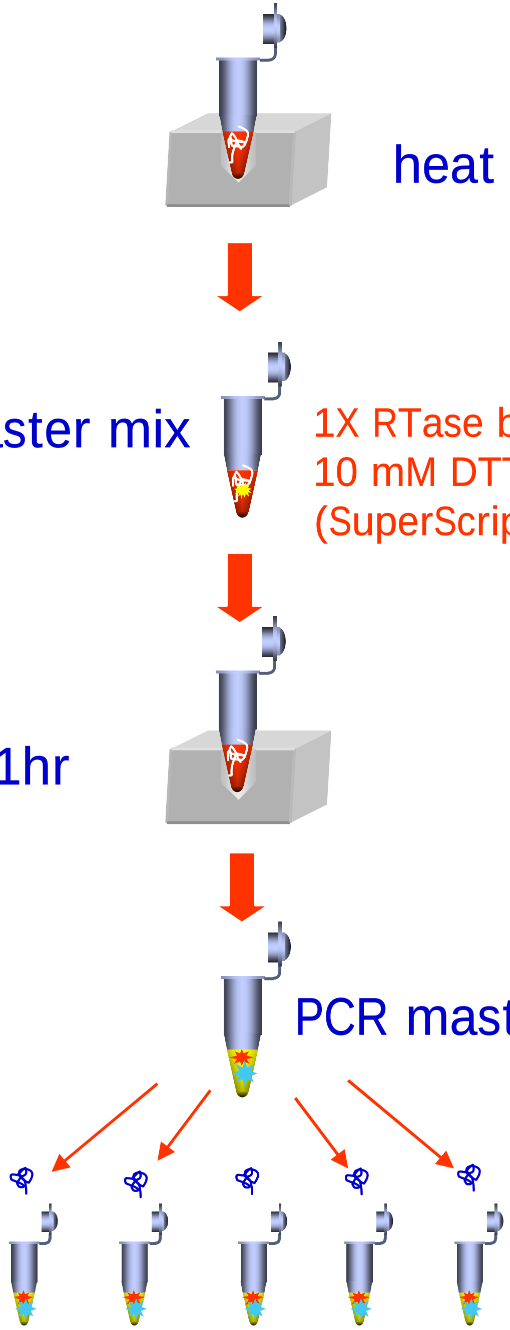
Nucleotide Sequencing

RNA Extraction

QIAamp® Viral RNA Kit



cDNA synthesis



heat viral RNA at 65^{o}C for 5min

add to cDNA master mix

1X RTase buffer, 0.5 mM dNTPs, 50 pmol primer, 10 mM DTT, 13 U RNaseOUT^{TM}, 200 U RTase (SuperScript^{TM} III), RNase-free water

incubate at 37^{o}C for 1hr

PCR mastermix

1X Phusion^{TM} HF buffer, 25 pmol primer, 0.2 mM dNTPs, 1 U Phusion^{TM} DNA polymerase, PCR-grade water

cDNA template

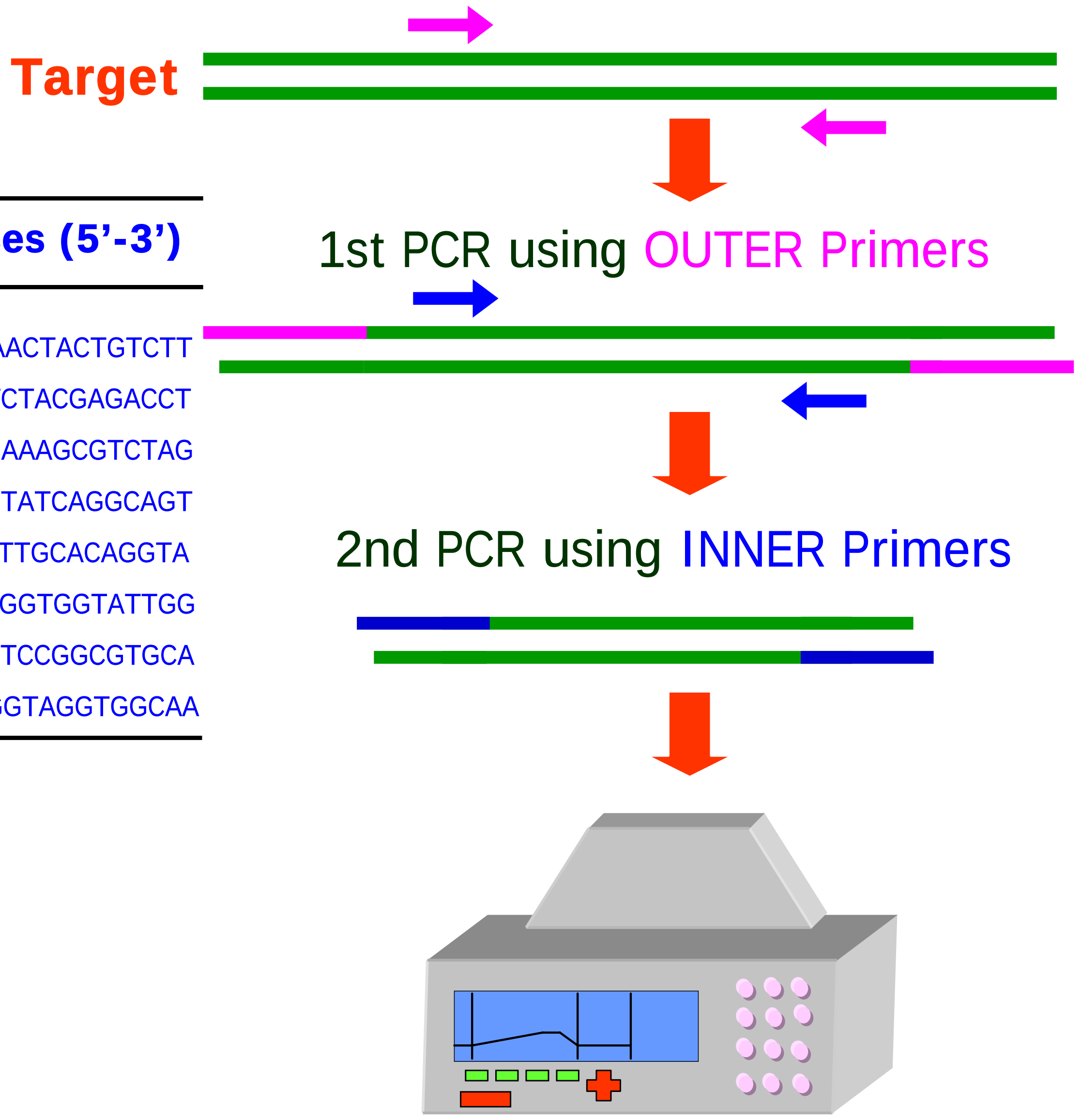
Nested RT-PCR

Primer	**Region**	**Sequences (5’-3’)**
OS	**5NCR**	CTGTGAGGAACTACTGTCTT
OAS	**5’NCR**	ATACTCGAGGTGCACGGTCTACGAGACCT
IS	**5’NCR**	TTCACGCAGAAAGCGTCTAG
IAS	**5’NCR**	CACTCTCGAGCACCCTATCAGGCAGT
OS	**NS5A**	TGGATGGAGTGCGGTTGCACAGGTA
OAS	**NS5A**	TCTTTCTCCGTGGAGGTGGTATTGG
IS	**NS5A**	TGTAAAACGACGGCCAGTCAGGTACGCTCCGGCGTGCA
IAS	**NS5A**	CAGGAAACAGCTATGACCGGGGCCTTGGTAGGTGGCAA

Reference

Chan S *et al.*, 1992. *J Gen Virol*. 73:1131-41

Enomoto N *et al.*, 1995. *N Engl J Med*. 334:77-81

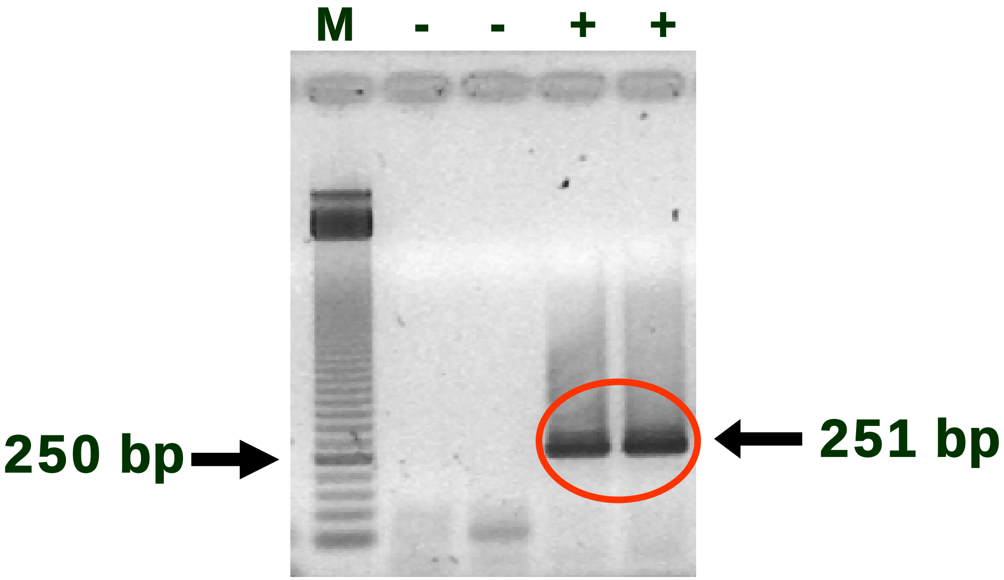


Target

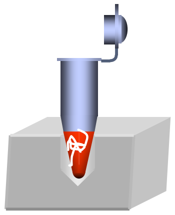
1st PCR using OUTER Primers

2nd PCR using INNER Primers

5’NCR PCR Amplification

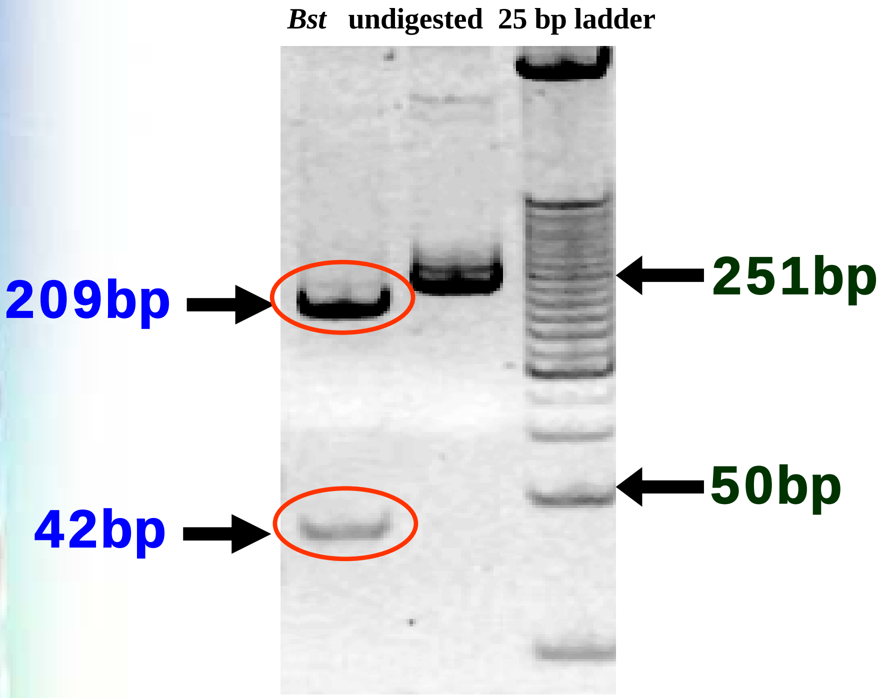


Restriction Enzyme Digestion

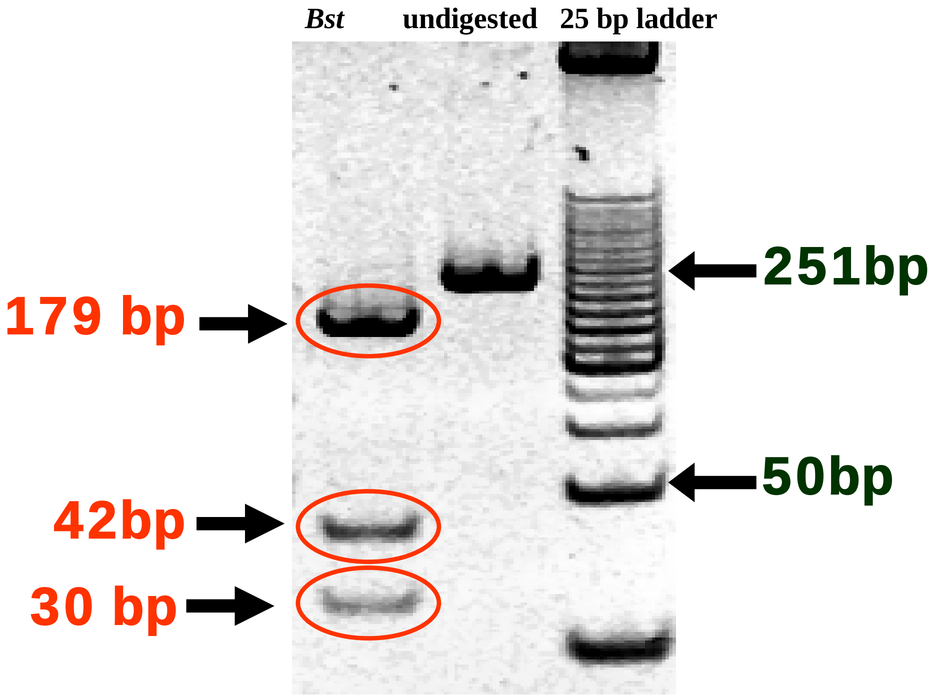


Bst UI

Bst UI Cleavage



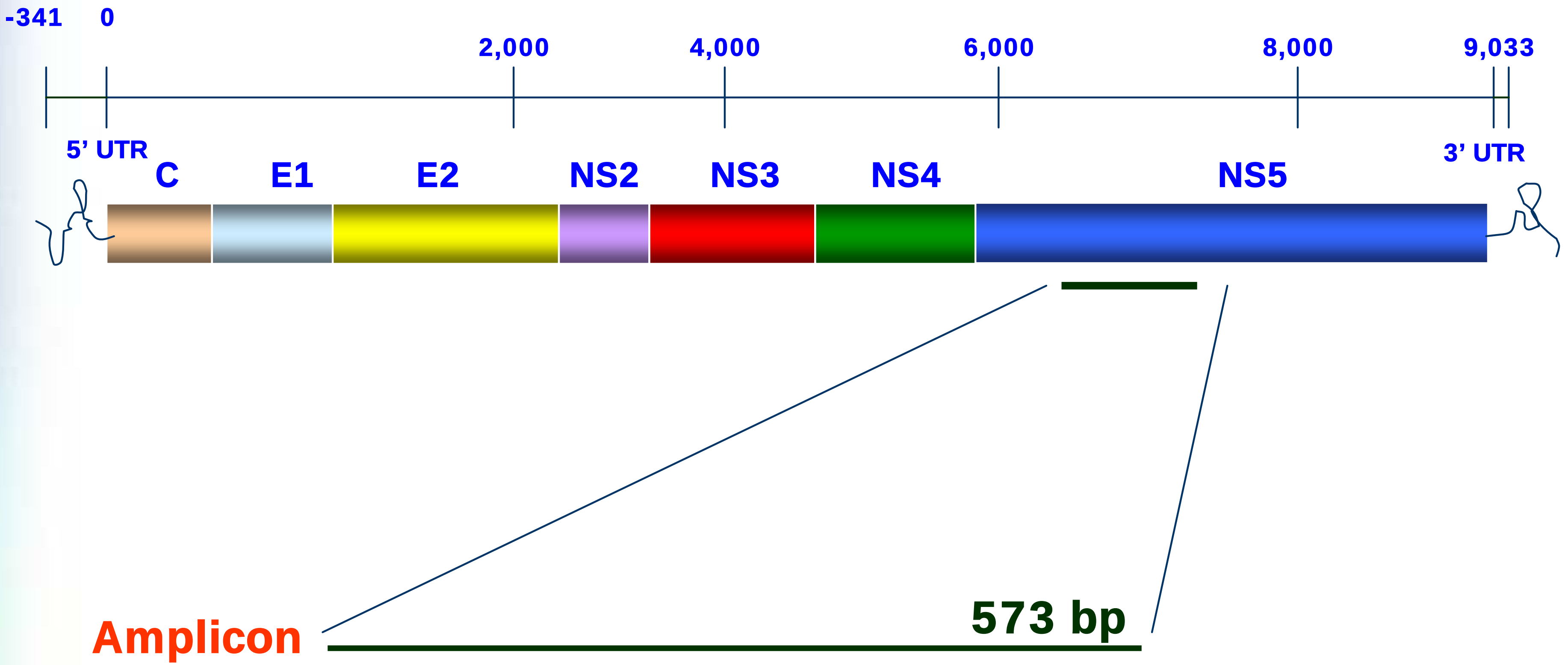
1a



1b

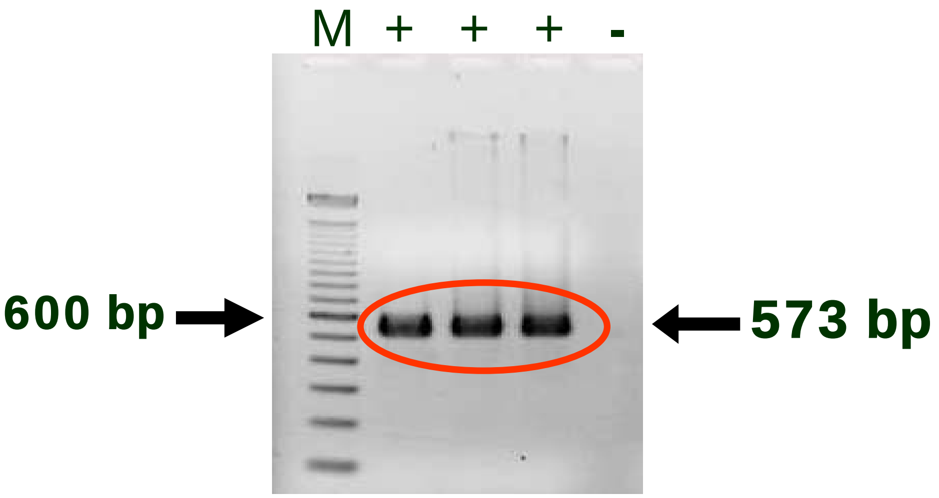
Type	Size (bp)
1a	209, 42
1b	179, 42, 30

Amplification of the NS5A

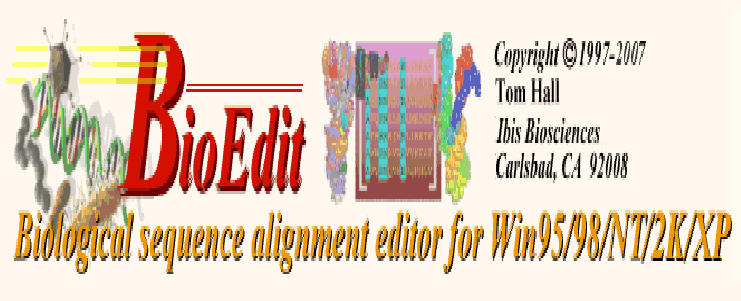


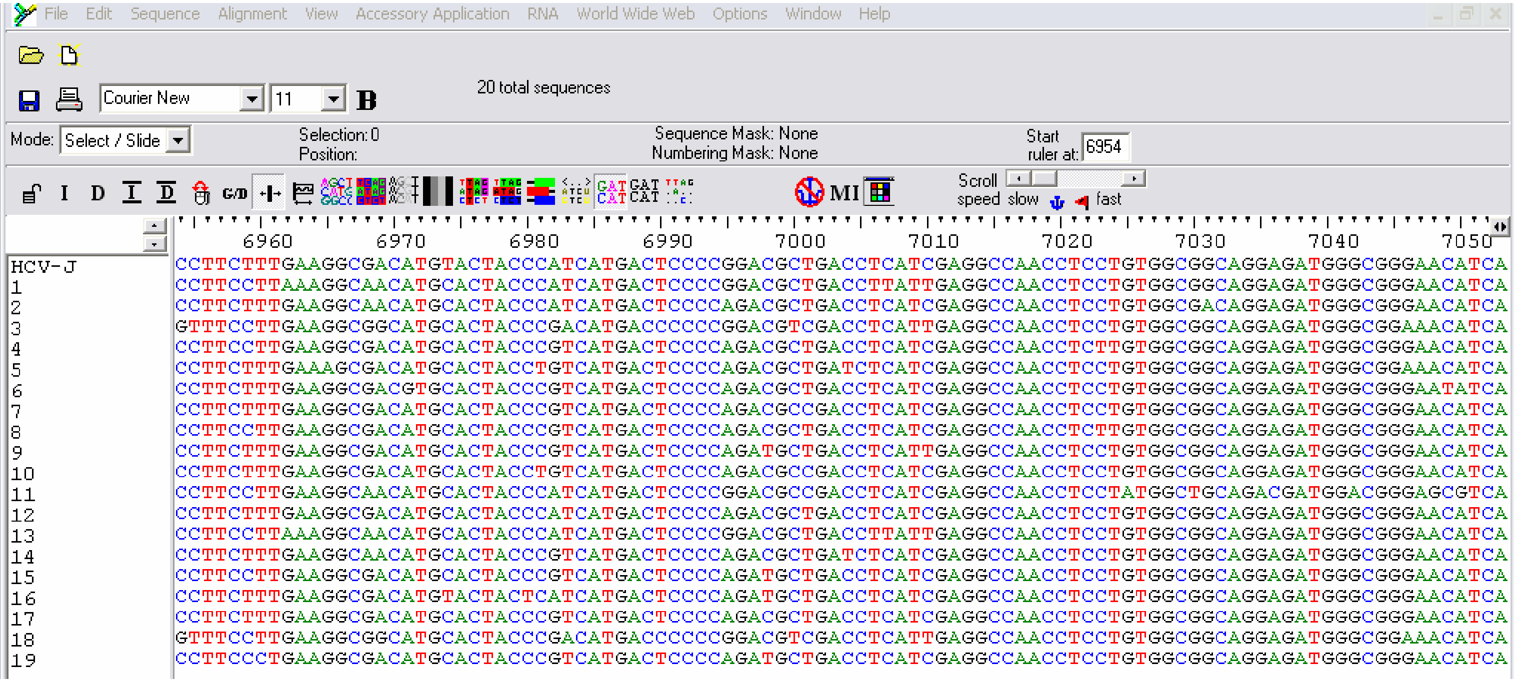
PCR Conditions

Initial Denaturation	98°C, 30 sec	
Denaturation	98°C, 10 sec	35 cycles
Annealing	56°C, 10 sec	
Extension	72°C, 30 sec	
Final extension	72°C, 10 min	



Bioinformatics Tools





20 total sequences

Mode: Select / Slide

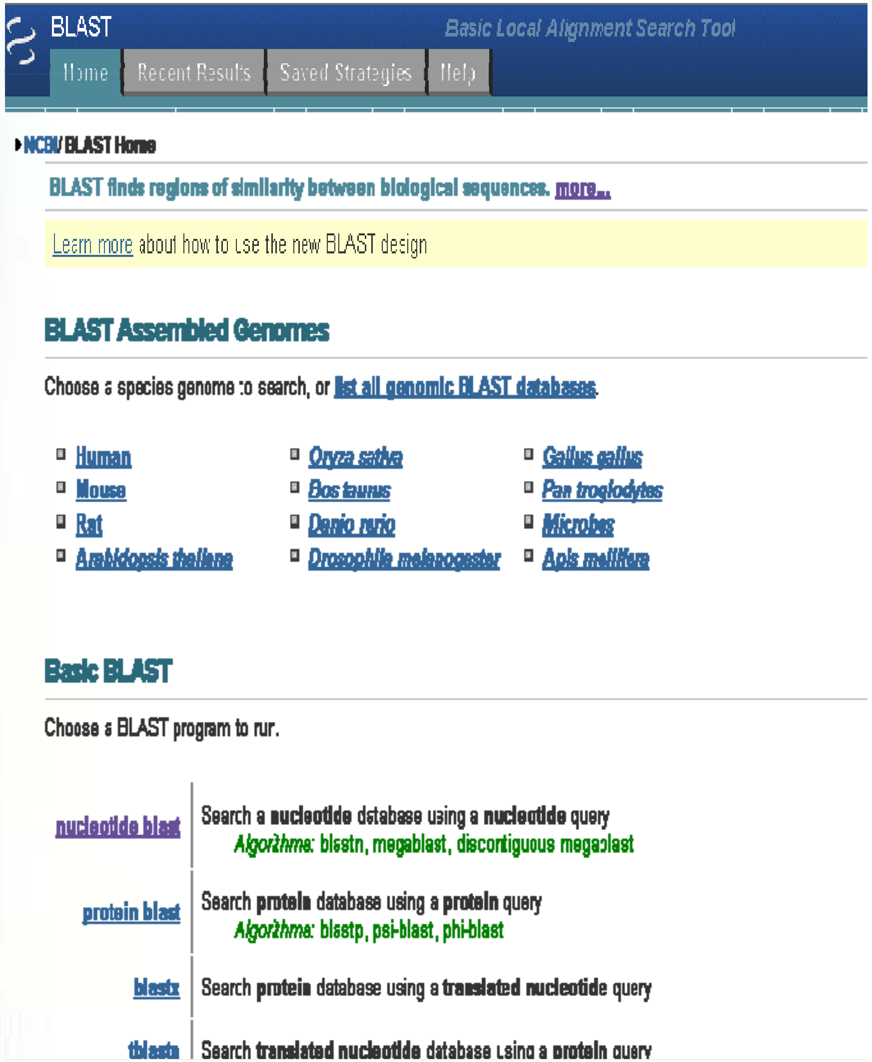
Selection: 0
Position:

Sequence Mask: None
Numbering Mask: None

Start ruler at: 6954

```
              6960      6970      6980      6990      7000      7010      7020      7030      7040      7050
HCV-J   CCTTCTTTGAAGGCGACATGTACTACCCATCATGACTCCCCGGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGGAACATCA
1       CCTTCCTTAAAGGCAACATGCACTACCCATCATGACTCCCCGGACGCTGACCTTATTGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGGAACATCA
2       CCTTCTTTGAAGGCAACATGCACTACCCATCATGACTCCCCAGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGACAGGAGATGGGCGGGAACATCA
3       GTTTCCTTGAAGGCGGCATGCACTACCCGACATGACCCCCCGGACGTCGACCTCATTGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGAAACATCA
4       CCTTCCTTGAAGGCGACATGCACTACCCGTCATGACTCCCCAGACGCTGACCTCATCGAGGCCAACCTCTTGTGGCGGCAGGAGATGGGCGGGAACATCA
5       CCTTCTTTGAAAGCGACATGCACTACCTGTCATGACTCCCCAGACGCTGATCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGAAACATCA
6       CCTTCTTTGAAGGCGACGTGCACTACCCGTCATGACTCCCCAGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGGAATATCA
7       CCTTCTTTGAAGGCGACATGCACTACCCGTCATGACTCCCCAGACGCCGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGGAACATCA
8       CCTTCCTTGAAGGCGACATGCACTACCCGTCATGACTCCCCAGACGCTGACCTCATCGAGGCCAACCTCTTGTGGCGGCAGGAGATGGGCGGGAACATCA
9       CCTTCTTTGAAGGCGACATGCACTACCCGTCATGACTCCCCAGATGCTGACCTCATTGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGGAACATCA
10      CCTTCTTTGAAGGCGACATGCACTACCTGTCATGACTCCCCAGACGCCGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGGAACATCA
11      CCTTCCTTGAAGGCAACATGCACTACCCATCATGACTCCCCGGACGCCGACCTCATCGAGGCCAACCTCCTATGGCTGCAGACGATGGACGGGAGCGTCA
12      CCTTCTTTGAAGGCGACATGCACTACCCATCATGACTCCCCAGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGGAACATCA
13      CCTTCCTTAAAGGCAACATGCACTACCCATCATGACTCCCCGGACGCTGACCTTATTGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGGAACATCA
14      CCTTCTTTGAAGGCAACATGCACTACCCGTCATGACTCCCCAGACGCTGATCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGGAACATCA
15      CCTTCCTTGAAGGCGACATGCACTACCCGTCATGACTCCCCAGATGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGGAACATCA
16      CCTTCTTTGAAGGCGACATGTACTACTCATCATGACTCCCCAGATGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGGAACATCA
17      CCTTCTTTGAAGGCGACATGCACTACCCGTCATGACTCCCCAGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGGAACATCA
18      GTTTCCTTGAAGGCGGCATGCACTACCCGACATGACCCCCCGGACGTCGACCTCATTGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGAAACATCA
19      CCTTCCCTGAAGGCGACATGCACTACCCGTCATGACTCCCCAGATGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGCGGGAACATCA
```

Blast Search



BLAST finds regions of similarity between biological sequences. more...

Learn more about how to use the new BLAST design

BLAST Assembled Genomes

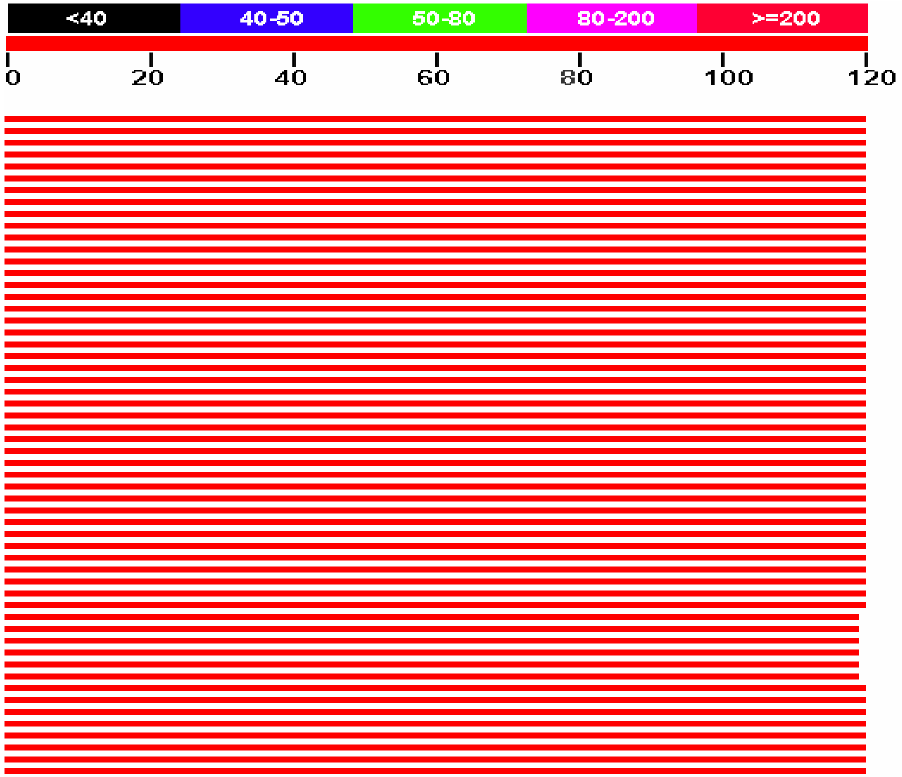
Choose a species genome to search, or list all genomic BLAST databases

- Human
- Mouse
- Rat
- Arabidopsis thaliana
- Oryza sativa
- Bos taurus
- Danio rerio
- Drosophila melanogaster
- Gallus gallus
- Pan troglodytes
- Microbes
- Apis mellifera

Basic BLAST

Choose a BLAST program to run.

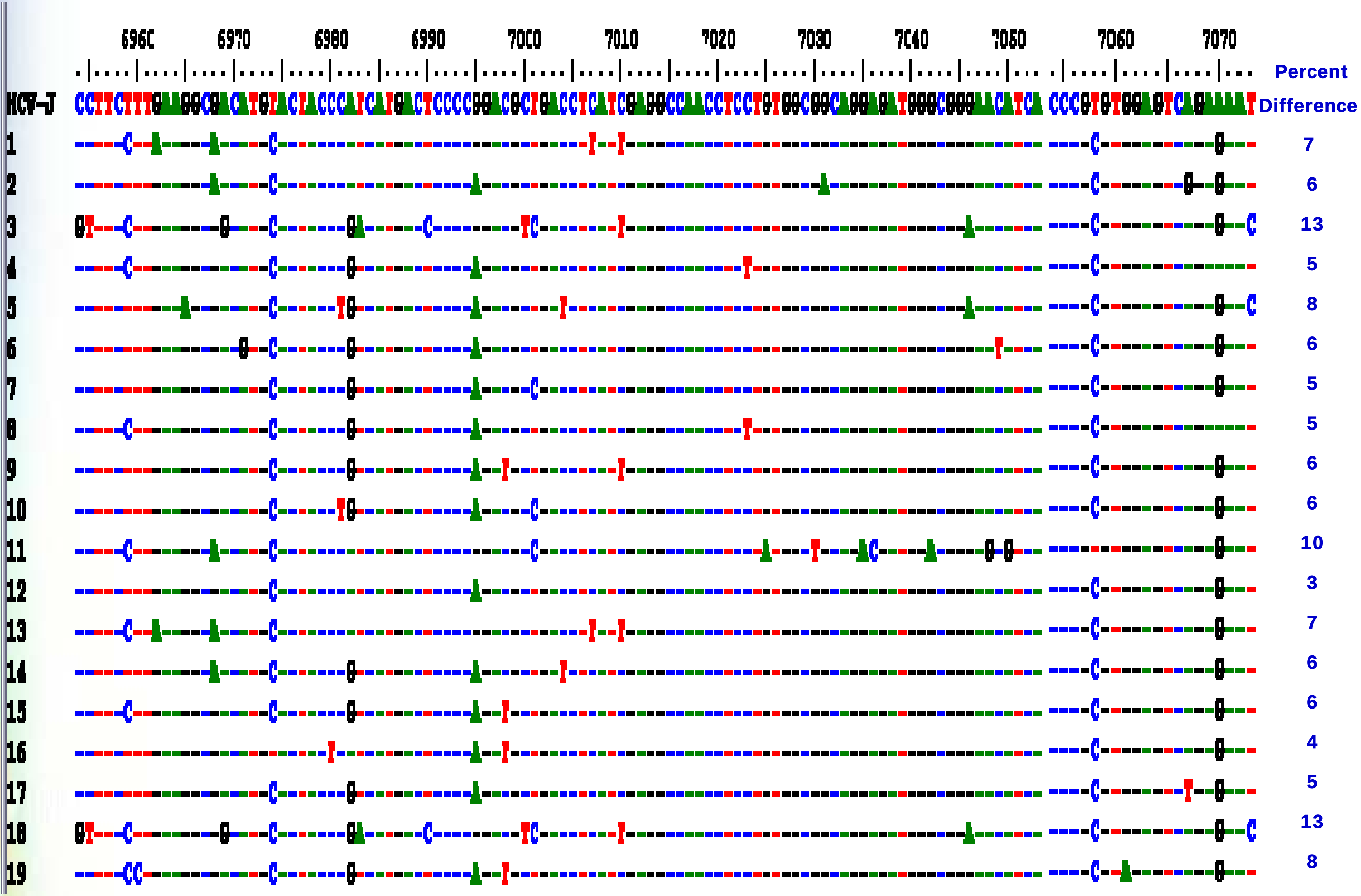
nucleotide blast	Search a **nucleotide** database using a **nucleotide** query *Algorithms:* blastn, megablast, discontiguous megablast
protein blast	Search **protein** database using a **protein** query *Algorithms:* blastp, psi-blast, phi-blast
blastx	Search **protein** database using a **translated nucleotide** query
tblastn	Search **translated nucleotide** database using a **protein** query

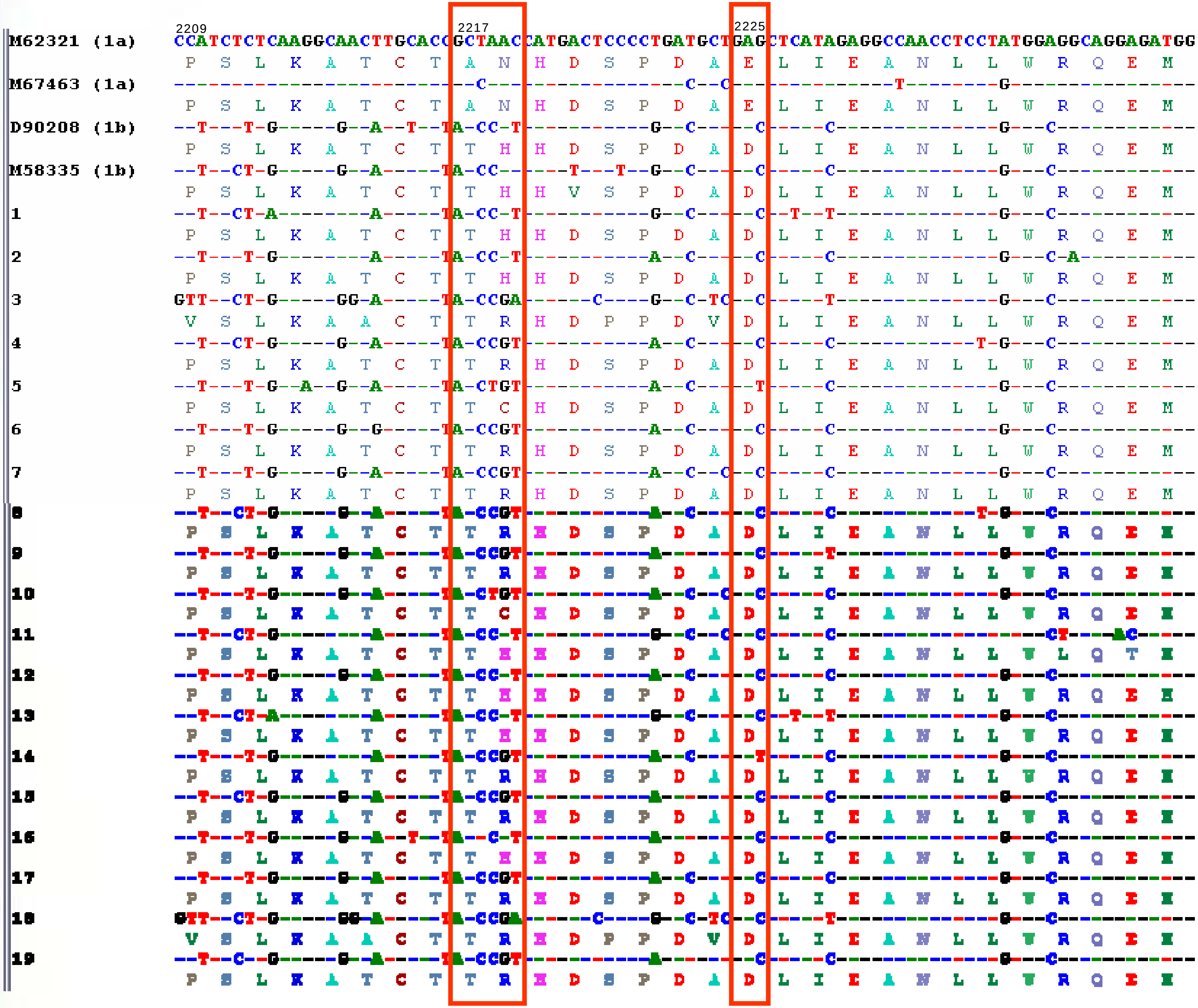


Results and Discussion

	PCR-RFLP	SEQUENCING	
Subtype	**5’NCR**	**5’NCR**	**NS5A**
1a	4	4	0
1b	15	4	19
Total	19	8	19

Alignment using the NS5A





```
              2209                    2217                    2225
M62321 (1a)   CCATCTCTCAAGGCAACTTGCACCGCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAGGAGATGG
               P  S  L  K  A  T  C  T  A  N  H  D  S  P  D  A  E  L  I  E  A  N  L  L  W  R  Q  E  M
M67463 (1a)   --------------------------C-----------------C--C--------------T--------G----------------
               P  S  L  K  A  T  C  T  A  N  H  D  S  P  D  A  E  L  I  E  A  N  L  L  W  R  Q  E  M
D90208 (1b)   --T---T-G-----G--A--T--TA-CC-T-----------G--C-----C-----C--------------G---C------------
               P  S  L  K  A  T  C  T  T  H  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
M58335 (1b)   --T--CT-G-----G--A-----TA-CC------T---T--G--C-----C-----C--------------G---C------------
               P  S  L  K  A  T  C  T  T  H  H  V  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
1             --T--CT-A--------A-----TA-CC-T-----------G--C-----C--T--T--------------G---C------------
               P  S  L  K  A  T  C  T  T  H  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
2             --T---T-G--------A-----TA-CC-T-----------A--C-----C-----C--------------G---C-A----------
               P  S  L  K  A  T  C  T  T  H  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
3             GTT--CT-G-----GG-A-----TA-CCGA------C----G--C-TC--C-----T--------------G---C------------
               V  S  L  K  A  A  C  T  T  R  H  D  P  P  D  V  D  L  I  E  A  N  L  L  W  R  Q  E  M
4             --T--CT-G-----G--A-----TA-CCGT-----------A--C-----C-----C------------T-G---C------------
               P  S  L  K  A  T  C  T  T  R  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
5             --T---T-G--A--G--A-----TA-CTGT-----------A--C-----T-----C--------------G---C------------
               P  S  L  K  A  T  C  T  T  C  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
6             --T---T-G-----G--G-----TA-CCGT-----------A--C-----C-----C--------------G---C------------
               P  S  L  K  A  T  C  T  T  R  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
7             --T---T-G-----G--A-----TA-CCGT-----------A--C--C--C-----C--------------G---C------------
               P  S  L  K  A  T  C  T  T  R  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
8             --T--CT-G-----G--A-----TA-CCGT-----------A--C-----C-----C------------T-G---C------------
               P  S  L  K  A  T  C  T  T  R  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
9             --T---T-G-----G--A-----TA-CCGT-----------A--------C-----T--------------G---C------------
               P  S  L  K  A  T  C  T  T  R  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
10            --T---T-G-----G--A-----TA-CTGT-----------A--C--C--C-----C--------------G---C------------
               P  S  L  K  A  T  C  T  T  C  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
11            --T--CT-G--------A-----TA-CC-T-----------G--C--C--C-----C------------------CT----AC-----
               P  S  L  K  A  T  C  T  T  H  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  L  Q  T  M
12            --T---T-G-----G--A-----TA-CC-T-----------A--C-----C-----C--------------G---C------------
               P  S  L  K  A  T  C  T  T  H  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
13            --T--CT-A--------A-----TA-CC-T-----------G--C-----C--T--T--------------G---C------------
               P  S  L  K  A  T  C  T  T  H  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
14            --T---T-G--------A-----TA-CCGT-----------A--C-----T-----C--------------G---C------------
               P  S  L  K  A  T  C  T  T  R  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
15            --T--CT-G-----G--A-----TA-CCGT-----------A--------C-----C--------------G---C------------
               P  S  L  K  A  T  C  T  T  R  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
16            --T---T-G-----G--A--T--TA--C-T-----------A--------C-----C--------------G---C------------
               P  S  L  K  A  T  C  T  T  H  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
17            --T---T-G-----G--A-----TA-CCGT-----------A--C-----C-----C--------------G---C------------
               P  S  L  K  A  T  C  T  T  R  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
18            GTT--CT-G-----GG-A-----TA-CCGA------C----G--C-TC--C-----T--------------G---C------------
               V  S  L  K  A  A  C  T  T  R  H  D  P  P  D  V  D  L  I  E  A  N  L  L  W  R  Q  E  M
19            --T--C--G-----G--A-----TA-CCGT-----------A--------C-----C--------------G---C------------
               P  S  L  K  A  T  C  T  T  R  H  D  S  P  D  A  D  L  I  E  A  N  L  L  W  R  Q  E  M
```

Accurate methods for both genotype and subtype classification are important tools to **optimize treatment type**, **duration**, and **dose**

Typing is important because **HCV-1 exhibits resistance** to combined interferon-α and ribavirin

*“Type-specific differences in response to new generation antiviral agents will be a major **research priority in the future”*** (Simmonds P. 2004)

Subtyping is **not currently used** to make clinical **treatment decisions**

Transfusion has been found to be a **risk factor** in the transmission of **subtype 1b**

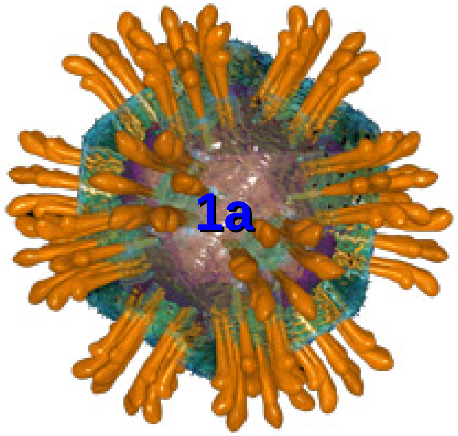
Subtype 1b is correlated with **increased risk of developing HCC**

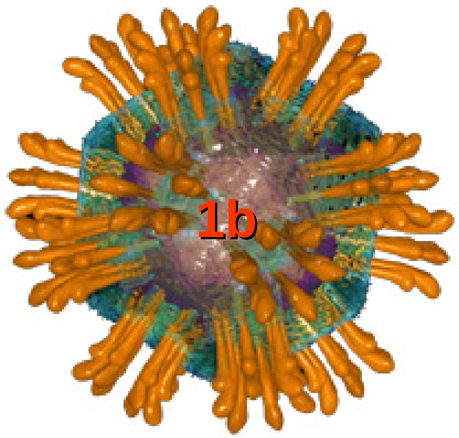
Alignments obtained with sequences from databases confirmed that amino acid positions **N2218** and **E2225** are widely distributed in HCV-1a subtype and can be considered as 1a markers (Punte *et al*., 2008)

Phylogenetic analysis of a coding region, is considered the ***“gold standard”*** for identifying HCV subtypes

Conclusion

- Sequence analysis of the **NS5A** region may be used in the identification of HCV-1 subtypes





	1a	**1b**
2217	**A**	**T**
2225	**E**	**D**

Liver Diseases Study Group

Acknowledgement