Comparative Genomics at PATRIC

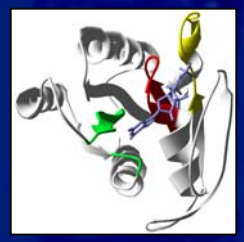
Targeted Comparison and a Focus on Coronavirus

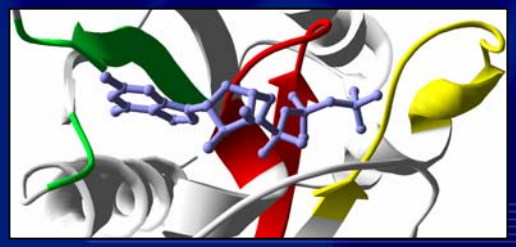
Stephen Cammer
PATRIC at V.B.I.
BRC Meeting 2

Virginia Tech

VBI

Genomic Comparison Tools for PATRIC





BLAST Suite

Smith-Waterman

MUMmer

CLUSTALW

Clustering

PHYLIP

Motif Analysis

SWISSPROT Annotation

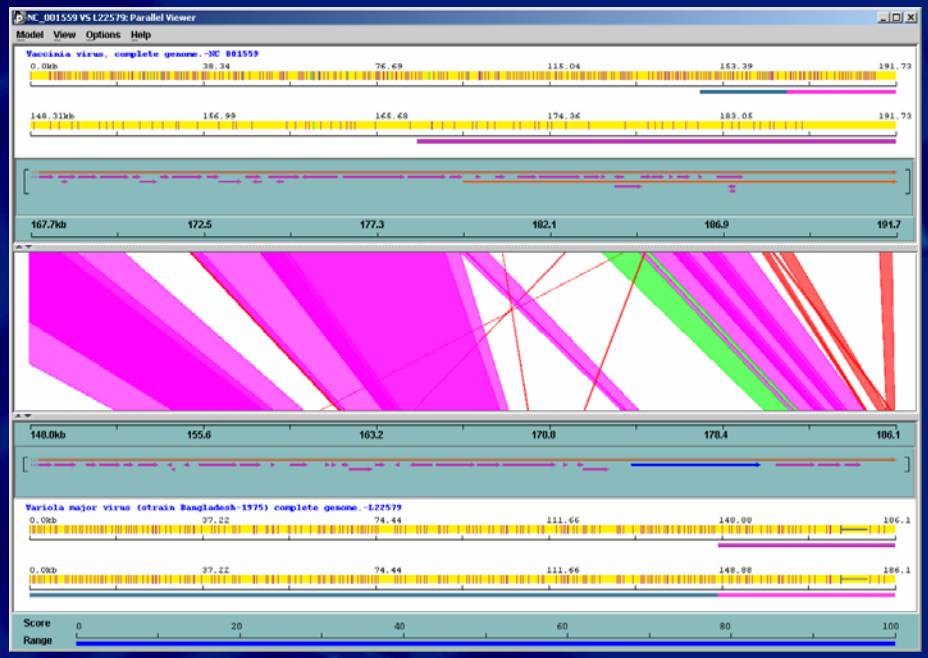
GO Annotation

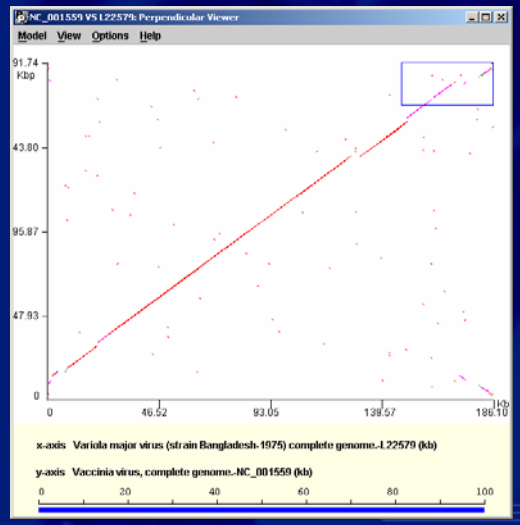
COGs

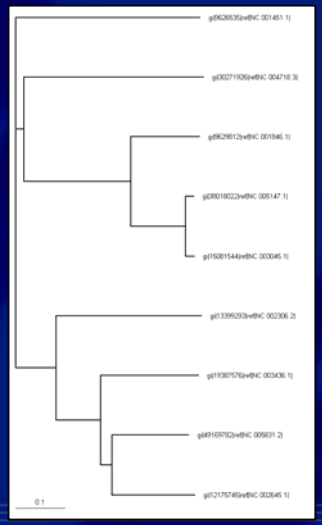
TIGRfam

3D Structure

Genome Comparison at PATRIC







Example Clustalw and MUMmer Output

<u>To Identify:</u>

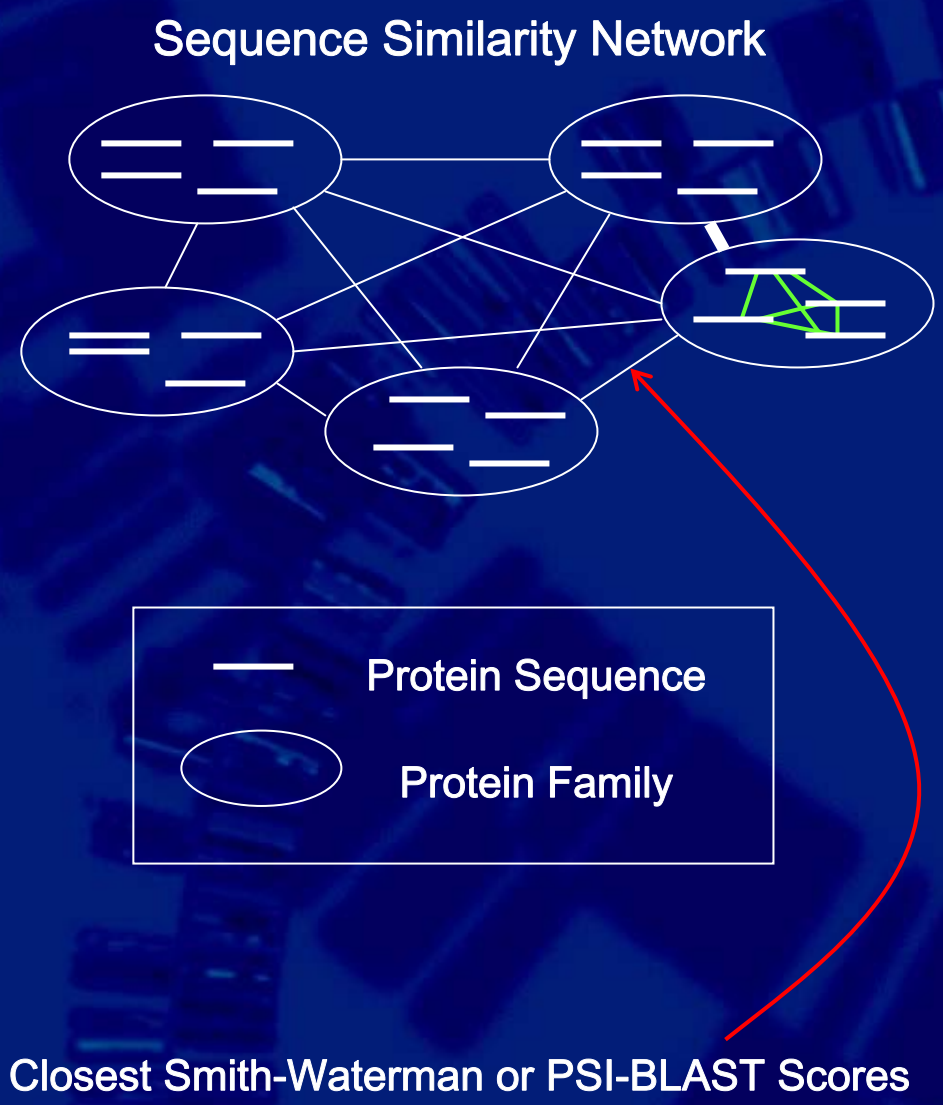
Conserved Syntenic Regions

Orthologous Groups

Unique Genes

SNPs

Protein Family Based Comparisons



Drug Target Space is Limited by Number of Protein Families

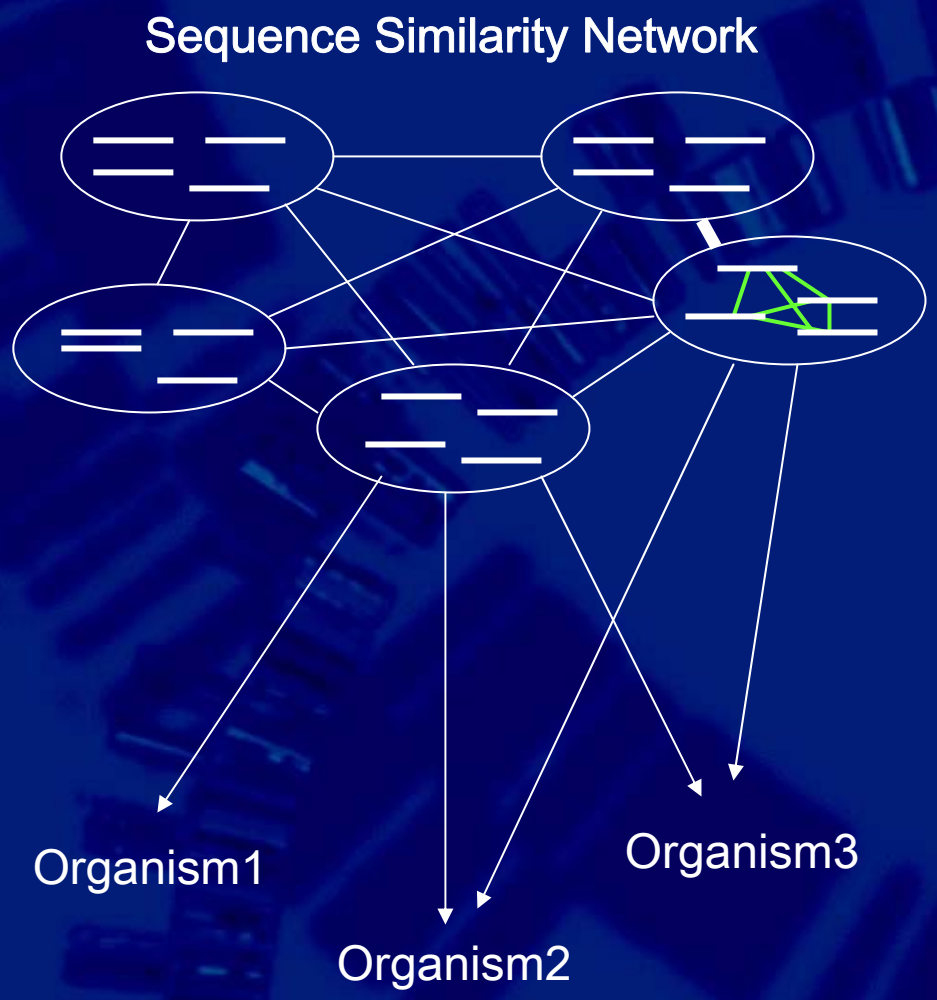
Map Protein Sequence Similarity Network

Exhaustive Pairwise Comparison

Generate BLink-Like Multiple Alignments Using PSI-BLAST

Estimate Trees Using Family-Specific Alignments

Protein Families Shared Across Organisms



Common Families:

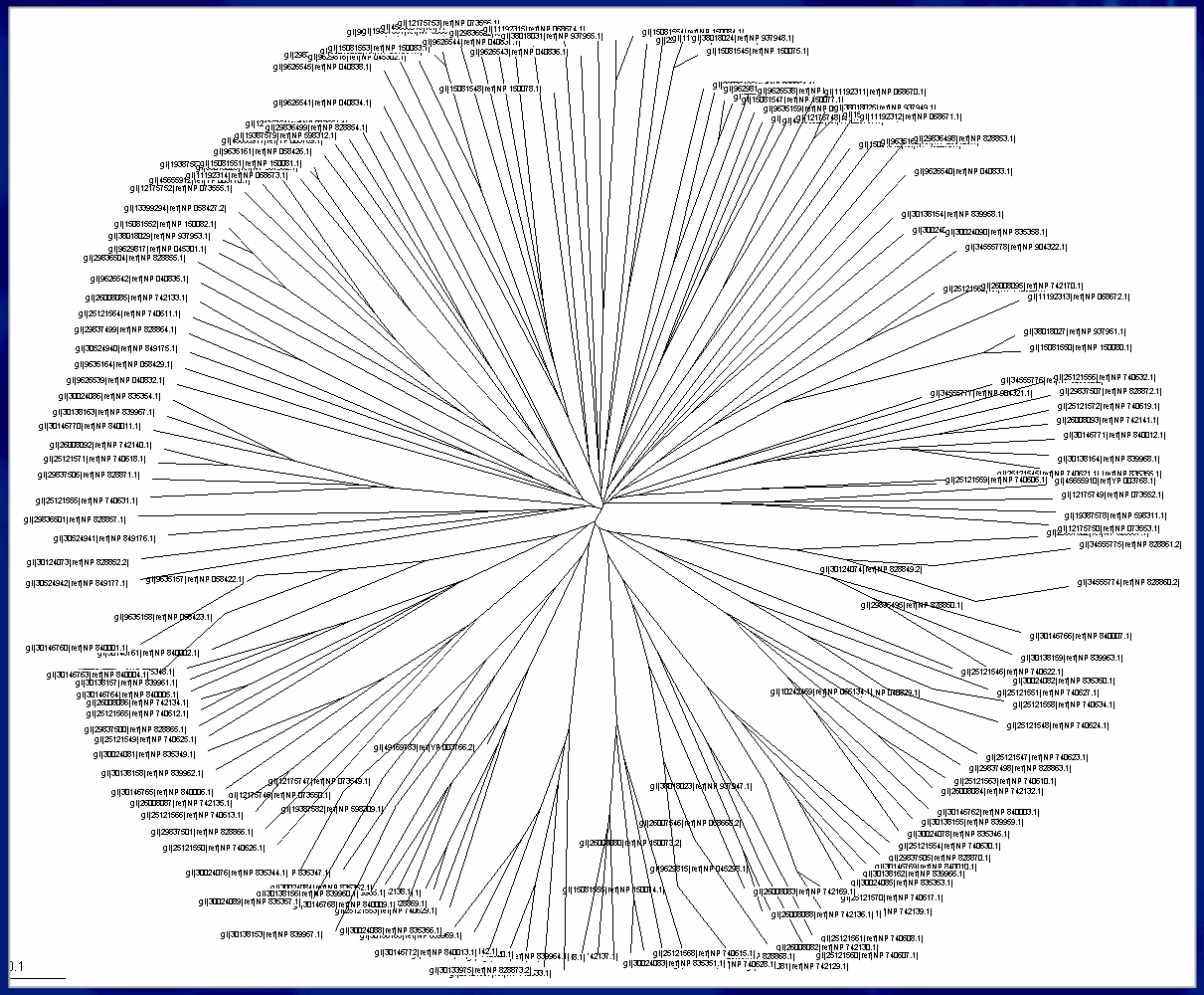
Aminoacyl-tRNA Synthetases

Ribosomal Proteins

Interesting Therapeutic Targets:
Organism-Specific Protein Variations

Transporters

Comparison of All Representative Proteins



Map All Proteome Variants

Identify Earliest Evolutionary Steps

Reveal Features Involved in Function

185 RefSeq Proteins for Coronavirus

PATRIC Organisms: Focus on Coronavirus

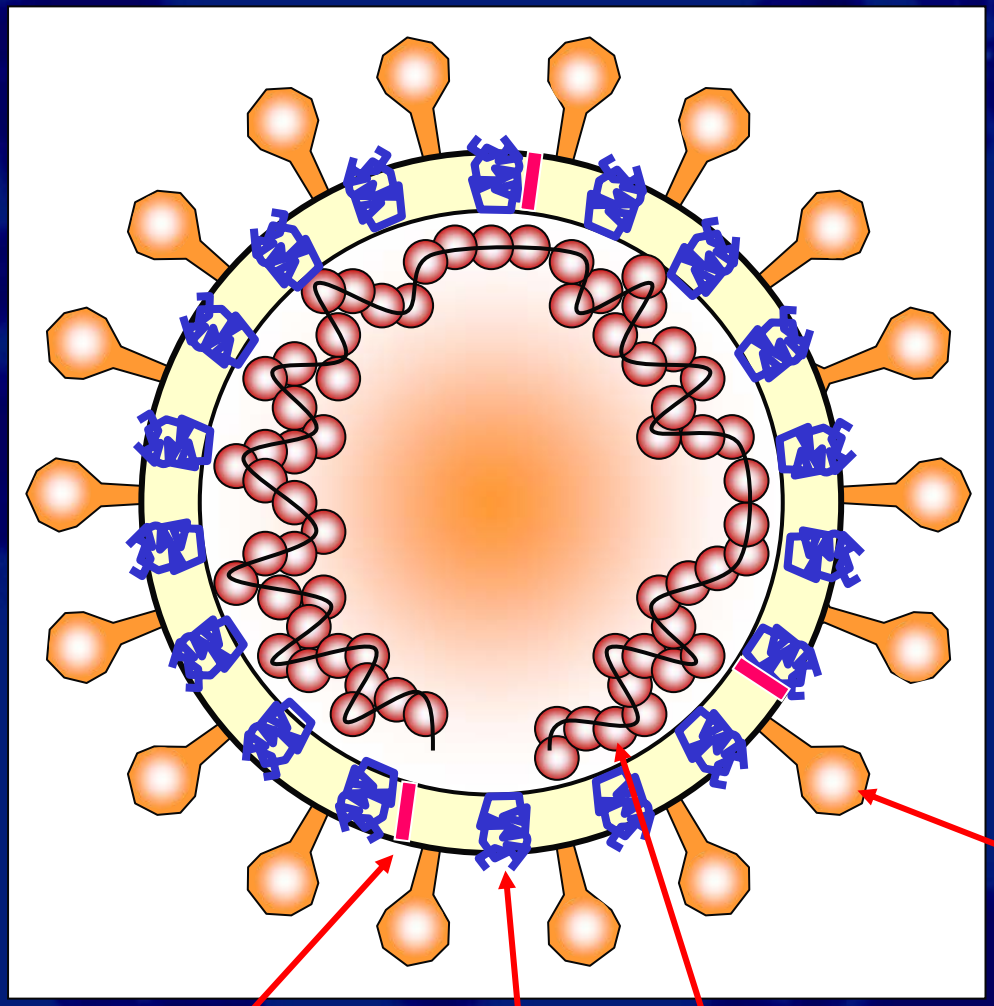
Bacteria

Rickettsia
Brucella
Coxiella burnetii

Viruses

Coronaviruses (SARS)
Caliciviruses
Rabies viruses
Hepatitis A viruses
Hepatitis E viruses

Coronavirus Morphology



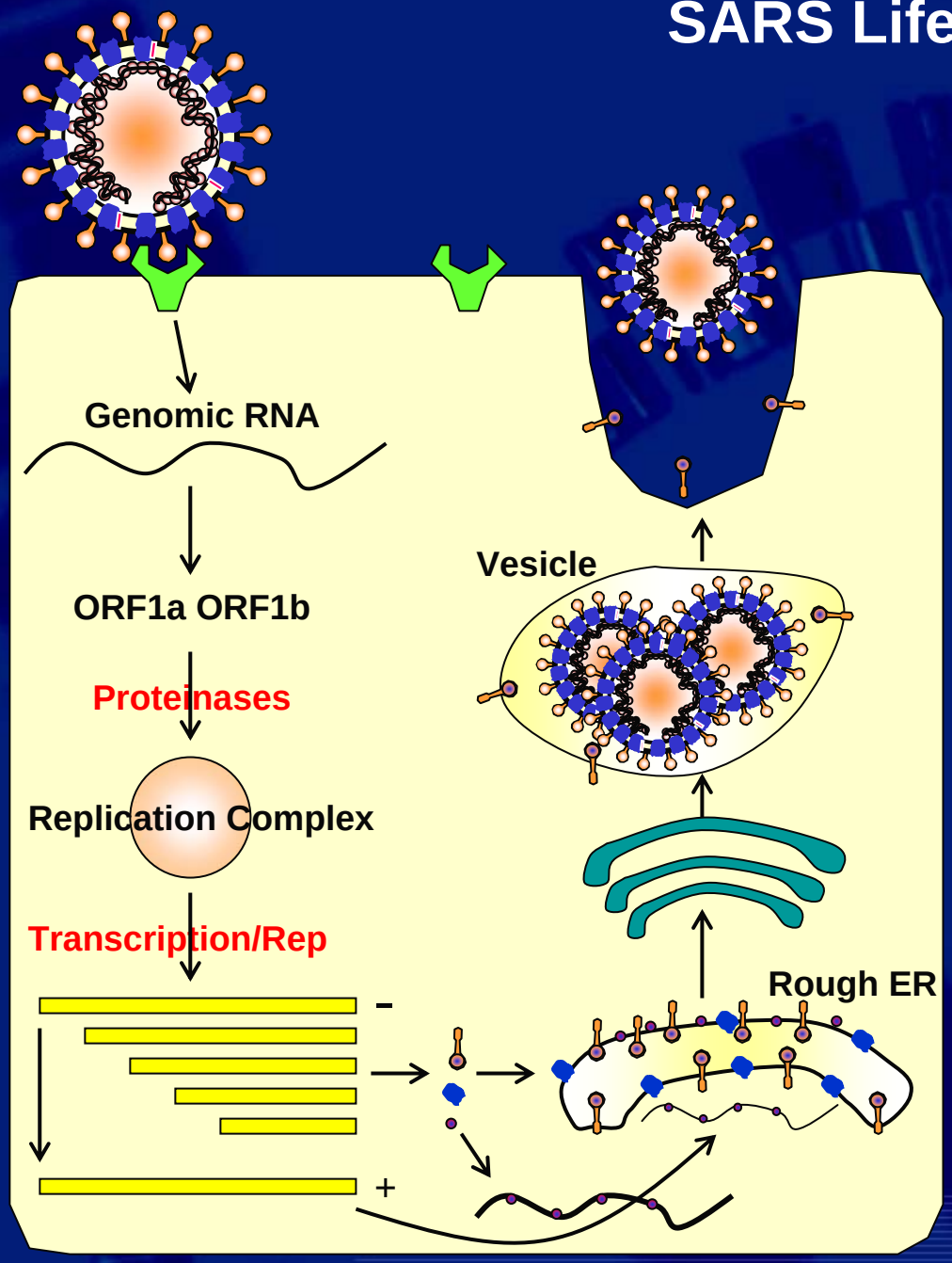
Lipid Bilayer From Intracellular Membrane

Spike Protein, Membrane Glycoprotein, and Envelope Protein

Helical Nucleocapsid With Genomic RNA

Stadler, et al, Nat. Rev. Micro. 1, Dec 2003

SARS Life Cycle



Receptor Binding, Entry, Uncoating

ORF1a & 1b Translated

Polyprotein Cleaved

Transcription Replication

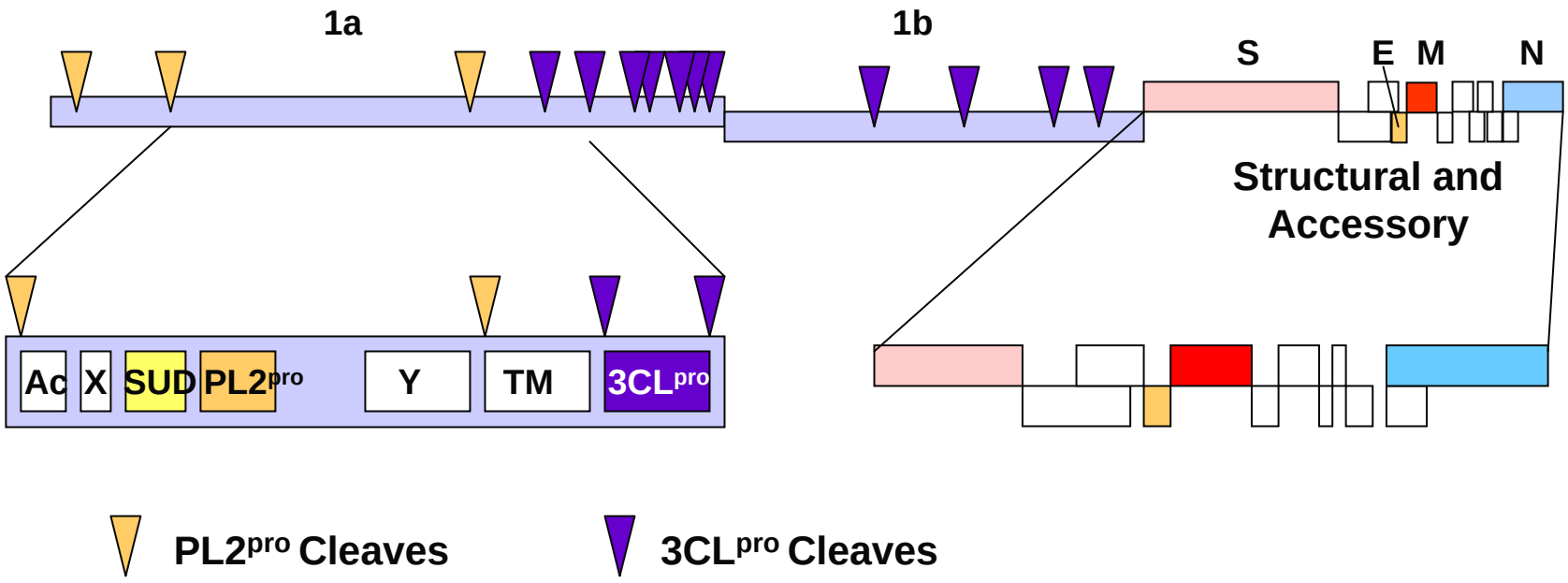
Translation

Processing/Targeting

Exocytosis and Release

Stadler, et al, Nat. Rev. Micro. 1, Dec 2003

Genome Organization Basics - SARS



SUD - SARS-Unique Domain
PL2^{pro} - Papain-like Cys Proteinase
3CL^{pro} – 3C-like Proteinase
TM2 – Transmembrane Domain

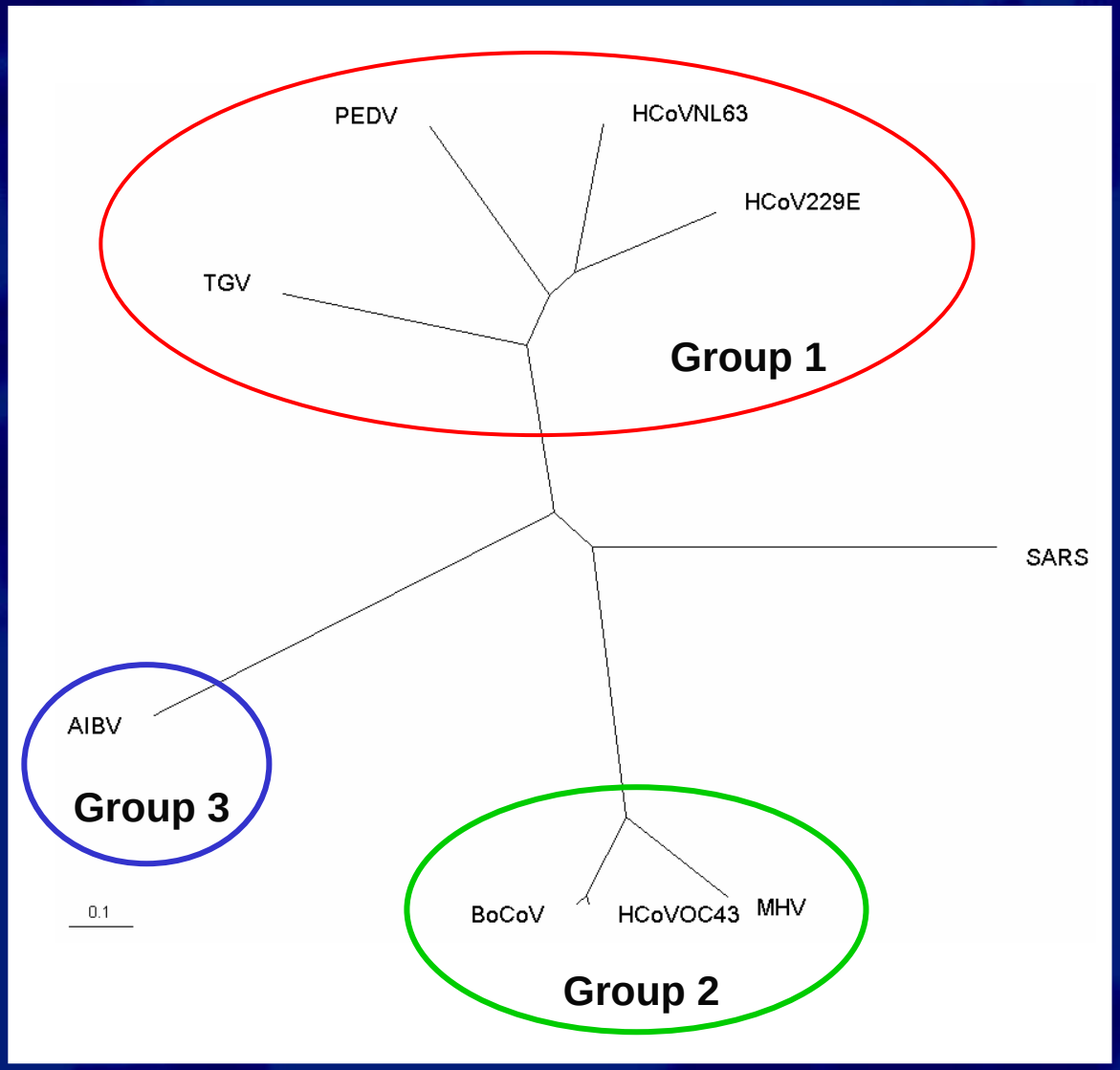
S - Spike Protein
E – Envelope Protein

M – Matrix Associated Protein
N – Nucleocapsid Protein

RNA-Dependent RNA Polymerase, Helicase w/Zn Finger,

Snijder, et al, J. Mol Biol (2003) 331

Coronavirus Phylogeny for 9 Whole Genomes



Maximum Likelihood

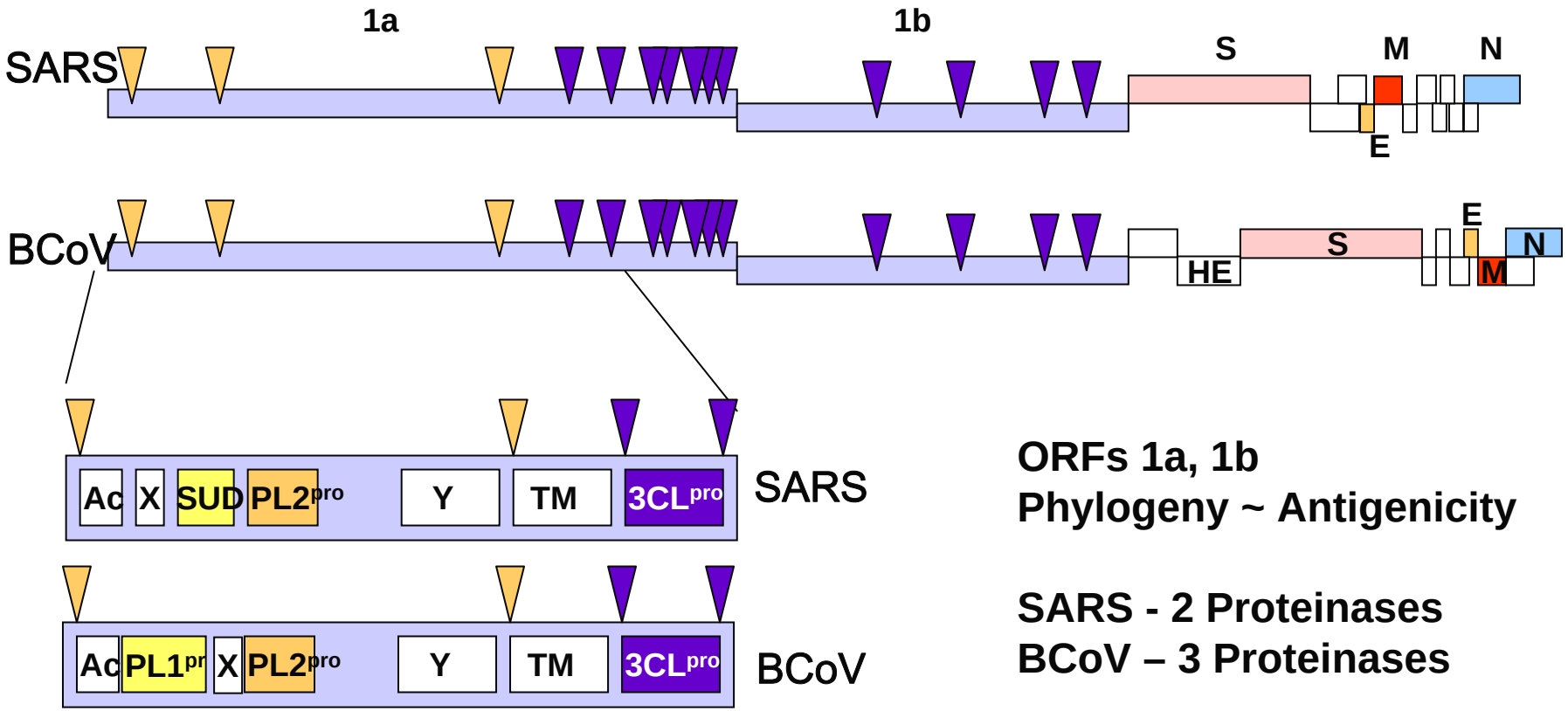
Tree Agrees With ORF1 Polyprotein Comparison

Similar to Nucleocapsid Comparison

3 Major Groups

SARS is Related But Distinct

Genome Organization – SARS vs. Bovine CoV

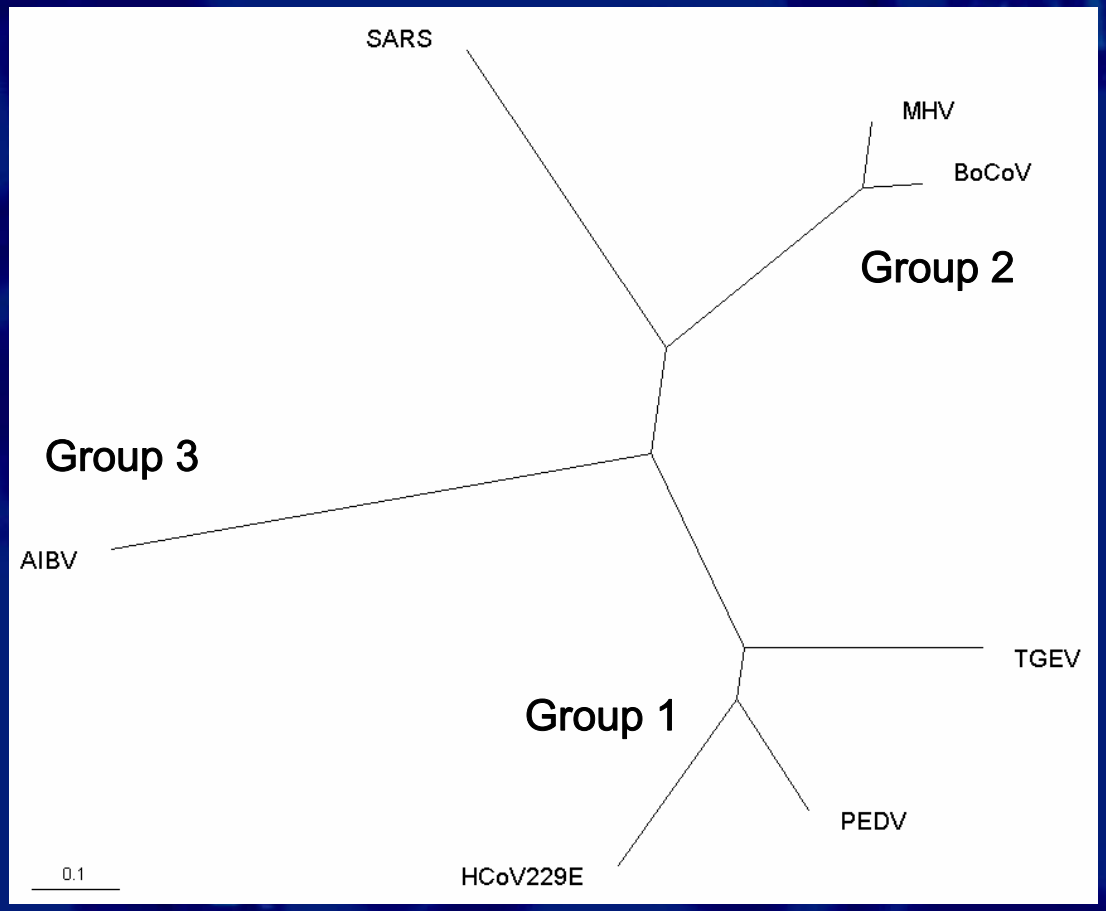


ORFs 1a, 1b
Phylogeny ~ Antigenicity

SARS - 2 Proteinases
BCoV – 3 Proteinases

Snijder, et al, J. Mol Biol (2003) 331

Comparison of Chymotrypsin-Like Proteases



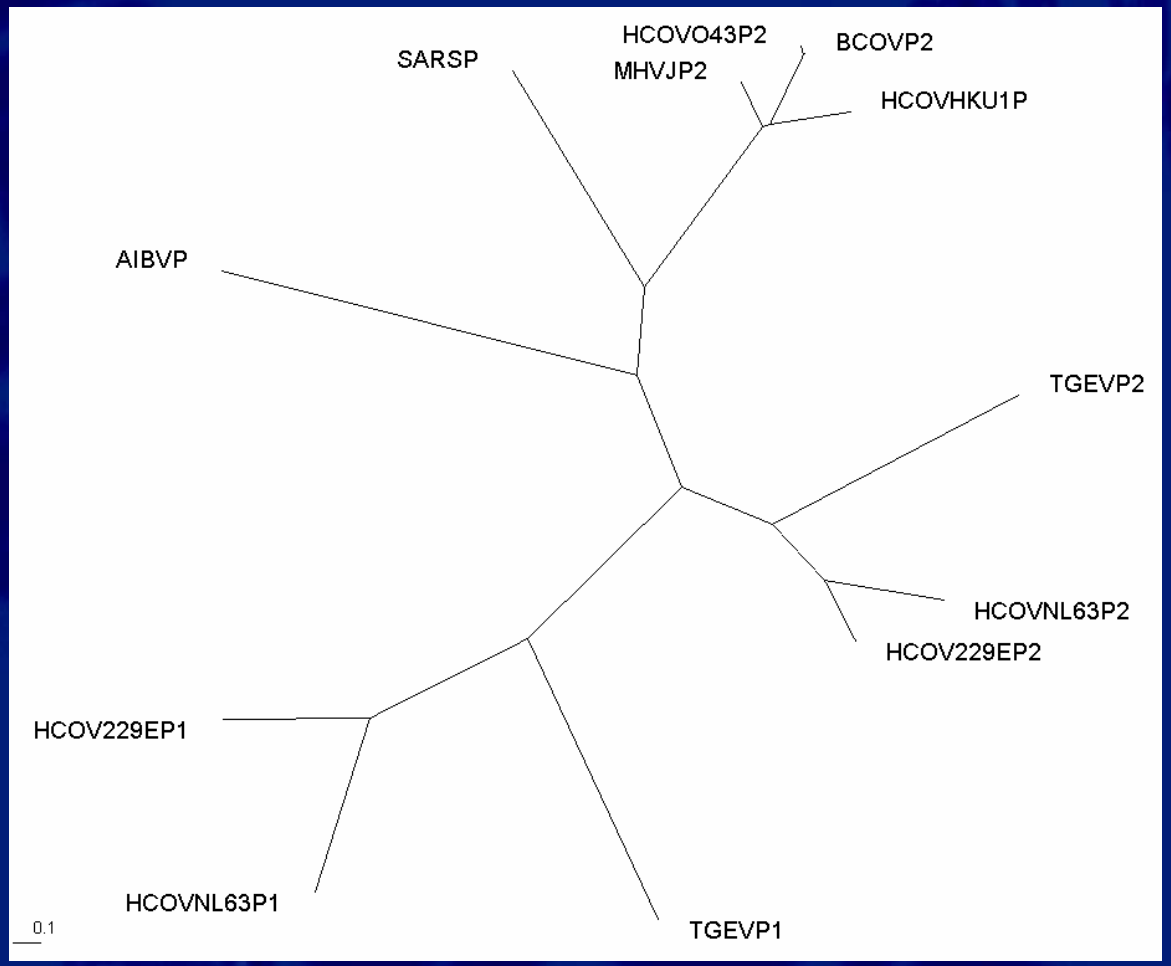
Maximum Likelihood

Agrees With Whole Genome Comparison

SARS CL Protease is Distinct from Others, But is Closer to Group 2

Avian Remains Most Distinct

Comparison of Papain-Like Proteases PL1 and PL2



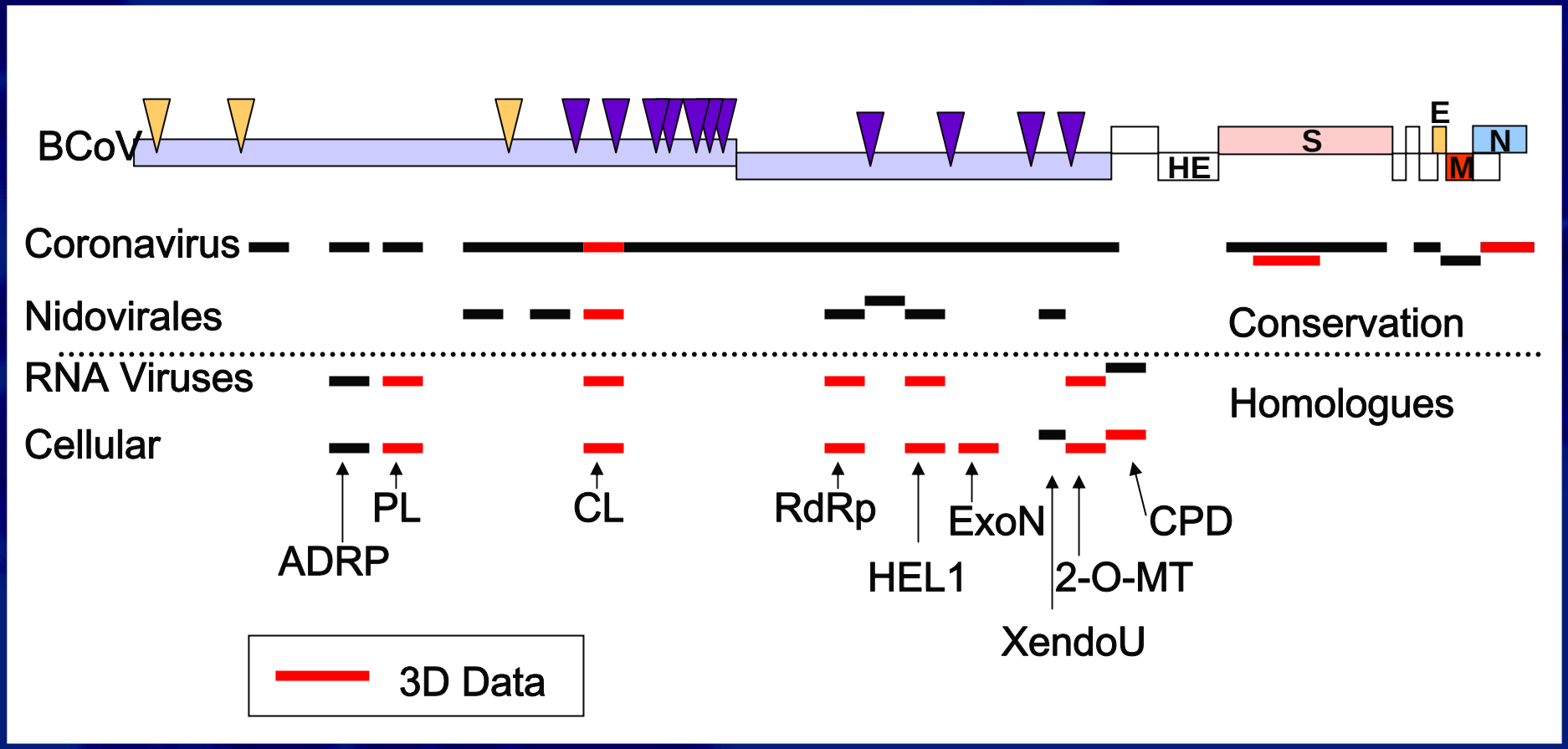
Maximum Likelihood

SARS Protease Resembles PL2 Types

HCoV HK Protease is Very Similar to Group 2 PL2

3 Groups/Types ?

Coronavirus Homologue Structures



ADRP – *Adenosine diphosphate-ribose phosphatase*
PL^{pro} - *Papain-like Cys Proteinase*
CL^{pro} – *Chymotrypsin-like Proteinase*
ExoN - *Exonuclease*
HEL1 - *Helicase w/Zn Finger*
CPD – *cyclic phosphodiesterase*
RdRp - *RNA-Dependent RNA Polymerase*
2'-O-MT – *SAM-dep ribose methyltransferase*

ORFs 1a, 1b, 2, 4, 5, 6, 9a Conserved in All Coronaviruses

Protein Family Active Site Profiles

Active Site Environment Map

Protein Sub-Site Fragments

KLGQGCFG

HRDLRAANI

KGSLLDF

IKTI

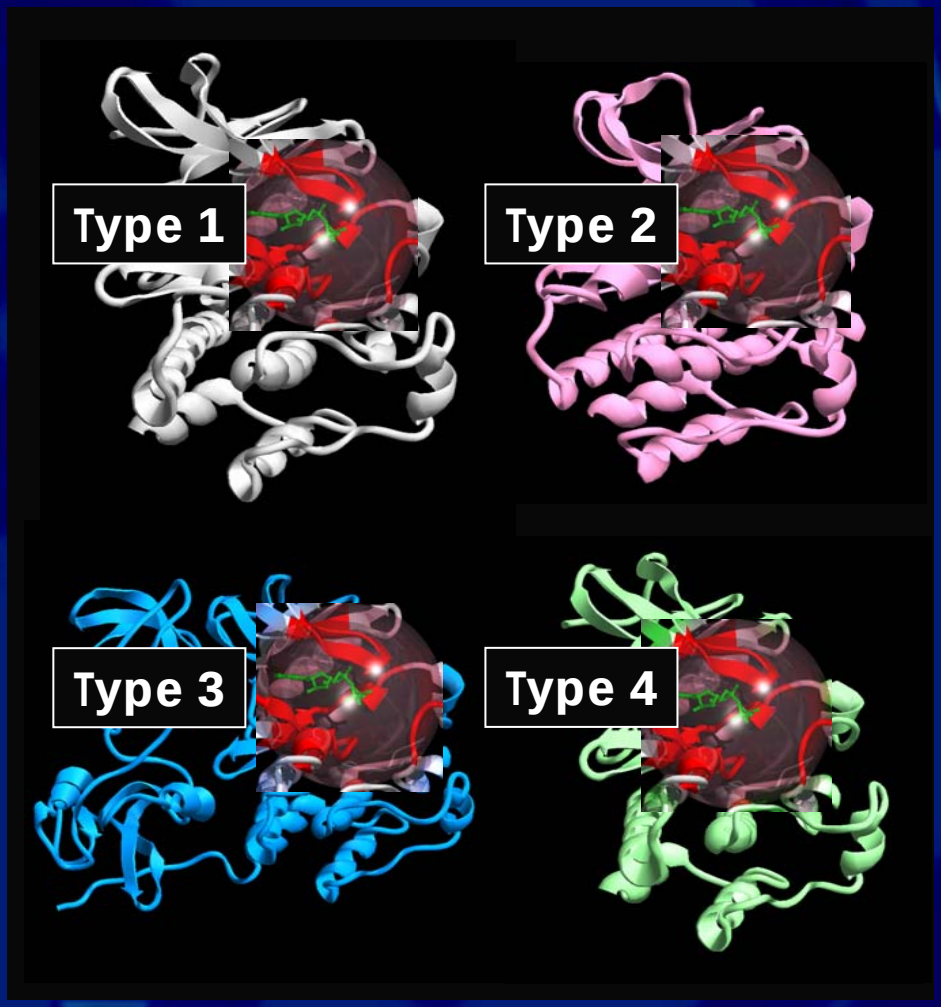
AKFPI

KLGQGCFGIKTIKGSLLDFHRDLRAANIAKFPI

Active Site Signature

Cammer, SA, et al, J. Mol. Biol. 2003, Nov 28, 334(3):387

Protein Family Active Site Comparison



Constant Features

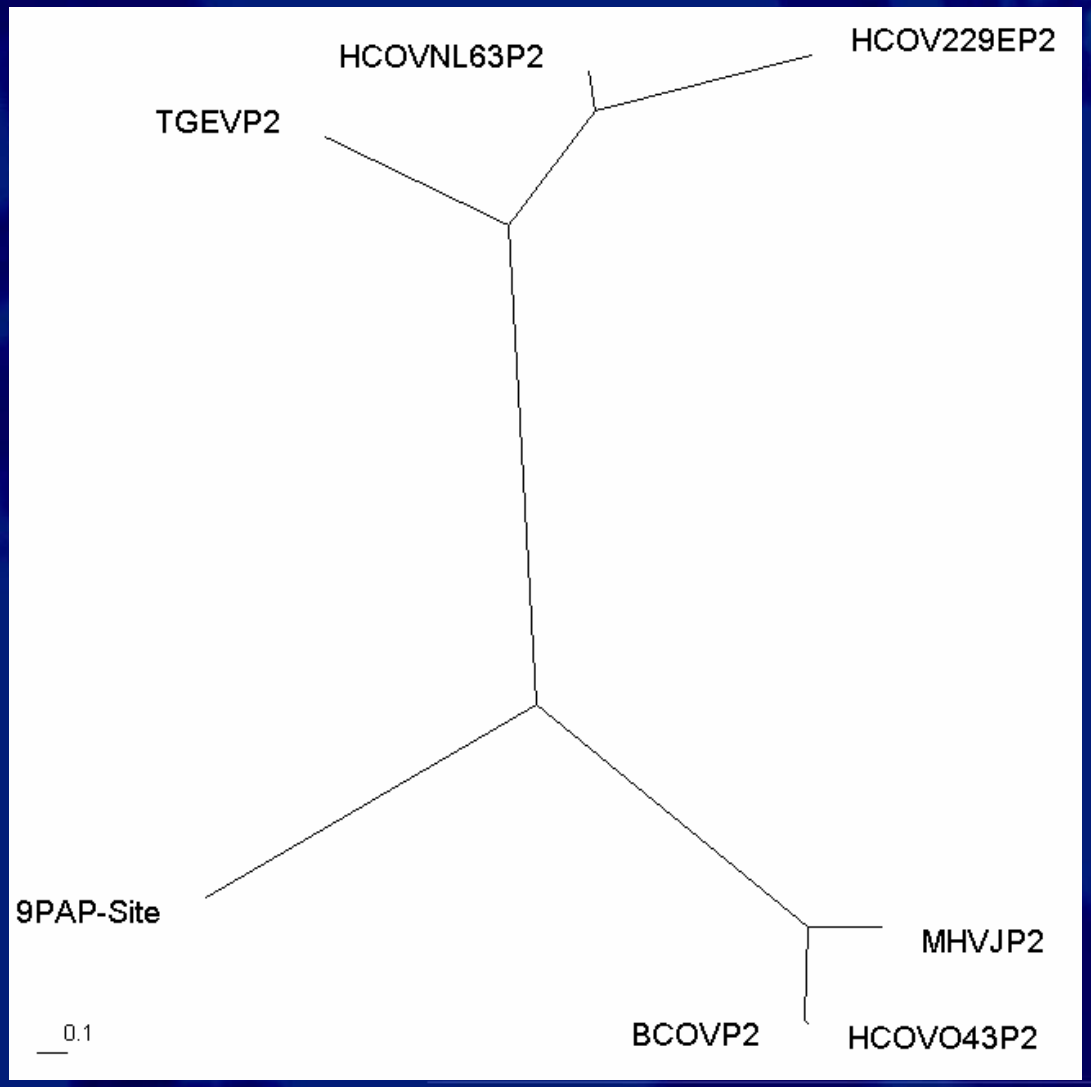
```
KLGQGCFGIKTIKGSLLDFHRDLRAANIAKFP
KLGQGSYGLKSVKGTIVDFHRDVRGAQI-KFP
KSGQ-GFGVKSIKGTLLDWHRDIRAG-VSRYP
KIAN-SYGIKTLKGSVFDYHRDLRAGNIGKWP
```

Variable Features

```
KLGQGCFGIKTIKGSLLDFHRDLRAANIAKFP
KLGQGSYGLKSVKGTIVDFHRDVRGAQI-KFP
KSGQ-GFGVKSIKGTLLDWHRDIRAG-VSRYP
KIAN-SYGIKTLKGSVFDYHRDLRAGNIGKWP
```

Cammer, SA, et al, J. Mol. Biol. 2003, Nov 28, 334(3):387

Active Site Comparisons: PL2 Protein Family



6 PL2 Protease Variants

Active Site for Papain Keyed Using Catalytic Site Database

Papain Site is Distinct from Coronavirus Variants

Next, Will Match to Viral Protease Variants and HAUSP