“CGG-CGG”: huella genética del origen de laboratorio del SARS-CoV-2

Antonio R. Romeu and Enric Ollé

AR: es Químico. Catedrático de Bioquímica y Biología Molecular de la Universidad Rovira i Virgili. Tarragona. España.

EO: es Veterinario y Bioquímico. Profesor Asociado del Departamento de Bioquímica y Biotecnología de la Universidad Rovira i Virgili. Tarragona. España.

El origen del virus es el gran debate, que se puede abordar desde muchos aspectos (1). En presente artículo lo abordamos exclusivamente desde la Ciencia, desde el análisis genético y de genomas.

El SARS-CoV-2 es un Coronavirus. En este grupo hay cuatro subgrupos (técnicamente, Géneros): Alfa-coronavirus, Beta-coronavirus, Gamma-coronavirus, Delta-coronavirus. El SARS-CoV-2 pertenece al grupo de Beta-coronavirus. Concretando más, dentro de los Beta-coronavirus, todavía hay cuatro subgrupos (subgéneros): Linaje A, Linaje B, Linaje C, y Linaje D. El SARS-CoV-2 pertenece al Linaje B (técnicamente, su subgénero se llama Sarbecovirus).

Es decir, el SARS-CoV-2 es un Coronavirus → Betacoronavirus → Sarbecovirus (2)

La característica fundamental del SARS-CoV-2 es que su proteína S, la que interacciona con las células humanas y produce la infección, ha adquirido un inserto con carga positiva en una región estratégica de la proteína, que facilita enormemente la infección y la propagación del virus de persona a persona (2,3,4). Es decir, un ancestro del virus, dentro del murciélago (o en otro animal intermedio) adquirió dicho inserto, y se convirtió en el actual SARS-CoV-2; después saltó del animal a las personas.

Así pues, de una forma reduccionista, se puede decir, que el origen del SARS-CoV-2 es el propio origen del inserto positivo en su proteína S (5).

Técnicamente, el inserto positivo se llama “sitio polibásico de escisión de furina”. La furina es una proteína de la membrana de nuestras células, que reconoce dicho inserto y colabora con los receptores correspondientes. En estas condiciones, la entrada del virus en nuestro organismo es por la puerta grande.

Estudiando el inserto del SARS-CoV-2, hay características que no encajan, y es lo que hizo pensar desde el principio que su origen pueda ser de laboratorio. El SARS-CoV-2, es el único miembro de su grupo (Sarbecovirus), que tiene este sitio de furina. Esto ya lo hace evolutivamente diferente, y desde el primer día, la pregunta del millón es ¿cómo le ha llegado?

En un análisis detallado de la composición del inserto, se observa que hay dos elementos positivos seguidos, llamados “arginina-arginina” (el símbolo de la arginina es una “R”; es decir, hay un dímero RR).

Un principio básico de la Biología o de la Genética es que cada elemento de las proteínas está codificado por 3 elementos (un triplete) del ADN o de los genes. Esto es el “Codigo Genético Universal”. En el caso concreto del elemento “arginina” (R), en los genes existen 6 grupos de 3 elementos, que codifican por igual a la R. Estos 6 elementos no se usan en la misma proporción en todas las especies. La Tabal 1 presenta los 6 tripletes que codifican R, y las frecuencias en que estos tripletes en que se utilizan en el SARS-CoV-2 (6) y en la especie humana (7). En la Tabla 1 se hace hincapié en que el uso de los 6 tripletes no es aleatorio como si se tratase de las caras de un dado.

Arginina (R)	AGA	AGG	CGA	CGC	CGG	CGT
Frecuencia aleatoria:	1/6	1/6	1/6	1/6	1/6	1/6
SARS-CoV-2 (%):	45	13	5	10	3	24
Genoma humano (%):	20	20	11	19	21	9

Tabla 1. Tripletes que codifican arginina, "R" (Código Genético Universal).

Volviendo al inserto postivo del SARS-CoV-2, su dímero RR está codificado por los tripletes CGG-CGG. Sorprendentemente, este triplete "CGG" sea el menos frecuente en el SARS-CoV-2 (3%) y el más utilizado en el genoma humano (21%). Aún sorprende más, que en el virus pandémico haya dos tripletes minoritarios seguidos. Por otra parte, y atendiendo a los datos de la Tabla 1, sea cual sea el origen de la secuencia CGGCGG, lo que es cierto es que está optimizado para las personas (8).

Es importante saber que el sitio de furina (inserto positivo), aunque no lo tienen los virus del grupo del SARS-CoV-2 (excepto él), es muy común en el mundo de los virus (9). Otros Coronavirus, lo tienen en su proteína S (por ejemplo, el MERS). Fuera del grupo de los Coronavirus, y dentro del grupo de los virus de la gripe; algunos virus de la gripe con la proteína H5, también tienen el inserto de furina. En general, muchos sitios de furina de los virus, comparten la presencia de un doblete RR (10).

En un estudio previo, estudiamos bioinformáticamente como estaban codificados los pares RR de los sitios de furina de toda clase de virus, en una muestra grande y representativa. Sorprendentemente, ningún par RR está codificado por la secuencia CGGCGG (6). De acuerdo con nuestros resultados, un estudio reciente describe que la secuencia CGG-CGG nunca se ha encontrado de forma natural en los Coronavirus (11).

En otro estudio (en proceso de publicación) hemos analizado los pares RR en las otras proteínas del SARS-CoV-2. El SARS-CoV-2 tiene 24 proteínas. En la proteína S, sólo hay un par RR (el del inserto). Pero, en las otras proteínas ¿tienen pares RR? Sólo en unas pocas, y no estan codificados por la secuencia CGGCGG. La Tabla 2 presenta las proteínas del SARS-CoV-2 con dímeros RR, sus posiciones, y como están codificados.

Gen/proteína	Position en el genoma	RR	Posición	Código genetico	Posición
nsp3	2720-8553	GV**RR**	1614	GGTGTT**AGAAGG**	4842
nsp4	8555-10054	RF**RR**	306	AGGTTT**AGAAGA**	918
		LK**RR**	401	CTAAAG**AGACGT**	1203
nsp6	10973-11842	GA**RR**	138	GGTGCT**AGGAGA**	414
nsp13	16237-18039	CI**RR**	22	TGCATA**CGTAGA**	66
		TC**RR**	443	ACTTGT**CGGCGT**	1329
		IP**RR**	595	ATTCCA**CGTAGG**	1785
nsp14A2	18040-19620	TY**RR**	53	ACCTAT**AGAAGA**	159
		CD**RR**	213	TGTGAT**AGACGT**	639
S	21563-25384	SP**RR**	683	TCTCCT**CGGCGG**	2049
ORF9 (N)	28274-29533	KQ**RR**	41	AAACAA**CGTCGG**	123
Nucleocapsida		YY**RR**	89	TACTAC**CGAAGA**	267
		AT**RR**	93	GCTACC**AGACGA**	279
		FG**RR**	277	TTCGGC**AGACGT**	831

Tabla 2. Proteínas del SARS-CoV-2 que tienen pares RR. La información relativa al inserto positivo de la proteína S se resalta en rojo.

La siguiente pregunta fue ¿los pares RR que se encuentran en el SARS-CoV-2 (a parte del RR del sitio de furina), se encuentran también en los virus de murciélago conocidos que son sus "parientes más cercanos"? La respuesta fue rotundamente sí. También fue sorprendente, que ninguno de ellos está codificado por la secuencia CGGCGG. Es más, muy pocos de los pares RR están codificados por el miso triplete dos veces seguidas. Solo en unos pocos dobletes se repite el códdigo "AGA" (AGAAGA), que es el mayoritario (45%) en el SARS-CoV-2 (Tabla 1). La Figura del Anexo 2 presenta, el detalle los dobletes RR de los Coronavirus estudiados y como están codificados.

Todo ellos nos lleva al origen del SARS-CoV-2. Frente a este gran enigma, hay que formularse la pregunta correcta: ¿cuál es el origen de la secuencia CGGCGG que codifica el inserto de la proteína S del Coronavirus?

Un origen natural debe contemplar dos posibles eventos (naturales): (*i*) mutaciones de inserción aleatorias en un antepasado próximo, en el murciélago (u otro animal), originando el SARS-CoV-2 que conocemos ; o (*ii*) relaciones de intercambio de material genético entre un antepasado pròximo y otros virus, en el murciélago (u otro animal), dando lugar, también, al SARS-CoV-2 que conocemos.

Una inserción al azar, relativamente larga, como la incluye la secuencia CGGCGG que codifica el dímero RR en cuestión, es muy improbable. En la reproducción de estos Coronavirus (replicación) hay mecanismos de corrección que evitan grandes errores, como hubiese sido la inserción que aparece en el SARS-CoV-2. Es muy poco probable que por azar, dos tripletes muy poco frecuentes del virus aparezcan, nada más ni nada menos, que seguidos. Serían dos procesos de inserción independientes que tendrían que haber ocurrido simultáneamente en el tiempo y en el espacio. Otro proceso independiente, ocurrido también en el mismo tiempo y espacio, es que la secuencia insertada en el curso de la evolución de un antepasado del SARS-CoV-2, sea la secuencia más frecuente en las personas (Tabla 1). En cuestión de origen de las especies, no hay nada imposible, pero este mecanismo de mutación aleatoria, como origen del virus, parte de la comunidad científica no lo acepta; de lo contrario, no habría este debate sobre el origen.

La segunda posibilidad de un origen natural se basa en un intercambio de material genético con otro virus (recombinación). Esta posibilidad es la más lógica. No obstante, si otro virus ha transfirido el inserto; este virus donador tendría que tener el código CGGCGG, y de momento, no se conoce ningún dímero RR de sitio de furina de otros virus con dicho código. Es decir, la recombinación, como mecanismo molecular, se acepta, pero hasta que no se descubra un virus donador, queda abierta.

Finalmente, un origen de laboratorio de la secuencia CGGCGG del SARS-CoV-2 no es imposible. Los detractores alegan que un origen de laboratorio siempre deja huella, de ser así, la propia secuencia CGGCGG es la huella.

Referencias

1. Nicholas Wade. Origin of Covid — Following the Clues. Acceso el 03/07/2021. https://nicholaswade.medium.com/origin-of-covid-following-the-clues-6f03564c038.

2. Philip V'kovski, Annika Kratzel, Silvio Steiner, Hanspeter Stalder, Volker Thiel. Coronavirus biology and replication: implications for SARS-CoV-2. Nat. Rev. Microbiol. Oct 28;1-16, 2020. PMID: 33116300. doi: 10.1038/s41579-020-00468-6.

3. Britt Glaunsinger. Coronavirus biology. The second lecture in the COVID-19, SARS-CoV-2 and the Pandemic Series. University of California, Berkeley. 2020. Accessed June 23, 2021. https://www.youtube.com/watch?v=r2mOU2qOCYs.

4. Kristian G. Andersen, Andrew Rambaut, W Ian Lipkin, Edward C Holmes, Robert F Garry. The proximal origin of SARS-CoV-2. Nat. Med. 26:450-452, 2020. PMID: 32284615. doi: 10.1038/s41591-020-0820-9.

5.Manuel Ansede, Artur GAlocha y Mariano Zafra. ccu cgg cgg gca, thw 12 leters that changed the world.El País, 19-mayo-2020. Acceso el 03/07/2021. https://elpais.com/elpais/2020/05/18/ciencia/1589818040_544543.html

6. Antonio R. Romeu, Enric Ollé. SARS-CoV-2 and the Secret of the Furin Site. Preprints 2021, 2021020264 (doi: 10.20944/preprints202102.0264.v1).

7. GenScript Codon Usage Frequency Table(chart) Tool. Accessed June 23, 2021. https://www.genscript.com/tools/codon-frequency-table.

8. Antonio R. Romeu, Enric Ollé. SARS-CoV-2 Origin: An Affair of Codons?. Preprints 2021, 2021060121 (doi: 10.20944/preprints202106.0121.v1).

9. Elisabeth Braun, Daniel Sauter. Furin-mediated protein processing in infectious diseases and cancer. Clin. Transl. Immunol. E1073, 2019. PMID: 31406574. doi.org/10.1002/cti2.1073.

10. Imène Kara, Marjorie Poggi, Bernadette Bonardo, Roland Govers, Jean-François Landrier, Sun Tian, Ingo Leibiger, Robert Day, John W M Creemers, Franck Peiretti. The Paired Basic Amino Acid-cleaving Enzyme 4 (PACE4) Is Involved in the Maturation of Insulin Receptor Isoform B. J. Biol. Chem. 290:2812-2821, 2015. PMID: 25527501. doi: 10.1074/jbc.M114.592543.

11. Steven Qauy, Richard Muller. The Science suggests a Wuhan lab leak. The Wall Street Journal, Monday, June 7, 2021. A17.

Anexos

Anexo 1. Metodología (lenguaje técnico)

Anexo 2 (en inglés). Figura de múltiples páginas

Se presentan fragmentos de los alineamientos múltiples de las secuencias de las proteínas y los genes de los Coronavirus del grupo del SARS-CoV-2, con presencia de dobletes RR.

Para cada proteína, en primer lugar se presenta una figura apaisada a fin de localizarla en el genoma del virus.

En los fragmentos de los alineamientos múltiples, los dobletes RR y sus códigos se resaltan en amarillo. Los números a la derecha indican las posiciones en la secuencia.

Anexo 1

Metodología

La fuente de información han sido las bases de datos NCBI GenBank (https://www.ncbi.nlm.nih.gov/nucleotide/) y GISAID (https://www.gisaid.org/).

El genoma de SARS-CoV-2 de referencia fue del Coronavirus WIV04, Wuhan-Hu-1 (NCBI, número de GenBank NC_045512.2 y GISAID, EPI_ISL_402124).

Coronavirus (Sarbecovirus) de la muestra de trabajo:

- Humanos: SARS-CoV-2 (MN996528.1, MT159709.2, MT066156.1), SARS-CoV (HKU-39849)
- De hurón: SARS-CoV (Tor2/FP1-10912)
- De murciélago: RatG13, RmYN02, RShsTT200, RShsTT182, RpYN06, RacCS203, PrC31, VZC45, ZXC21, BM48-31 y BtKY72.
- De pangolín: MP789, GX-P5L, GX-P4L, GX-P1E, GX-P5E y GX-P2V.

Plan de trabajo:

- Primero, identificación de las proteínas de SARS-CoV-2 con presencia de dímeros RR.
- Segundo, creación de una base de datos interna con dichas proteínas y sus genes, de los Coronavirus de la muestra de trabajo.
- Tercero, alineamientos múltiples, tanto de las secuencias de proteínas como de genes, mediante el programa EMBL Clustal Omega (v.1.2.4) (https://www.ebi.ac.uk/Tools/msa/clustalo/).

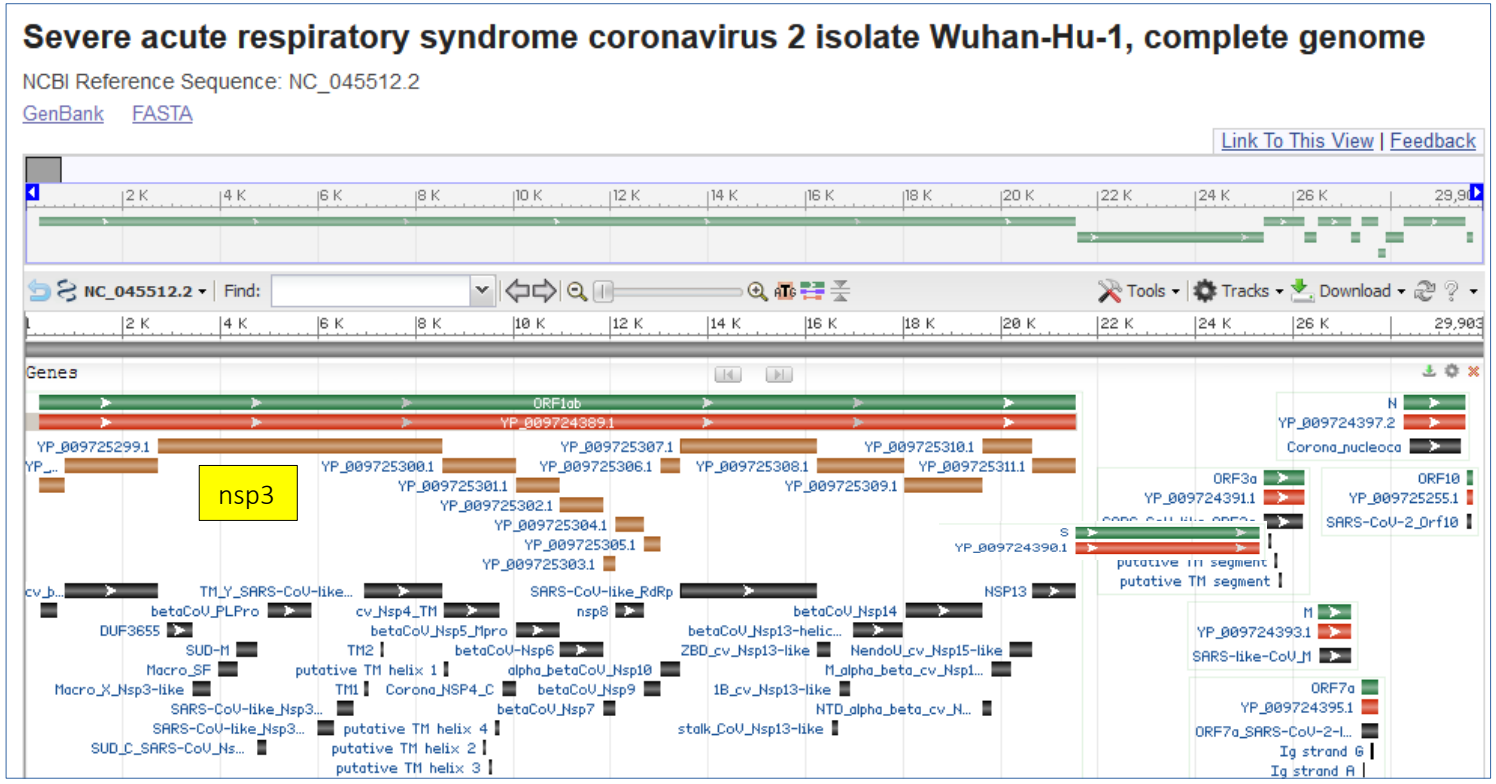
Dado que no todos los genomas se anotaron completamente, las secuencias de genes se obtuvieron mediante análisis BLASTn sucesivos, utilizando los genes de referencia del SARS-CoV-2 como query contra los genomas de Sarbecovirus de la muestra.

Anexo 2

Figura (varias páginas) de fragmentos de los alineaminetos múltiles de las proteínas y genes de los Coronavirus de la muestra, siguiendo el orden de la Tabla 2.

nsp3

Papain-like proteinase. Double-membrane vesicles formation. Macro domains



Protein nsp3 (the RR pair)

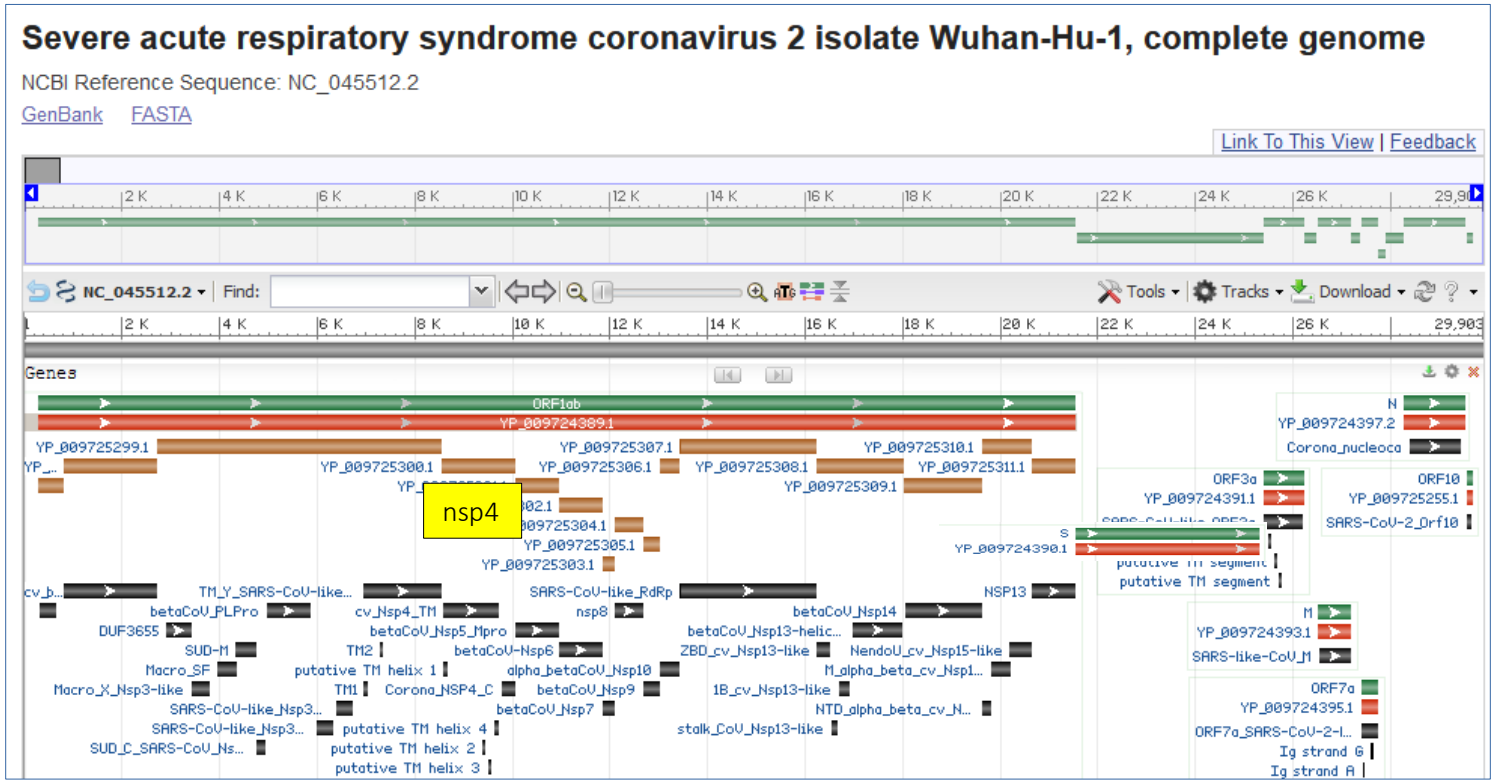
```
PCoV_GX-P2V             YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARD      1449
PCoV_GX-P1E             YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARD      1449
PCoV_GX-P5L             YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARD      1449
PCoV_GX-P4L             YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARD      1449
PCoV_GX-P5E             YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARD      1449
ZC45                    YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCINCDTFCAGSTFISDEVARD      1650
ZXC21                   YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCINCDTFCAGSTFISDEVARD      1630
PCoV-MP789              YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARD      1471
RShSTT200               YKRNRATRVECTTIVNGVKRSFYVYANGGNGFCKLHNWNCVNCDTFCAGSTFISDEVARD      1647
RShSTT182               YKRNRATRVECTTIVNGVKRSFYVYANGGNGFCKLHNWNCVNCDTFCAGSTFISDEVARD      1647
RacCS203                YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCTGSTFISDEVARD      1671
RaTG13                  YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARD      1653
INMI1                   YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARD      1654
CDC-CruiseA-12          YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARD      1654
WIV04                   YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARD      1654
Wuhan-Hu-1              YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARD      1654
RpYN06                  YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARD      1653
RmYN02                  YKRNRATRVECTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARD      1652
HKU-39849               YKRNRATRVECTTIVNGMKRSFYVYANGGRGFCKTHNWNCLNCDTFCTGSTFISDEVARD      1243
Tor2/FP1-10912          YKRNRATRVECTTIVNGMKRSFYVYANGGRGFCKTHNWNCLNCDTFCTGSTFISDEVARD      1243
BM48-31                 YKRNRATRVECTTIVNGMKKSFYVYANGGQGFCKLHNWNCLNCDTFCSGSTFISDEVARD      1151
                        ******************:::*********.**** *****:******:************
```

Gene nsp3 (the encoding of the RR pair)

```
PCoV_GX-P2V             AAACGTAATAGAGCTACAAGGGTTGAATGCACAACCATTGTAAATGGTGTAAGAAGATCA 4230
PCoV_GX-P4L             AAACGTAATAGAGCTACAAGGGTTGAATGCACAACCATTGTAAATGGTGTAAGAAGATCA 4230
PCoV_GX-P5E             AAACGTAATAGAGCTACAAGGGTTGAATGCACAACCATTGTAAATGGTGTAAGAAGATCA 4230
PCoV_GX-P1E             AAACGTAATAGAGCTACAAGGGTTGAATGCACAACCATTGTAAATGGTGTAAGAAGATCA 4230
PCoV_GX-P5L             AAACGTAATAGAGCTACAAGGGTTGAATGCACAACCATTGTAAATGGTGTAAGAAGATCA 4230
PCoV-MP789              AAGCGTAATAGAGCAACAAGAGTTGAATGTACAACTATTGTTAATGGTGTTAGAAGATCC 4297
ZC45                    AAACGTAATAGAGCAACAAGAGTTGAATGTACAACTATTGTTAATGGTGTTAGAAGGTCC 4833
ZXC21                   AAACGTAATAGAGCAACAAGAGTTGAATGTACAACTATTGTTAATGGTGTTAGAAGGTCC 4739
RShSTT200               AAACGCAATAGGGCAACAAGAGTTGAATGCACAACTATTGTTAATGGTGTCAAAAGGTCC 4824
RShSTT182               AAACGCAATAGGGCAACAAGAGTTGAATGCACAACTATTGTTAATGGTGTCAAAAGGTCC 4824
RacCS203                AAACGTAATAGAGCAACAAGAGTTGAGTGTACAACTATTGTTAATGGTGTTAGAAGGTCT 4816
RaTG13                  AAACGTAATAGAGCAACAAGAGTTGAATGTACAACTATTGTTAATGGTGTTAGAAGGTCC 4842
INMI1                   AAACGTAATAGAGCAACAAGAGTCGAATGTACAACTATTGTTAATGGTGTTAGAAGGTCC 4845
CDC-CruiseA-12          AAACGTAATAGAGCAACAAGAGTCGAATGTACAACTATTGTTAATGGTGTTAGAAGGTCC 4845
WIV04                   AAACGTAATAGAGCAACAAGAGTCGAATGTACAACTATTGTTAATGGTGTTAGAAGGTCC 4845
Wuhan-Hu-1              AAACGTAATAGAGCAACAAGAGTCGAATGTACAACTATTGTTAATGGTGTTAGAAGGTCC 4845
RpYN06                  AAACGTAATAGAGCAACAAGAGTTGAATGTACAACTATTGTTAATGGTGTTAGAAGGTCC 4842
RmYN02                  AAACGTAATAGAGCAACAAGAGTTGAATGTACAACTATTGTTAATGGTGTTAGAAGGTCC 4839
HKU-39849               AAGCGCAATCGTGCCACACGCGTTGAGTGTACAACTATTGTTAATGGCATGAAGAGATCT 3612
Tor2/FP1-10912          AAGCGCAATCGTGCCACACGCGTTGAGTGTACAACTATTGTTAATGGCATGAAGAGATCT 3612
BM48-31                 AAGCGTAATCGTGCAACACGTGTTGAATGCACCACCATTGTCAACGGCATGAAGAAGTCC 3338
                        ** ** *** * ** *** * ** ** ** ** ** ***** ** **  * *  *  **
```

nsp4

Transmembrane domains. Double-membrane vesicles formation



Protein nsp4 (the RR pair 1)

```
BM48-31              FMKFRRAFGEYNHVVFANALLFLLSFTILCLTPAYTFLPGIYSLLYLYLTFYFTNDVSFL     298
PCoV_GX-P1E          FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVFYLYLTFYLTNDVSFL     337
PCoV_GX-P4L          FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVFYLYLTFYLTNDVSFL     337
PCoV_GX-P2V          FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVFYLYLTFYLTNDVSFL     337
PCoV_GX-P5L          FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVFYLYLTFYLTNDVSFL     337
PCoV_GX-P5E          FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVFYLYLTFYLTNDVSFL     337
ZC45                 FMRFRRAFGEYSHVVAFNTLLFFMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFL     360
ZXC21                FMRFRRAFGEYSHVVAFNTLLFFMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFL     360
RacCS203             FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVVYLYLTFYLTNDVSFL     360
PCoV-MP789           FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFL     360
RShSTT200            FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFL     360
RShSTT182            FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFL     360
INMI1                FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFL     360
CDC-CruiseA-12       FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFL     360
WIV04                FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFL     360
Wuhan-Hu-1           FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFL     360
RmYN02               FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFL     360
RaTG13               FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFL     360
RpYN06               FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFL     360
                     **:********.***  *:***::***:*****.*:****:**:.*******:*******
```

Gene nsp4 (the encoding of the RR pair 1)

```
BM48-31              TTTATGAAGTTTAGGCGTGCGTTTGGAGAGTACAACCATGTTGTCTTTGCTAATGCACTT 775
PCoV_GX-P5E          TTTATGAGGTTCAGGAGAGCTTTTGGCGAATACAGTCATGTAGTTGCCTTTAATACTCTC 891
PCoV_GX-P1E          TTTATGAGGTTCAGGAGAGCTTTTGGCGAATACAGTCATGTAGTTGCCTTTAATACTCTC 891
PCoV_GX-P4L          TTTATGAGGTTCAGGAGAGCTTTTGGCGAATACAGTCATGTAGTTGCCTTTAATACTCTC 891
PCoV_GX-P2V          TTTATGAGGTTCAGGAGAGCTTTTGGCGAATACAGTCATGTAGTTGCCTTTAATACTCTC 891
PCoV_GX-P5L          TTTATGAGGTTCAGGAGAGCTTTTGGCGAATACAGTCATGTAGTTGCCTTTAATACTCTC 891
ZC45                 TTCATGAGGTTTAGAAGAGCTTTTGGTGAATACAGTCATGTAGTTGCCTTTAACACTCTA 960
ZXC21                TTCATGAGGTTTAGAAGAGCTTTTGGTGAATACAGTCATGTAGTTGCCTTTAACACTCTA 960
PCoV-MP789           TTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGTCATGTAGTTGCCTTCAATACTCTG 960
RacCS203             TTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGTCATGTAGTTGCCTTTAATACTCTG 960
RaTG13               TTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGCCATGTAGTTGCCTTTAATACTTTA 960
RShSTT200            TTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGCCATGTAGTTGCCTTTAATACTTTA 960
RShSTT182            TTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGCCATGTAGTTGCCTTTAATACTTTA 960
RmYN02               TTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGTCATGTAGTTGCCTTTAATACTTTA 960
INMI1                TTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGTCATGTAGTTGCCTTTAATACTTTA 960
CDC-CruiseA-12       TTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGTCATGTAGTTGCCTTTAATACTTTA 960
WIV04                TTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGTCATGTAGTTGCCTTTAATACTTTA 960
Wuhan-Hu-1           TTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGTCATGTAGTTGCCTTTAATACTTTA 960
RpYN06               TTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGTCATGTAGTTGCCTTTAATACTTTA 960
                     ** **** *** **  * ** ***** ** ****  ***** **       **  *  *
```

Protein nsp4 (the RR pair 2)

```
BM48-31              AHLQWLAMFSPIVPFWITVTYVVCISIKHCHWFFSNYLKKRVVFNGVTFSTFEEAALCTF  358
PCoV_GX-P1E          AHVQWMVMFTPLVPFWITIVYVICISTKHCYWFFSNYLRRRVVFNGTSFSTFEEAALCTF  397
PCoV_GX-P4L          AHVQWMVMFTPLVPFWITIVYVICISTKHCYWFFSNYLRRRVVFNGTSFSTFEEAALCTF  397
PCoV_GX-P2V          AHVQWMVMFTPLVPFWITIVYVICISTKHCYWFFSNYLRRRVVFNGTSFSTFEEAALCTF  397
PCoV_GX-P5L          AHVQWMVMFTPLVPFWITIVYVICISTKHCYWFFSNYLRRRVVFNGTSFSTFEEAALCTF  397
PCoV_GX-P5E          AHVQWMVMFTPLVPFWITIVYVICISTKHCYWFFSNYLRRRVVFNGTSFSTFEEAALCTF  397
ZC45                 AHIQWMVMFTPLVPFWMTIVYVICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF  420
ZXC21                AHIQWMVMFTPLVPFWMTIVYVICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF  420
RacCS203             AHIQWMVMFTPLVPFWITIVYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF  420
PCoV-MP789           AHIQWMVMFTPLVPFWITIVYVICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF  420
RShSTT200            AHIQWMVMFTPLVPFWITTVYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF  420
RShSTT182            AHIQWMVMFTPLVPFWITTVYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF  420
INMI1                AHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF  420
CDC-CruiseA-12       AHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF  420
WIV04                AHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF  420
Wuhan-Hu-1           AHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF  420
RmYN02               AHIQWMVMFTPLVPFWITIVYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF  420
RaTG13               AHIQWMVMFTPLVPFWITIVYVICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF  420
RpYN06               AHIQWMVMFTPLVPFWITIVYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF  420
                     **:**:.**:*:****:* .*::*** ** :*******::******.:************
```

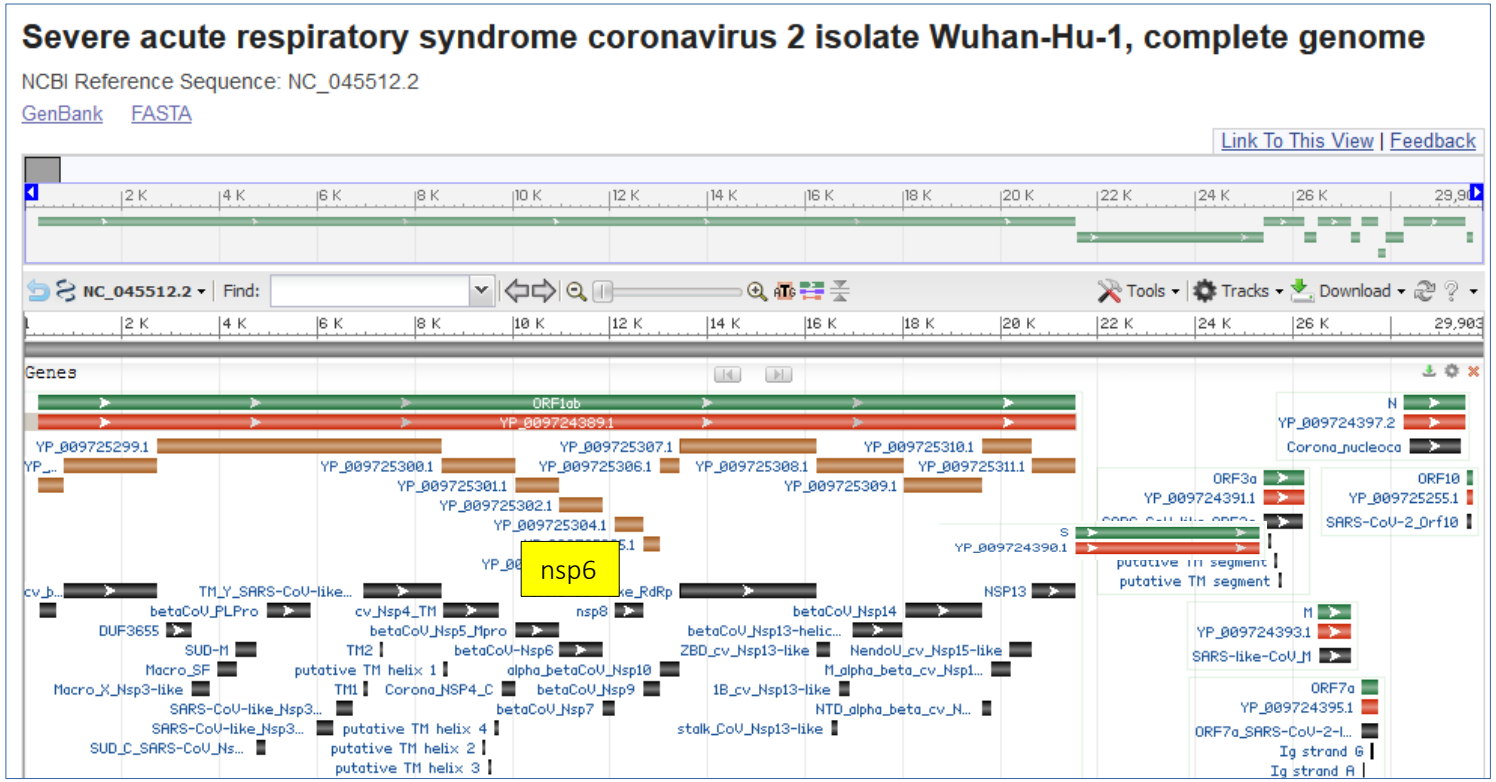
Gene nsp4 (the encoding of the RR pair 2)

```
BM48-31              TATGTTGTCTGTATTTCTATTAAGCATTGCCATTGGTTCTTTAGTAATTACCTCAAGAAG 1015
PCoV_GX-P5E          TATGTCATTTGTATATCTACTAAGCATTGTTACTGGTTCTTTAGTAATTACCTTAGACGT 1131
PCoV_GX-P1E          TATGTCATTTGTATATCTACTAAGCATTGTTACTGGTTCTTTAGTAATTACCTTAGACGT 1131
PCoV_GX-P4L          TATGTCATTTGTATATCTACTAAGCATTGTTACTGGTTCTTTAGTAATTACCTTAGACGT 1131
PCoV_GX-P2V          TATGTCATTTGTATATCTACTAAGCATTGTTACTGGTTCTTTAGTAATTACCTTAGACGT 1131
PCoV_GX-P5L          TATGTCATTTGTATATCTACTAAGCATTGTTACTGGTTCTTTAGTAATTACCTTAGACGT 1131
ZC45                 TATGTCATTTGCATTTCCACAAAGCATTTTTATTGGTTCTTTAGTAACTACCTAAAGAGA 1200
ZXC21                TATGTCATTTGCATTTCCACAAAGCATTTTTATTGGTTCTTTAGTAACTACCTAAAGAGA 1200
PCoV-MP789           TATGTCATTTGTATTTCCACAAAGCATTTTTATTGGTTCTTTAGCAACTACCTAAAGAGA 1200
RacCS203             TATATCATTTGTATTTCCACAAAGCACTTTTATTGGTTCTTTAGCAATTACCTAAAGAGA 1200
RaTG13               TATGTCATTTGTATTTCCACAAAGCATTTTTACTGGTTCTTTAGTAACTACCTAAAAAGG 1200
RShSTT200            TATATCATTTGTATTTCCACAAAGCATTTTTACTGGTTCTTCAGCAACTACCTAAAGAGA 1200
RShSTT182            TATATCATTTGTATTTCCACAAAGCATTTTTACTGGTTCTTCAGCAACTACCTAAAGAGA 1200
RmYN02               TATATCATTTGTATTTCCACAAAGCATTTTTATTGGTTCTTCAGCAACTACCTAAAGAGA 1200
INMI1                TATATCATTTGTATTTCCACAAAGCATTTCTATTGGTTCTTTAGTAATTACCTAAAGAGA 1200
CDC-CruiseA-12       TATATCATTTGTATTTCCACAAAGCATTTCTATTGGTTCTTTAGTAATTACCTAAAGAGA 1200
WIV04                TATATCATTTGTATTTCCACAAAGCATTTCTATTGGTTCTTTAGTAATTACCTAAAGAGA 1200
Wuhan-Hu-1           TATATCATTTGTATTTCCACAAAGCATTTCTATTGGTTCTTTAGTAATTACCTAAAGAGA 1200
RpYN06               TATATCATTTGTATTTCCACAAAGCATTTCTATTGGTTCTTTAGTAATTACCTAAAGAGA 1200
                     *** *  * ** ** ** *  ***** *   * ******** ** ** ***** *
```

```
BM48-31              AGAGTTGTTTTTAATGGAGTTACATTTAGCACTTTTGAGGAGGCTGCTCTGTGTACCTTT 1075
PCoV_GX-P5E          AGAGTTGTCTTTAATGGTACTTCCTTTAGCACTTTTGAAGAAGCAGCTTTGTGTACATTC 1191
PCoV_GX-P1E          AGAGTTGTCTTTAATGGTACTTCCTTTAGCACTTTTGAAGAAGCAGCTTTGTGTACATTC 1191
PCoV_GX-P4L          AGAGTTGTCTTTAATGGTACTTCCTTTAGCACTTTTGAAGAAGCAGCTTTGTGTACATTC 1191
PCoV_GX-P2V          AGAGTTGTCTTTAATGGTACTTCCTTTAGCACTTTTGAAGAAGCAGCTTTGTGTACATTC 1191
PCoV_GX-P5L          AGAGTTGTCTTTAATGGTACTTCCTTTAGCACTTTTGAAGAAGCAGCTTTGTGTACATTC 1191
ZC45                 CGTGTAGTCTTTAATGGTGTTTCCTTTAGTACATTTGAGGAGGCTGCATTATGTACCTTT 1260
ZXC21                CGTGTAGTCTTTAATGGTGTTTCCTTTAGTACATTTGAGGAGGCTGCATTATGTACCTTT 1260
PCoV-MP789           CGTGTTGTCTTTAATGGTGTTTCCTTTAGTACATTTGAAGAGGCTGCTTTATGCACCTTT 1260
RacCS203             CGTGTAGTCTTTAATGGTGTTTCCTTTAGTACTTTTGAAGAAGCTGCTCTATGCACCTTT 1260
RaTG13               CGTGTAGTCTTTAATGGTGTTTCCTTTAGTACTTTTGAAGAAGCTGCGTTATGCACTTTC 1260
RShSTT200            CGTGTAGTCTTTAATGGTGTTTCCTTTAGTACTTTTGAAGAAGCTGCGCTGTGCACCTTT 1260
RShSTT182            CGTGTAGTCTTTAATGGTGTTTCCTTTAGTACTTTTGAAGAAGCTGCGCTGTGCACCTTT 1260
RmYN02               CGTGTAGTCTTTAATGGTGTTTCCTTTAGTACTTTTGAAGAAGCTGCGCTTTGCACCTTT 1260
INMI1                CGTGTAGTCTTTAATGGTGTTTCCTTTAGTACTTTTGAAGAAGCTGCGCTGTGCACCTTT 1260
CDC-CruiseA-12       CGTGTAGTCTTTAATGGTGTTTCCTTTAGTACTTTTGAAGAAGCTGCGCTGTGCACCTTT 1260
WIV04                CGTGTAGTCTTTAATGGTGTTTCCTTTAGTACTTTTGAAGAAGCTGCGCTGTGCACCTTT 1260
Wuhan-Hu-1           CGTGTAGTCTTTAATGGTGTTTCCTTTAGTACTTTTGAAGAAGCTGCGCTGTGCACCTTT 1260
RpYN06               CGTGTAGTCTTTAATGGTGTTTCCTTTAGTACTTTTGAAGAAGCTGCGCTGTGCACCTTT 1260
                      * ** ** ********   * * ***** ** ***** ** ** **  * ** ** **
```

nsp6

Transmembrane domains. Double-membrane vesicles formation



Protein nsp6 (the RR pair)

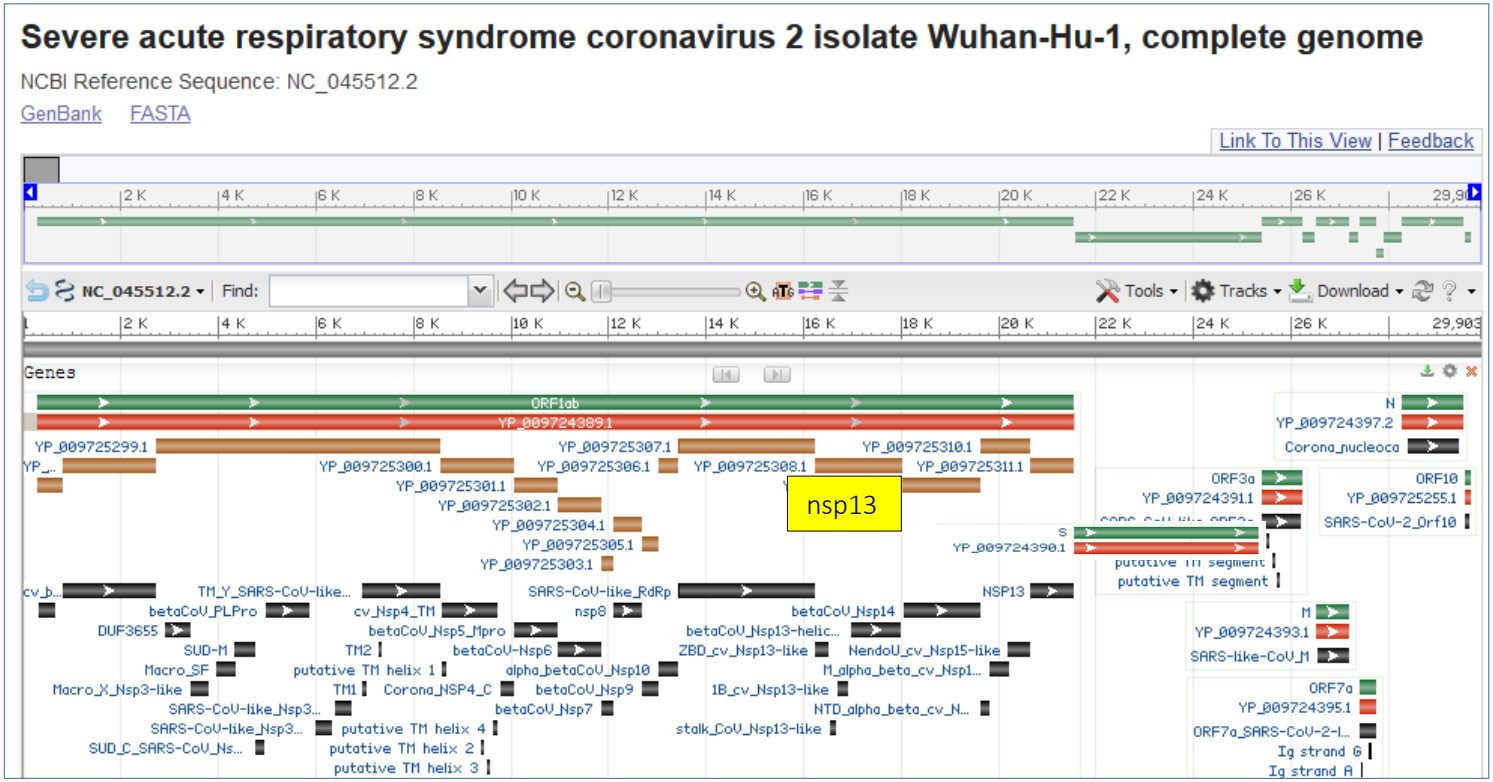
```
BtKY72                ILLILMTARTVYDDATRRVWTFMNVITLVYKVYYGNVLDQAIAMWALVISVTSNYSGVVT     173
ZC45                  VLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT     180
ZXC21                 VLLILMTARTVYDDSARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT     180
RmYN02                VLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT     180
RShSTT200             VLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT     180
RShSTT182             VLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT     180
PCoV-MP789            VLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT     180
RacCS203              VLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT     180
WIV04                 VLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT     180
Wuhan-Hu-1            VLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT     180
INMI1                 VLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT     180
CDC-CruiseA-12        VLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT     180
RpYN06                VLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT     180
RaTG13                VLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT     180
                      :*************.:*****:***:**********.*****:****:************
```

Gene nsp6 (the encoding RR pair)

```
BtKY72                ATACTGCTCATCCTCATGACAGCTAGAACTGTCTATGACGATGCCACTAGACGAGTTTGG 399
ZC45                  GTGTTATTAATCCTCATGACAGCAAGAACCGTATATGATGATGGTGCTAGAAGAGTTTGG 420
ZXC21                 GTGTTATTAATCCTCATGACAGCAAGAACTGTATATGATGATAGTGCTAGAAGAGTTTGG 420
PCoV-MP789            GTGTTATTAATCCTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGAAGAGTTTGG 420
RaTG13                GTGTTACTAATCCTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGGAGGGTGTGG 420
RpYN06                GTCTTACTAATCCTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGGAGAGTGTGG 420
RmYN02                GTCTTACTAATCCTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGGAGAGTGTGG 420
WIV04                 GTGTTACTAATCCTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGGAGAGTGTGG 420
Wuhan-Hu-1            GTGTTACTAATCCTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGGAGAGTGTGG 420
INMI1                 GTGTTACTAATCCTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGGAGAGTGTGG 420
CDC-CruiseA-12        GTGTTACTAATCCTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGGAGAGTGTGG 420
RacCS203              GTGTTACTAATCCTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGGAGAGTGTGG 420
RShSTT200             GTGTTACTAATCCTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGGAGAGTGTGG 420
RShSTT182             GTGTTACTAATCCTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGGAGAGTGTGG 420
                       *  *  * ***** ******** ***** ** ***** ***    *****  * ** ***
```

nsp13

Zinc-binding domain. Helicase, RNA 5 triphosphatase



Protein nsp13 (the RR pair 1)

```
BtKY72                AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
BM48                  AVGACVLCNSQTSLRCGSCIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
PCoV_GX-P2V           AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNATGCDVTDV      60
PCoV_GX-P1E           AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNATGCDVTDV      60
PCoV_GX-P4L           AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNATGCDVTDV      60
PCoV_GX-P5L           AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNATGCDVTDV      60
PCoV_GX-P5E           AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNATGCDVTDV      60
ZXC21                 AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNALGCDVTDV      60
ZC45                  AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
RacCS203              AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
RmYN02                AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
RpYN06                AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
HKU                   AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
Tor2/FP1-10912        AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
MP789                 AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
INMI1                 AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
CDC-CruiseA-12        AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
WIV04                 AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
Wuhan-Hu-1            AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
RShSTT200             AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
RShSTT182             AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
RaTG13                AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDV      60
                      ****************:**********************************  *******
```

Gene nsp13 (the encoding of the RR pair1)

```
PCoV_GX-P2V           CGTAGACCCTTCTTATGCTGTAAATGTTGTTATGACCATGTCATATCAACATCTCATAAA 120
PCoV_GX-P5E           CGTAGACCCTTCTTATGCTGTAAATGTTGTTATGACCATGTCATATCAACATCTCATAAA 120
PCoV_GX-P4L           CGTAGACCCTTCTTATGCTGTAAATGTTGTTATGACCATGTCATATCAACATCTCATAAA 120
PCoV_GX-P5L           CGTAGACCCTTCTTATGCTGTAAATGTTGTTATGACCATGTCATATCAACATCTCATAAA 120
PCoV_GX-P1E           CGTAGACCCTTCTTATGCTGTAAATGTTGTTATGACCATGTCATATCAACATCTCATAAA 120
MP789                 CGCAGACCATTCTTATGTTGTAAATGCTGTTATGACCATGTCATATCAACATCACATAAA 120
RpYN06                CGTAGACCATTCTTATGTTGTAAATGCTGTTACGACCACGTCATATCAACATCACATAAA 120
RmYN02                CGTAGACCATTCTTATGCTGTAAATGCTGTTACGACCACGTCATATCAACATCACATAAA 120
RacCS203              CGTAGACCATTCTTATGTTGTAAATGCTGTTATGACCATGTCATATCAACATCACATAAA 120
RShSTT200             CGTAGACCATTTTTATGTTGTAAATGCTGTTACGACCATGTCATATCAACATCACATAAA 120
RShSTT182             CGTAGACCATTTTTATGTTGTAAATGCTGTTACGACCATGTCATATCAACATCACATAAA 120
INMI1                 CGTAGACCATTCTTATGTTGTAAATGCTGTTACGACCATGTCATATCAACATCACATAAA 120
CDC-CruiseA-12        CGTAGACCATTCTTATGTTGTAAATGCTGTTACGACCATGTCATATCAACATCACATAAA 120
WIV04                 CGTAGACCATTCTTATGTTGTAAATGCTGTTACGACCATGTCATATCAACATCACATAAA 120
Wuhan-Hu-1            CGTAGACCATTCTTATGTTGTAAATGCTGTTACGACCATGTCATATCAACATCACATAAA 120
RaTG13                CGTAGACCATTCTTATGCTGTAAATGCTGTTACGACCACGTCATATCTACATCACATAAA 120
HKU                   AGGAGACCATTCCTATGTTGCAAGTGCTGCTATGACCATGTCATTTCAACATCACACAAA 120
Tor2/FP1-10912        AGGAGACCATTCCTATGTTGCAAGTGCTGCTATGACCATGTCATTTCAACATCACACAAA 120
ZXC21                 AGGAGACCTTTCCTTTGTTGTAAGTGCTGCTATGATCATGTCATTTCAACATCACATAAG 120
ZC45                  AGGAGACCTTTCCTTTGTTGCAAGTGCTGCTATGATCACGTCATCTCAACATCACATAAA 120
BM48                  AGACGACCATTCCTGTGTTGCAAGTGCTGCTATGACCATGTCATTTCGACTTCGCATAAA 120
BtKY72                AGGCGACCATTCCTGTGTTGCAAGTGCTGCTATGATCATGTTATTTCTACTTCACATAAA 120
                       *  ***** **  * ** ** ** ** ** ** ** ** ** ** ** ** ** ** **
```

Protein nsp13 (the RR pair 2)

```
BtKY72                YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSTLVYDNKLKAHKEKSSQCFKIFYKGVI     480
BM48                  YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLRAHKGKSSQCFKMFYKGVI     480
PCoV_GX-P2V           YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLRAHKDKSQQCFKMFYKGVI     480
PCoV_GX-P1E           YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLRAHKDKSQQCFKMFYKGVI     480
PCoV_GX-P4L           YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLRAHKDKSQQCFKMFYKGVI     480
PCoV_GX-P5L           YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLRAHKDKSQQCFKMFYKGVI     480
PCoV_GX-P5E           YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLRAHKDKSQQCFKMFYKGVI     480
ZXC21                 YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKEKSAQCFKMFYKGVI     480
ZC45                  YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKEKSAQCFKMFYKGVI     480
RacCS203              YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVI     480
RmYN02                YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVI     480
RpYN06                YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVI     480
HKU                   YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVI     480
Tor2/FP1-10912        YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVI     480
MP789                 YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKEKSAQCFKMFYKGVI     480
INMI1                 YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVI     480
CDC-CruiseA-12        YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVI     480
WIV04                 YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVI     480
Wuhan-Hu-1            YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVI     480
RShSTT200             YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVI     480
RShSTT182             YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVI     480
RaTG13                YFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVI     480
                      ************************:*******:*** ** ****:******
```

Gene nsp13 (the encoding of the RR pair 2)

```
PCoV_GX-P2V           TGTAGAAGATGTCCTGCTGAAATAGTTGACACTGTAAGTGCTCTAGTTTATGATAATAAA1380
PCoV_GX-P5E           TGTAGAAGATGTCCTGCTGAAATAGTTGACACTGTAAGTGCTCTAGTTTATGATAATAAA1380
PCoV_GX-P4L           TGTAGAAGATGTCCTGCTGAAATAGTTGACACTGTAAGTGCTCTAGTTTATGATAATAAA1380
PCoV_GX-P5L           TGTAGAAGATGTCCTGCTGAAATAGTTGACACTGTAAGTGCTCTAGTTTATGATAATAAA1380
PCoV_GX-P1E           TGTAGAAGATGTCCTGCTGAAATAGTTGACACTGTAAGTGCTCTAGTTTATGATAATAAA1380
MP789                 TGTCGTCGCTGTCCTGCTGAAATAGTCGACACTGTAAGTGCTCTAGTTTATGACAATAAG1380
RpYN06                TGTCGGCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTAGTTTATGATAATAAG1380
RmYN02                TGTCGGCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTGGTTTATGATAATAAG1380
RacCS203              TGTCGGCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTGGTTTACGATAATAAG1380
RShSTT200             TGTCGGCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTGGTTTATGATAATAAG1380
RShSTT182             TGTCGGCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTGGTTTATGATAATAAG1380
INMI1                 TGTCGGCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTGGTTTATGATAATAAG1380
CDC-CruiseA-12        TGTCGGCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTGGTTTATGATAATAAG1380
WIV04                 TGTCGGCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTGGTTTATGATAATAAG1380
Wuhan-Hu-1            TGTCGGCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTGGTTTATGATAATAAG1380
RaTG13                TGTCGGCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTGGTTTATGATAATAAG1380
HKU                   TGTCGCCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTAGTTTATGACAATAAG1380
Tor2/FP1-10912        TGTCGCCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTAGTTTATGACAATAAG1380
ZXC21                 TGTCGTCGTTGTCCTGCTGAAATTGTTGACACAGTGAGTGCTTTAGTTTATGATAATAAG1380
ZC45                  TGTCGTCGTTGTCCCGCTGAAATTGTTGACACAGTGAGTGCTTTAGTTTATGATAATAAG1380
BM48                  TGTCGCCGATGTCCTGCTGAAATAGTCGACACTGTCAGTGCTTTAGTTTATGATAATAAA1380
BtKY72                TGCCGTAGGTGTCCAGCTGAAATAGTTGACACTGTTAGTACTTTGGTTTATGATAATAAG1380
                      **  *  * ***** ******** ** ***** ** *** ** * ***** ** *****
```

Protein nsp13 (the RR pair 3)

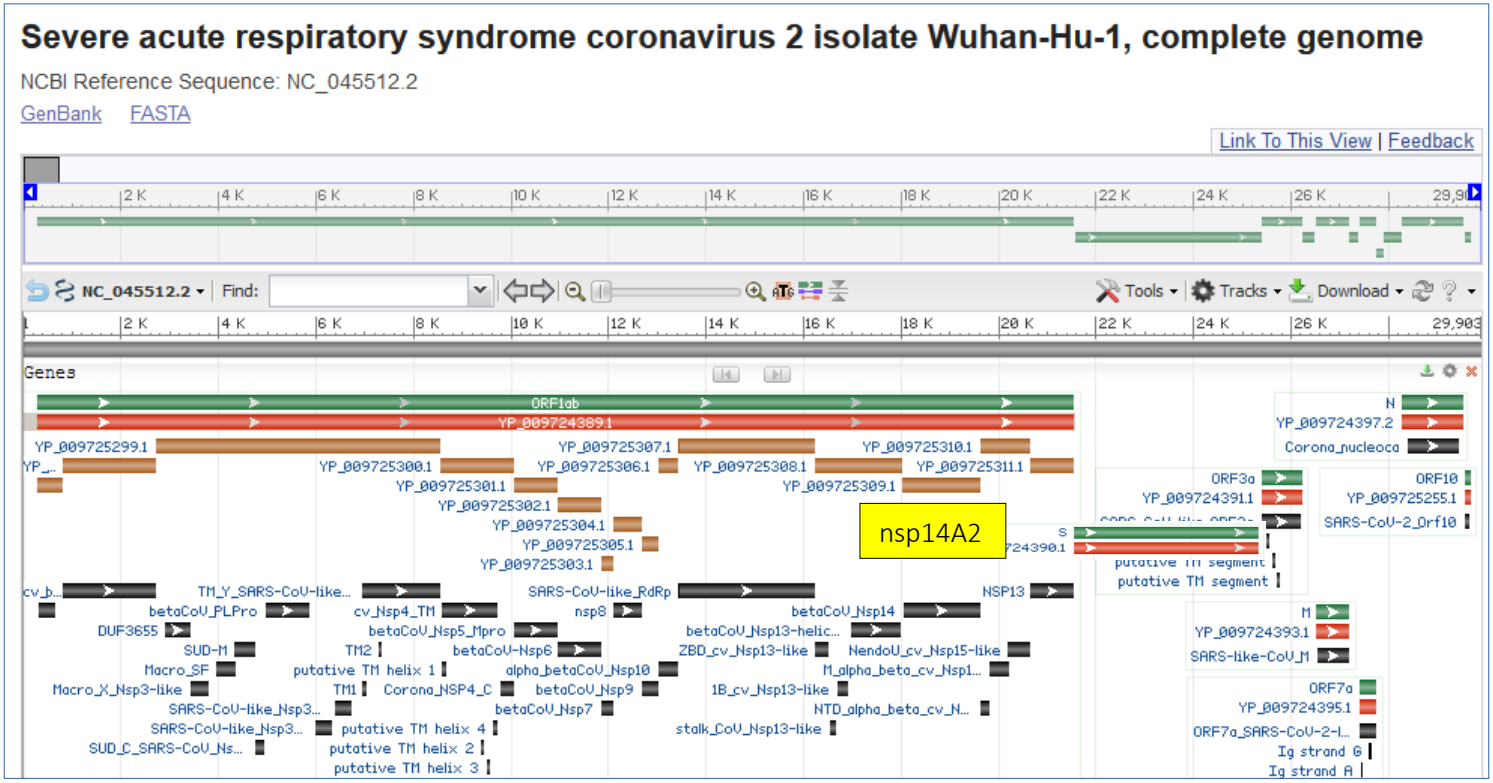
```
BtKY72                YDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFASLEVPRKSVA--      598
BM48                  YDYVIFAQTTETAHSCNVNRFNVAITRAKVGILCIMSDKDLYDKLQFTSLEVPRRSVA--      598
PCoV_GX-P2V           YDYVIFTQTTETAHSCNVNRFNVAITRAKIGILCIMSDRDLYDKLQFTSLEVPRRNVATL      600
PCoV_GX-P1E           YDYVIFTQTTETAHSCNVNRFNVAITRAKIGILCIMSDRDLYDKLQFTSLEVPRRNVATL      600
PCoV_GX-P4L           YDYVIFTQTTETAHSCNVNRFNVAITRAKIGILCIMSDRDLYDKLQFTSLEVPRRNVATL      600
PCoV_GX-P5L           YDYVIFTQTTETAHSCNVNRFNVAITRAKIGILCIMSDRDLYDKLQFTSLEVPRRNVATL      600
PCoV_GX-P5E           YDYVIFTQTTETAHSCNVNRFNVAITRAKIGILCIMSDRDLYDKLQFTSLEVPRRNVATL      600
ZXC21                 YDYVIFTQTTETAHSCNVNRFNVAITRAKIGILCIMSDRDLYDKLQFTSLEVPRRNVATL      600
ZC45                  YDYVIFTQTTETAHSCNVNRFNVAITRAKIGILCIMSDRDLYDKLQFMSLEVPRRNVATL      600
RacCS203              YDYVIFTQTTETSHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTSLEIPRRNVVTL      600
RmYN02                YDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTSLVIPRRNVATL      600
RpYN06                YDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTSLEIPRRNVATL      600
HKU                   YDYVIFTQTTETAHSCNVNRFNVAITRAKIGILCIMSDRDLYDKLQFTSLEIPRRNVATL      600
Tor2/FP1-10912        YDYVIFTQTTETAHSCNVNRFNVAITRAKIGILCIMSDRDLYDKLQFTSLEIPRRNVATL      600
MP789                 YDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTSLEIPRRNVATL      600
INMI1                 YDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTSLEIPRRNVATL      600
CDC-CruiseA-12        YDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTSLEIPRRNVATL      600
WIV04                 YDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTSLEIPRRNVATL      600
Wuhan-Hu-1            YDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTSLEIPRRNVATL      600
RShSTT200             YDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTSLEIPRRNVATL      600
RShSTT182             YDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTSLEIPRRNVATL      600
RaTG13                YDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTSLEIPRRNVATL      600
                      ******:*****:****************:********:******** ** :**:.*.
```

Gene nsp13 (the encoding of the RR pair 3)

```
PCoV_GX-P2V           CTTTATGACAAATTACAATTTACAAGCCTTGAAGTTCCACGTCGAAACGTGGCAACCTTA1800
PCoV_GX-P5E           CTTTATGACAAATTACAATTTACAAGCCTTGAAGTTCCACGTCGAAACGTGGCAACCTTA1800
PCoV_GX-P4L           CTTTATGACAAATTACAATTTACAAGCCTTGAAGTTCCACGTCGAAACGTGGCAACCTTA1800
PCoV_GX-P5L           CTTTATGACAAATTACAATTTACAAGCCTTGAAGTTCCACGTCGAAACGTGGCAACCTTA1800
PCoV_GX-P1E           CTTTATGACAAATTACAATTTACAAGCCTTGAAGTTCCACGTCGAAACGTGGCAACCTTA1800
MP789                 CTTTATGACAAGTTGCAATTTACAAGTCTTGAAATTCCACGTAGAAATGTGGCAACTTTA1800
RpYN06                CTTTATGACAAGTTGCAATTTACAAGTCTTGAAATTCCACGTAGGAATGTGGCAACTTTA1800
RmYN02                CTTTATGACAAGTTGCAATTTACAAGTCTTGTGATTCCACGTAGGAATGTGGCAACTTTA1800
RacCS203              CTTTATGACAAGTTGCAATTTACAAGTCTTGAGATTCCACGTAGGAATGTGGTAACTTTA1800
RShSTT200             CTTTATGACAAGTTGCAATTTACAAGTCTTGAGATTCCACGTAGGAATGTGGCAACTTTA1800
RShSTT182             CTTTATGACAAGTTGCAATTTACAAGTCTTGAGATTCCACGTAGGAATGTGGCAACTTTA1800
INMI1                 CTTTATGACAAGTTGCAATTTACAAGTCTTGAAATTCCACGTAGGAATGTGGCAACTTTA1800
CDC-CruiseA-12        CTTTATGACAAGTTGCAATTTACAAGTCTTGAAATTCCACGTAGGAATGTGGCAACTTTA1800
WIV04                 CTTTATGACAAGTTGCAATTTACAAGTCTTGAAATTCCACGTAGGAATGTGGCAACTTTA1800
Wuhan-Hu-1            CTTTATGACAAGTTGCAATTTACAAGTCTTGAAATTCCACGTAGGAATGTGGCAACTTTA1800
RaTG13                CTTTATGACAAGTTGCAATTTACAAGTCTTGAGATTCCACGTAGGAATGTGGCAACTTTA1800
HKU                   CTTTATGACAAACTGCAATTTACAAGTCTAGAAATACCACGTCGCAATGTGGCTACATTA1800
Tor2/FP1-10912        CTTTATGACAAACTGCAATTTACAAGTCTAGAAATACCACGTCGCAATGTGGCTACATTA1800
ZXC21                 CTTTATGACAAGCTGCAATTTACGAGTCTAGAAGTACCGCGTCGTAATGTGGCTACTTTA1800
ZC45                  CTTTATGACAAGCTTCAATTTATGAGTCTAGAAGTACCGCGTCGAAATGTGGCTACTTTA1800
BM48                  CTCTATGATAAATTACAATTTACTAGTCTGGAAGTCCCACGTAGAAGTGTGGC-------1793
BtKY72                CTTTATGACAAACTCCAATTTGCTAGTCTAGAAGTCCCACGTAAAAGTGTGGC-------1793
                      ** ***** **  * ******   ** ** *   * ** ***   *  ****
```

nsp14A2

Guanosine N7- methyltransferase. 3 to 5 exoribonuclease, proofreading



Protein:nsp14A2 (the RR pair 1)

```
PCoV_GX-P2V             AENVTGLFKDCSKVITGLHPTQAPTYLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
PCoV_GX-P5E             AENVTGLFKDCSKVITGLHPTQAPTYLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
PCoV_GX-P1E             AENVTGLFKDCSKVITGLHPTQAPTYLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
PCoV_GX-P4L             AENVTGLFKDCSKVITGLHPTQAPTYLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
PCoV_GX-P5L             AENVTGLFKDCSKVITGLHPTQAPTYLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
PCoV-MP789              AENVTGLFKDCSKVINGLHPTQALTHLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
RShSTT200               AENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
RShSTT182               AENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
RacCS203                AENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIPKDLTYRRLISMMGF      60
RaTG13                  AENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
RmYN02                  AENVTGLFKDCSKVITGLHPTQAPTHISVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
INMI1                   AENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
CDC-CruiseA-12          AENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
WIV04                   AENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
Wuhan-Hu-1              AENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
RpYN06                  AENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
BtKY72                  AENVTGLFKDCSKIINGLHPTQSPTYLSVDTKFKTEGLCVDIPSIPKDMTYRRLISMMGF      60
BM48-31                 -ENVTGLFKDCSKLITGLHPTQAPTYLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      59
Tor2_FP1-10912          AENVTGLFKDCSKIITGLHPTQAPTHLSVDIKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
HKU-39849               AENVTGLFKDCSKIITGLHPTQAPTHLSVDIKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
ZXC21                   AENVTGLFKDCSKIITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
ZC45                    AENVTGLFKDCSKIITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGF      60
                         ************:*.******: *::*** ************.****:**********
```

Gene nsp14A2 (the encoding of the RR pair 1)

```
PCoV_GX-P2V             GACATACCAGGAATACCAAAAGACATGACCTATAGGAGACTCATCTCTATGATGGGTTTC 180
PCoV_GX-P5E             GACATACCAGGAATACCAAAAGACATGACCTATAGGAGACTCATCTCTATGATGGGTTTC 180
PCoV_GX-P1E             GACATACCAGGAATACCAAAAGACATGACCTATAGGAGACTCATCTCTATGATGGGTTTC 180
PCoV_GX-P4L             GACATACCAGGAATACCAAAAGACATGACCTATAGGAGACTCATCTCTATGATGGGTTTC 180
PCoV_GX-P5L             GACATACCAGGAATACCAAAAGACATGACCTATAGGAGACTCATCTCTATGATGGGTTTC 180
PCoV-MP789              GACATACCAGGTATACCCAAGGACATGACCTATAGGAGACTCATTTCCATGATGGGTTTC 180
RShSTT200               GACATACCTGGCATACCTAAGGACATGACTTATAGAAGACTCATCTCTATGATGGGTTTC 180
RShSTT182               GACATACCTGGCATACCTAAGGACATGACTTATAGAAGACTCATCTCTATGATGGGTTTC 180
RacCS203                GACATACCTGGCATACCTAAGGACTTGACCTATAGAAGACTCATCTCTATGATGGGTTTC 180
RpYN06                  GACATACCTGGCATACCCAAGGACATGACCTATAGAAGACTCATCTCTATGATGGGTTTT 180
RaTG13                  GACATACCTGGTATACCTAAGGACATGACCTATAGAAGACTCATCTCTATGATGGGTTTC 180
INMI1                   GACATACCTGGCATACCTAAGGACATGACCTATAGAAGACTCATCTCTATGATGGGTTTT 180
CDC-CruiseA-12          GACATACCTGGCATACCTAAGGACATGACCTATAGAAGACTCATCTCTATGATGGGTTTT 180
WIV04                   GACATACCTGGCATACCTAAGGACATGACCTATAGAAGACTCATCTCTATGATGGGTTTT 180
Wuhan-Hu-1              GACATACCTGGCATACCTAAGGACATGACCTATAGAAGACTCATCTCTATGATGGGTTTT 180
RmYN02                  GACATACCTGGCATACCTAAGGACATGACCTATAGAAGACTCATCTCCATGATGGGTTTC 180
BtKY72                  GACATACCTAGCATACCTAAAGACATGACTTATCGTAGACTCATCTCTATGATGGGCTTC 180
BM48-31                 GACATACCAGGAATACCAAAGGACATGACCTATCGTAGGCTCATCTCTATGATGGGTTTT 177
Tor2/FP1-10912          GACATACCAGGCATACCAAAGGACATGACCTACCGTAGACTCATCTCTATGATGGGTTTC 180
HKU-39849               GACATACCAGGCATACCAAAGGACATGACCTACCGTAGACTCATCTCTATGATGGGTTTC 180
ZXC21                   GACATACCAGGAATACCAAAAGACATGACCTATCGTAGACTCATCTCTATGATGGGTTTT 180
ZC45                    GACATACCAGGAATACCAAAGGACATGACCTATCGTAGACTCATCTCTATGATGGGCTTT 180
                        ********  * ***** ** *** **** **  * ** ***** ** ******** **
```

Protein:nsp14A2 (the RR pair 2)

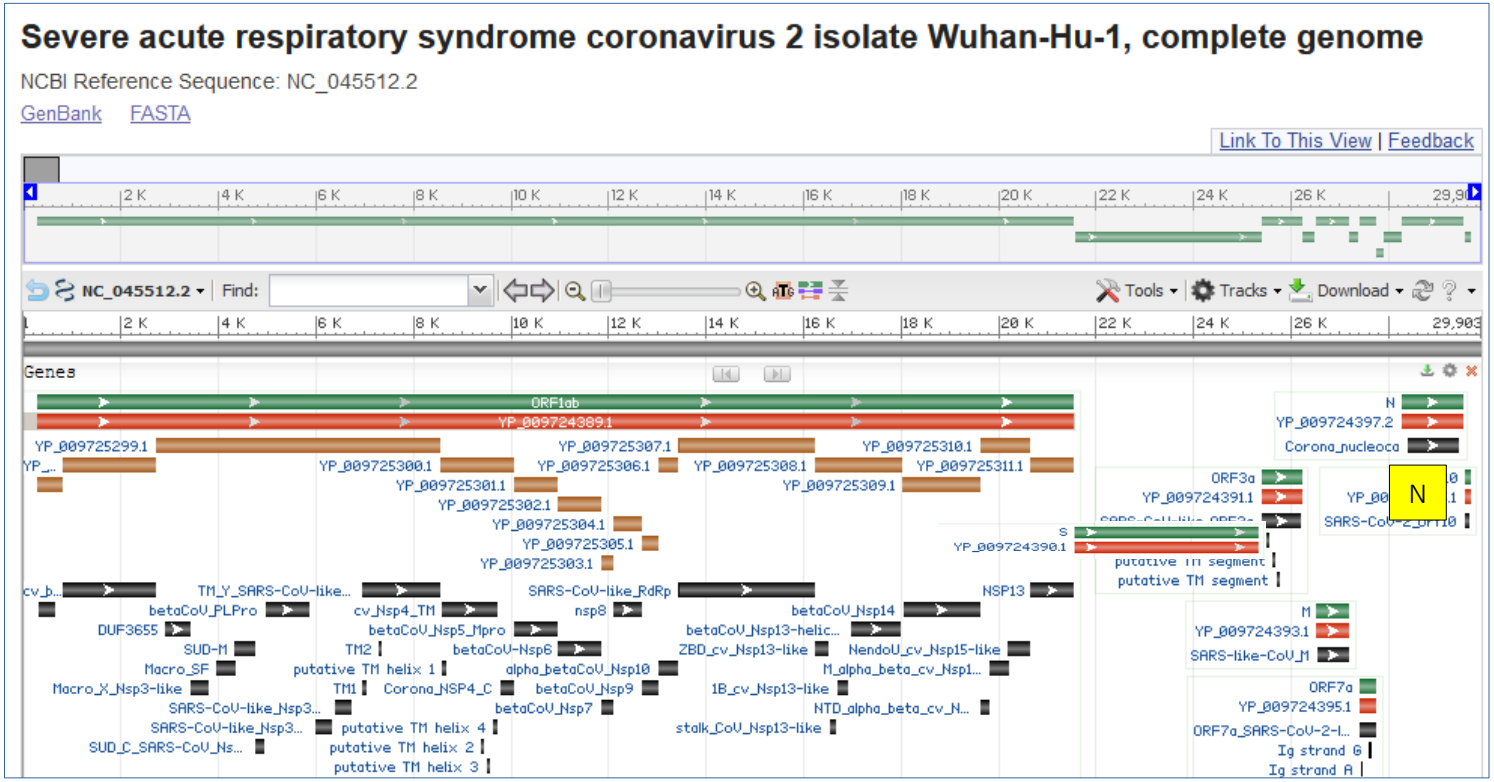
```
PCoV_GX-P2V           VVFVLWAHGXELTSMKYFVKIGPERTCCLCDKRATCFCTASDTYACWHHSVGFDYVYNPF     240
PCoV_GX-P5E           VVFVLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFCTASDTYACWHHSVGFDYVYNPF     240
PCoV_GX-P1E           VVFVLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFCTASDTYACWHHSVGFDYVYNPF     240
PCoV_GX-P4L           VVFVLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFCTASDTYACWHHSVGFDYVYNPF     240
PCoV_GX-P5L           VVFVLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFCTASDTYACWHHSVGFDYVYNPF     240
PCoV-MP789            VVFVLWAHGFELTSMKYFVKIGPERTCCLCDRRATCFSTASDTYACWHHSIGFDYVYNPF     240
RShSTT200             VVFVLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFSTASDTYACWHHSIGFDYVYNPF     240
RShSTT182             VVFVLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFSTASDTYACWHHSIGFDYVYNPF     240
RacCS203              VVFVLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFSTASDTYACWHHSIGFDYVYNPF     240
RaTG13                VVFVLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFSTASDTYACWHHSIGFDYVYNPF     240
RmYN02                VVFVLWAHGFELTSMKYFVKIGPERTCCLCDRRATCFSTASDTYACWHHSIGFDYVYNPF     240
INMI1                 VVFVLWAHGFELTSMKYFVKIGPERTCCLCDRRATCFSTASDTYACWHHSIGFDYVYNPF     240
CDC-CruiseA-12        VVFVLWAHGFELTSMKYFVKIGPERTCCLCDRRATCFSTASDTYACWHHSIGFDYVYNPF     240
WIV04                 VVFVLWAHGFELTSMKYFVKIGPERTCCLCDRRATCFSTASDTYACWHHSIGFDYVYNPF     240
Wuhan-Hu-1            VVFVLWAHGFELTSMKYFVKIGPERTCCLCDRRATCFSTASDTYACWHHSIGFDYVYNPF     240
RpYN06                VVFVLWAHGFELTSMKYFVKIGPERTCCLCDRRATCFSTASDTYACWHHSIGFDYVYNPF     240
BtKY72                VVFVLWAHGFELTSMKYFVKIGPERTCCLCDRRATCFSTSADTYACWHHSVGFDYVYNPF     240
BM48-31               VVFVLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFSTSSDTYACWHHSVGFDYVYNPF     239
Tor2_FP1-10912        VVFVLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFSTSSDTYACWNHSVGFDYVYNPF     240
HKU-39849             VVFVLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFSTSSDTYACWNHSVGFDYVYNPF     240
ZXC21                 VVFVLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFSTSSDTYACWNHSVGFDYVYNPF     240
ZC45                  VVFVLWAHGFELTSMKYFVKIGPERTCCLCDKRATCFSTSSDTYACWNHSVGFDYVYNPF     240
                      ********* ********************:*****.*::******:**:*********
```

Gene nsp14A2 (the encoding of the RR pair 2)

```
PCoV_GX-P2V           ATAGGACCTGAACGCACTTGTTGCTTATGTGACAAACGTGCTACCTGTTTTTGCACAGCA 660
PCoV_GX-P5E           ATAGGACCTGAACGCACTTGTTGCTTATGTGACAAACGTGCTACCTGTTTTTGCACAGCA 660
PCoV_GX-P1E           ATAGGACCTGAACGCACTTGTTGCTTATGTGACAAACGTGCTACCTGTTTTTGCACAGCA 660
PCoV_GX-P4L           ATAGGACCTGAACGCACTTGTTGCTTATGTGACAAACGTGCTACCTGTTTTTGCACAGCA 660
PCoV_GX-P5L           ATAGGACCTGAACGCACTTGTTGCTTATGTGACAAACGTGCTACCTGTTTTTGCACAGCA 660
PCoV-MP789            ATAGGTCCTGAACGCACTTGCTGTCTCTGTGACAGACGTGCTACCTGTTTTTCCACAGCT 660
RShSTT200             ATAGGACCTGAGCGCACTTGTTGTCTATGTGATAAACGTGCCACATGTTTTTCCACTGCT 660
RShSTT182             ATAGGACCTGAGCGCACTTGTTGTCTATGTGATAAACGTGCCACATGTTTTTCCACTGCT 660
RacCS203              ATAGGACCTGAGCGCACTTGTTGTCTATGTGATAAACGTGCCACATGTTTTTCCACTGCT 660
RpYN06                ATAGGACCTGAGCGCACTTGTTGTCTATGTGATAGACGTGCCACATGCTTTTCCACTGCT 660
RaTG13                ATAGGACCTGAGCGCACTTGTTGTCTATGTGATAAACGTGCCACATGCTTTTCCACTGCT 660
INMI1                 ATAGGACCTGAGCGCACCTGTTGTCTATGTGATAGACGTGCCACATGCTTTTCCACTGCT 660
CDC-CruiseA-12        ATAGGACCTGAGCGCACCTGTTGTCTATGTGATAGACGTGCCACATGCTTTTCCACTGCT 660
WIV04                 ATAGGACCTGAGCGCACCTGTTGTCTATGTGATAGACGTGCCACATGCTTTTCCACTGCT 660
Wuhan-Hu-1            ATAGGACCTGAGCGCACCTGTTGTCTATGTGATAGACGTGCCACATGCTTTTCCACTGCT 660
RmYN02                ATAGGACCTGAGCGCACTTGTTGTCTATGTGATAGACGTGCCACATGCTTTTCCACTGCT 660
BtKY72                ATAGGGCCTGAACGCACTTGCTGTCTTTGTGACAGACGTGCTACTTGCTTTTCTACTTCA 660
BM48-31               ATTGGACCAGAACGGACGTGTTGTCTGTGTGACAAGCGCGCAACTTGCTTTTCAACTTCA 657
Tor2/FP1-10912        ATTGGACCTGAAAGAACGTGTTGTCTGTGTGACAAACGTGCAACTTGCTTTTCTACTTCA 660
HKU-39849             ATTGGACCTGAAAGAACGTGTTGTCTGTGTGACAAACGTGCAACTTGCTTTTCTACTTCA 660
ZXC21                 ATCGGACCAGAAAGAACGTGTTGTCTGTGTGACAAACGCGCGACTTGCTTCTCTACTTCA 660
ZC45                  ATCGGACCAGAAAGAACGTGTTGTCTGTGTGACAAACGCGCGACTTGCTTCTCTACTTCA 660
                      ** ** ** **  * ** ** **  * ***** *  ** ** ** ** ** *  **  *
```

Orf9 N. Nucleocapsid

Packages the positive strand viral genome RNA. Virion assembly



Protein N orf9 Nucleoprotein (the RR pair 1)

```
HKU-39849               MSDNGPQSNQRSAPRITFGGPTDSTDNNQNGGRNGARPKQRRPQGLPNNTASWFTALTQH   60
Tor2/FP1-10912          MSDNGPQSNQRSAPRITFGGPTDSTDNNQNGGRNGARPKQRRPQGLPNNTASWFTALTQH   60
PcoV_GX-P5E             MSDNGPQ-N--RAPRITFGGPSDSTDNNQNGDRSGARPKQRRPQGLPNSTASWFTALTQH   57
PcoV_GX-P2V             MSDNGPQ-X--RAPRITFGGPSDSTDNNQNGDRSGARPKQRRPQGLPNNTASWFTALTQH   57
PcoV_GX-P4L             MSDNGPQ-N--RAPRITFGGPSDSTDNNQNGDRSGARPKQRRPQGLPNNTASWFTALTQH   57
PCoV_GX-P5L             MSDNGPQ-N--RAPRITFGGPSDSTDNNQNGDRSGARPKQRRPQGLPNNTASWFTALTQH   57
PcoV_GX-P1E             MSDNGPQ-N--RAPRITFGGPSDSTDNNQNGDRSGARPKQRRPQGLPNNTASWFTALTQH   57
ZC45                    MSDNGPQ-NQRSAPRITFGGPSDSSDNSKNGERNGARPKQRRPQGLPNNTASWFTALTQH   59
ZXC21                   MSDNGPQ-NQSSAPRITFGGPSDSSDNSQNGERNGARPKQRRPQGLPNNTASWFTALTQH   59
RShSTT182               MSDNGPQ-NQRNAPRITFGGPSDSSDNNQNGERSGARPKQRRPQGLPNNTASWFTALTQH   59
RShSTT200               MSDNGPQ-NQRNAPRITFGGPSDSSDNNQNGERSGARPKQRRPQGLPNNTASWFTALTQH   59
RacCS203                MSDNGPQ-NQRSAPRITFGGPSDSSDNNQNGERSGARPKQRRPQGLPNNTASWFTALTQH   59
PanCoV-MP789            MSDNGPQ-NQRNAPRITFGGPSDSAGSNQNGERSGARPKQRRPQGLPNNTASWFTALTQH   59
RmYN02                  MSDNGHQ-SQRNAPRITFGGPSDSTGSNQNGERSGARPKQRRPQGLPNNTASWFTALTQH   59
RaTG13                  MSDNGPQ-NQRNAPRITFGGPSDSTGSNQNGERSGARPKQRRPQGLPNNTASWFTALTQH   59
WIV04                   MSDNGPQ-NQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQH   59
CDC-CruiseA-12          MSDNGPQ-NQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQH   59
WIV04                   MSDNGPQ-NQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQH   59
INMI1                   MSDNGPQ-NQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQH   59
RpYN06                  MSDNGPQ-SQRNAPRITFGGPSDSTGSNQNGERSGARPKQRRPQGLPNNTASWFTALTQH   59
BM48-31                 MTDNGQ-SNSRNAPRITFGV-SDTSDNNQNAERAGARPKQRRPQGPPNNTASWFTALTQH   58
BtKY72                  MTDNGQ-QGPRNAPRITFGV-SDNFDNNQNGDRTGARPKHRRPQGPPNNTASWFTALTQH   58
                        *:***       *******  :*. ...:*. * *** *:***** **.***********
```

Gene N orf9 Nucleoprotein (the encoding of the RR pair 1)

```
HKU-39849               CGCCGACCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACAGCTCTCACTCAGCAT 180
Tor2/FP1-10912          CGCCGACCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACAGCTCTCACTCAGCAT 180
PcoV_GX-P5E             CGAAGGCCCCAGGGATTACCCAATAGTACTGCGTCTTGGTTCACCGCTCTCACTCAACAT 171
PcoV_GX-P2V             CGAAGGCCCCAGGGATTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACAT 171
PcoV_GX-P4L             CGAAGGCCCCAGGGATTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACAT 171
PcoV_GX-P5L             CGAAGGCCCCAGGGATTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACAT 171
PcoV_GX-P1E             CGAAGGCCCCAGGGATTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACAT 171
ZC45                    CGTCGACCCCAAGGCTTACCCAATAATACTGCATCTTGGTTCACCGCTCTCACTCAACAT 177
ZXC21                   CGTCGACCCCAAGGCTTACCCAATAATACTGCATCTTGGTTCACCGCTCTCACTCAACAT 177
PanCoV-MP789            CGTCGTCCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACAT 177
Wuhan-Hu-1              CGTCGGCCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACAT 177
CDC-CruiseA-12          CGTCGGCCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACAT 177
WIV04                   CGTCGGCCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACAT 177
INMI1                   CGTCGGCCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACAT 177
Bat-CoV-RatG13          CGTCGGCCTCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACAT 177
RmYN02                  CGTCGTCCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACAT 177
RpYN06                  CGTCGTCCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACAT 177
RShSTT200               CGTCGTCCACAAGGCCTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAGCAT 177
RShSTT182               CGTCGTCCACAAGGCCTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAGCAT 177
RacCS203                CGTCGTCCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAGCAT 177
BM48-31                 AGAAGACCGCAAGGCCCTCCTAACAACACAGCATCCTGGTTCACAGCTCTCACTCAGCAT 174
BtKY72                  AGAAGACCGCAAGGCCCTCCTAACAACACGGCATCCTGGTTCACCGCTCTCACTCAACAT 174
                         *  * ** ** **    ** ** *  ** ** ** ******** ********** ***
```

Protein N orf9 Nucleoprotein (the RR pair 2 and 3)

```
HKU-39849               GKEELRFPRGQGVPINTNSGPDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTGPEA  120
Tor2/FP1-10912          GKEELRFPRGQGVPINTNSGPDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTGPEA  120
PcoV_GX-P5E             GKEDLRFPRGQGVPINTNSTKDDQIGYYRRATRRVRGGDGKMKDLSPRWYFYYLGTGPEA  117
PcoV_GX-P2V             GKEDLRFPRGQGVPINTNSTKDDQIGYYRRATRRVRGGDGKMKDLSPRWYFYYLGTGPEA  117
PcoV_GX-P4L             GKEDLRFPRGQGVPINTNSTKDDQIGYYRRATRRVRGGDGKMKDLSPRWYFYYLGTGPEA  117
PCoV_GX-P5L             GKEDLRFPRGQGVPINTNSTKDDQIGYYRRATRRVRGGDGKMKDLSPRWYFYYLGTGPEA  117
PcoV_GX-P1E             GKEDLRFPRGQGVPINTNSTKDDQIGYYRRATRRVRGGDGKMKDLSPRWYFYYLGTGPEA  117
ZC45                    GKENLTFPRGQGVPINTNSSKDDQIGYYRRATRRIRGGDGKMKELSPRWYFYYLGTGPEA  119
ZXC21                   GKENLTFPRGQGVPINTNSSKDDQIGYYRRATRRIRGGDGKMKELSPRWYFYYLGTGPEA  119
RShSTT182               GKENLTFPRGQGVPINTNSTKDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEA  119
RShSTT200               GKENLTFPRGQGVPINTNSTKDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEA  119
RacCS203                GKENLTFPRGQGVPINTNSTKDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEA  119
PanCoV-MP789            GKEDLRFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEA  119
RmYN02                  GKEDLKFPRGQGVPINTNSSRDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEA  119
RaTG13                  GKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEA  119
WIV04                   GKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEA  119
CDC-CruiseA-12          GKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEA  119
WIV04                   GKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEA  119
INMI1                   GKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEA  119
RpYN06                  GKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEA  119
BM48-31                 GKEGLSFPRGQGVPVNTNSTRDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTGPEA  118
BtKY72                  GKETLTFPRGQGVPINTNSGKDDQIGYYRRASRRVRGGDGKMKELSPRWYFYYLGTGPEA  118
                        *** * ********:****  **********:**:********:****************
```

Gene N orf9 Nucleoprotein (the encoding of the RR pair 2 and 3)

```
HKU-39849               CCAGATGACCAAATTGGCTACTACCGAAGAGCTACCCGACGAGTTCGTGGTGGTGACGGC 300
Tor2/FP1-10912          CCAGATGACCAAATTGGCTACTACCGAAGAGCTACCCGACGAGTTCGTGGTGGTGACGGC 300
PcoV_GX-P5E             AAAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAGTTCGTGGTGGTGACGGT 291
PcoV_GX-P2V             AAAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAGTTCGTGGTGGTGACGGT 291
PcoV_GX-P4L             AAAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAGTTCGTGGTGGTGACGGT 291
PcoV_GX-P5L             AAAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAGTTCGTGGTGGTGACGGT 291
PcoV_GX-P1E             AAAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAGTTCGTGGTGGTGACGGT 291
ZC45                    AAAGATGACCAAATTGGCTACTACCGTAGAGCTACCAGACGAATTCGTGGTGGTGACGGT 297
ZXC21                   AAAGATGACCAAATTGGCTACTACCGTAGAGCTACCAGACGAATTCGTGGCGGTGACGGT 297
PanCoV-MP789            CCAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGT 297
Wuhan-Hu-1              CCAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGT 297
CDC-CruiseA-12          CCAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGT 297
WIV04                   CCAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGT 297
INMI1                   CCAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGT 297
Bat-CoV-RatG13          CCAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGT 297
RmYN02                  CGAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGT 297
RpYN06                  CCAGATGATCAAATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGT 297
RShSTT200               AAAGATGACCAGATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGT 297
RShSTT182               AAAGATGACCAGATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGT 297
RacCS203                AAAGATGACCAGATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGT 297
BM48-31                 AGGGACGACCAAATTGGCTACTATCGCAGAGCTACCCGACGAGTTCGTGGTGGTGATGGT 294
BtKY72                  AAAGATGACCAAATTGGCTACTATAGAAGAGCTTCCCGACGAGTTCGTGGTGGTGACGGA 294
                           ** ** ** ***********  * ******* ** ***** ******* ***** **
```

Protein N orf9 Nucleoprotein (the RR pair 4)

```
HKU-39849               QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQDLIRQGTDYK   300
Tor2/FP1-10912          QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQDLIRQGTDYK   300
PcoV_GX-P5E             QQQSQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEKTQGNFGDKELIRQGTEYK   297
PcoV_GX-P2V             QQQSQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTEYK   297
PcoV_GX-P4L             QQQSQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTEYK   297
PCoV_GX-P5L             QKQSQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTEYK   297
PcoV_GX-P1E             QQQSQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTEYK   297
ZC45                    QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK   299
ZXC21                   QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK   299
RShSTT182               QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK   299
RShSTT200               QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK   299
RacCS203                QQQGQTVTKKSAAEASKKPRQKRTATRQYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK   299
PanCoV-MP789            QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK   299
RmYN02                  QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK   299
RaTG13                  QQQSQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK   299
WIV04                   QQQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK   299
CDC-CruiseA-12          QQQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK   299
WIV04                   QQQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK   299
INMI1                   QQQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK   299
RpYN06                  QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK   299
BM48-31                 Q-QSQVVTKKTAAEASKKPRQKRTATKAYNVTQAFGRRGPEPTQGNFGDQELIRLGTDYK   296
BtKY72                  QQQPQVVTRKSASEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRLGTDYK   298
                        * * *.**:*:*:************: ********** ******::*** **:**
```

Gene N orf9 Nucleoprotein (the encoding of the RR pair 4)

```
HKU-39849               CAAAAACGTACTGCCACAAAACAGTACAACGTCACTCAAGCATTTGGGAGACGTGGTCCA 840
Tor2/FP1-10912          CAAAAACGTACTGCCACAAAACAGTACAACGTCACTCAAGCATTTGGGAGACGTGGTCCA 840
PcoV_GX-P5E             CAAAAACGTACTGCCACTAAACAATACAATGTCACTCAGGCATTTGGCAGACGTGGTCCG 831
PcoV_GX-P2V             CAAAAACGTACTGCCACTAAACAATACAATGTCACTCAGGCATTTGGCAGACGTGGTCCT 831
PcoV_GX-P4L             CAAAAACGTACTGCCACTAAACAATACAATGTCACTCAGGCATTTGGCAGACGTGGTCCT 831
PcoV_GX-P5L             CAAAAACGTACTGCCACTAAACAATACAATGTCACTCAGGCATTTGGCAGACGTGGTCCT 831
PcoV_GX-P1E             CAAAAACGTACTGCCACTAAACAATACAATGTCACTCAGGCATTTGGCAGACGTGGTCCT 831
ZC45                    CAAAAACGTACTGCTACAAAACAGTACAACGTCACTCAAGCATTTGGGAGACGTGGTCCA 837
ZXC21                   CAAAAACGTACTGCCACAAAACAGTACAACGTCACTCAAGCATTTGGGAGACGTGGTCCA 837
PanCoV-MP789            CAAAAACGTACTGCCACCAAACAATACAATGTAACACAAGCTTTTGGCAGACGTGGTCCA 837
Wuhan-Hu-1              CAAAAACGTACTGCCACTAAAGCATACAATGTAACACAAGCTTTCGGCAGACGTGGTCCA 837
CDC-CruiseA-12          CAAAAACGTACTGCCACTAAAGCATACAATGTAACACAAGCTTTCGGCAGACGTGGTCCA 837
WIV04                   CAAAAACGTACTGCCACTAAAGCATACAATGTAACACAAGCTTTCGGCAGACGTGGTCCA 837
INMI1                   CAAAAACGTACTGCCACTAAAGCATACAATGTAACACAAGCTTTCGGCAGACGTGGTCCA 837
Bat-CoV-RatG13          CAAAAACGTACTGCCACCAAACAATACAATGTAACACAAGCTTTTGGCAGACGTGGTCCA 837
RmYN02                  CAAAAACGTACTGCCACTAAACAATACAATGTAACACAAGCTTTTGGCAGACGTGGTCCA 837
RpYN06                  CAAAAACGTACTGCCACTAAACAATACAATGTAACACAAGCTTTTGGCAGACGTGGTCCT 837
RShSTT200               CAAAAACGTACTGCCACCAAACAATACAACGTAACACAAGCTTTTGGCAGACGTGGTCCA 837
RShSTT182               CAAAAACGTACTGCCACCAAACAATACAACGTAACACAAGCTTTTGGCAGACGTGGTCCA 837
RacCS203                CAAAAACGTACTGCCACCAGACAATACAATGTAACACAAGCTTTTGGCAGACGTGGTCCT 837
BM48-31                 CAGAAAAGAACAGCTACCAAAGCCTATAATGTTACTCAGGCTTTTGGTAGGCGAGGTCCC 828
BtKY72                  CAAAAACGAACGGCTACCAAGCAGTACAATGTCACGCAAGCTTTTGGTAGACGAGGACCC 834
                        ** *** * ** ** ** *     ** ** ** ** ** ** ** ** ** ** ** **
```