UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCIÊNCIAS
PROGRAMA DE PÓS GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

TMPRSS2 E OUTRAS SERINO PROTEASES TRANSMEMBRANA TIPO II (TTSPS): SINAL DE UMA CORRIDA ARMAMENTISTA DE LONGO PERIODO ENTRE VÍRUS E MAMÍFEROS PLACENTÁRIOS

Dissertação para a obtenção de título de mestre em Genética e Biologia Molecular

Matheus de Castro Nóbrega

Orientação: Maria Cátira Bortolini

Brasília/Porto Alegre, Junho de 2022

Matheus de Castro Nóbrega

TMPRSS2 E OUTRAS SERINO PROTEASES TRANSMEMBRANA TIPO II (TTSPS): SINAL DE UMA CORRIDA ARMAMENTISTA DE LONGO PERIODO ENTRE VÍRUS E MAMÍFEROS PLACENTÁRIOS

Dissertação apresentada ao Programa de Pós-Graduação *Strictu Sensu* em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de mestre em Genética e Biologia molecular.

Brasília/Porto Alegre, Junho de 2022

Brasília/Porto Alegre, Junho de 2022

Este trabalho foi realizado no Laboratório de Evolução Humana e Molecular do Departamento de Genética da Universidade Federal do Rio Grande do Sulcom o fomento do Conselho Nacional de Pesquisa e Desenvolvimento (CNPq) e pela CAPES.

Agradecimentos

Agradeço a minha família em especial aos meus pais João e Elânia por todo o apoio e suporte durante essa etapa da minha vida em particularmente durante a pandemia que me desanimou muito a princípio e também ao meu irmão Gustavo por todo o interesse quando decidia explicar ou mostrar alguma coisa do trabalho, seu entusiasmo com cada nova figura me deixou mais animado com meu próprio desenvolvimento. Agradeço também a todos os colegas do Laboratório de Evolução e Humana e Molecular pelo acolhimento, mesmo a distância, em particular á Bibiana Fam, pela disponibilidade, ao Yuri Yépes, por toda a paciência e pelas diversas aulas em que me ensinou sobre novas metodologias com doses iguais de coleguismo e experiência, e principalmente a minha orientadora Maria Cátira, pela atenção, interesse e por sua visão peculiar sobre o projeto, sem a colaboração dos três esse trabalho não teria sido realizado. Também agradeço às instituições de fomento Capes e CNPq pela bolsa, e a UFRGS e ao PPGBM pela oportunidade de realizar esse mestrado.

“O que sabemos é uma gota,
O que ignoramos é um oceano”
Sir Isaac Newton

Contents

Resumo

O gene da Serina Protease Transmembrana 2 (*TMPRSS2*), localizado no cromossomo 21q22.3, codifica uma proteína com o mesmo nome, sendo membro da família das Serino Proteases Transmembrana Tipo II (TTSPs). A TMPRSS2 humana é normalmente relacionada a resposta a andrógenos, mas foi cooptada pelo SARS-CoV-2 para clivar a glicoproteína viral Spike, e assim permitir a infecção da célula hospedeira. Apesar da existência de outras proteases cooptadas pelo SARS-CoV-2 para realizar essa função, recentes estudos funcionais mostram uma ativação e penetração mais velozes do SARS-CoV-2 em células expressando *TMPRSS2* que naquelas dependentes de outras proteases. De forma a avaliar os padrões evolutivos que moldaram a relação entre a TMPRSS2 do *Homo sapiens* e a Spike do SARS-CoV-2, buscamos identificar todas as TTSPs presentes no genoma do *Homo sapiens*, visto que o último estudo sobre o tema foi publicado em 2009. Foram encontradas 18 TTSPs pertencentes a 4 subfamílias, resgatando a relação filogenética original das subfamílias TTSPs. Porém quando foram analizados somente os 30 sítios que interagem com a Spike do SARS-CoV-2 um padrão filogenético distinto foi encontrado. Também investigamos a região codificante dos ortólogos do *TMPRSS2* de 182 espécies de mamíferos placentários. A variabilidade interespecífica em 33 sítios pode ser explicada por seleção positiva de acordo com a análise no pacote MEME, sendo que seis desses sítios (299, 340, 389, 413, 431, 438), ou seja, 15%, são reconhecidos como importantes para a interação com o vírus. Esses resultados podem indicar um sinal de uma corrida armamentista entre os coronavirus e os seus potenciais hospedeiros mamíferos. Em outras palavras, esse padrão de variação sugere que a maior

parte da variação entre espécies seja resultados de pressões seletivas que potencialmente vêm moldando a função normal da TMPRSS2 humana e de seus ortólogos nas células das espécies correspondentes. Por outro lado, a Spike viral estaria sendo moldada evolutivamente para se ligar em posições nas proteases dos hospedeiros mamíferos com menos propensão a terem variação promovida por ação de seleção positiva, o que conferiria vantagem ao vírus, pois ele teria menos chance de perder afinidade com o hospedeiro ao mesmo tempo que daria mais chances para saltos zoonóticos.

Abstract

The Transmembrane Serine Protease 2 (TMPRSS2) gene, located at human chromosome 21q22.3, encodes a protein with the same name member from the type II transmembrane serine proteases (TTSPs). TMPRSS2 is usually related to the response to androgens but is co-opted by the SARS-CoV2 to cleave the viral Spike glycoprotein to infect the host cell. Despite the existence of other proteases co-opted by SARS-CoV2 to that function, recent functional studies show a more rapid activation and penetration of SARS-CoV2 in cells expressing TMPRSS2 than with those in which infection depends on other proteases. In order to assess the evolutionary patterns that shaped the relationship between the Homo sapiens TMPRSS2 and SARS-CoV2 Spike, we aimed to identify all TTSPs present in the *Homo sapiens* genome since the previous study with this purpose was published in 2009. One of our goals is to understand better why TMPRSS2 has been evolutionarily co-opted and is preferentially used to cleave Spike. Eighteen canonic *Homo sapiens* TTSPs, grouped in 4 clades were found, rescuing the original phylogenetic relationship of the TTSPs subfamilies. However, when only 30 sites that interact with the Spike of SARS-CoV-2 were used, a distinct phylogenetic pattern was found. We also investigated the coding region of TMPRSS2 orthologs of the 182 species of placental mammals. Using the MEME package, our evolutionary analysis shows that the interspecific variability in 33 sites can be explained by positive selection, six of them (299, 340, 389, 413, 431, 438), that is, 15%, with importance for the interaction with SARS-CoV-2. These results may be a sign of the biological arms race between coronaviruses and their potential mammalian hosts. In other words, this pattern of variation suggests that most of the variation between species is the result of selective pressures that have been shaping the normal function of human TMPRSS2 and its orthologs in the cells of the corresponding species. On the other hand, the virtal Spike would be evolutionarily shaped to bind at positions in the proteases of mammalian hosts that are less likely to have variation promoted by positive selection action, which would give the virus an advantage, as it would have less chance of losing affinity with the host while giving more chances for zoonotic jumps.

1. Introdução

1.1 Contexto geral

No contexto atual de pandemia da COVID-19 um esforço científico global foi empreendido para compreender os diversos aspectos que levaram um vírus a promover uma infecção tão ampla e com um número tão grande de mortes. Como exemplo, uma rápida busca no Google Acadêmico com a palavra chave "COVID-19" resgata ~466.000 publicações (busca no dia 15 de maio de 2022). Visando ajudar nesse esforço, esse trabalho teve como objetivo principal elucidar a história evolutiva da Serino Protease Transmembrana 2 (do inglês Transmembrane Serine Protease 2 or TMPRSS2) uma das proteínas envolvidas na infecção pelo β-coronavírus SARS-CoV-2 causador da COVID-19, e que está envolvida também em outras infecções virais.

A TMPRSS2 é uma proteína codificada por um gene de mesmo nome e faz parte da família das Serino proteases transmembrana tipo 2 (*Type II Transmembrane Serine Protease* - TTSP) com três domínios funcionais: um receptor LDL classe A, um domínio scavenger (*Scavenger Receptor Cysteine-Rich* - SRCR) e um dominio peptidase extracelular S1 (Tripsina) (Hussain *et al.*, 2020).

As TTSPs compõem uma família de proteínas identificada recentemente, em 2001 por Hooper *et al.* Os autores identificaram, à época, 17 membros em mamíferos placentários (Hooper et a., 2001). Em 2009, Bugge *et al.* aumentou o número de proteínas identificadas em humanos e as classificou em 4 subfamílias usando como critério os domínios que cada proteína possui: HAT (human airway trypsin – Tripsina humana de vias aéreas) /DESC (7 genes/proteínas), Hepsina/TMPRSS (7), Matriptase (4), and Corina (1). *TMPRSS2*é membro da subfamília Hepsin/TMPRSS junto com *TMPRSS3, TMPRSS4, TMPRSS5/Spinesin*, MSP (Mosaic serine protease) e Enteropeptidase.

No entanto, nenhum trabalho mais específico sobre a história evolutiva da TMPRSS2 e sua família pode ser encontrado. Desse modo, o presente estudo busca contribuir para preencher essa lacuna de conhecimento, visando, dentre outras coisas,

avaliar uma potencial e quem sabe antiga guerra armamentista biológica entre vírus e seus potenciais hospedeiros mamíferos.

Nos itens abaixo poderá ser encontrado uma revisão mais detalhada sobre o tema SARS-CoV-2 e COVID-19. Vale destacar, contudo, que uma introdução abrangente, porém focada mais na TMPRSS2, e em proteínas parálogas no *Homo sapiens*, bem como aquelas ortólogas em outros mamíferos placentários, também poderá ser encontrada no item **Resultados- Manuscrito** (página 21), de modo que alguma repetição é inevitável.

1.2 A pandemia da COVID-19

A COVID-19 (Coronavirus Disease 2019) é uma doença causada pela infecção do β-coronavírus SARS-CoV-2 sendo caracterizada por um amplo espectro de sintomas. Os indivíduos infectados podem permanecer assintomáticos, apresentar sintomas brandos, apresentar pneumonia de graus leve a grave, bem como apresentar a síndrome respiratória aguda grave, caracterizada por um comprometimento respiratório severo que pode levar a óbito (WHO, 2022). Dados no primeiro ano na pandemia indicavam que os quadros mais severos e as taxas de mortalidade mais altas dessa doença estavam associados a idosos (PADHAN & PRABHEESH, 2021), pacientes imunossuprimidos (ISER *et al.* 2020) e a pacientes com condições preexistentes como hipertensão e problemas cardíacos (ROBINSON *et al.*, 2020, CHEN *et al.*, 2020). Como outros vírus respiratórios, entre os sintomas mais comuns encontravam-se a febre, tosse, dificuldade respiratória, dor muscular e fadiga, bem como uma pouco comum perda severa de olfato e paladar. No entanto, já era perceptível no início da pandemia que um ou mais desses sintomas poderiam estar ausentes (ISER *et al.* 2020). Além do pulmão, logo foi possível detectar que o SARS-CoV-2 também podia afetar coração, sistema digestivo, rins e cérebro (CHEN *et al.*, 2020, NERSISYAN *et al.*, 2020). Devido ao amplo espectro de sintomas e complicações a COVID-19 tem sido reportada desde sua emergência como uma doença multi-sistêmica (Temgoua *et al.*, 2020; Mir *et al.*, 2021). Ainda, por ter sintomas e quadros clínicos semelhantes aos da gripe comum e outros resfriados, juntamente com a presença de portadores assintomáticos, a COVID-19 apresentava um quadro difícil de ser diagnosticado de forma clínica imediata, o que ajudou, inicialmente, sua rápida transmissão pelo mundo.

Em 31 de dezembro de 2019, a Organização Mundial da Saúde (OMS) foi alertada sobre vários casos incomuns de pneumonia grave na cidade de Wuhan, na China. Tratava-se de uma nova cepa de coronavírus que não havia sido identificada antes em seres humanos. Em 30 de janeiro de 2020, depois de um surto que começava a assustar a Itália e a Europa inteira, a OMS decretou que o caso constituia uma "Emergência de Saúde Pública de Importância Internacional". Desde então a pandemia da COVID-19 tem movimentado tanto os poderes políticos dos vários países, quanto a sociedade de modo geral, causando um impacto negativo nos sistemas de saúde globais, forçando os governos a fechar as fronteiras, restringir viagens, promover "lockdowns' e tomar precauções contra colapsos econômicos (BIEBER, 2020; NICOLA *et al.*, 2020; PADHAN & PRABHEESH, 2021; MEYER *et al.*, 2022). Para se ter uma ideia, o setor primário foi impactado principalmente pela falta de pessoal e pela insegurança em relação aos empregos. Além disso, a deficiência de pessoal para o transporte e verificação dos produtos trouxe grandes consequências para o setor (NICOLA *et al.*, 2020). O setor secundário foi igualmente afetado pela falta de pessoal e o fechamento das fronteiras. A quebra na cadeia de suprimentos , em uma escala nunca vista antes (ROZHKOV *et al.*, 2022; MEYER *et al.*, 2022), e o afastamento de pessoal levou a um declínio da indústria (NICOLA *et al.*, 2020) assim como a redução na demanda por produtos do setor (MEYER *et al.*, 2022). O setor terciário, por sua vez, foi o setor que provavelmente teve mais áreas afetadas. Os setores de turismo, hoteleiro, aviação, tiveram profunda queda na demanda (MEYER *et al.*, 2022) enquanto os de saúde, farmacêutico e alimentício tiveram um aumento na demanda, seguido de baixa nas vagas para pacientes ou no estoque de produtos e adaptações significativas seja na infraestrutura ou na entrega dos produtos e serviços (NICOLA *et al.*, 2020). De forma ampla os três setores foram afetados pela redução no número de empregos, que consequentemente afeta a distribuição de capital, pelo aumento nos custos de transações internacionais, pelo declínio do turismo e dos serviços de modo amplo e geral (PADHAN & PRABHEESH, 2021). Mais recentemente, de acordo com o relatório econômico da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) de setembro de 2021 (OCDE, 2021), o PIB global atual é superior ao nível anterior à pandemia. No entanto, a produção em meados de 2021 ainda era 3,5% menor do que o esperado antes da disseminação do COVID-19, resultando em perdas de empregos e renda (MARTINEZ *et al.*, 2022).

Do ponto de vista social, as medidas de distanciamento, como o fechamento das escolas, elevaram a preocupação com o aumento nos níveis de violência doméstica, que inclui abuso físico, emocional e sexual, inclusive de crianças (NICOLA *et al.*, 2020; ABRAMS *et al.*, 2022). Para os especialistas, um período maior de confinamento significa que pessoas vulneráveis estão mais expostas ao abuso e tem mais dificuldade para buscar ajuda (NICOLA *et al.*, 2020; ABRAMS *et al.*, 2022). O prolongamento da pandemia também levou ao aumento da insegurança alimentar (ABRAMS *et al.*, 2022) e problemas mentais como depressão e ansiedade, adventos que tornaram a situação de vida ainda pior para aqueles em condições sociais de vulnerabilidade (ABRAMS *et al.*, 2022; WIRKNER *et al.*, 2022). Soma-se a esse cenário a discussão intensa e disseminação em escala global de teorias da conspiração, caracterizadas por idéias pseudocientíficos e fantasiosas, que levaram a efeitos sociais bastante negativos como diminuição da confiança nas instituições públicas e ao apoio das medidas de contenção como o distanciamento social e aplicação de vacinas (PUMMERER *et al.*, 2022).

Por outro lado, a emergência mundial também mobilizou a comunidade científico-acadêmica de um modo nunca antes visto. Os esforços têm buscado avançar no conhecimento sobre o coronavírus e como combatê-lo. Para se ter uma idéia da amplitude desse esforço coletivo da comunidade científico-acadêmica, uma rápida busca no PubMed (dia 25/09) com as palavras chaves “SARS-CoV2 e COVID-19” identifica 35.338 artigos publicados em 2020. A mesma pesquisa feita em 2022 (04/05) revelou que no período de 2019 a 2022 o número se elevou para 209.982 artigos científicos. Já no Google Acadêmico, como já comentado, o número salta para ~466.000 publicações. Dentre esses milhares de estudos muitos, incluindo de nosso grupo de pesquisa (FAM *et al.*, 2020; YÉPEZ *et al.*, 2022), focaram na compreensão de como o SARS-CoV-2 era tão eficiente evolutivamente, ou seja, quais eram os meios moleculares que este organismo se apropriou para infectar seres humanos (ver outros exemplos em SONG *et al.*, 2019; ANDERSEN *et al.*, 2020; CAO *et al.*, 2020; LI *et al.*, 2020; MOUSAVIZADEH & GHASEMI,2020; GHEWARE *et al.*, 2022; HOROWITZ *et al.*, 2022).

1.2 Origem do SARS-CoV-2

Até 2003 os coronavírus (família Coronaviridae) eram conhecidos por causar apenas, de modo geral, doenças mais leves em humanos, como o resfriado comum sazonal (vírus HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU1). No entanto, desde então, três coronavírus com capacidade para causar impactos graves na saúde humana foram identificados. No ano de 2003 foi identificado o β-coronavírus SARS-CoV, responsável pela epidemia da Síndrome Respiratória Aguda Grave (SARS "Severe Acute Respiratory Syndrome"). O SARS-CoV estava sendo transmitido, inicialmente, entre pacientes de hospitais na China, e foi posteriormente identificado como um coronavírus originário da espécie de morcego, *Rhinolophus sinicus*. No entanto, teria sido transmitido ao homem por espécie intermediária a civeta da palmeira (*Paguma larvata*), um pequeno carnívoro asiático (SONG *et al.*, 2019). Passado cerca de 10 anos, outro coronavírus da mesma família foi isolado inicialmente na Arábia Saudita. O β-coronavírus chamado de MERS-CoV foi o responsável pela epidemia da Síndrome Respiratória do Oriente Médio (MERS, "Middle East Respiratory Syndrome"), que acometeu diferentes países, sendo transmitido para humanos inicialmente por dromedários (*Camelus dromedarius*) (AZHAR et 2014, SONG *et al.* 2019; LAMMERS & HAAGMANS *et al.*, 2022). Estudos demonstram que o reservatório natural do MERS-CoV também seriam morcegos da família Rhinolophidae, tendo os dromedários como hospedeiros intermediários, antes do transbordamento zoonótico para o *Homo sapiens* (SONG *et al.* 2019).

Vale destacar que semelhante ao SARS-CoV, o SARS-CoV-2 utiliza a Enzima Conversora de Angiotensina 2 (ACE2 da sigla do inglês para Angiotensin-Converting Enzyme 2) como portal de entrada para as células hospedeiras (ver revisão em FAM *et al.*, 2020). Ainda, de acordo com a OMS e vários pesquisadores (ver revisão em YÈREZ *et al.*, 2022), a taxa de mortalidade média estimada, considerando casos detectáveis/notificados, para COVID-19 é menor (2,72%) do que a doença causada por MERS-CoV (34,4%) e SARS- CoV (9,6%). Esse número permanece baixo mesmo considerando que o número de mortes causadas pela COVID-19 pode estar subestimado. Apesar dessa taxa de mortalidade relativamente baixa, em 15 de maio de 2022 foi estimado que a infecção por SARS-CoV-2 já tinha levado a mais de 518 milhões de casos confirmados e mais de seis milhões de mortes de acordo com OMS (https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---18-may-2022). Comparativamente, SARS-CoV e

MERS-CoV infectaram 8.098 e 2.566 pessoas e mataram 774 e 866 pessoas, respectivamente (YÈPEZ *et al.*, 2022).

Desse modo é bem conhecido que os genomas virais mais semelhante ao SARS-CoV-2 são oriundos de linhagens de coronavírus de morcego da espécie *Rhinolophus affinis*, tais como "RaTG13" e "bat-SL-CoVZC45" (ZHOU *et al.*, 2020, SINGH & YI, 2021). Inicialmente, SONG *et al.* (2019) postularam que esses vírus de morcegos não possuíam domínios de ligação iguais aos do SARS-CoV-2, o que sugeria que provavelmente não se conectariam bem com o seu receptor celular em humanos (ACE2), indicando a possibilidade de existir hospedeiro intermediário na origem do SARS-CoV-2. O genoma com o domínio de ligação mais semelhante ao SARS-CoV-2 foi encontrado em um coronavírus do pangolim malaio (*Manis javanica*) denominado Pangolin-CoV (ZHANG *et al.*, 2020). No entanto, tanto o bat-SL-CoVZC45 quanto o Pangolin-CoV encontrados nesses animais não possuem os sítios de clivagens presentes no SARS-CoV-2, o que sugere que nenhum dos dois é o progenitor direto do SARS-CoV-2 (SONG *et al.*, 2019). A recombinação entre linhagens virais também tem sido proposta há algum tempo (SALLARD *et al.*, 2020, SINGH & YI, 2021). Porém, é importante notar que nenhuma dessas espécies de morcegos e pangolins foi amostrada o suficiente para abarcar toda a variedade de coronavírus potencialmente presente nelas. Para adquirir os sítios de clivagem e os domínios de ligação necessários para infectar seres humanos o animal que serve de reservatório para o vírus precisaria ter uma população suficientemente grande onde esses sítios/domínios poderiam aparecer por meio de mutações aleatórias com posterior ação da seleção natural para um efetivo salto zoonótico bem sucedido (ANDERSEN *et al.*, 2020).

Mais recentemente, TEMMAN *et al.* (2022) salientam que o reservatório animal do SARS-CoV-2 ainda é controverso, apesar de relatos de similaridade com vírus de morcegos *Rhinolophus affinis*, como já comentado acima. Os autores reforçaram a proposta de que o SARS-CoV-2 teria um genoma em mosaico, sugerindo a contribuição de diferentes progenitores. TEMMAN et al (2022) sustentam que o SARS-CoV-2 seria resultado de uma recombinação de sequências pré-existentes em espécies de morcegos *Rhinolophus* que vivem nos extensos sistemas de cavernas de calcário do Sudeste Asiático e do sul da China, incluindo *R. malayanus* e *R. pusillus*. Além disso, as distribuições de *R. marshalli*, *R. malayanus* e *R. pusillus* se sobrepõem na sub-região da Indochina, o que significa que eles podem compartilhar cavernas como abrigos e habitats de forrageamento.

Suas descobertas, portanto, indicariam que vírus semelhantes ao SARS-CoV-2 transmitidos por morcegos potencialmente infecciosos para humanos circulam em *Rhinolophus spp*. na península da Indochina (TEMMAN *et al.*, 2022). Os últimos autores comentam, no entanto, que nenhum desses coronavírus de morcegos apresentavam o sítio com quatro aminoácidos (PRRA) na junção S1/S2 da proteína Spike que é clivado pela Furina. Esse sítio de clivagem por Furina é importante na determinação da infectividade viral do SARS-CoV-2. Com base na comparação das sequências em torno do local de clivagem entre os sítios S1 e S2 da proteína Spike dos morcegos foi sugerido que o local de clivagem pela Furina no SARS-CoV-2 poderia ter se originado por eventos de recombinação entre os coronavírus BANAL-116, BANAL-247, bat RmYN02 e bat RacCS203. Alternativamente, o sítio de clivagem pela Furina no SARS-CoV-2 poderia ter emergido através de passagens do vírus em um hospedeiro alternativo ou mesmo durante uma circulação precoce pouco sintomática e não relatada em humanos (TEMMAN *et al.*, 2022).

É oportuno salientar que assim como outros Coronaviridae, SARS-CoV-2 é um vírus envelopado com genoma de fita simples de RNA (~ 30 kb), de sentido positivo, o que significa que seu RNA genômico serve tanto como um modelo direto para tradução quanto para a replicação. SARS-CoV-2 apresenta duas "fases de leitura abertas" (em inglês *open reading frame* ou *ORF*) sobrepostas, *ORF1ae ORFb* que geram polipeptídios contínuos que são clivados em um total de 16 proteínas não estruturais. A tradução de *ORF1b* é mediada por mudança na matriz de leitura que permite que a tradução continue além do códon de parada de *ORF1a*. Além de ORF1a e ORF1b, todos os outros ORFs virais canônicos são traduzidos de RNAs sub genômicos. Esses RNAs subgenômico codificam as 4 proteínas estruturais conservadas - Spike (S), membranares (M), de envelope (E) e nucleocapsídeo (N), - e vários polipeptídeos acessórios, tais como as denominadas ORF3a, ORF6, ORF7a, ORF7b, ORF8 e ORF10 que desempenhariam papéis importantes no ciclo de vida viral e poderiam contribuir para sua patogênese e virulência (HASSAN *et al.*, 2021). Ainda em 2020 FINKEL *et al.* (2020) identificaram um número maior de *ORFs* do que aquelas propostas inicialmente, e sugeriram que essas *ORFs* adicionais, quando traduzidas, resultariam em polipeptídios com papel regulatório. Atualmente, são reconhecidos pelo menos 13 *ORFs* (MALONE *et al.*, 2022).

Finalmente, é sabido que cinco dos seis aminoácidos críticos nos respectivos domínios de ligação ao receptor (RBD) da Spike viral, são diferentes entre SARS-CoV-2 e SARS-CoV (L455Y, F486L, Q493N, S494D, N501T, respectivamente; TANG *et al.*, 2020).

1.3 Infecção celular por SARS-CoV-2

Como já mencionado, o SARS-CoV-2 utiliza a peptidase de membrana ACE2 como receptor celular para infectar as células humanas (GHEWARE *et al.*, 2022; HOROWITZ *et al.*,2022). Essa enzima foi anteriormente descrita como receptor de entrada para outros dois coronavírus humanos: o β-coronavírus SARS-CoV e o α-coronavirus NL63 (HOFMANN *et al.*, 2005, HOFMANN & PÖHLMANN 2004). E está associada à severidade da doença no caso da COVID 19 (GHEWARE *et al.*, 2022; HOROWITZ *et al.*,2022).

A ACE2 é uma enzima transmembrana responsável pela clivagem de angiotensina tipos 1 e 2 (Omim número de referência *300335) e foi identificada como sítio de ligação para a proteína Spike do coronavírus SARS-CoV-2 (LI *et al.*, 2020; LAMMERS & HAAGMANS *et al.*, 2022).

Além de ACE2, há outros elementos genéticos chaves no sucesso da infecção. Dentre estes, o receptor Serino Protease Transmembrana 2 (no inglês Transmembrane Serine Protease 2 ou TMPRSS2), cujo papel é fundamental para uma bem sucedida infecção pelo SARS-CoV-2 (HOFFMANN *et al.*, 2020; LAMMERS & HAAGMANS *et al.*, 2022). O papel da TMPRSS2 (Omim número de referência * 602060) no hospedeiro normalmente está relacionado à resposta à sinalização de andrógenos, mas foi evolutivamente recrutada pelo SARS-CoV-2 para processar a proteína Spike (HOFFMANN *et al.*, 2020; LAMMERS & HAAGMANS *et al.*, 2022).

A proteína Spike do SARS-CoV-2 possui 2 subunidades: a subunidade S1 que se liga ao ACE2 e a subunidade S2 que media a fusão de membranas (JACKSON *et al.*, 2022; LAMMERS & HAAGMANS *et al.*, 2022). As duas subunidades são separadas pelo sítio S1/S2 que contém um motivo de clivagem para Furina (JACKSON *et al.*, 2022; LAMMERS & HAAGMANS *et al.*, 2022). Para a infecção ocorrer com sucesso, ambas as proteínas são necessárias, após se ligar a ACE2 da célula alvo, a proteína Spike é clivada

pela TMPRSS2 no sítio S2'. A clivagem ativa a subunidade S2 promovendo a fusão entre a célula hospedeira e o vírus (JACKSON *et al.*, 2022; LAMMERS & HAAGMANS *et al.*, 2022). Outra entrada que pode ser utilizada pelo vírus é através dos endossomos. Nessa via a Catepsina pode clivar a proteína Spike, mas estudos recentes mostram que essa via não é eficiente considerando células epiteliais primárias (JACKSON *et al.*, 2022; LAMMERS & HAAGMANS *et al.*, 2022). Outros co-receptores como a Neuropilina 1 e proteases como Catepsina L, os parálogos de T TMPRSS2, MPRSS11d e TMPRSS13 podem também estar envolvidas com a infecção por SARS-CoV-2, mas não está claro ainda como essas vias moleculares alternativas operam (JACKSON *et al.*, 2022; LAMMERS & HAAGMANS *et al.*, 2022).

1.4 O Gene *TMPRSS2* e seu produto como mediador de infecção viral

O gene *TMPRSS2*, também conhecido como *PRSS10*, codifica a proteína TMPRSS2 pertencente à família das Serina-Proteases ("Type II Transmembrane Serine Protease" – TTSP). O gene possui 43.854 nucleotídeos, 15 éxons e está localizado no cromossomo 21 humano, na região 21q22.3 (PAOLINI-GIACOBINO *et al.*, 1997). TMPRSS2 contém três domínios funcionais: um receptor LDL classe A, um domínio scavenger ("Scavenger Receptor Cysteine-Rich" - SRCR) e um domínio peptidase extracelular S1 (Tripsina) (HUSSAIN *et al.*, 2020).

Baseando-se na sua estrutura, foi especulado que a proteína deveria funcionar como mediadora de sinais entre o ambiente extracelular e a célula (VAARALA *et al.*, 2001). Análises posteriores *in vitro* relataram que essa proteína regulava a atividade de canais de sódio, estando ela associada a processos fisiológicos e patológicos como a digestão, remodelamento tecidual, invasão celular tumoral, apoptose e dor (THUNDERS & DELAHUNT, 2020).

A expressão do *TMPRSS2* também é observada durante o desenvolvimento e aumenta com o envelhecimento. *TMPRSS* tem alta transcrição no cérebro fetal, mas baixa no cérebro adulto e uma baixa transcrição também no pulmão fetal quando comparado ao pulmão adulto (THUNDERS & DELAHUNT, 2020). O gene *TMPRSS2* possui expressão diferenciada em outros órgãos, tais como fígado, coração, e trato gastrointestinal, órgãos que podem ser afetados pelo SARS-CoV-2 durante a infecção. Esse gene também é

altamente expresso no tecido epitelial do lúmen da próstata (VAARALA *et al.*, 2001; THUNDERS & DELAHUNT, 2020). Na próstata ele contribui para uma cascata proteolítica que resulta na ativação do antígeno próstata-específico, a protease presente no fluido seminal com atividade enzimática análoga à coagulação do sangue (THUNDERS & DELAHUNT, 2020). *TMPRSS2* possui elementos responsivos a andrógenos na sua região 5' UTR, portanto, dentro do seu promotor. A testosterona e a di-hidrotestosterona regulam a transcrição desse gene através do estímulo do receptor andrógeno presente nas células epiteliais do lúmen da próstata, um dos locais onde o gene é expresso (THUNDERS & DELAHUNT, 2020).

Devido a sua atividade proteolítica, TMPRSS2 foi ao longo da evolução cooptada por vírus para facilitar suas entradas nas células hospedeiras. Porém, não se sabe desde quando isso vem acontecendo, pois estudos com outros animais e seus respectivos vírus são praticamente inexistentes. Inicialmente foi descrito o papel de TMPRSS2 para o sucesso das infecções por cepas do vírus influenza, considerando populações humanas. A replicação do vírus influenza humano é iniciada pela glicoproteína de superfície hemaglutinina (HA) que media a ligação aos receptores de superfície celulares contendo ácido salicílico e fusão do envelope viral com a membrana celular (BÖTTCHER-FRIEBERTSHÄUSER *et al.*, 2010). HA é sintetizada como uma molécula precursora que precisa ser clivada pela célula hospedeira nas subunidades HA1 e HA2 para adquirir sua capacidade de fusão com a membrana da célula hospedeira (BÖTTCHER-FRIEBERTSHÄUSER *et al.*, 2010). A clivagem induz mudanças conformacionais em pH baixo que expõe o peptídeo de fusão hidrofóbico N-terminal da subunidade HA2 e inicia a fusão (BÖTTCHER-FRIEBERTSHÄUSER *et al.*, 2010). Em seres humanos a clivagem pode ser feita por duas serino-proteases: a tripsina humana de vias aéreas – (human air way trypsin-like protease, HAT também conhecida como TMPRSS11D) ou pela TMPRSS2 (BÖTTCHER-FRIEBERTSHÄUSER *et al.*, 2010). Estudos subsequentes mostraram que TMPRSS2 e a protease relacionada TMPRSS4 clivam a HA do vírus influenza H1N1 (família Orthomyxoviridae), responsável pela pandemia de 1918. Além disso, TMPRSS2 também ativaria a proteína de fusão do Metapneumovirus (MPV; família Paramyxoviridae) humano identificado em 2001 (BÖTTCHER-FRIEBERTSHÄUSER *et al.*, 2010).

Analises *in vitro* com plasmídeos realizadas por GLOWACKA *et al.* (2011) demonstraram que as proteínas Spike do SARS-CoV eram clivadas de duas formas pela

TMPRSS2: pela clivagem em cis e pela clivagem em trans. Na clivagem em cis, TMPRSS2 e a proteína Spike do SARS-CoV eram expressos juntos na mesma célula e a clivagem resultava na quebra da proteína Spike e na liberação dos fragmentos desta proteína no sobrenadante. Os fragmentos funcionam como chamarizes de anticorpos e provocam a inibição da resposta imune mediada por anticorpos. Ainda, os mesmos testes *in vitro* indicaram que a clivagem em trans ocorria quando uma célula expressava a proteína Spike viral na membrana, enquanto outra célula expressava TMPRSS2. Quando ocorria o encontro de ambas moléculas ocorria a clivagem da proteína Spike e a posterior fusão das células em questão. Isso indica que a proteína Spike precisa de um processamento para permitir a entrada do vírus na célula, e em seres humanos esse processamento ocorre através do recrutamento da TMPRSS2 (GLOWACKA *et al.*, 2011). Posteriormente, BERTRAM *et al.* (2013) demonstraram que TMPRSS2 estava envolvida com a infecção pelo α-coronavírus humano CoV-229E, enquanto IWATA-YOSHIKAWA *et al.* (2019) corroboraram que TMPRSS2 ativava SARS-CoV, sendo também indispensável para a infecção bem sucedida do β-coronavírus MERS-CoV. Com o advento da COVID-19, os estudos se voltaram para identificar se TMPRSS2 também exercia um papel crucial na infecção por SARS-CoV-2, o que de fato foi demonstrado por vários autores (HOFFMANN *et al.*,2020; MATSUYAMA *et al.*, 2020; JACKSON *et al.*, 2022; LAMMERS & HAAGMANS *et al.*, 2022). Por outro lado, como já comentado, o papel de ortólogos de TMPRSS2 nas infecções de vírus presentes em outras espécies de mamíferos ainda precisa ser investigado.

Como já comentado, a TMPRSS2 não é a única capaz de processar Spike, sendo a Catepsina L pode realizar esse processo. No entanto, a rota da TMPRSS2 aparenta ser mais veloz, e portanto mais eficiente (KOCH *et al.*, 2021; JACKSON *et al.*, 2022; LAMMERS & HAAGMANS *et al.*, 2022). Além disso, outras proteínas da mesma família da TMPRSS2 podem ser utilizadas. Esses achados indicam que a proteína TMPRSS2 além de outras da mesma família vem sendo cooptadas muito antes da interação entre SARS-CoV-2 e o *Homo sapiens.*

1.5 As Serino Proteases Transmembrana Tipo 2 (“Type II Transmembrane Serine Protease” - TTSP).

As TTSPs são uma família identificada recentemente em 2001 por HOOPER et al, consistindo à época de 17 membros em mamíferos placentários. Em 2009 BUGGE et al aumentou o número de proteínas identificadas em humanos e as classificou em 4 subfamílias usando como critério os domínios que cada proteína possui: HAT/DESC (7 genes/proteínas), Hepsina/TMPRSS (7), Matriptase (4), and Corina (1). *TMPRSS2*é membro da subfamília Hepsin/TMPRSS junto com *TMPRSS3, TMPRSS4, TMPRSS5/Spinesin,* MSPL e Enteropeptidase.

Desde então, estudos sobre esta família gênica não têm sido publicados, pelo menos considerando parálogos humanos e ortólogos presentes em outras espécies de mamíferos.

2. Objetivo

Tendo em vista que a protease TMPRSS2 é a molécula preferencial para a clivagem da Spike do SARS-CoV-2, buscamos neste trabalho atingir o seguinte objetivo de identificar eventuais padrões de diversidade na região codificadora dessa molécula e seus parálogos no *Homo sapiens*, bem como de ortólogos em mamíferos, com intuito de responder, basicamente, duas principais perguntas: 1) O padrão de diversidade genética encontrado poderia ser resultado, mesmo que parcialmente, de pressão seletiva exercida por uma corrida armamentista biológica entre vírus-hospedeiro? e 2) Porque TMPRSS2 foi preferencialmente cooptada evolutivamente pelo SARS-CoV-2 para o processamento da Spike, enquanto outras TTSPs não o foram?

3. Resultados- Manuscrito

Os resultados podem ser encontrados no manuscrito a seguir:

Castro-Nóbrega M, Yépez Y, Fam B, and Bortolini MC. TMPRSS2 AND OTHER TYPE II TRANSMEMBRANE SERINE PROTEASES (TTSPS): SIGNATURE OF A LONG-TERM ARMS RACE BETWEEN VIRUSES AND PLACENTAL MAMMALS.

1. Introduction

1.1 SARS-CoV-2 and its ability to infect humans

The β-coronavirus SARS-CoV-2 has a lower estimated average mortality rate (2.72%) when it is compared with other β-coronaviruses that infect humans, MERS-CoV (34.4%) and SARS-CoV (9.6%) (Xiao *et al.*, 2020; ECDC, 2021a, b; Krishnamoorthy *et al.*, 2021; Awadasseid *et al.*, 2021). Despite that, SARS-CoV-2 is responsible for the current COVID-19 pandemic, which has already killed millions of people, a stark contrast to the number of deaths reported from infections of MERS-CoV (866) and SARS-CoV (774) (WHO, 2003, 2020; Alfaraj *et al.*, 2019; Petersen *et al.*, 2020; Peng *et al.*, 2021). This extraordinary evolutionary success of the original (Wuhan) SARS-CoV-2 and its derived lineages arose from some characteristics in addition to its relatively low lethality: stronger tropism with the host cell, high transmissibility, high transmission rate from asymptomatic individuals, and high viral load (Peng *et al.*, 2021; Yépez *et al.*, 2022).

Coronavirus (COV) originated millions of years ago (Wertheim *et al.*, 2013), so zoonotic spillover's success, *i.e.*, transmitting a pathogen from a vertebrate animal to a human and *vice-versa*, is a common phenomenon. As a result, it can be assumed that an evolutionary arms race has been established between CoVs and their potential vertebrate hosts (Fam *et al.*, 2020; Yépez *et al.*, 2022). The CoV-RaTG13 of the brown bat

(*Rhinolophus affinis*) is the Wuhan-SARS-CoV-2 potential ancestor since their genomes have 97.41% of identity (Malaiyan *et al.*, 2021). At least five amino acid (aa) substitutions (F486L, Q493Y, S494R, N501D, e Y505H) at critical sites of the Spike (S) glycoprotein receptor-binding domain (RBD) of CoV-RaTG13 were crucial to break the interspecific barrier and to endow the Wuhan-SARS-CoV-2 of a high tropism for some human cells that have the peptidase named Angiotensin-Converting Enzyme (ACE2) in their cell-membranes (Wan *et al.*, 2020).

Wuhan-SARS-CoV-2 also has five critical amino acid differences in its RBD (domain responsible for binding to host receptor angiotensin-converting enzyme 2, ACE2) when compared with SARS-CoV (L455Y, F486L, Q493N, S494D, N501T; SARS-CoV-2 and SARS-CoV aa respectively (Wan *et al.*, 2020; Andersen *et al.*, 2020; Tang *et al.*, 2020). Later, successive mutations in key Spike residues appeared in derived SARS-CoV-2 lineages, including in the most critical concern variants (VOCs: Alpha, Beta, Gamma, and Delta), increasing their transmission capacity when compared with the Wuhan lineage but maintaining a similar mortality rate (Yépez *et al.*, 2022 and references therein). This scenario changed with the arrival of the VOC Omicron (BA.1).

Viruses do not inevitably evolve toward being less virulent, although it is a commonly observed phenomenon. Evolution selects those that excel at multiplying. In the case of COVID-19, in which the vast majority of transmission occurs before any symptomatology, the reduced severity may not be directly selected for all lineages (Bhattacharyya and Hanage, 2022). Despite these circumstances, Omicron has increased transmissibility compared to other variants but reduced the risk of hospitalization and severe disease (WHO, 2022; Madhi *et al.*, 2022; Wolter *et al.*, 2022). Derived Omicron lineages are even more contagious and have already swept the globe in recent months, but it makes people no sicker than the original Omicron. In other words, Omicron and its sub-variants are less virulent, even in settings where vaccination is ineffective. Nevertheless, some authors point out that this trend can be reversed since the ongoing rapid antigenic evolution will likely produce new variants that may escape immunity and be more severe (Markov *et al.*, 2022).

Viruses enter host cells through highly complex molecular mechanisms shaped by natural selection over thousands of years. Thus, similar to other viruses, all SARS-CoV-2 variants need to co-opt genetic elements from host cells for effective infection and

replication. The process starts through the viral fusion glycoprotein SARS-CoV-2 Spike, and ACE2, an aminopeptidase, present in the host cell membranes of several organs and tissues, including the respiratory system. ACE2 as the gateway for SARS-CoV-2 into human cells was identified concomitantly by several research groups soon after the COVID-19 outbreak (Zhou *et al.*, 2020; Walls *et al.*, 2020; Hoffmann *et al.*, 2020; review in Peng *et al.*, 2021). SARS-CoV-2 Spike, a class I viral fusion glycoprotein, presents the S1 and S2 subunits, the first contains the RBD (Peng *et al.*, 2021). Three copies of the RBD are found, and they exhibit some level of mobility. So, the three RBDs can adopt different conformations at the prefusion state, exposed for access by ACE2 (open conformation) but not synchronously (Walls *et al.*, 2020; Ke *et al.*, 2020; Peng *et al.*, 2021). Besides, the binding of ACE2 to an open RBD can promote the conformational transition of the other closed RBDs and make them accessible by the receptor. Thus, the Spike can bind to 1–3 ACE2 molecules, depending on the conformation of each RBD (Walls *et al.*, 2020; Ke *et al.*, 2020; Peng *et al.*, 2021). Interestingly, Omicron RBD has a ~3-fold enhanced binding affinity for ACE2 relative to the Wuhan-SARS-CoV-2 and Delta RBDs (Meng *et al.*, 2022).

After that link between RDB and ACE2, the conformation of the S1 subunit is affected, disrupting its interactions with the S2 that contains the machinery that promotes fusion between the cell and viral membranes (Chi *et al.*, 2020; Hoffmann *et al.*, 2020; Mousavizadeh and Ghasemi, 2020; Vargas-Alarcón *et al.*, 2020; Wettstein *et al.*, 2021). Among the domains of S2, there is the fusion peptide (FP), but, in contrast to most typical class I viral fusion glycoproteins, FP of the coronavirus Spike is shielded by the upstream helix region of S2, thus requiring a second cleavage (at site S2´) to expose FP. This last event activates a molecular cascade, triggering irreversible S2 conformational changes essential to initiate host cell-SARS-CoV-2 membrane fusion (Peng *et al.*, 2021).

1.2 Transmembrane serine protease 2 (TMPRSS2)

As seen above, viral fusion glycoproteins need to be cleavage by proteases co-opted from the host for their functional processing. For instance, Furin cleaves the Spike at site S1/S2, a region with a multibasic motif (Bestle *et al.*, 2020; Coutard *et al.*, 2020; Hoffmann *et al.*, 2020a; Koch *et al.*, 2021; Peng *et al.*, 2021). It is known that VOCs Alpha

and Delta Spike proteins confer more efficient cell-cell fusion kinetics compared to Wuhan-SARS-CoV-2 due to mutations in the Furin cleavage site region that increase S1/S2 cleavage and promote syncytia formation, which is associated with pathogenesis (Saito *et al.*, 2021; Meng *et al.*, 2022).

The second major step of proteolysis of SARS-CoV-2 Spike is executed mainly by the transmembrane serine protease 2 (TMPRSS2), which cleaves the S2 at the S2′ site to expose the FP (Bestle *et al.*, 2020; Hoffmann *et al.*, 2020; Wettstein *et al.*, 2021; Peng *et al.*, 2021; Koch *et al.*, 2021). It has been known for decades that TMPRSS2 is expressed in epithelial cells of the prostate gland (Lin *et al.*, 1999). However, its presence is currently recognized in organs and tissues that also express ACE2, including the nasal mucosa of the olfactory cleft, olfactory bulb, and lung epithelium (Lechien *et al.*, 2020).

The *TMPRSS2* gene, located at human chromosome 21q22.3, presents 14 exons. TMPRSS2 is an androgen-regulated member composed of 492 amino acids. It was initially classified as a Type II Transmembrane Serine Protease (TTSP) that contains the following main domains: Transmembrane domain (TMD; 84–106), an LDL Receptor Class A domain (LDLRA; 113–148), a Scavenger Receptor Cysteine-Rich (SRCR; 149–242) domain and an Extracellular Serine Protease domain (SPD; 255–492) of the S1 family (Paoloni-Giacobino *et al.*, 1997; Lucas *et al.*, 2008). More recently, just three main functional domains are recognized: an N-terminal LDLRA domain (113–148), followed by SRCR (153–246), and finally, the Extracellular C-terminal Peptidase S1 domain (SPD; 256-487) (Hussain *et al.*, 2020), in which is the central catalytic domain of the molecule. The named catalytic triad of human TMPRSS2 is composed of H296, D345, and S441 residues (Sgrignani & Cavalli, 2021). S441 was demonstrated as the most critical residue for the proteolytic cleavage of viral Spike glycoprotein (Böttcher *et al.*, 2006; Shirogane *et al.*, 2008). Hussain *et al.* (2020) found S441 interacting with several flanking residues of other cleavage sites found in SARS-CoV-2 Spike, indicating the importance of neighbor residues of the catalytic triad in the establishment of the molecular cascade involving cleavage between TMPRSS2 and Spike.

The S1/S2 multibasic cleavage site of SARS-CoV-2 Spike harbors several arginine (R) residues (a characteristic not present in RaTG13). The precise positioning of these proteolytic cleavage sites has been mapped. For instance, Hussain *et al.* (2020) showed that S1/S2 and the S2' cleavage sites are located at the junction of R685/S686 (first cleavage

site) and R815/S816 (second cleavage site) of the SARS-CoV-2 Spike, respectively. Other authors identified the critical amino acids of the S1/S2 cleavage site motif: PRRAR685↓ (Hoffmann *et al.*, 2020; Tang *et al.*,2020), while Essalmani *et al.*, (2021) predicted S2' motif sequence of SARS-CoV-2 as KPSKR815↓SF.

The Histidine (H) residue at position 296, a catalytic site of TMPRSS2, forms a hydrogen bond and electrostatic interaction with R682 of the SARS-CoV-2 Spike, potentially crucial to the first cleavage site R685/S686. Besides, both H296 and S441 establish hydrogen bond interactions with P809, K814, and S810 of the SARS-CoV-2 Spike protein, a connection potentially critical to the second cleavage site R815/S816 (Hussain *et al.*, 2020).

Recent studies with serine protease inhibitor α1AT showed that it inhibits TMPRSS2 action, suppressing SARS-CoV-2 replication in human airway epithelial cell cultures (Wettstein *et al.*, 2021). Other studies, however, showed that when TMPRSS2 was blocked, cathepsin L, an endolysosomal cysteine protease, would not take over and subsequently process SARS-CoV-2 (Koch *et al.*, 2021). Besides, the last authors suggest that SARS-CoV-2 uses two distinct routes to enter cells, one fast (˜10 min), corresponding to the timing of TMPRSS2 activation, and one slower (40–50 min), corresponding to cathepsin L priming. In other words, several host cell proteases can prepare the Spike protein for fusion with the host cell, but it is not yet well known whether this phenomenon requires specific proteases or a coordinated, spatiotemporal combination of multiple proteases (Koch *et al.*, 2021).

Other studies show that infected lung cells (known to express endogenous TMPRSS2) presented more excellent viral replication for Delta than Omicron (Meng *et al.*, 2022). The last authors also observed a lower intracellular S1/S2 cleavage for Omicron versus Delta Spike proteins and that Omicron is inefficient in utilizing TMPRSS2 for its entry into host cells (Meng *et al.*, 2022). The Omicron contains six unique residues in S2 (764K, 796Y, 856K, 954H, 969K, and 981F) that were not previously detected in other VOCs (Meng *et al.*, 2022). Given the knowledge of S1/S2 cleavage, the most likely explanation is that TMPRSS2 sub optimally cleaves S1/S2 in Omicron, resulting in inefficient S2 processing and FP exposure (Meng *et al.*, 2022). Because of this, the Cathepsin B/L molecules are preferentially used by this lineage (Cathepsin B/L pathway (Padmanabhan and Dixit, 2022). Nonetheless the higher rate of transmission detected in

this variant can be explained by a higher affinity for ACE2 compared to Delta and by its hability to evade antybodies (Meng *et al.*, 2022).

It is noteworthy that TMPRSS2 has a critical role in influenza viruses and other coronaviruses (Shen *et al.*, 2017; Liu *et al.*, 2021). So, the evolutionary cooptation of this molecule by viruses goes far beyond the history of SARS-CoV-2 and its host, the *Homo sapiens*.

1.3 Type II Transmembrane Serine Proteases (TTSPs)

Hooper *et al.* (2001) identified that the family of type II transmembrane serine proteases (TTSPs) in placental mammals consists of 17 members, of which seven are found in *Homo sapiens*. Since then, the number of known TTSPs has increased. For instance, Bugge *et al.* (2009) identify that the human TTSP family has 17 members. Some of them were identified in mice (*Mus musculus*). Bugge *et al.* (2009) also showed that TTSPs belong to one of four subfamilies: HAT/DESC (7 genes/proteins), Hepsin/TMPRSS (7), Matriptase (4), and Corin (1). *TMPRSS2* is a member of the Hepsin/TMPRSS family together with *TMPRSS3*, *TMPRSS4*, *TMPRSS5/Spinesin*, MSPL, and Enteropeptidase.

In the present study, we aim to investigate the evolutionary patterns of TMPRSS2 and other members of the TTSP family in placental mammals to explore the amino acid (aa) differences between species and their potential evolutionary significance.

2. Material and Methods

2.1 TTSP family

To have a better panorama of the TTSP family, a protein alignment was made using the algorithm Clustal Omega in the Uniprot server, and a tree was generated by the same algorithm based on the alignment. The *Homo sapiens* paralogous protein sequences were retrieved from the Uniprot database. All the structures of the proteins were assessed using Pfam (Protein Families database), and the similarity between the proteins was checked using the Treefam database. Another alignment was made only with the 30 sites that interact with SARS-CoV-2 to evaluate which proteins are more similar to TMPRSS2 in this region. These 30 sites were described in a study made by Hussain *et al.* (2020) that used docking simulations to provide this information.

2.2 DNA Sequences

The *TMPRSS2* coding sequences were retrieved from the curated library of orthologs from NCBI (n=141, https://www.ncbi.nlm.nih.gov, 20/08/2020) and Ensembl (n= 13, https://www.ensembl.org 20/08/2020, Yates *et al.*, 2020) and also through a blast to retrieve sequences from non-curated primate genomes (n=28) from NCBI. (Supplementary Material, Table S1). We considered the analysis of the set of Primates (n = 56) initially, and later we analyzed it together with the other mammal orders, including six species of Afrotheria, 29 of Rodentia, 28 of Artiodactyla, 28 of Carnivora, 15 of Chiroptera and 20 others (Table S1). All the placental mammals found on the mentioned libraries were used on the study.

Both sets, primates and placental mammals, were aligned using the GUIDANCE2 server (Landan & Graur, 2007; Penn *et al.*, 2010; Sela *et al.*, 2015). The algorithm for multiple sequence alignment Prank was used once it is more accurate to recognize insertions and deletions and thus leads to a better alignment (Loytynoja & Goldman 2008).

To assess human variation, the Ensembl and Uniprot (UniProt: the universal protein knowledgebase in 2021) databases were used.

2.3 Evolutionary Analysis

To assess the evolutionary patterns of the *TMPRSS2* orthologs, we used several tests to estimate the rate of nonsynonymous to synonymous substitutions defined by equation $\omega = dN/dS$ (dN = nonsynonymous and dS = synonymous). The ω values serve as an indicator of negative selection ($\omega < 1$), neutral or relaxed selection ($\omega \approx 1$), or positive selection ($\omega > 1$) acting on sites and/or specific phylogenetic tree branches (Nei & Kumar, 2000; Yang, 2006). We performed a Site Model (NsSite) test with *Codeml* package in PAML 4.9 software (Yang, 2007), which allows ω to vary across sites in the alignment and fit neutral (M1a *and* M7) or positive selection (M2a *and* M8) models to the data using maximum likelihood (Yang, 2007). M2a and M8 models admit positive selection. In cases in which tests indicate models of positive selection were significantly more likely, we used *post hoc* Bayes Empirical Bayes (BEB) to infer individual sites with a high

probability ($p \leq 0.05$) of being under positive selection (Yang *et al.*, 2005), being a Bayesian model the P value was onbtained from de likehood ratio test (LRT) by using the program WINPEPI (Abramson, 2011).

We also used another tool from the Datamonkey server to compare with PAML results and to find more discrete patterns of positive selection: MEME (Mixed Effects Model of Evolution; Murrell *et al.*, 2012), a method to detect pervasive and episodic positive selection (diversifying). In other words, MEME can identify sites where only some branches have experienced selective pressure, with an extensive data set it provides the most power in Datamonkey (Weaver *et al.*, 2018). The MEME test also uses ω as a metric. MEME is also a conceptual advance over the first generation of random effects models designed to detect episodic selection (called "branch-site models" in the literature) (Murrell *et al.*, 2012). We established a *p*-value ≤ 0.1 for MEME for statistical significance thresholds (Spielman *et al.*, 2019).

To assess the impact of substitutions both on family and between species, we used the Grantham score (Grantham, 1974) to compare the substitutions and the classification of Li *et al.* (1984).

2.4 TTSPs bibliography reviewing and classification according to the interaction with SARS-CoV-2

In order to understand our results regarding the TTSPs, we have performed a bibliographic review over the studies involving the family and the SARS-CoV-2. All the articles released involving the proteins of the family and the virus were considered.

All proteins were researched separately and all the studies released after 2019 containing the names of each protein on its content were pre-selected. After careful selection, all the articles related to the theme were read.

After reading all the articles we have classified the TTSPs in three categories regarding its relationship with SARS-CoV-2 infection: Not used; Not definitive evidence; and Strong evidence. Being classified as not definitive evidences the ones that have not involved in vitro cell tests or have controversial results in literature; Not used the ones that have at leas one in vitro test indicating that is not used by the virus, with no controversial results; And Strong evidence the ones that have two or more studies attesting that the protein is used by the virus, also with no controversial results.

The studies used and the classification can be found on table S4.

3. Results

3.1. TTSP family

Here, we identified potential 18 *Homo sapiens* protein coding genes belonging to five TTSP subfamilies (Figure 1): HAT/DESC: TMPRSS11A, TMPRSS11B, TMPRSS11D, TMPRSS11E and TMPRSS11F (TMPRSSC and G were identified only in Rodentia). Hepsin/TMPRSS: TMPRSS1, TMPRSS2, TMPRSS3, TMPRSS4, TMPRSS5, , TMPRSS13 and TMPRSS15; Matriptase: TMPRSS6, TMPRSS7, TMPRSS9 and TMPRSS12, TMPRSS14; and Corin: TMPRSS10.

Noteworthy that we classified TMPRSS12 and TMPRSS14 as belonging to the Matriptase subfamily based on information presented in Uniprot and NCBI GenBank databases (Uniprot, 2021, Sayers *et al.*, 2021, *Homo sapiens* coding sequences).

When the structure is observed, it is possible to notice that TMPRSS12 has lost domains that characterize the other members of the Matriptase subfamily (Figure 2).

We also compared the potential 18 *Homo sapiens* TTSPs (Table 1), considering just the 30 critical human TMPRSS2 sites (all located at the Extracellular Serine Protease domain; SPD; Figure 3) that interact with SARS-CoV-2 to promote the cleavage of its Spike glycoprotein (Hussain *et al.* 2020). Positions 296, 441, 460, 462, and 465 are the most conserved among the *Homo sapiens* TTSPs (Table S2). Two of them (296 and 441) are part of the catalytic triad of human TMPRSS2 (Sgrignani & Cavalli, 2021), with an almost complete fixation in the paralogs of the amino acids histidine (H) and serine (S) (the exception is TMPRSS1). The H296 and S441 are key residues to form a hydrogen bond and electrostatic interaction with specific sites of the SARS-CoV-2 Spike, potentially crucial to the first cleavage site R685/S686 and the second cleavage site R815/S816 (Hussain *et al.*, 2020).

Besides, although ~83% of the 30 sites have different residues among paralogs, changes predicted as radical are rare (< 3%; Table S3).

Three *Homo sapiens* proteins of the family (TMPRSS2, TMPRSS11D, and TMPRSS13) present strong evidence related to the infection, taking into consideration *in vitro* tests with human cells, among others. Seven other proteins (TMPRSS1, TMPRSS3, TMPRSS4, TMPRSS5, TMPRSS7, TMPRSS9, TMPRSS11A, TMPRSS11E,

TMPRSS11F, TMPRSS12, and TMPRSS15) have controversial (not definitive) *in vitro* pieces of evidence of having been co-opted by the SARS-CoV-2 for its processing and subsequent entry into the host cell. In the case of TMPRSS 9 and TMPRSS12, no article was found studying these proteins in the context of this particular cleavage viral mechanism. Considering the four remaining proteins (TMPRSS6, TMPRSS10, TMPRSS11b, and TMPRSS14), *in vitro* evidence indicates that they are not used to process the Spike glycoprotein.

The shreds of evidence considered not definitive used by us were: co-expression with ACE2 (Meng *et al.*, 2020; Salas Orozco *et al.*, 2021; Wruck & Adjaye, 2020), homology with TMPRSS2 (Zhang *et al.*, 2021), and allelic frequency of variants (Vargas-Alarcón *et al.*, 2020). In contrast, the evidence that we considered to be strong came from studies that used *in vitro* human cell lines (Ou *et al.*, 2020; Zang *et al.*, 2020; Tharappel *et al.*, 2020; Laporte *et al.*, 2021; Hoffmann *et al.*, 2021; Kishimoto *et al.*, 2021). The particularities of each protein classification and references are in Table S4.

Figure 4 illustrates the amino acid identity between the human TMPRSS paralogs considering these 30 selected sites. Noteworthy, the tree topology has some discrepancies from Figure 1. There are two clades, and one of them contains all TTSPs with solid evidence that is co-opted by the viruses for its cleavage (TMPRSS2, TMPRSS13, and TMPRSS11D; the first two from the Hepsin subfamily and the last from the Hat/Desc subfamily) while the other contains three of the four described as not co-opted (TMPRSS6, TMPRSS10, TMPRSS11A, all of the different subfamilies). Another TMPRSSs with non-definitive evidence of having been co-opted by the virus are distributed in both groups (Table S4) (Meng *et al.*, 2020; Salas Orozco *et al.*, 2021; Wruck & Adjaye, 2020; Zhang *et al.*, 2021; Vargas-Alarcón *et al.*, 2020; Ou *et al.*, 2020; Zang *et al.*, 2020; Tharappel *et al.*, 2020; Laporte *et al.*, 2021; Hoffmann *et al.*, 2021; Kishimoto *et al.*, 2021). This result indicates that it can have a minimum amino acid framework necessary for evolutionary co-opting of TMPRSS members by viruses since there is identity between members of different subfamilies. However, the TMPRSS2 and its closest phylogenetically paralog, TMPRSS3 (Figure 1), remain together in the tree considering the 30 key sites to promote the cleavage of the Spike (Figure 4), although studies with the latter show no definitive evidence of it had been co-opted (Table S4). Noteworthy, TMPRSS3 has an LDL receptor

absent when compared with TMPRSS2, which probably interferes with its ability to be used by the SARS-CoV-2 to cleave the Spike.

These results highlight the complexity of the scenario involving the critical elements for the virus's evolutionary co-optation of the host molecules. Table 1 shows the amino acid classification based on side chains, considering these 30 sites. No variation is found in the two-site members of the catalytic triad of TMPRSS2 (296 and 441; Sgrignani & Cavalli, 2021). Besides these, 460 and 465 also have the same amino acid in all paralogs, contrasting with a notable difference in the others. Interestingly at position 387, all the TTSPs with solid evidence that are possibly co-opted have an amino acid with a polar neutral side chain, while a hydrophobic side chain is prevalent in those with no signs of being used by the virus (Table 1).

3.2 *TMPRSS2* evolution in placental mammals and primates

Considering the mammal set (MS) and Primates (PS) as two subsets, the PALM model that best fits in both cases is M8 ($p < 0.0001$) in comparison to M7 (Table 2), indicating that the action of positive selection can explain the variation found between species. Under the M8 model, the BEB algorithm identified only two sites with high probability to are under positive selection (Tables 3, S5 and S6): PS:173 *(p*=0,01) and 358 (*p*=0,03); MS: 214 *(p*=0,0) and 360 (*p*=0,05). Sites 173 and 214 are located in the SRCR domain, an ancient and highly conserved protein module (Sarrias *et al.*, 2004), while 358 and 360 at the Extracellular Serine Peptidase domain (SPD). These sites are in regions that form disulfide bonds (172 ↔ 231, 244 ↔ 365) or nest to a glycosylation site (213) (UniProt Consortium, 2018).

As expected, MEME indicated much more sites under positive selection, possibly because the test can detect episodic selection in specific branches of a given phylogeny (Tables 3, S5, and S6). MEME matches the performance of older approaches, such as PALM, when natural selection is pervasive but possesses greater power to identify sites where episodes of positive selection are confined to a small subset of branches in a phylogenetic tree (Murrell *et al.*, 2012).

Note that the variation at site 173 in primates can be explained by a positive selection action detected by PALM and MEME tests. Site 173 is not listed as a key to

Spike's cleavage (Hussain *et al.*, 2020), but as already commented, it is next to a site that forms a disulfide bond. A consultation of public databases (https://www.ncbi.nlm.nih.gov/gene/7113), considering the most studied primate of all, *Homo sapiens*, indicates that only one missense polymorphism at site 173 was identified (a minor allele frequency of ~0.04% in Africa; https://www.ncbi.nlm.nih.gov/snp/rs1189756287). No clinical relevance has been reported regarding this variation. This finding suggests that some evolutionary constriction operates within our species, reinforcing that amino acid changes in primates may be taxon-specific with potential functional and evolutionary implications.

When MS is considered, six of the TMPRSS2 sites under positive selection are known to be critical for the proteolytic cleavage of SARS-CoV-2 Spike glycoprotein according to Hussain *et al.* (2020): 299 (glutamic acid, E, in human sequence), 340 (K), 389 (E), 413 (R), 431 (Q) and 438 (Q). Across the species, these six sites presented many variations, particularly at positions 340, 413, and 431. For instance, at position 340, the amino acid residues differ from *Homo sapiens* in 129 species, the most common R>K (n=126), but it is a conservative substitution. Other sites at positions 299, 389, and 438 have few or unique changes across all species. Thus, this last result is probably due to an episodic case of positive selection when only a few species or branches of the phylogeny have amino acid variants.

On the other hand, no important site for SARS-CoV-2 Spike cleavage is under positive selection considering PS.

.

4. Discussion

The last update regarding the TTSP family was published in 2009 (Bugge *et al.*, 2009). A new protein belonging to the family was identified since that publication, TMPRSS12 (The MGC Project Team, 2004). However, although 33 articles were published in NCBI and 123 on Google scholar (access 09/12/2021), none of them explores evolutionary issues regarding TTSP family members.

We detected that the phylogenetic tree using 18 *Homo sapiens* TMPRSS coding genes (Figure 1) has a different topology than that constructed using only the data of the 30

key sites that interact with the SARS-CoV-2. For instance, Figure 1 shows clades with group members belonging to each subfamily. In contrast, Figure 4 presents the proteins already associated with the infection TMPRSS2 and TMPRSS11D (Hepsin subfamily), as well as TMPRSS13 (Hat/Desc subfamily) into a clade along with others whose evolutionary information on co-option (or not) by the virus is not yet well defined. Another clade contains three of the four described as not co-opted (TMPRSS6, TMPRSS10, TMPRSS11A), all different subfamilies. These findings suggest that it could have standard components (*e.g.*, a combination of amino acids with specific profiles) in the different subfamilies that are potential of "interest" to the virus in its evolutionary strategy of co-opting host molecules to promote the infection and its replication.

Coronaviridae (COVs) may have originated millions of years ago (Wertheim *et al.*, 2013); thus, their mammalian hosts have been under attack for an equivalent time. We can assume that strategies of co-opting molecules from the host by COVs, concomitant with host strategies to circumvent this kind of infection, have been shaped by natural selection for much longer than the recent history of the SARS-CoV-2 and of the *Homo sapiens*. The TMPRSS2 evolution is not driven only by the long history of coexistence of the *Homo* genus and the *Coronaviridae viruses* but also by a potential ancient arms race between potential mammalian hosts and several viruses and other virus families (like Orthomyxoviridae; which includes the four genera of Influenza; Shen *et al.*, 2017; Liu *et al.*, 2021). Besides, the natural pressure under the gene's normal function is also expected. Thus, it is not surprising that our tests showed sites whose amino the action of positive selection can explain acid variation between species (33 sites considering mammalian data set, and 18 considering Primates data set; Table 3).

However, it is notable that of all these sites with a signature of the diversifying positive selection, including episodic, only 20% of those (299, 340, 389, 413, 431, and 438) are co-opted for SARS-CoV-2 Spike (and potentially other COVs; Shen *et al.*, 2017) cleavage, considering placental mammals data set. On the other hand, none of them appear when Primate's data set is considered. In this scenario and considering only the 30 sites mentioned by Hussain *et al.* (2020) as key to the cleavage of Spike glycoprotein, primates would be more susceptible to SARS-CoV-2 (or some new lineages or even a COV derived from it), as they do not present variation in any of these sites, at least considering variation promoted by selective action. This pattern of variation suggests that perhaps most of the

variation between species is the result of selective pressures that have been shaping the normal function of TMPRSS2 and its orthologs in the cells of the corresponding species. On the other hand, it also indicates that the virus, through its Spike glycoprotein (that needs cleavage), would be evolutionarily shaped to bind at positions in potential mammalian host proteases that are less likely to have variation promoted by positive selection action due to demand for host cell functions. This would give the virus an advantage, as it would, for example, have less chance of losing affinity with the host while giving more chances for zoonotic jumps.

Functional studies (Darrell *et al.*, 2022) demonstrate that SARS-CoV-2 could replicate to high levels in cell lines expressing the *TMPRSS2* gene from cat (*Felis catus*), goat (*Capra hircus*) and golden hamster (*Mesocricetus auratus*), despite of some differences found at positions 299, 340, 389, 413, 431, and 438 of the respective orthologs, regarding TMPRSS2 *Homo sapiens* TMPRSS2 (Table S5).

Another aspect that must be considered is that the action of a protein can be superimposed by its paralogs, as in the case of the TTSP family members. For instance, Kim *et al.* (2006) generated *TMPRSS2*−/− mice and compared it to wild-type littermates. The authors described that *TMPRSS2*−/− mice usually developed, survived to adulthood with no differences in protein levels of prostatic secretions, and exhibited no discernible abnormalities in organ histology or function. As a result, they conclude that the lack of a discernible phenotype in *TMPRSS2*−/− mice suggest functional redundancy involving one or more of the TTSPs members or other serine proteases in specific branches of a given phylogeny. Abnormal expression of human TMPRSS2 was closely related to tumor growth, invasion, metastasis, and prognosis in various cancers, especially prostate cancer (Liu *et al.*, 2021). Alternatively, some authors postulated that TMPRSS2 might contribute a specialized but non vital function apparent only in stress, disease, or other systemic perturbation (Kim *et al.*, 2006). Nevertheless, if this is the case for tumors after reproductive age, changes in this gene may not necessarily have evolutionary relevance to the host organisms, at least considering humans. On the other hand, examples such as position 173 of TMPRSS2 in primates make it difficult to suppose that there is no functional and evolutionary relevance of this enzyme, at least in humans or other primates.

Our thought-provoking findings may reflect some pattern shaped by an ancient and continued biological arms race between viruses (COVs and other viruses) and their

potential mammalian hosts. Nevertheless, as evidenced by the studies mentioned in the paragraphs above, it is hard to define precisely the causes behind the positive selection pattern found for us regarding placental mammals. In other words, it is due just to a long-term evolutionary arms race between viruses and their hosts or the normal environmental pressure considering the TMPRSS2 performance in each potential host/clade. Furthermore, both scenarios are not mutually exclusive.

We must also compute the current stage of SARS-CoV-2 evolution, illustrated by the fact that Omicron primarily uses the Cathepsin B/L pathway (TMPRSS2 independent) for its entry into the host cell (Meng *et al.*, 2022; Padmanabhan and Dixit, 2022) since there is a suboptimal S1/S2 cleavage in Omicron when TMPRSS2 is triggered (Meng *et al.*, 2022). Furthermore, there are the "*Homo sapiens* reactions". For instance, most individuals affected by COVID-19 survive because their immune responses adequately fight the infection, probably reflecting the long history of COVs attacks. The immune response is also induced by the current large-scale SARS-CoV-2 vaccine deployment, a new and potent element in the effective host reaction. In other words, old and new elements are also part of this complex scenario.

Despite these difficulties, a clear evolutionary pattern can be observed in other cases (Fam *et al.*, 2020; Landau *et al.*, 2022). In other words, variable sites due to the positive selection are always a minority among those in the host molecules that the virus co-opts for its efficient infection and replication.

From the point of view of viral evolution, we can assume that there is a preference for co-opting conserved sites intra and inter-species, thus preserving affinity over time and facilitating the jump across potentially host mammalian species (spillover).In summary, SARS-CoV-2 is a specialist for infecting humans due to the particular evolutionary events keeping its relatively low lethality, high transmissibility, and evolvability (Yépez *et al.*, 2022). In this context, the successful co-option of all host molecules needs to promote the infection, and viral replication is a crucial element. One of these molecules is TMPRSS2, whose evidence of being used by processing the SARS-CoV-2 Spike glycoprotein is consolidated. In the present study, we detect the signature of the positive selection action modulating TMPRSS2 diversity among mammalian species, including primates, which

may reflect adaptation for normal enzyme functions in the host organism and a long-term evolutionary arms race between viruses and these mammalian hosts.

5.References

Alfaraj, S. H., Al-Tawfiq, J. A., Assiri, A. Y., Alzahrani, N. A., Alanazi, A. A., & Memish, Z. A. (2019). Clinical predictors of mortality of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) infection: A cohort study. *Travel Medicine and Infectious Disease*, *29*, 48–50. https://doi.org/10.1016/j.tmaid.2019.03.004

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature Medicine*, *26*(450–452). https://doi.org/10.1038/s41591-020-0820-9

Asselta, R., Paraboschi, E. M., Mantovani, A., & Duga, S. (2020). ACE2 *and* TMPRSS2 *variants and expression as candidates to sex and country differences in COVID-19 severity in Italy*. https://doi.org/10.1101/2020.03.30.20047878

Awadasseid, A., Wu, Y., Tanaka, Y., & Zhang, W. (2021). SARS-CoV-2 variants evolved during the early stage of the pandemic and effects of mutations on adaptation in Wuhan populations. *International Journal of Biological Sciences*, *17*(1), 97–106. https://doi.org/10.7150/ijbs.47827

Berman, H. M. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. https://doi.org/10.1093/nar/28.1.235

Bestle, D., Heindl, M. R., Limburg, H., Van Lam van, T., Pilgram, O., Moulton, H., Stein, D. A., Hardes, K., Eickmann, M., Dolnik, O., Rohde, C., Klenk, H.-D., Garten, W., Steinmetzer, T., & Böttcher-Friebertshäuser, E. (2020). TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Science Alliance*, *3*(9), e202000786. https://doi.org/10.26508/lsa.202000786

Bhattacharyya, C., Das, C., Ghosh, A., Singh, A. K., Mukherjee, S., Majumder, P. P., Basu, A., & Biswas, N. K. (2020). *Global Spread of SARS-CoV-2 Subtype with Spike Protein Mutation D614G is Shaped by Human Genomic Variations that Regulate Expression of* TMPRSS2 *and* MX1 *Genes*. https://doi.org/10.1101/2020.05.04.075911

Bhattacharyya, R. P., & Hanage, W. P. (2022). Challenges in Inferring Intrinsic Severity of the SARS-CoV-2 Omicron Variant. *New England Journal of Medicine*. https://doi.org/10.1056/nejmp2119682

BöttcherE., Matrosovich, T., Beyerle, M., Klenk, H.-D., Garten, W., & Matrosovich, M. (2006). Proteolytic Activation of Influenza Viruses by Serine Proteases TMPRSS2 and HAT from Human Airway Epithelium. *Journal of Virology*, *80*(19), 9896–9898. https://doi.org/10.1128/jvi.01118-06

Bugge, T. H., Antalis, T. M., & Wu, Q. (2009). Type II Transmembrane Serine Proteases. *Journal of Biological Chemistry*, *284*(35), 23177–23181. https://doi.org/10.1074/jbc.r109.021006

Cao, B., Zhang, L., Liu, H., Ma, S., & Mi, K. (2021). The Dynamic Expression of Potential Mediators of Severe Acute Respiratory Syndrome Coronavirus 2 Cellular Entry in Fetal, Neonatal, and Adult Rhesus Monkeys. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.607479

Chen, Z., Song, X., Li, Q., Xie, L., Guo, T., Su, T., Tang, C., Chang, X., Liang, B., & Huang, D. (2019). Androgen Receptor-Activated Enhancers Simultaneously Regulate Oncogene TMPRSS2 and lncRNA PRCAT38 in Prostate Cancer. *Cells*, *8*(8). https://doi.org/10.3390/cells8080864

Chi, X., Yan, R., Zhang, J., Zhang, G., Zhang, Y., Hao, M., Zhang, Z., Fan, P., Dong, Y., Yang, Y., Chen, Z., Guo, Y., Zhang, J., Li, Y., Song, X., Chen, Y., Xia, L., Fu, L., Hou, L., & Xu, J. (2020). A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science*, *369*(6504), 650–655. https://doi.org/10.1126/science.abc6952

Coden, M. E., Loffredo, L. F., Abdala-Valencia, H., & Berdnikovs, S. (2021). Comparative Study of SARS-CoV-2, SARS-CoV-1, MERS-CoV, HCoV-229E and Influenza Host Gene Expression in Asthma: Importance of Sex, Disease Severity, and Epithelial Heterogeneity. *Viruses*, *13*(6), 1081. https://doi.org/10.3390/v13061081

Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N. G., & Decroly, E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Research*, *176*, 104742. https://doi.org/10.1016/j.antiviral.2020.104742

ECDC. European Centre for Disease Prevention and Control. (2021a, January 21). *Risk Assessment: Risk related to the spread of new SARS-CoV-2 variants of concern in the EU/EEA – first update*. European Centre for Disease Prevention and Control. https://www.ecdc.europa.eu/en/publications-data/covid-19-risk-assessment-spread-new-variants-concern-eueea-first-update

ECDC. European Centre for Disease Prevention and Control. (2021b, July 5). *Geographical distribution of 2019-nCov cases globally*. European Centre for Disease Prevention and Control. https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases

Essalmani, R., Jain, J., Susan-Resiga, D., Andréo, U., Evagelidis, A., Derbali, R. M., Huynh, D. N., Dallaire, F., Laporte, M., Delpal, A., Sutto-Ortiz, P., Coutard, B., Mapa, C., Wilcoxen, K., Decroly, É., Pham, T. N., Cohen, É. A., & Seidah, N. G. (2021). Implications of Spike-glycoprotein processing at S1/S2 by Furin, at S2' by Furin and/or TMPRSS2 and shedding of ACE2: cell-to-cell fusion, cell entry and infectivity of SARS-CoV-2. *BioRiv*. https://doi.org/10.1101/2021.07.02.450896

Fagerberg, L., Hallström, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjöstedt, E., Lundberg, E., Szigyarto, C. A.-K., Skogs, M., Takanen, J. O., Berling, H., Tegel, H., Mulder, J., & Nilsson, P. (2013). Analysis of the Human Tissue-specific Expression by Genome-wide Integration of Transcriptomics and Antibody-based Proteomics. *Molecular & Cellular Proteomics*, *13*(2), 397–406. https://doi.org/10.1074/mcp.m113.035600

Fam, B. S. O., Vargas-Pinilla, P., Amorim, C. E. G., Sortica, V. A., & Bortolini, M. C. (2020). ACE2 diversity in placental mammals reveals the evolutionary strategy of SARS-CoV-2. *Genetics and Molecular Biology*, *43*(2). https://doi.org/10.1590/1678-4685-gmb-2020-0104

Fraser, B., Beldar, S., Hutchinson, A., Li, Y., Seitova, A., Edwards, A. M., Benard, F., Arrowsmith, C. H., & Halabelian, L. (2021). *Crystal structure of human TMPRSS2 in complex with Nafamostat*. https://doi.org/10.2210/pdb7meq/pdb

Google. (2021). *Google Scholar*. Google Scholar; Google. https://scholar.google.com.br/

Grantham, R. (1974). Amino acid difference formula to help explain protein evolution. *Science (New York, N.Y.)*, *185*(4154), 862–864. https://doi.org/10.1126/science.185.4154.862

Guaita Martínez, J. M., Carracedo, P., Gorgues Comas, D., & Siemens, C. H. (2022). An analysis of the blockchain and COVID-19 research landscape using a bibliometric study. *Sustainable Technology and Entrepreneurship*, *1*(1), 100006. https://doi.org/10.1016/j.stae.2022.100006

Harrison, S. C. (2015). Viral membrane fusion. *Virology*, *0*, 498–507. https://doi.org/10.1016/j.virol.2015.03.043

Hartenian, E., Nandakumar, D., Lari, A., Ly, M., Tucker, J. M., & Glaunsinger, B. A. (2020). The molecular virology of Coronaviruses. *Journal of Biological Chemistry*, *295*(37), jbc.REV120.013930. https://doi.org/10.1074/jbc.rev120.013930

Hoffmann, M., Hofmann-Winkler, H., Smith, J. C., Krüger, N., Arora, P., Sørensen, L. K., Søgaard, O. S., Hasselstrøm, J. B., Winkler, M., Hempel, T., Raich, L., Olsson, S., Danov, O., Jonigk, D., Yamazoe, T., Yamatsuta, K., Mizuno, H., Ludwig, S., Noé, F., & Kjolby, M. (2021). Camostat mesylate inhibits SARS-CoV-2 activation by TMPRSS2-related proteases and its metabolite GBPA exerts antiviral activity. *EBioMedicine*, *65*(103255). https://doi.org/10.1016/j.ebiom.2021.103255

Hoffmann, M., Kleine-Weber, H., & Pöhlmann, S. (2020). A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Molecular Cell*, *78*(4). https://doi.org/10.1016/j.molcel.2020.04.022

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N.-H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, *181*(2). https://doi.org/10.1016/j.cell.2020.02.052

Hooper, J. D., Clements, J. A., Quigley, J. P., & Antalis, T. M. (2001). Type II Transmembrane Serine Proteases. *Journal of Biological Chemistry*, *276*(2), 857–860. https://doi.org/10.1074/jbc.r000020200

Huang, N., Perez, P., Kato, T., Mikami, Y., Okuda, K., Gilmore, R. C., Domínguez Conde, C., Gasmi, B., Stein, S., Beach, M., Pelayo, E., Maldonado, J., LaFont, B., Padilla, R., Murrah, V., Maile, R., Lovell, W., Wallet, S., Bowman, N. M., & Meinig, S. L.

(2020). Integrated Single-Cell Atlases Reveal an Oral SARS-CoV-2 Infection and Transmission Axis. *MedRxiv*. https://doi.org/10.1101/2020.10.26.20219089

Hulo, N. (2006). The PROSITE database. *Nucleic Acids Research*, *34*(90001), D227–D230. https://doi.org/10.1093/nar/gkj063

Hunt, S. E., McLaren, W., Gil, L., Thormann, A., Schuilenburg, H., Sheppard, D., Parton, A., Armean, I. M., Trevanion, S. J., Flicek, P., & Cunningham, F. (2018). Ensembl variation resources. *Database*, *2018*. https://doi.org/10.1093/database/bay119

Hussain, M., Jabeen, N., Amanullah, A., Ashraf Baig, A., Aziz, B., Shabbir, S., Raza, F., & Uddin, N. (2020). Molecular docking between human TMPRSS2 and SARS-CoV-2 spike protein: conformation and intermolecular interactions. *AIMS Microbiology*, *6*(3), 350–360. https://doi.org/10.3934/microbiol.2020021

Kapczynski, D. R., Sweeney, R., Spackman, E., Pantin-Jackwood, M., & Suarez, D. L. (2022). Development of an in vitro model for animal species susceptibility to SARS-CoV-2 replication based on expression of ACE2 and TMPRSS2 in avian cells. *Virology*, *569*, 1–12. https://doi.org/10.1016/j.virol.2022.01.014

Ke, Z., Oton, J., Qu, K., Cortese, M., Zila, V., McKeane, L., Nakane, T., Zivanov, J., Neufeldt, C. J., Cerikan, B., Lu, J. M., Peukes, J., Xiong, X., Kräusslich, H.-G., Scheres, S. H. W., Bartenschlager, R., & Briggs, J. A. G. (2020). Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature*, *588*(498-502), 1–5. https://doi.org/10.1038/s41586-020-2665-2

Kim, T. S., Heinlein, C., Hackman, R. C., & Nelson, P. S. (2006). Phenotypic Analysis of Mice Lacking the Tmprss2-Encoded Protease. *Molecular and Cellular Biology*, *26*(3), 965–975. https://doi.org/10.1128/mcb.26.3.965-975.2006

Kishimoto, M., Uemura, K., Sanaki, T., Sato, A., Hall, W. W., Kariwa, H., Orba, Y., Sawa, H., & Sasaki, M. (2021). TMPRSS11D and TMPRSS13 Activate the SARS-CoV-2 Spike Protein. *Viruses*, *13*(3), 384. https://doi.org/10.3390/v13030384

Koch, J., Uckeley, Z. M., Doldan, P., Stanifer, M., Boulant, S., & Lozach, P. (2021). TMPRSS2 expression dictates the entry route used by SARS-CoV-2 to infect host cells. *The EMBO Journal*, *40*(16). https://doi.org/10.15252/embj.2021107821

Krishnamoorthy, P., Raj, A. S., Roy, S., Kumar, N. S., & Kumar, H. (2021). Comparative transcriptome analysis of SARS-CoV, MERS-CoV, and SARS-CoV-2 to identify

potential pathways for drug repurposing. *Computers in Biology and Medicine*, *128*, 104123. https://doi.org/10.1016/j.compbiomed.2020.104123

Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution*, *34*(7), 1812–1819. https://doi.org/10.1093/molbev/msx116

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., & Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, *581*, 215–220. https://doi.org/10.1038/s41586-020-2180-5

Landan, G., & Graur, D. (2007). Local relianility measures from sets of co-optimal multiple sequence alignments. *Biocomputing 2008*. https://doi.org/10.1142/9789812776136_0003

Laporte, M., Raeymaekers, V., Van Berwaer, R., Vandeput, J., Marchand-Casas, I., Thibaut, H.-J., Van Looveren, D., Martens, K., Hoffmann, M., Maes, P., Pöhlmann, S., Naesens, L., & Stevaert, A. (2021). The SARS-CoV-2 and other human coronavirus spike proteins are fine-tuned towards temperature and proteases of the human airways. *PLOS Pathogens*, *17*(4), e1009500. https://doi.org/10.1371/journal.ppat.1009500

Latini, A. O., Souza, C. L., Resende, D. C., Carvalho, D. B. F., & Lopes, C. F. G. (2020). Obtenção On-line (https://covid-19.ufsj.tech) do Número de Leitos Municipais Demandados por COVID-19: popularização da ciência e apoio a políticas públicas. *Cadernos de Prospecção*, *13*(4), 907. https://doi.org/10.9771/cp.v13i4.37135

Lechien, J. R., Radulesco, T., Calvo-Henriquez, C., Chiesa-Estomba, C. M., Hans, S., Barillari, M. R., Cammaroto, G., Descamps, G., Hsieh, J., Vaira, L., De Riu, G., Sowerby, L., Gengler, I., Michel, J., & Saussez, S. (2020). ACE2 & TMPRSS2 Expressions in Head & Neck Tissues: A Systematic Review. *Head and Neck Pathology*, *15*(1), 225–235. https://doi.org/10.1007/s12105-020-01212-5

Li, F., Han, M., Dai, P., Xu, W., He, J., Tao, X., Wu, Y., Tong, X., Xia, X., Guo, W., Zhou, Y., Li, Y., Zhu, Y., Zhang, X., Liu, Z., Aji, R., Cai, X., Li, Y., Qu, D., & Chen, Y. (2021). Distinct mechanisms for TMPRSS2 expression explain organ-specific inhibition of SARS-CoV-2 infection by enzalutamide. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-21171-x

Li, W.-H., Wu, C.-I., & Luo, C.-C. (1984). Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *Journal of Molecular Evolution*, *21*(1), 58–71. https://doi.org/10.1007/bf02100628

Lin, B., Ferguson, C., White, J. T., Wang, S., Vessela, R., True, L. D., Hood, L., & Nelson, P. S. (1999). Prostate-localized and Androgen-regulated Expression of the Membrane-bound Serine Protease TMPRSS2. *Cancer Research*, *59*(17), 4180–4184. https://aacrjournals.org/cancerres/article/59/17/4180/505436/Prostate-localized-and-Androgen-regulated

Lin, B., White, J. T., Utleg, A. G., Wang, S., Ferguson, C., True, L. D., Vessella, R., Hood, L., & Nelson, P. S. (2003). Isolation and characterization of human and mouse WDR19,a novel WD-repeat protein exhibiting androgen-regulated expression in prostate epithelium☆. *Genomics*, *82*(3), 331–342. https://doi.org/10.1016/s0888-7543(03)00151-4

Liu, L., Qin, J., Zuo, M., & Zhou, Q. (2021). Multi-omics of the expression and clinical outcomes of *TMPRSS2* in human various cancers: A potential therapeutic target for COVID-19. *Journal of Cellular and Molecular Medicine*, *26*(3), 709–724. https://doi.org/10.1111/jcmm.17090

Loytynoja, A., & Goldman, N. (2008). Phylogeny-Aware Gap Placement Prevents Errors in Sequence Alignment and Evolutionary Analysis. *Science*, *320*(5883), 1632–1635. https://doi.org/10.1126/science.1158395

Lu, S., Zhao, Y., Yu, W., Yang, Y., Gao, J., Wang, J., Kuang, D., Yang, M., Yang, J., Ma, C., Xu, J., Qian, X., Li, H., Zhao, S., Li, J., Wang, H., Long, H., Zhou, J., Luo, F., & Ding, K. (2020). Comparison of nonhuman primates identified the suitable model for COVID-19. *Signal Transduction and Targeted Therapy*, *5*(1), 1–9. https://doi.org/10.1038/s41392-020-00269-6

Lucas, J., True, L., Hawley, S., Matsumura, M., Morrissey, C., Vessella, R., & Nelson, P. (2008). The androgen-regulated type II serine protease TMPRSS2 is differentially expressed and mislocalized in prostate adenocarcinoma. *The Journal of Pathology*, *215*(2), 118–125. https://doi.org/10.1002/path.2330

Madhi, S. A., Kwatra, G., Myers, J. E., Jassat, W., Dhar, N., Mukendi, C. K., Nana, A. J., Blumberg, L., Welch, R., Ngorima-Mabhena, N., & Mutevedzi, P. C. (2021). South

African Population Immunity and Severe Covid-19 with Omicron Variant. *MedRxiv*. https://doi.org/10.1101/2021.12.20.21268096

Malaiyan, J., Arumugam, S., Mohan, K., & Gomathi Radhakrishnan, G. (2020). An update on the origin of SARS-CoV-2: Despite closest identity, bat (RaTG13) and pangolin derived coronaviruses varied in the critical binding site and O-linked glycan residues. *Journal of Medical Virology*, *93*, 499–505. https://doi.org/10.1002/jmv.26261

Markov, P. V., Katzourakis, A., & Stilianakis, N. I. (2022). Antigenic evolution will lead to new SARS-CoV-2 variants with unpredictable severity. *Nature Reviews Microbiology*, *20*, 251–252. https://doi.org/10.1038/s41579-022-00722-z

Meng, B., Abdullahi, A., Ferreira, I. A. T. M., Goonawardane, N., Saito, A., Kimura, I., Yamasoba, D., Gerber, P. P., Fatihi, S., Rathore, S., Zepeda, S. K., Papa, G., Kemp, S. A., Ikeda, T., Toyoda, M., Tan, T. S., Kuramochi, J., Mitsunaga, S., Ueno, T., & Shirakawa, K. (2022). Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts tropism and fusogenicity. *Nature*. https://doi.org/10.1038/s41586-022-04474-x

Meng, T., Cao, H., Zhang, H., Kang, Z., Xu, D., Gong, H., Wang, J., Li, Z., Cui, X., Xu, H., Wei, H., Pan, X., Zhu, R., Xiao, J., Zhou, W., Cheng, L., & Liu, J. (2020). The insert sequence in SARS-CoV-2 enhances spike protein cleavage by TMPRSS. *BioRxiv*. https://doi.org/10.1101/2020.02.08.926006

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1), D412–D419. https://doi.org/10.1093/nar/gkaa913

Mousavizadeh, L., & Ghasemi, S. (2020). Genotype and phenotype of COVID-19: Their roles in pathogenesis. *Journal of Microbiology, Immunology and Infection*, *54*(2). https://doi.org/10.1016/j.jmii.2020.03.022

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S. L. (2012). Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLoS Genetics*, *8*(7), e1002764. https://doi.org/10.1371/journal.pgen.1002764

Murza, A., Dion, S. P., Boudreault, P.-L., Désilets, A., Leduc, R., & Marsault, É. (2020). Inhibitors of type II transmembrane serine proteases in the treatment of diseases of

the respiratory tract – A review of patent literature. *Expert Opinion on Therapeutic Patents*, *30*(11), 807–824. https://doi.org/10.1080/13543776.2020.1817390

National Center for Biotechnology Information. (2021). *Home - PMC - NCBI*. Nih.gov; U.S. National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc

NCBI. (2019). *National Center for Biotechnology Information*. Nih.gov. https://www.ncbi.nlm.nih.gov/

Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., Guo, L., Guo, R., Chen, T., Hu, J., Xiang, Z., Mu, Z., Chen, X., Chen, J., Hu, K., Jin, Q., Wang, J., & Qian, Z. (2020). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-15562-9

Padmanabhan, P., & Dixit, N. M. (2022). Evidence of increased Cathepsin B/L and decreased TMPRSS2 usage for cell entry by the SARS-CoV-2 Omicron variant. *BioRxiv*. https://doi.org/10.1101/2022.01.13.476267

Peng, R., Wu, L.-A., Wang, Q., Qi, J., & Gao, G. F. (2021). Cell entry by SARS-CoV-2. *Trends in Biochemical Sciences*, *46*(10), 848–860. https://doi.org/10.1016/j.tibs.2021.06.001

Penn, O., Privman, E., Ashkenazy, H., Landan, G., Graur, D., & Pupko, T. (2010). GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Research*, *38*(Web Server issue), W23–W28. https://doi.org/10.1093/nar/gkq443

Petersen, E., Koopmans, M., Go, U., Hamer, D. H., Petrosillo, N., Castelli, F., Storgaard, M., Al Khalili, S., & Simonsen, L. (2020). Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *The Lancet Infectious Diseases*, *20*(9). https://doi.org/10.1016/s1473-3099(20)30484-9

Rigden, D. J., & Fernández, X. M. (2020). The 2021 Nucleic Acids Research database issue and the online molecular biology database collection. *Nucleic Acids Research*, *49*(D1), D1–D9. https://doi.org/10.1093/nar/gkaa1216

Saito, A., Irie, T., Suzuki, R., Maemura, T., Nasser, H., Uriu, K., Kosugi, Y., Shirakawa, K., Sadamasu, K., Kimura, I., Ito, J., Wu, J., Iwatsuki-Horimoto, K., Ito, M., Yamayoshi, S., Loeber, S., Tsuda, M., Wang, L., Ozono, S., & Butlertanaka, E. P.

(2021). Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. *Nature*, *602*, 300–306. https://doi.org/10.1038/s41586-021-04266-9

Salas Orozco, M. F., Niño-Martínez, N., Martínez-Castañón, G.-A., Patiño Marín, N., Sámano Valencia, C., Dipp Velázquez, F. A., Sosa Munguía, P. del C., & Casillas Santana, M. A. (2021). Presence of SARS-CoV-2 and Its Entry Factors in Oral Tissues and Cells: A Systematic Review. *Medicina*, *57*(6), 523. https://doi.org/10.3390/medicina57060523

Sayers, E. W., Cavanaugh, M., Clark, K., Pruitt, K. D., Schoch, C. L., Sherry, S. T., & Karsch-Mizrachi, I. (2020). GenBank. *Nucleic Acids Research*, *49*(D1), D92–D96. https://doi.org/10.1093/nar/gkaa1023

Sehnal, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., Velankar, S., Burley, S. K., Koča, J., & Rose, A. S. (2021). Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Research*, *49*(W1). https://doi.org/10.1093/nar/gkab314

Sela, I., Ashkenazy, H., Katoh, K., & Pupko, T. (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research*, *43*(W1), W7–W14. https://doi.org/10.1093/nar/gkv318

Sgrignani, J., & Cavalli, A. (2021). Computational Identification of a Putative Allosteric Binding Pocket in TMPRSS2. *Frontiers in Molecular Biosciences*, *8*. https://doi.org/10.3389/fmolb.2021.666626

Shen, L. W., Mao, H. J., Wu, Y. L., Tanaka, Y., & Zhang, W. (2017). TMPRSS2: A potential target for treatment of influenza virus and coronavirus infections. *Biochimie*, *142*, 1–10. https://doi.org/10.1016/j.biochi.2017.07.016

Shirogane, Y., Takeda, M., Iwasaki, M., Ishiguro, N., Takeuchi, H., Nakatsu, Y., Tahara, M., Kikuta, H., & Yanagi, Y. (2008). Efficient Multiplication of Human Metapneumovirus in Vero Cells Expressing the Transmembrane Serine Protease TMPRSS2. *Journal of Virology*, *82*(17), 8942–8946. https://doi.org/10.1128/jvi.00676-08

Singh, D., & Yi, S. V. (2021). On the origin and evolution of SARS-CoV-2. *Experimental & Molecular Medicine*, *53*. https://doi.org/10.1038/s12276-021-00604-z

Song, J., Li, Y., Huang, X., Chen, Z., Li, Y., Liu, C., Chen, Z., & Duan, X. (2020). Systematic Analysis of ACE2 and TMPRSS2 Expression in Salivary Glands Reveals Underlying Transmission Mechanism Caused by SARS-CoV-2. *Journal of Medical Virology*, *92*(11). https://doi.org/10.1002/jmv.26045

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–2729. https://doi.org/10.1093/molbev/mst197

Tang, T., Bidon, M., Jaimes, J. A., Whittaker, G. R., & Daniel, S. (2020). Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Research*, *178*, 104792. https://doi.org/10.1016/j.antiviral.2020.104792

Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., Duan, Y., Zhang, H., Wang, Y., Qian, Z., Cui, J., & Lu, J. (2020). On the origin and continuing evolution of SARS-CoV-2. *National Science Review*, *7*(6). https://doi.org/10.1093/nsr/nwaa036

Tharappel, A. M., Samrat, S. K., Li, Z., & Li, H. (2020). Targeting Crucial Host Factors of SARS-CoV-2. *ACS Infectious Diseases*, *6*(11), 2844–2865. https://doi.org/https://dx.doi.org/10.1021/acsinfecdis.0c00456?ref=pdf

The MGC Project Team. (2004). The Status, Quality, and Expansion of the NIH Full-Length cDNA Project: The Mammalian Gene Collection (MGC). *Genome Research*, *14*(10b), 2121–2127. https://doi.org/10.1101/gr.2596504

Tomlins, S. A., Rhodes, D. R., Perner, S., Dhanasekaran, S. M., Mehra, R., Sun, X.-W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., Lee, C., Montie, J. E., Shah, R. B., Pienta, K. J., Rubin, M. A., & Chinnaiyan, A. M. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science (New York, N.Y.)*, *310*(5748), 644–648. https://doi.org/10.1126/science.1117679

UniProt Consortium, T. (2018). UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, *46*(5), 2699–2699. https://doi.org/10.1093/nar/gky092

Vargas-Alarcón, G., Posadas-Sánchez, R., & Ramírez-Bello, J. (2020). Variability in genes related to SARS-CoV-2 entry into host cells (ACE2, TMPRSS2, TMPRSS11A, ELANE, and CTSL) and its potential use in association studies. *Life Sciences*, *260*, 118313. https://doi.org/10.1016/j.lfs.2020.118313

Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, *181*(2). https://doi.org/10.1016/j.cell.2020.02.058

Wan, Y., Shang, J., Graham, R., Baric, R. S., & Li, F. (2020). Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. *Journal of Virology*, *94*(7). https://doi.org/10.1128/JVI.00127-20

Wang, Q., Qiu, Y., Li, J.-Y., Zhou, Z.-J., Liao, C.-H., & Ge, X.-Y. (2020). A Unique Protease Cleavage Site Predicted in the Spike Protein of the Novel Pneumonia Coronavirus (2019-nCoV) Potentially Related to Viral Transmissibility. *Virologica Sinica*, *35*, 337–339. https://doi.org/10.1007/s12250-020-00212-7

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Molecular Biology and Evolution*, *35*(3), 773–777. https://doi.org/10.1093/molbev/msx335

Wertheim, J. O., Chu, D. K. W., Peiris, J. S. M., Kosakovsky Pond, S. L., & Poon, L. L. M. (2013). A Case for the Ancient Origin of Coronaviruses. *Journal of Virology*, *87*(12), 7039–7045. https://doi.org/10.1128/JVI.03273-12

Wettstein, L., Weil, T., Conzelmann, C., Müller, J. A., Groß, R., Hirschenberger, M., Seidel, A., Klute, S., Zech, F., Prelli Bozzo, C., Preising, N., Fois, G., Lochbaum, R., Knaff, P. M., Mailänder, V., Ständker, L., Thal, D. R., Schumann, C., Stenger, S., & Kleger, A. (2021). Alpha-1 antitrypsin inhibits TMPRSS2 protease activity and SARS-CoV-2 infection. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-21972-0

Word Health Organization. (2020). *MERS Situation Update*. https://applications.emro.who.int/docs/WHOEMCSR326E-eng.pdf?ua=1

World Health Organization. (2020). *Consensus document on the epidemiology of severe acute respiratory syndrome (SARS) DEPARTMENT OF COMMUNICABLE DISEASE SURVEILLANCE AND RESPONSE*. https://www.who.int/csr/sars/en/WHOconsensus.pdf

World Health Organization. (2022, February 8). *Weekly epidemiological update on COVID-19 - 8 February 2022*. Www.who.int.

https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---8-february-2022

Wruck, W., & Adjaye, J. (2020). SARS-CoV-2 receptor ACE2 is co-expressed with genes related to transmembrane serine proteases, viral entry, immunity and cellular stress. *Scientific Reports*, *10*(1), 1–14. https://doi.org/10.1038/s41598-020-78402-2

Xiao, J., Fang, M., Chen, Q., & He, B. (2020). SARS, MERS and COVID-19 among healthcare workers: A narrative review. *Journal of Infection and Public Health*, *13*(6), 843–848. https://doi.org/10.1016/j.jiph.2020.05.019

Yang, Z. (2005). Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. *Molecular Biology and Evolution*, *22*(4), 1107–1118. https://doi.org/10.1093/molbev/msi097

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugán, J. C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., & Giron, C. G. (2019). Ensembl 2020. *Nucleic Acids Research*, *48*(D1). https://doi.org/10.1093/nar/gkz966

Yépez, Y., Marcano-Ruiz, M., Bezerra, R. S., Fam, B., Ximenez, J. P., Silva Jr, W. A., & Bortolini, M. C. (2022). Evolutionary history of the SARS-CoV-2 Gamma variant of concern (P.1): a perfect storm. *Genetics and Molecular Biology*, *45*(1). https://doi.org/10.1590/1678-4685-gmb-2021-0309

Zang, R., Gomez Castro, M. F., McCune, B. T., Zeng, Q., Rothlauf, P. W., Sonnek, N. M., Liu, Z., Brulois, K. F., Wang, X., Greenberg, H. B., Diamond, M. S., Ciorba, M. A., Whelan, S. P. J., & Ding, S. (2020). TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes. *Science Immunology*, *5*(47), eabc3582. https://doi.org/10.1126/sciimmunol.abc3582

Zhang, Q., Wadgaonkar, P., Xu, L., Thakur, C., Fu, Y., Bi, Z., Qiu, Y., Almutairy, B., Zhan, W., Stemmer, P., & Chen, F. (2021). Environmentally-induced mdig contributes to the severity of COVID-19 through fostering expression of SARS-CoV-2 receptor NRPs and glycan metabolism. *Theranostics*, *11*(16), 7970–7983. https://doi.org/https://dx.doi.org/10.7150%2Fthno.62138

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., & Zheng, X.-S. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798). https://doi.org/10.1038/s41586-020-2012-7

Zhu, S., Liu, Y., Zhou, Z., Zhang, Z., Xiao, X., Liu, Z., Chen, A., Dong, X., Tian, F., Chen, S., Xu, Y., Wang, C., Li, Q., Niu, X., Pan, Q., Du, S., Xiao, J., Wang, J., & Wei, W. (2021). Genome-wide CRISPR activation screen identifies candidate receptors for SARS-CoV-2 entry. *Science China Life Sciences*. https://doi.org/10.1007/s11427-021-1990-5

6. Figures and Tables

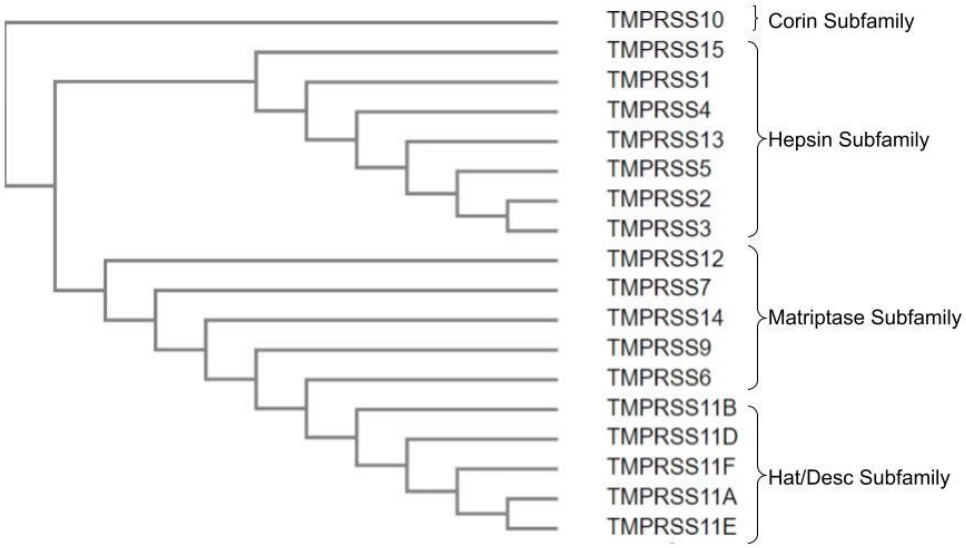


Figure 1. Phylogenetic alignment of the 18 TTSP proteins of the *Homo sapiens.*

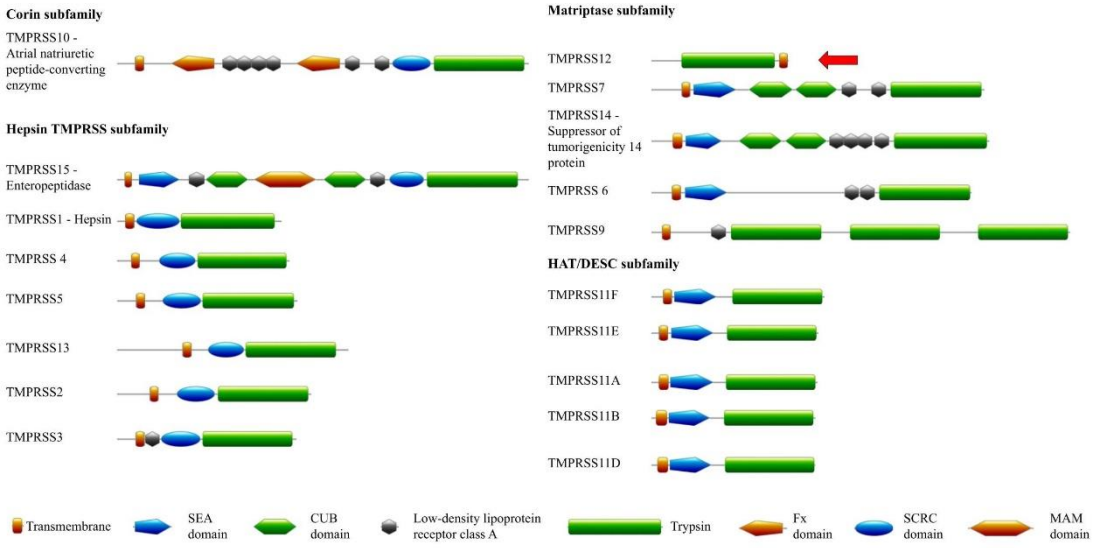


Figure 2. *Homo sapiens* TTSP subfamilies. The member highlighted with a red arrow was classified in this study.

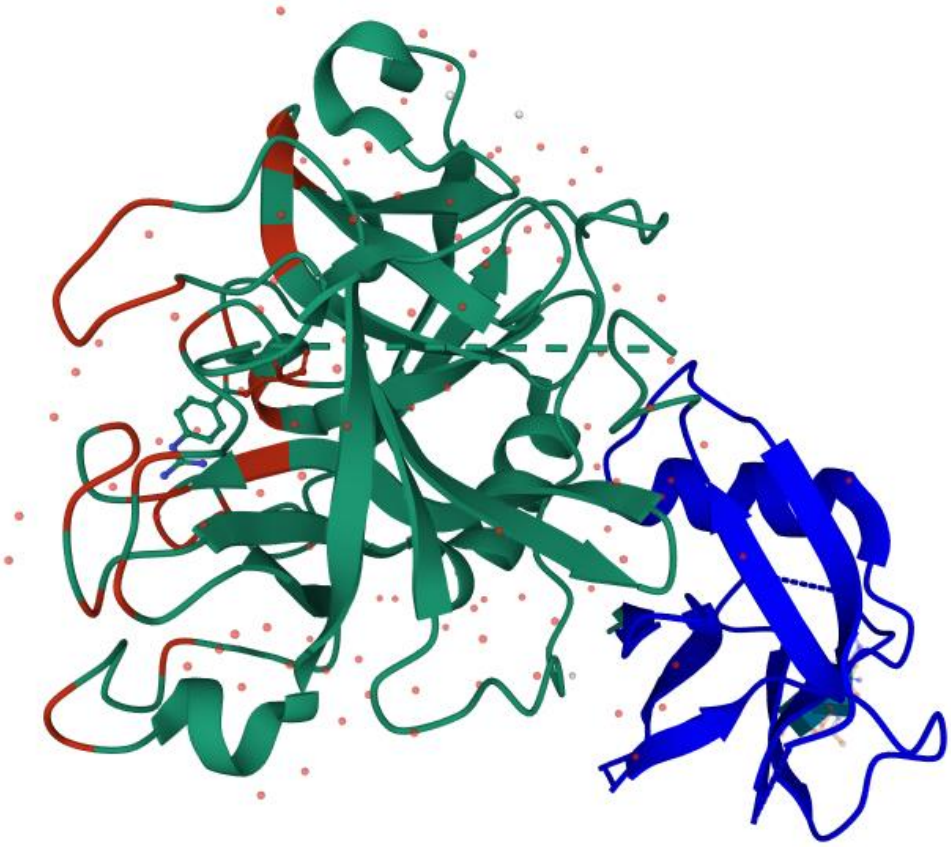


Figure 3. In the 3D structure of TMPRSS2 (Frasser *et al.* 2021), in blue is the scavenger receptor domain and in green is the peptidase domain. The red stripes in the peptidase domain are the 30 sites that interact with SARS-CoV2 spike protein, according to Hussain *et al.* (2020).

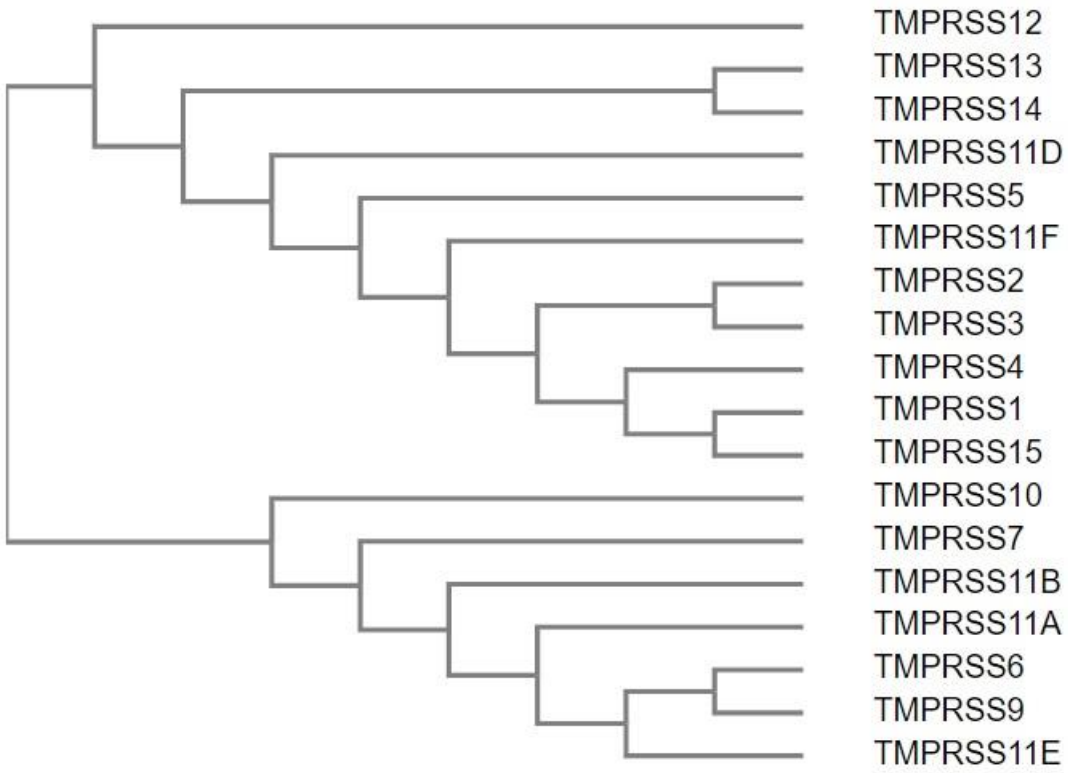
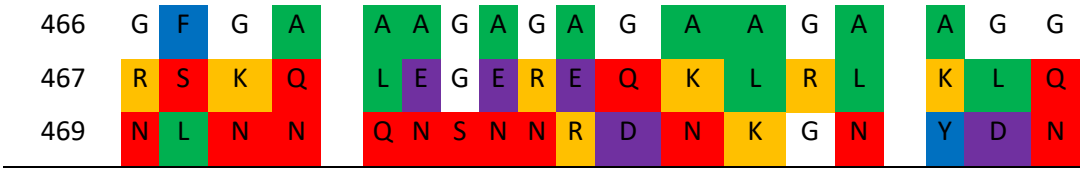


Figure 4. Phylogenetic alignment of the 18 *Homo sapiens* TMPRSSs considering just 30 sites that are critical for the proteolytic cleavage of SARS-CoV-2 Spike glycoprotein.

Table 1. *Homo sapiens* TTSPs are aligned and classified according to amino acid classification based on side chains.

Amino acid position	Not Used				Not Definitive Evidences											Strong Evidences		
	6	10	11B	14	1	3	4	5	7	9	11A	11E	11F	12	15	2	11D	13
275	V	S	W	A	Y	F	Y	L	F	E	Y	W	L	I	Y	V	L	F
276	R	E	K	L	D	Q	D	G	V	N	D	D	I	K	G	Q	N	G
278	R	G	R	G	A	Y	Q	R	S	E	I	S	G	V	R	V	A	T
280	I	I	Y	I	L	L	V	T	Y	F	Q	R	Q	V	L	V	H	I
296	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
299	Q	E	A	I	P	Y	R	H	H	N	Q	T	W	K	Y	E	R	F
300	E	G	K	D	E	D	K	S	G	E	K	T	K	D	G	K	S	V
338	E	S	S	N	L	K	P	S	N	N	R	K	H	I	N	D	K	T
340	D	A	P	F	E	K	Y	Q	Q	D	A	P	E	E	R	K	A	E
341	S	V	G	T	E	R	P	N	T	T	A	S	T	S	R	T	T	E
342	H	V	L	F	N	L	K	H	F	A	R	H	N	Y	K	K	H	D
387	L	M	L	T	T	T	T	T	R	L	L	L	I	T	V	T	Q	T
388	R	G	Y	Q	Q	E	K	H	H	K	Y	K	V	K	V	E	E	R
389	E	-	M	Y	Y	D	Q	P	E	E	Y	N	D	E	Y	E	Y	E
390	G	-	N	G	Y	G	N	S	A	D	G	D	D	E	Q	K	A	T
391	P	N	S	T	G	G	G	T	N	L	E	Y	P	N	G	G	H	D
392	I	K	F	G	Q	D	K	Y	K	V	S	S	I	A	T	K	T	K
413	V	Y	S	L	A	R	D	S	T	L	P	P	K	E	Q	R	P	Y
414	Y	F	Y	L	D	D	D	C	Y	Y	Q	Q	D	R	M	Y	H	L
419	Q	T	F	Q	Q	I	E	A	I	S	D	A	L	I	N	L	A	Y
431	K	S	S	S	E	T	E	D	S	D	E	E	E	D	E	Q	Q	R
433	K	T	E	G	G	G	G	R	K	K	I	K	K	A	G	N	G	G
438	Q	M	Q	Q	Q	Q	Q	Q	K	Q	R	Q	K	R	Q	Q	Q	Q
441	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
460	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
462	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
465	C	C	C	C	C	C	C	C	S	C	C	C	C	C	C	C	C	C



The amino acids observed in 30 human TMPRSS2 sites on the left column interact with SARS-CoV-2 Spike protein (Hussain *et al.*, 2020). The two site members of the catalytic triad of TMPRSS2 (Sgrignani & Cavalli, 2021) are in bold. Green is for hydrophobic side chain – aliphatic; Blue is hydrophobic side chain – aromatic; Red is polar neutral side chain; Purple is electrically charged side chain – acidic; Yellow is electrically charged side chain – basic; and white is for unique amino acids that do not belong to no group, proline, and glycine**.**

Table 2. PAML results for the Primate and placental mammal data sets.

Set Primates				Set Mammals			
LN		LNR	*p*- value	LN		LNR	*p*- value
Model 1	Model 2			Model 1	Model 2		
-10217,3	-10217,3	0	1	-53200,9	-53200,9	0	1
Model 7	Model 8			Model 7	Model 8		
-10198,6	-10184,4	28,35355	0,000 [7.0E-0007]	-52302	-52292,7	9,310079	0,002 [2.3E-0003]

Table 3. Sites under positive selection. The six sites that interact with the virus are highlighted in blue.

Mammals				Primates			
MEME	PAML			MEME	PAML		
3	214			36	173		
4	360			66	358		
5				85			
66				114			
85				115			
111				116			
117				117			
122				173			
137				214			
166				247			
181				250			
189				324			
217				378			
253				399			
263				401			
264				415			
				448			
299							
318							
319							
340							
360							
389							
409							
413							
423							
431							
438							

448
471
483
484
485

In **bold,** TMPRSS2 critical sites to the cleavage of SARS-CoV-2 Spike glycoprotein according to Hussain *et al.* (2020).

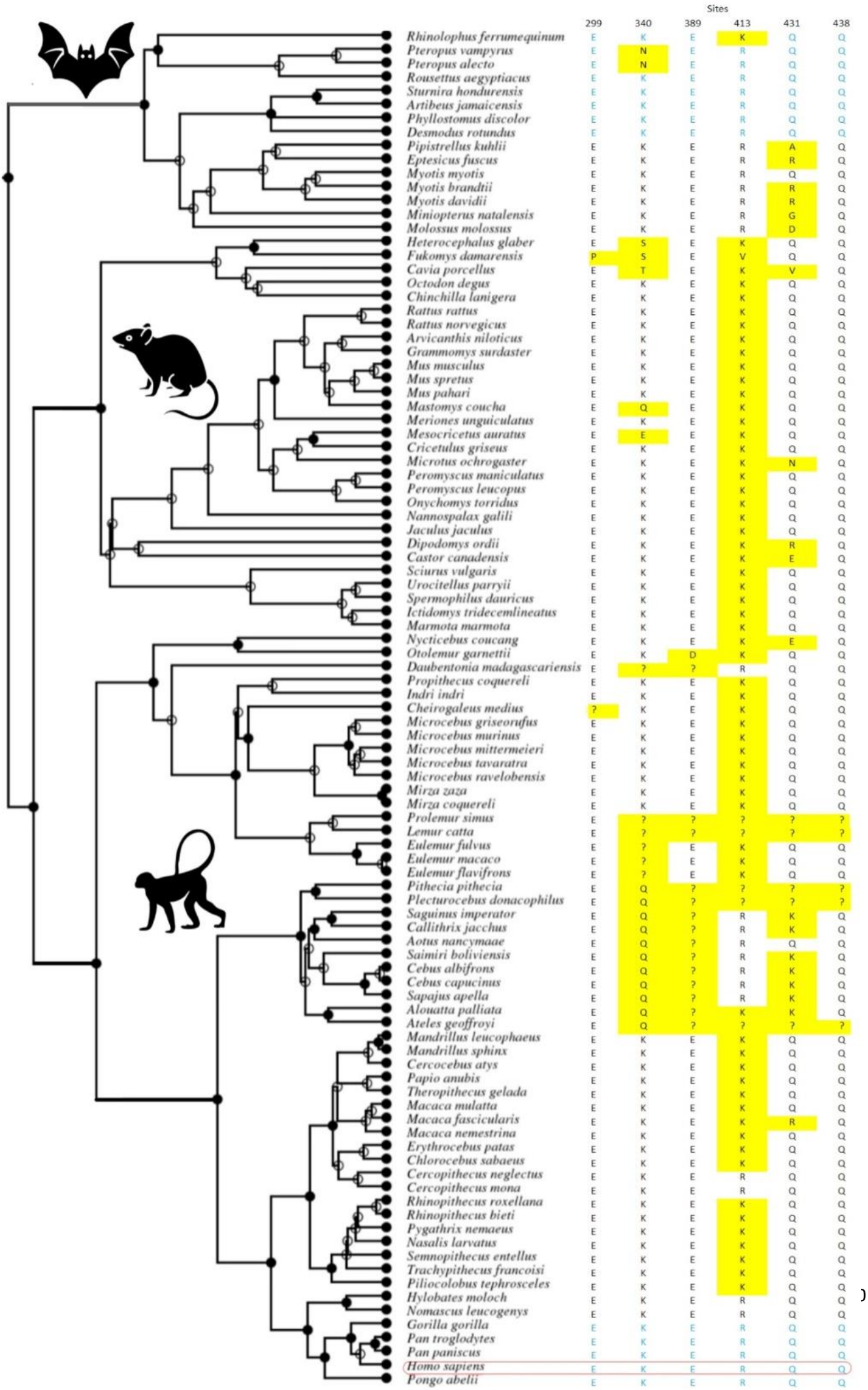

Sites
299 340 389 413 431 438
Rhinolophus ferrumequinum E K E K Q Q
Pteropus vampyrus E N E R Q Q
Pteropus alecto E N E R Q Q
Rousettus aegyptiacus E K E R Q Q
Sturnira hondurensis E K E R Q Q
Artibeus jamaicensis E K E R Q Q
Phyllostomus discolor E K E R Q Q
Desmodus rotundus E K E R Q Q
Pipistrellus kuhlii E K E R A Q
Eptesicus fuscus E K E R R Q
Myotis myotis E K E R Q Q
Myotis brandtii E K E R R Q
Myotis davidii E K E R R Q
Miniopterus natalensis E K E R G Q
Molossus molossus E K E R D Q
Heterocephalus glaber E S E K Q Q
Fukomys damarensis P S E V Q Q
Cavia porcellus E T E K V Q
Octodon degus E K E K Q Q
Chinchilla lanigera E K E K Q Q
Rattus rattus E K E K Q Q
Rattus norvegicus E K E K Q Q
Arvicanthis niloticus E K E K Q Q
Grammomys surdaster E K E K Q Q
Mus musculus E K E K Q Q
Mus spretus E K E K Q Q
Mus pahari E K E K Q Q
Mastomys coucha E Q E K Q Q
Meriones unguiculatus E K E K Q Q
Mesocricetus auratus E E E K Q Q
Cricetulus griseus E K E K Q Q
Microtus ochrogaster E K E K N Q
Peromyscus maniculatus E K E K Q Q
Peromyscus leucopus E K E K Q Q
Onychomys torridus E K E K Q Q
Nannospalax galili E K E K Q Q
Jaculus jaculus E K E K Q Q
Dipodomys ordii E K E K R Q
Castor canadensis E K E K E Q
Sciurus vulgaris E K E K Q Q
Urocitellus parryii E K E K Q Q
Spermophilus dauricus E K E K Q Q
Ictidomys tridecemlineatus E K E K Q Q
Marmota marmota E K E K Q Q
Nycticebus coucang E K E K E Q
Otolemur garnettii E K D K Q Q
Daubentonia madagascariensis E ? ? R Q Q
Propithecus coquereli E K E K Q Q
Indri indri E K E K Q Q
Cheirogaleus medius ? K E K Q Q
Microcebus griseorufus E K E K Q Q
Microcebus murinus E K E K Q Q
Microcebus mittermeieri E K E K Q Q
Microcebus tavaratra E K E K Q Q
Microcebus ravelobensis E K E K Q Q
Mirza zaza E K E K Q Q
Mirza coquereli E K E K Q Q
Prolemur simus E ? ? ? ? ?
Lemur catta E ? ? ? ? ?
Eulemur fulvus E ? E K Q Q
Eulemur macaco E ? E K Q Q
Eulemur flavifrons E ? E K Q Q
Pithecia pithecia E Q ? ? ? ?
Plecturocebus donacophilus E Q ? ? ? ?
Saguinus imperator E Q ? R K Q
Callithrix jacchus E Q ? R K Q
Aotus nancymaae E Q ? R Q Q
Saimiri boliviensis E Q ? R K Q
Cebus albifrons E Q ? R K Q
Cebus capucinus E Q ? R K Q
Sapajus apella E Q ? R K Q
Alouatta palliata E Q ? K K Q
Ateles geoffroyi E Q ? ? ? ?
Mandrillus leucophaeus E K E K Q Q
Mandrillus sphinx E K E K Q Q
Cercocebus atys E K E K Q Q
Papio anubis E K E K Q Q
Theropithecus gelada E K E K Q Q
Macaca mulatta E K E K Q Q
Macaca fascicularis E K E K R Q
Macaca nemestrina E K E K Q Q
Erythrocebus patas E K E K Q Q
Chlorocebus sabaeus E K E K Q Q
Cercopithecus neglectus E K E R Q Q
Cercopithecus mona E K E R Q Q
Rhinopithecus roxellana E K E K Q Q
Rhinopithecus bieti E K E K Q Q
Pygathrix nemaeus E K E K Q Q
Nasalis larvatus E K E K Q Q
Semnopithecus entellus E K E K Q Q
Trachypithecus francoisi E K E K Q Q
Piliocolobus tephrosceles E K E K Q Q
Hylobates moloch E K E R Q Q
Nomascus leucogenys E K E R Q Q
Gorilla gorilla E K E R Q Q
Pan troglodytes E K E R Q Q
Pan paniscus E K E R Q Q
Homo sapiens E K E R Q Q
Pongo abelii E K E R Q Q

Figure 5. Phylogeny of TMPRSS2 considering the six sites under positive selection that interact with SARS-CoV-2 Spike glycoprotein, in yellow are marked the sites that differ from *Homo sapiens* (highlited in red). Primates, Rodentia, and Chiroptera species are shown. Monkey by bmijnlieff, Rat by Vectors Market, and Bat by Md Saiful Alam Saif from NounProject.com.

7. Supplementary Material

Table S1. Data sets used in present study.

ORDER	*SPECIES*	REFERENCE TMPRSS2	DATA SET
Afrotheria	*Chrysochloris asiatica*	XM_006870856.1	Mammalian set
Afrotheria	*Echinops telfairi*	XM_030888670.1	Mammalian set
Afrotheria	*Loxodonta africana*	XM_023559111.1	Mammalian set
Afrotheria	*Orycteropus afer*	XM_007947316.1	Mammalian set
Afrotheria	*Procavia capensis*	ENSPCAG00000001830.1	Mammalian set
Afrotheria	*Trichechus manatus*	XM_023739122.1	Mammalian set
Artiodactyla	*Balaenoptera acutorostrata*	XM_007172707.2	Mammalian set
Artiodactyla	*Balaenoptera musculus*	XM_036850886.1	Mammalian set
Artiodactyla	*Bison bison*	XM_010834939.1	Mammalian set
Artiodactyla	*Bos grunniens*	ENSBGRG00000014644.1	Mammalian set

Artiodactyla	*Bos indicus*	XM_019962885.1	Mammalian set
Artiodactyla	*Bos mutus*	XM_005893019.2	Mammalian set
Artiodactyla	*Bos taurus*	NM_001081585.1	Mammalian set
Artiodactyla	*Bubalus bubalis*	XM_025292133.1	Mammalian set
Artiodactyla	*Camelus dromedarius*	XM_010994865.2	Mammalian set
Artiodactyla	*Camelus ferus*	XM_032461838.1	Mammalian set
Artiodactyla	*Capra hircus*	XM_005675629.3	Mammalian set
Artiodactyla	*Catagonus wagneri*	ENSCWAG00000012074.1	Mammalian set
Artiodactyla	*Cervus elaphus*	ENSCHYG00000026573.1	Mammalian set
Artiodactyla	*Delphinapterus leucas*	XM_022552961.1	Mammalian set
Artiodactyla	*Globicephala melas*	XM_030835403.1	Mammalian set
Artiodactyla	*Lagenorhynchus obliquidens*	XM_027122628.1	Mammalian set
Artiodactyla	*Lipotes vexillifer*	XM_007452880.1	Mammalian set
Artiodactyla	*Monodon monoceros*	XM_029214475.1	Mammalian set
Artiodactyla	*Moschus moschiferus*	ENSMMSG00000009350.1	Mammalian set

Artiodactyla	*Neophocaena phocaenoides*	XM_024758324.1	Mammalian set
Artiodactyla	*Odocoileus virginianus*	XM_020907938.1	Mammalian set
Artiodactyla	*Orcinus orca*	XM_033418343.1	Mammalian set
Artiodactyla	*Ovis aries*	XM_027960704.1	Mammalian set
Artiodactyla	*Phocoena sinus*	XM_032631437.1	Mammalian set
Artiodactyla	*Physeter catodon*	XM_024120360.1	Mammalian set
Artiodactyla	*Sus scrofa*	NM_001386131.1	Mammalian set
Artiodactyla	*Tursiops truncatus*	XM_033856384.1	Mammalian set
Artiodactyla	*Vicugna pacos*	XM_031684299.1	Mammalian set
Carnivora	*Acinonyx jubatus*	XM_027041574.1	Mammalian set
Carnivora	*Ailuropoda melanoleuca*	XM_034655405.1	Mammalian set
Carnivora	*Callorhinus ursinus*	XM_025863512.1	Mammalian set
Carnivora	*Canis lupus*	XM_038443561.1	Mammalian set
Carnivora	*Enhydra lutris*	XM_022515963.1	Mammalian set

Carnivora	*Eumetopias jubatus*	XM_028126498.1	Mammalian set
Carnivora	*Felis catus*	XM_023238709.1	Mammalian set
Carnivora	*Halichoerus grypus*	XM_036108025.1	Mammalian set
Carnivora	*Lontra canadensis*	XM_032865299.1	Mammalian set
Carnivora	*Lynx canadensis*	XM_030330003.1	Mammalian set
Carnivora	*Mirounga leonina*	XM_035028851.1	Mammalian set
Carnivora	*Mustela erminea*	XM_032355539.1	Mammalian set
Carnivora	*Mustela putorius*	NM_001386127.1	Mammalian set
Carnivora	*Neomonachus schauinslandi*	XM_021679201.1	Mammalian set
Carnivora	*Neovison vison*	ENSNVIG00000005330.1	Mammalian set
Carnivora	*Odobenus rosmarus*	XM_012564540.1	Mammalian set
Carnivora	*Panthera leo*	ENSPLOG00000013265.1	Mammalian set
Carnivora	*Panthera pardus*	XM_019420396.1	Mammalian set
Carnivora	*Panthera tigris*	XM_015541202.1	Mammalian set
Carnivora	*Phoca vitulina*	XM_032392537.1	Mammalian set

Carnivora	*Puma concolor*	XM_025913088.1	Mammalian set
Carnivora	*Suricata suricatta*	XM_029938936.1	Mammalian set
Carnivora	*Ursus americanus*	ENSUAMG00000026733.1	Mammalian set
Carnivora	*Ursus arctos*	XM_026499970.1	Mammalian set
Carnivora	*Ursus maritimus*	XM_040621677.1	Mammalian set
Carnivora	*Ursus thibetanus*	ENSUTTG00000017173.1	Mammalian set
Carnivora	*Vulpes vulpes*	XM_025983380.1	Mammalian set
Carnivora	*Zalophus californianus*	XM_027587844.2	Mammalian set
Chiroptera	*Artibeus jamaicensis*	XM_037144605.1	Mammalian set
Chiroptera	*Desmodus rotundus*	XM_024561824.1	Mammalian set
Chiroptera	*Eptesicus fuscus*	XM_028151885.1	Mammalian set
Chiroptera	*Miniopterus natalensis*	XM_016220949.1	Mammalian set
Chiroptera	*Molossus molossus*	XM_036243106.1	Mammalian set
Chiroptera	*Myotis brandtii*	XM_005885695.2	Mammalian set
Chiroptera	*Myotis davidii*	XM_006754329.2	Mammalian set

Chiroptera	*Myotis myotis*	XM_036350387.1	Mammalian set
Chiroptera	*Phyllostomus discolor*	XM_028504691.2	Mammalian set
Chiroptera	*Pipistrellus kuhlii*	XM_036446834.1	Mammalian set
Chiroptera	*Pteropus alecto*	NM_001386133.1	Mammalian set
Chiroptera	*Pteropus vampyrus*	XM_023538199.1	Mammalian set
Chiroptera	*Rhinolophus ferrumequinum*	XM_033088803.1	Mammalian set
Chiroptera	*Rousettus aegyptiacus*	XM_016134588.2	Mammalian set
Chiroptera	*Sturnira hondurensis*	XM_037042998.1	Mammalian set
Dasyuromorphia	*Sarcophilus harrisii*	XM_003766371.4	Mammalian set
Dermoptera	*Galeopterus variegatus*	XM_008571221.1	Mammalian set
Didelphimorphia	*Monodelphis domestica*	XM_007493067.2	Mammalian set
Diprotodontia	*Phascolarctos cinereus*	XM_020976261.1	Mammalian set
Diprotodontia	*Trichosurus vulpecula*	XM_036744151.1	Mammalian set
Diprotodontia	*Vombatus ursinus*	XM_027843625.1	Mammalian set
Eulipotyphla	*Erinaceus europaeus*	XM_016194345.1	Mammalian set

Eulipotyphla	*Sorex araneus*	XM_004602515.1	Mammalian set
Eulipotyphla	*Talpa occidentalis*	XM_037524481.1	Mammalian set
Lagomorpha	*Ochotona princeps*	XM_004594171.2	Mammalian set
Lagomorpha	*Oryctolagus cuniculus*	NM_001386128.1	Mammalian set
Monotremata	*Ornithorhynchus anatinus*	XM_029046634.1	Mammalian set
Perissodactyla	*Ceratotherium simum*	XM_004429710.2	Mammalian set
Perissodactyla	*Equus asinus*	XM_014839273.1	Mammalian set
Perissodactyla	*Equus caballus*	XM_005606160.3	Mammalian set
Perissodactyla	*Equus przewalskii*	XM_008538229.1	Mammalian set
Pholidota	*Manis javanica*	XM_037014836.1	Mammalian set
Pholidota	*Manis pentadactyla*	XM_036876908.1	Mammalian set
Primates	*Alouatta palliata*	GCA_004027835.1	Mammalian setandPrimates set
Primates	*Aotus nancymaae*	XM_021669490.1	Mammalian setandPrimates set
Primates	*Ateles geoffroyi*	GCA_004024785.1	Mammalian setandPrimates set

Primates	*Callithrix jacchus*	XM_008986722.3	Mammalian setandPrimates set
Primates	*Carlito syrichta*	XM_008068563.1	Mammalian setandPrimates set
Primates	*Cebus albifrons*	GCA_004027755.1	Mammalian setandPrimates set
Primates	*Cebus capucinus*	XM_017538612.2	Mammalian setandPrimates set
Primates	*Cercocebus atys*	XM_012037329.1	Mammalian setandPrimates set
Primates	*Cercopithecus mona*	GCA_014849445.1	Mammalian setandPrimates set
Primates	*Cercopithecus neglectus*	GCA_004027615.1	Mammalian setandPrimates set
Primates	*Cheirogaleus medius*	GCA_008086735.1	Mammalian setandPrimates set
Primates	*Chlorocebus sabaeus*	XM_037985428.1	Mammalian setandPrimates set
Primates	*Daubentonia madagascariensis*	GCA_004027145.1	Mammalian setandPrimates set
Primates	*Erythrocebus patas*	GCA_004027335.1	Mammalian setandPrimates set

Primates	*Eulemur flavifrons*	GCA_001262665.1	Mammalian setandPrimates set
Primates	*Eulemur fulvus*	GCA_004027275.1	Mammalian setandPrimates set
Primates	*Eulemur macaco*	GCA_001262655.1	Mammalian setandPrimates set
Primates	*Gorilla gorilla*	XM_004062839.3	Mammalian setandPrimates set
Primates	*Homo sapiens*	NM_005656.4	Mammalian setandPrimates set
Primates	*Hylobates moloch*	XM_032174257.1	Mammalian setandPrimates set
Primates	*Indri indri*	GCA_004363605.1	Mammalian setandPrimates set
Primates	*Lemur catta*	GCA_004024665.1	Mammalian setandPrimates set
Primates	*Macaca fascicularis*	XM_005548645.2	Mammalian setandPrimates set
Primates	*Macaca mulatta*	XM_015132845.2	Mammalian setandPrimates set
Primates	*Macaca nemestrina*	XM_011725990.2	Mammalian setandPrimates set

Primates	*Mandrillus leucophaeus*	XM_011984191.1	Mammalian setandPrimates set
Primates	*Mandrillus sphinx*	GCA_004802615.1	Mammalian setandPrimates set
Primates	*Microcebus griseorufus*	GCA_008750995.1	Mammalian setandPrimates set
Primates	*Microcebus mittermeieri*	GCA_008750955.1	Mammalian setandPrimates set
Primates	*Microcebus murinus*	XM_012749836.2	Mammalian setandPrimates set
Primates	*Microcebus ravelobensis*	GCA_008750975.1	Mammalian setandPrimates set
Primates	*Microcebus tavaratra*	GCA_008750935.1	Mammalian setandPrimates set
Primates	*Mirza coquereli*	GCA_004024645.1	Mammalian setandPrimates set
Primates	*Mirza zaza*	GCA_008750895.1	Mammalian setandPrimates set
Primates	*Nasalis larvatus*	GCA_000772465.1	Mammalian setandPrimates set
Primates	*Nomascus leucogenys*	XM_030806307.1	Mammalian setandPrimates set

Primates	*Nycticebus coucang*	GCA_004027815.1	Mammalian setandPrimates set
Primates	*Otolemur garnettii*	XM_023518095.1	Mammalian setandPrimates set
Primates	*Pan paniscus*	XM_034947921.1	Mammalian setandPrimates set
Primates	*Pan troglodytes*	XM_016938565.2	Mammalian setandPrimates set
Primates	*Papioanubis*	XM_009202216.4	Mammalian setandPrimates set
Primates	*Piliocolobus tephrosceles*	XM_023191885.1	Mammalian setandPrimates set
Primates	*Pithecia pithecia*	GCA_004026645.1	Mammalian setandPrimates set
Primates	*Plecturocebus donacophilus*	GCA_004027715.1	Mammalian setandPrimates set
Primates	*Pongo abelii*	XM_024239539.1	Mammalian setandPrimates set
Primates	*Prolemur simus*	GCA_003258685.1	Mammalian setandPrimates set
Primates	*Propithecus coquereli*	XM_012640067.1	Mammalian setandPrimates set

Primates	*Pygathrix nemaeus*	GCA_004024825.1	Mammalian setandPrimates set
Primates	*Rhinopithecus bieti*	XM_017880226.1	Mammalian setandPrimates set
Primates	*Rhinopithecus roxellana*	XM_010354509.2	Mammalian setandPrimates set
Primates	*Saguinus imperator*	GCA_004024885.1	Mammalian setandPrimates set
Primates	*Saimiri boliviensis*	XM_003927594.3	Mammalian setandPrimates set
Primates	*Sapajus apella*	XM_032276822.1	Mammalian setandPrimates set
Primates	*Semnopithecus entellus*	GCA_004025065.1	Mammalian setandPrimates set
Primates	*Theropithecus gelada*	XM_025380760.1	Mammalian setandPrimates set
Primates	*Trachypithecus francoisi*	XM_033232534.1	Mammalian setandPrimates set
Rodentia	*Arvicanthis niloticus*	XM_034514887.1	Mammalian set
Rodentia	*Castor canadensis*	XM_020181818.1	Mammalian set
Rodentia	*Cavia porcellus*	XM_013146070.1	Mammalian set

Rodentia	*Chinchilla lanigera*	XM_005375855.2	Mammalian set
Rodentia	*Cricetulus griseus*	NM_001386129.1	Mammalian set
Rodentia	*Dipodomys ordii*	XM_013027332.1	Mammalian set
Rodentia	*Fukomys damarensis*	XM_010636600.3	Mammalian set
Rodentia	*Grammomys surdaster*	XM_028752540.1	Mammalian set
Rodentia	*Heterocephalus glaber*	XM_004842457.3	Mammalian set
Rodentia	*Ictidomys tridecemlineatus*	XM_021725256.2	Mammalian set
Rodentia	*Jaculus jaculus*	XM_004654601.1	Mammalian set
Rodentia	*Marmota marmota*	XM_015478793.1	Mammalian set
Rodentia	*Mastomyscoucha*	XM_031364839.1	Mammalian set
Rodentia	*Meriones unguiculatus*	XM_021660850.1	Mammalian set
Rodentia	*Mesocricetus auratus*	XM_013116227.2	Mammalian set
Rodentia	*Microtus ochrogaster*	XM_013354818.1	Mammalian set
Rodentia	*Mus musculus*	NM_015775.2	Mammalian set
Rodentia	*Mus pahari*	XM_021209684.1	Mammalian set

Rodentia	*Mus spretus*	MGP_SPRETEiJ_G0021917.1	Mammalian set
Rodentia	*Nannospalax galili*	XM_008827408.3	Mammalian set
Rodentia	*Octodon degus*	XM_023715938.1	Mammalian set
Rodentia	*Onychomystorridus*	XM_036204162.1	Mammalian set
Rodentia	*Peromyscus leucopus*	XM_037209598.1	Mammalian set
Rodentia	*Peromyscus maniculatus*	XM_015991310.1	Mammalian set
Rodentia	*Rattus norvegicus*	NM_130424.3	Mammalian set
Rodentia	*Rattus rattus*	XM_032900351.1	Mammalian set
Rodentia	*Sciurus vulgaris*	ENSSVLG00005015845.1	Mammalian set
Rodentia	*Spermophilus dauricus*	ENSSDAG00000013476.1	Mammalian set
Rodentia	*Urocitellus parryii*	XM_026379435.1	Mammalian set
Scandentia	*Tupaia belangeri*	ENSTBEG00000008474.1	Mammalian set
Xenarthra	*Dasypus novemcinctus*	XM_004466304.3	Mammalian set

Note: The references starting with ENS and MGP belong to Ensembl. the others are from NCBI and the GCA refers to the non curated primate sequences.

Table S2. Comparison between human TMPRSS2 and other TMPRSSs according to the Grantham Score. considering the 30 sites where TMPRSS2 interacts with SARS-CoV-2 Spike protein to its cleavage (Hussain *et al.* 2020).

Site	TMPRSS Identical	Conservative	Moderately Conservative	Moderately Radical	Radical	Gap
VAL275	1	7	6	3	0	0
GLN276	1	4	9	3	0	0
VAL278	1	1	8	6	0	1
VAL280	2	7	7	0	0	1
HIS296*	16	1	0	0	0	0
GLU299	1	6	4	5	1	0
LYS300	4	0	4	8	1	0
ASP338	0	4	8	4	1	0
LYS340	1	1	6	9	0	0
THR341	5	1	11	0	0	0
LYS342	2	5	3	7	0	0
THR387	6	1	9	1	0	0

GLU388	2	3	8	4	0	0
GLU389	5	5	1	5	0	1
LYS390	0	0	6	10	0	1
GLY391	4	2	8	3	0	0
LYS392	4	0	6	7	0	0
ARG413	1	2	6	7	1	0
TYR414	4	4	4	1	4	0
LEU419	1	6	4	4	2	0
GLN431	1	7	9	0	0	0
ASN433	0	1	14	2	0	0
GLN438	11	2	3	1	0	0
SER441*	16	0	1	0	0	0
SER460	16	0	1	0	0	0
GLY462	16	0	0	1	0	0
CYS465	15	0	0	1	1	0

ALA466	7	0	9	1	0	0
LYS467	3	3	6	4	1	0
TYR469	0	1	2	11	2	1

*296 and 441 are part of the catalytic triad of *Homo sapiens* TMPRSS2

Table S3. Comparison of TMPRSSs according to Grantham Score regarding the 30 sites where TMPRSS2 interacts with SARS-CoV-2 Spike glycoprotein.

	Conservative	Moderately Conservative	Moderately Radical	Radical	Gap	Identical
TMPRSS1	4	11	8	4	3	0
TMPRSS3	5	4	7	1	0	13
TMPRSS4	3	10	4	1	0	12
TMPRSS5	4	13	4	1	0	8

TMPRSS6	6	10	5	0	0	9
TMPRSS7	6	10	6	0	0	8
TMPRSS9	4	8	8	0	0	10
TMPRSS10	4	10	6	0	2	8
TMPRSS11A	5	10	7	2	0	6
TMPRSS11B	2	11	8	0	0	9
TMPRSS11D	4	9	7	1	0	9
TMPRSS11E	3	11	8	0	0	8
TMPRSS11F	5	8	6	2	0	9
TMPRSS13	5	13	3	0	0	9
TMPRSS12	4	12	3	1	0	10
TMPRSS14	4	5	12	0	0	9
TMPRSS15	6	8	6	1	0	9

Table S5. Sites under positive selection across all the evolutive analysis and its respective p-value.

Set Mammals			
MEME		PAML	
Site	*p*-value	Site	*p*-value
3	0.00	214	0.00
4	0.00	**360**	0.05
5	0.01		
66	0.01		
85	0.01		
111	0.01		
117	0.01		
122	0.00		
137	0.02		
166	0.04		
181	0.00		
189	0.01		
217	0.04		
253	0.04		
263	0.00		
264	0.00		
284	0.01		
299	0.01		
318	0.05		
319	0.01		
340	0.01		
360	0.05		
389	0.02		
409	0.01		
413	0.05		
423	0.01		
431	0.01		
438	0.04		

448	0.01
471	0.05
483	0.00
484	0.00
485	0.00

Set Primates			
MEME		PAML	
Site	*p*-value	Site	*p*-value
36	0.05	**173**	0.01
66	0.05	358	0.03
85	0.00		
114	0.03		
115	0.03		
116	0.03		
117	0.03		
173	0.02		
214	0.04		
247	0.05		
250	0.04		
324	0.00		
378	0.05		
399	0.02		
401	0.03		
415	0.05		
448	0.00		

Table S6. Sites under positive selection that interacts with SARS-CoV-2 spike glycoprotein (Hussain *et al.* 2020). The sites that differs from *Homo sapiens* are highlighted in yellow

	Sites					
Specie	299	340	389	413	431	438
Sarcophilus harrisii	E	K	E	R	K	Q
Phascolarctos cinereus	E	E	E	R	R	Q
Vombatus ursinus	E	K	E	R	R	Q
Trichosurus vulpecula	E	K	E	R	R	Q
Erinaceus europaeus	E	T	E	R	Q	Q
Sorex araneus	E	T	E	-	-	P
Talpa occidentalis	E	K	E	K	Q	Q
Suricata suricatta	E	K	E	K	Q	Q
Panthera tigris	-	K	E	K	Q	Q
Panthera pardus	E	K	E	K	Q	Q
Panthera leo	E	K	E	K	Q	Q
Lynx canadensis	E	K	E	K	Q	Q
Felis catus	E	K	E	K	Q	Q
Puma concolor	-	K	-	K	Q	Q
Acinonyx jubatus	E	K	E	K	Q	Q
Canis lupus	E	K	E	K	R	Q
Vulpes vulpes	E	K	E	K	R	Q
Enhydra lutris	E	K	E	K	R	Q
Lontra canadensis	E	K	E	K	R	Q
Neovison vison	E	K	E	-	-	-
Mustela putorius	E	K	E	K	R	Q
Mustela erminea	E	K	E	K	R	Q
Callorhinus ursinus	E	K	E	K	Q	Q

Monodelphis domestica	E	N	E	R	R	Q
Zalophus californianus	E	K	E	K	Q	Q
Eumetopias jubatus	E	K	E	K	Q	Q
Odobenus rosmarus	E	G	E	K	Q	Q
Mirounga leonina	E	K	E	K	Q	Q
Monachusschauinslandi	E	K	E	K	R	Q
Phocavitulina	E	K	E	K	Q	Q
Halichoerus grypus	E	K	E	K	Q	Q
Ailuropoda melanoleuca	E	K	E	K	Q	Q
Ursus thibetanus	E	K	E	K	Q	Q
Ursus americanus	E	K	E	K	Q	Q
Ursus arctos	E	K	E	K	Q	Q
Ursus maritimus	E	K	E	K	Q	Q
Manis javanica	E	Q	E	K	Q	Q
Manis pentadactyla	E	Q	E	K	Q	Q
Equus caballus	E	K	E	K	E	Q
Equus przewalskii	E	K	E	K	E	Q
Equus asinus	E	K	E	K	E	Q
Ceratotherium simum	E	N	E	M	Q	Q
Vicugna pacos	E	K	E	K	Q	Q
Camelus dromedarius	E	K	E	K	Q	Q
Camelus ferus	E	K	E	K	Q	Q
Balaenoptera acutorostrata	E	K	E	K	Q	Q
Balaenoptera musculus	E	K	E	K	Q	Q
Physeter catodon	E	K	E	K	Q	Q
Lipotesvexillifer	E	K	E	K	Q	Q
Neophocaenaphocaenoides	E	R	E	K	Q	Q
Phocoena sinus	E	R	E	K	Q	Q
Delphinapterus leucas	E	K	E	K	Q	Q
Monodon monoceros	E	K	E	K	Q	Q
Globicephala melas	E	K	E	K	Q	Q
Tursiops truncatus	E	K	E	K	Q	Q
Lagenorhynchusobliquidens	E	K	E	K	Q	Q

Orcinus orca	E	K	E	K	Q	Q
Moschus moschiferus	E	K	E	K	R	Q
Bubalus bubalis	E	K	E	K	R	Q
Bos taurus	E	K	E	K	R	Q
Bos indicus	E	K	E	K	Q	Q
Bos mutus	E	K	E	K	R	Q
Bos grunniens	E	K	E	K	R	Q
Bison bison	E	K	E	K	R	Q
Capra hircus	E	N	E	K	R	Q
Ovis aries	E	K	E	K	Q	Q
Odocoileus virginianus	E	K	E	K	Q	Q
Cervus elaphus	E	K	E	K	Q	Q
Catagonuswagneri	E	K	E	K	Q	Q
Sus scrofa	E	K	E	K	Q	Q
Rhinolophus ferrumequinum	E	K	E	K	Q	Q
Pteropus vampyrus	E	N	E	R	Q	Q
Pteropus alecto	E	N	E	R	Q	Q
Rousettus aegyptiacus	E	K	E	R	Q	Q
Sturnirahondurensis	E	K	E	R	Q	Q
Artibeusjamaicensis	E	K	E	R	Q	Q
Phyllostomus discolor	E	K	E	R	Q	Q
Desmodusrotundus	E	K	E	R	Q	Q
Pipistrellus kuhlii	E	K	E	R	A	Q
Eptesicus fuscus	E	K	E	R	R	Q
Myotis myotis	E	K	E	R	Q	Q
Myotis brandtii	E	K	E	R	R	Q
Myotis davidii	E	K	E	R	R	Q
Miniopterusnatalensis	E	K	E	R	G	Q
Molossus molossus	E	K	E	R	D	Q
Oryctolagus cuniculus	E	K	E	K	Q	Q
Ochotona princeps	E	K	E	R	E	Q
Heterocephalus glaber	E	S	E	K	Q	Q
Fukomys damarensis	P	S	E	V	Q	Q

Cavia porcellus	E	T	E	K	V	Q
Octodon degus	E	K	E	K	Q	Q
Chinchilla lanigera	E	K	E	K	Q	Q
Rattus rattus	E	K	E	K	Q	Q
Rattus norvegicus	E	K	E	K	Q	Q
Arvicanthisniloticus	E	K	E	K	Q	Q
Grammomyssurdaster	E	K	E	K	Q	Q
Mus musculus	E	K	E	K	Q	Q
Mus spretus	E	K	E	K	Q	Q
Mus pahari	E	K	E	K	Q	Q
Mastomyscoucha	E	Q	E	K	Q	Q
Merionesunguiculatus	E	K	E	K	Q	Q
Mesocricetus auratus	E	E	E	K	Q	Q
Cricetulus griseus	E	K	E	K	Q	Q
Microtus ochrogaster	E	K	E	K	N	Q
Peromyscus maniculatus	E	K	E	K	Q	Q
Peromyscus leucopus	E	K	E	K	Q	Q
Onychomys torridus	E	K	E	K	Q	Q
Nannospalax galili	E	K	E	K	Q	Q
Jaculus jaculus	E	K	E	K	Q	Q
Dipodomys ordii	E	K	E	K	R	Q
Castor canadensis	E	K	E	K	E	Q
Sciurus vulgaris	E	K	E	K	Q	Q
Urocitellus parryii	E	K	E	K	Q	Q
Spermophilusdauricus	E	K	E	K	Q	Q
Ictidomystridecemlineatus	E	K	E	K	Q	Q
Marmota marmota	E	K	E	K	Q	Q
Nycticebus coucang	E	K	E	K	E	Q
Otolemur garnettii	E	K	D	K	Q	Q
Daubentonia madagascariensis	E	X	X	R	Q	Q
Propithecus coquereli	E	K	E	K	Q	Q
Indri indri	E	K	E	K	Q	Q

Cheirogaleus medius	X	K	E	K	Q	Q
Microcebus griseorufus	E	K	E	K	Q	Q
Microcebus murinus	E	K	E	K	Q	Q
Microcebus mittermeieri	E	K	E	K	Q	Q
Microcebus tavaratra	E	K	E	K	Q	Q
Microcebus ravelobensis	E	K	E	K	Q	Q
Mirza zaza	E	K	E	K	Q	Q
Mirza coquereli	E	K	E	K	Q	Q
Prolemur simus	E	X	X	X	X	X
Lemur catta	E	X	X	X	X	X
Eulemur fulvus	E	X	E	K	Q	Q
Eulemur macaco	E	X	E	K	Q	Q
Eulemur flavifrons	E	X	E	K	Q	Q
Pithecia pithecia	E	Q	-	-	-	-
Plecturocebusdonacophilus	E	Q	-	-	-	-
Saguinus imperator	E	Q	-	R	K	Q
Callithrix jacchus	E	Q	-	R	K	Q
Aotusnancymaae	E	Q	-	R	Q	Q
Saimiri boliviensis	E	Q	-	R	K	Q
Cebus albifrons	E	Q	-	R	K	Q
Cebus capucinus	E	Q	-	R	K	Q
Sapajus apella	E	Q	-	R	K	Q
Alouatta palliata	E	Q	-	K	K	Q
Ateles geoffroyi	E	Q	-	-	-	-
Mandrillusleucophaeus	E	K	E	K	Q	Q
Mandrillus sphinx	E	K	E	K	Q	Q
Cercocebus atys	E	K	E	K	Q	Q
Papio anubis	E	K	E	K	Q	Q
Theropithecus gelada	E	K	E	K	Q	Q
Macaca mulatta	E	K	E	K	Q	Q
Macaca fascicularis	E	K	E	K	R	Q
Macaca nemestrina	E	K	E	K	Q	Q
Erythrocebus patas	E	K	E	K	Q	Q

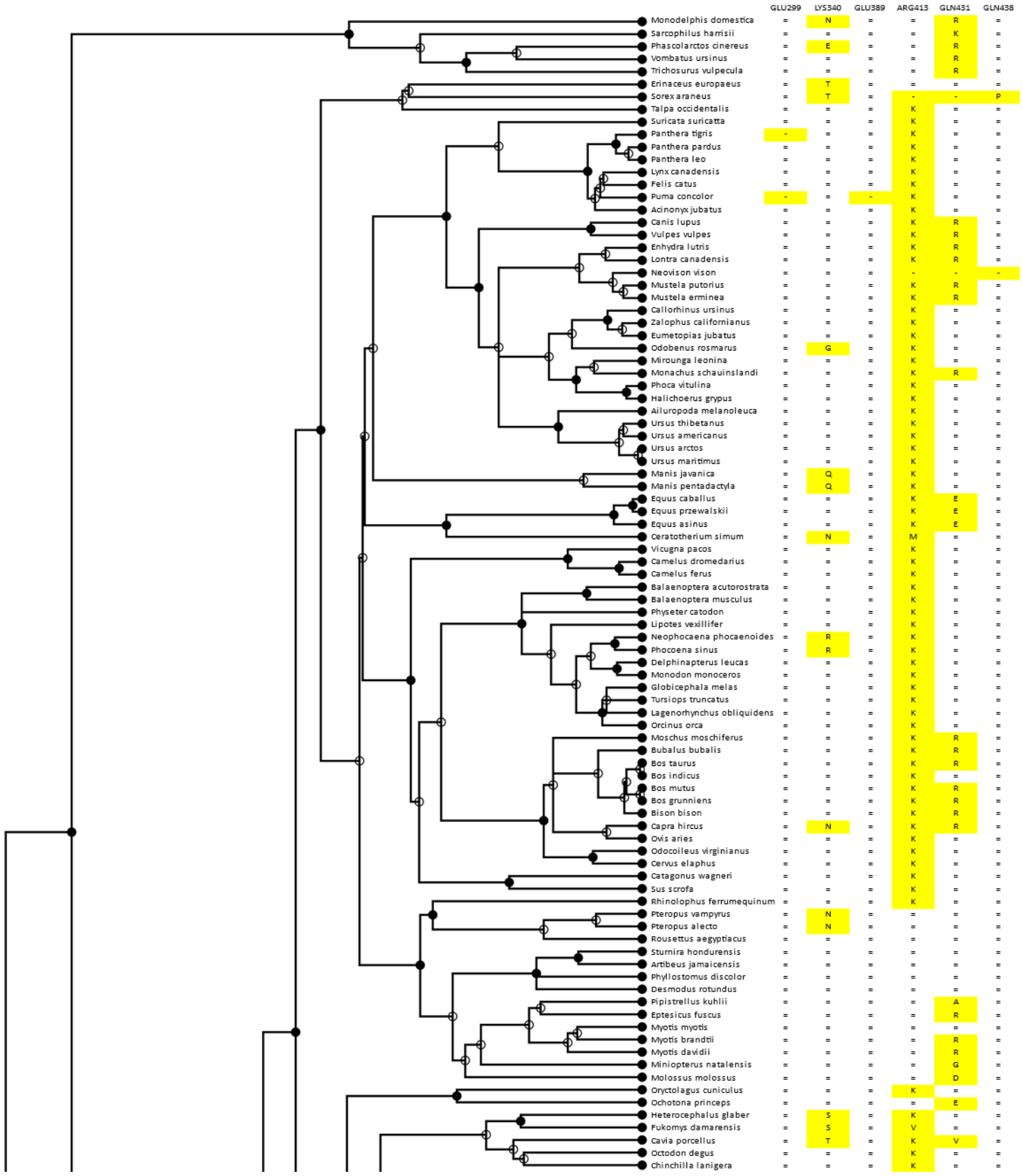
Chlorocebus sabaeus	E	K	E	K	Q	Q
Cercopithecus neglectus	E	K	E	R	Q	Q
Cercopithecus mona	E	K	E	R	Q	Q
Rhinopithecus roxellana	E	K	E	K	Q	Q
Rhinopithecus bieti	E	K	E	K	Q	Q
Pygathrix nemaeus	E	K	E	K	Q	Q
Nasalis larvatus	E	K	E	K	Q	Q
Semnopithecus entellus	E	K	E	K	Q	Q
Trachypithecus francoisi	E	K	E	K	Q	Q
Piliocolobus tephrosceles	E	K	E	K	Q	Q
Hylobates moloch	E	K	E	R	Q	Q
Nomascus leucogenys	E	K	E	R	Q	Q
Gorilla gorilla	E	K	E	R	Q	Q
Pan troglodytes	E	K	E	R	Q	Q
Pan paniscus	E	K	E	R	Q	Q
Homo sapiens	E	K	E	R	Q	Q
Pongo abelii	E	K	E	R	Q	Q
Carlito syrichta	E	K	E	R	Q	Q
Galeopterus variegatus	E	-	E	K	Q	Q
Tupaia belangeri	E	N	E	-	N	Q
Dasypus novemcinctus	E	K	E	R	Q	Q
Trichechus manatus	E	K	E	S	Q	Q
Procavia capensis	E	K	P	R	Q	Q
Loxodonta africana	E	K	E	R	E	Q
Echinops telfairi	E	K	E	R	Q	Q
Chrysochloris asiatica	E	K	E	K	Q	Q
Orycteropus afer	E	K	E	R	E	Q
Ornithorhynchus anatinus	E	K	E	R	K	Q

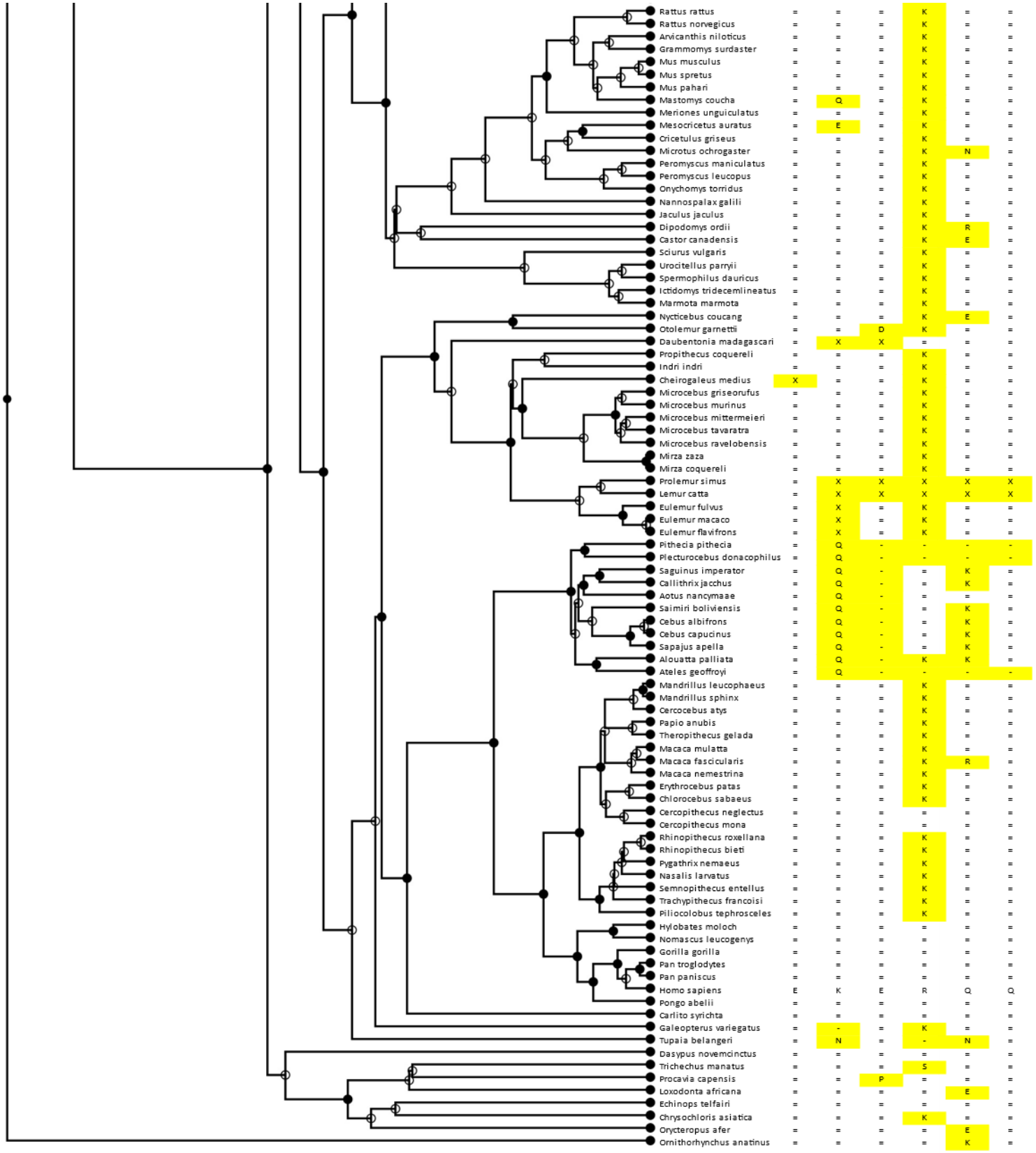
Table S4. Functional studies with TMPRSSs.

Type of Evidence	TMPRSS																	Article
	1	2	3	4	5	6	7	9	10	11A	11B	11D	11E	11F	12	13	14	15	
Coexpression with ACE2	yes	yes																	Meng *et al.*, 2020
Coexpression with ACE2			yes	yes	yes		yes					yes							Salas Orozco *et al.*, 2021
Coexpression with ACE2		yes		yes						yes		yes							Wruck & Adjaye, 2020
Homology with TMPRSS2		*															no	yes	Zhang *et al.*, 2021
Allelic frequency		yes								yes									Vargas-Alarcón *et al.*, 2020
In vitro HEK293 cell line		yes		yes						yes		yes	yes						Ou *et al.*, 2020
In vitro HEK293 cell line		yes		yes															Zang *et al.*, 2020
In vitro HEK293 cell line		yes		yes								yes					no		Tharappel *et al.*, 2020
In vitro HEK293 cell line		yes										yes	no	no		yes			Laporte *et al.*, 2021
In vitro BHK21 cell line		yes	no	no					no	no	no	yes	yes	yes		yes			Hoffmann *et al.*, 2021
In vitro E6 Cell line	no	yes	no	no	no	no			no	no		yes	no			yes	no		Kishimoto *et al.*, 2021
Table 1-classification	nd	se	nd	nd	nd	nu	nd	nd	nu	nd	nu	se	nd	nd	nd	se	nu	nd	

Note: Studies considered to build the groups presented at the Table, where nd represents "not definitive evidence", Se represents "strong evidence", and nu represents "not used". The evidences highlighted in yellow were considered by us as "not definitive". Yes and no refer to proteins tested or not tested, respectively. Proteins with controversial results or not found in any article were also classified by us as "not definitive".

FigureS1. Phylogeny of all the species and respective sites under positive selection that interacts with SARS-CoV-2 spike protein (Hussain *et al.* 2020). The sites that differ from *Homo sapiens* are highlighted in yellow.


GLU299 LYS340 GLU389 ARG413 GLN431 GLN438
Monodelphis domestica = N = = R =
Sarcophilus harrisii = = = = K =
Phascolarctos cinereus = E = = R =
Vombatus ursinus = = = = R =
Trichosurus vulpecula = = = = R =
Erinaceus europaeus = T = = = =
Sorex araneus = T = - - P
Talpa occidentalis = = = K = =
Suricata suricatta = = = K = =
Panthera tigris - = = K = =
Panthera pardus = = = K = =
Panthera leo = = = K = =
Lynx canadensis = = = K = =
Felis catus = = = K = =
Puma concolor - = - K = =
Acinonyx jubatus = = = K = =
Canis lupus = = = K R =
Vulpes vulpes = = = K R =
Enhydra lutris = = = K R =
Lontra canadensis = = = K R =
Neovison vison = = = - - -
Mustela putorius = = = K R =
Mustela erminea = = = K R =
Callorhinus ursinus = = = K = =
Zalophus californianus = = = K = =
Eumetopias jubatus = = = K = =
Odobenus rosmarus = G = K = =
Mirounga leonina = = = K = =
Monachus schauinslandi = = = K R =
Phoca vitulina = = = K = =
Halichoerus grypus = = = K = =
Ailuropoda melanoleuca = = = K = =
Ursus thibetanus = = = K = =
Ursus americanus = = = K = =
Ursus arctos = = = K = =
Ursus maritimus = = = K = =
Manis javanica = Q = K = =
Manis pentadactyla = Q = K = =
Equus caballus = = = K E =
Equus przewalskii = = = K E =
Equus asinus = = = K E =
Ceratotherium simum = N = M = =
Vicugna pacos = = = K = =
Camelus dromedarius = = = K = =
Camelus ferus = = = K = =
Balaenoptera acutorostrata = = = K = =
Balaenoptera musculus = = = K = =
Physeter catodon = = = K = =
Lipotes vexillifer = = = K = =
Neophocaena phocaenoides = R = K = =
Phocoena sinus = R = K = =
Delphinapterus leucas = = = K = =
Monodon monoceros = = = K = =
Globicephala melas = = = K = =
Tursiops truncatus = = = K = =
Lagenorhynchus obliquidens = = = K = =
Orcinus orca = = = K = =
Moschus moschiferus = = = K R =
Bubalus bubalis = = = K R =
Bos taurus = = = K R =
Bos indicus = = = K = =
Bos mutus = = = K R =
Bos grunniens = = = K R =
Bison bison = = = K R =
Capra hircus = N = K R =
Ovis aries = = = K = =
Odocoileus virginianus = = = K = =
Cervus elaphus = = = K = =
Catagonus wagneri = = = K = =
Sus scrofa = = = K = =
Rhinolophus ferrumequinum = = = K = =
Pteropus vampyrus = N = = = =
Pteropus alecto = N = = = =
Rousettus aegyptiacus = = = = = =
Sturnira hondurensis = = = = = =
Artibeus jamaicensis = = = = = =
Phyllostomus discolor = = = = = =
Desmodus rotundus = = = = = =
Pipistrellus kuhlii = = = = A =
Eptesicus fuscus = = = = R =
Myotis myotis = = = = = =
Myotis brandtii = = = = R =
Myotis davidii = = = = R =
Miniopterus natalensis = = = = G =
Molossus molossus = = = = D =
Oryctolagus cuniculus = = = K = =
Ochotona princeps = = = = E =
Heterocephalus glaber = S = K = =
Fukomys damarensis = S = V = =
Cavia porcellus = T = K V =
Octodon degus = = = K = =
Chinchilla lanigera = = = K = =


Rattus rattus = = = K = =
Rattus norvegicus = = = K = =
Arvicanthis niloticus = = = K = =
Grammomys surdaster = = = K = =
Mus musculus = = = K = =
Mus spretus = = = K = =
Mus pahari = = = K = =
Mastomys coucha = Q = K = =
Meriones unguiculatus = = = K = =
Mesocricetus auratus = E = K = =
Cricetulus griseus = = = K = =
Microtus ochrogaster = = = K N =
Peromyscus maniculatus = = = K = =
Peromyscus leucopus = = = K = =
Onychomys torridus = = = K = =
Nannospalax galili = = = K = =
Jaculus jaculus = = = K = =
Dipodomys ordii = = = K R =
Castor canadensis = = = K E =
Sciurus vulgaris = = = K = =
Urocitellus parryii = = = K = =
Spermophilus dauricus = = = K = =
Ictidomys tridecemlineatus = = = K = =
Marmota marmota = = = K = =
Nycticebus coucang = = = K E =
Otolemur garnettii = = D K = =
Daubentonia madagascari = X X = = =
Propithecus coquereli = = = K = =
Indri indri = = = K = =
Cheirogaleus medius X = = K = =
Microcebus griseorufus = = = K = =
Microcebus murinus = = = K = =
Microcebus mittermeieri = = = K = =
Microcebus tavaratra = = = K = =
Microcebus ravelobensis = = = K = =
Mirza zaza = = = K = =
Mirza coquereli = = = K = =
Prolemur simus = X X X X X
Lemur catta = X X X X X
Eulemur fulvus = X = K = =
Eulemur macaco = X = K = =
Eulemur flavifrons = X = K = =
Pithecia pithecia = Q - - - -
Plecturocebus donacophilus = Q - - - -
Saguinus imperator = Q - = K =
Callithrix jacchus = Q - = K =
Aotus nancymaae = Q - = = =
Saimiri boliviensis = Q - = K =
Cebus albifrons = Q - = K =
Cebus capucinus = Q - = K =
Sapajus apella = Q - = K =
Alouatta palliata = Q - K K =
Ateles geoffroyi = Q - - - -
Mandrillus leucophaeus = = = K = =
Mandrillus sphinx = = = K = =
Cercocebus atys = = = K = =
Papio anubis = = = K = =
Theropithecus gelada = = = K = =
Macaca mulatta = = = K = =
Macaca fascicularis = = = K R =
Macaca nemestrina = = = K = =
Erythrocebus patas = = = K = =
Chlorocebus sabaeus = = = K = =
Cercopithecus neglectus = = = = = =
Cercopithecus mona = = = = = =
Rhinopithecus roxellana = = = K = =
Rhinopithecus bieti = = = K = =
Pygathrix nemaeus = = = K = =
Nasalis larvatus = = = K = =
Semnopithecus entellus = = = K = =
Trachypithecus francoisi = = = K = =
Piliocolobus tephrosceles = = = K = =
Hylobates moloch = = = = = =
Nomascus leucogenys = = = = = =
Gorilla gorilla = = = = = =
Pan troglodytes = = = = = =
Pan paniscus = = = = = =
Homo sapiens E K E R Q Q
Pongo abelii = = = = = =
Carlito syrichta = = = = = =
Galeopterus variegatus = - = K = =
Tupaia belangeri = N = - N =
Dasypus novemcinctus = = = = = =
Trichechus manatus = = = S = =
Procavia capensis = = P = = =
Loxodonta africana = = = = E =
Echinops telfairi = = = = = =
Chrysochloris asiatica = = = K = =
Orycteropus afer = = = = E =
Ornithorhynchus anatinus = = = = K =

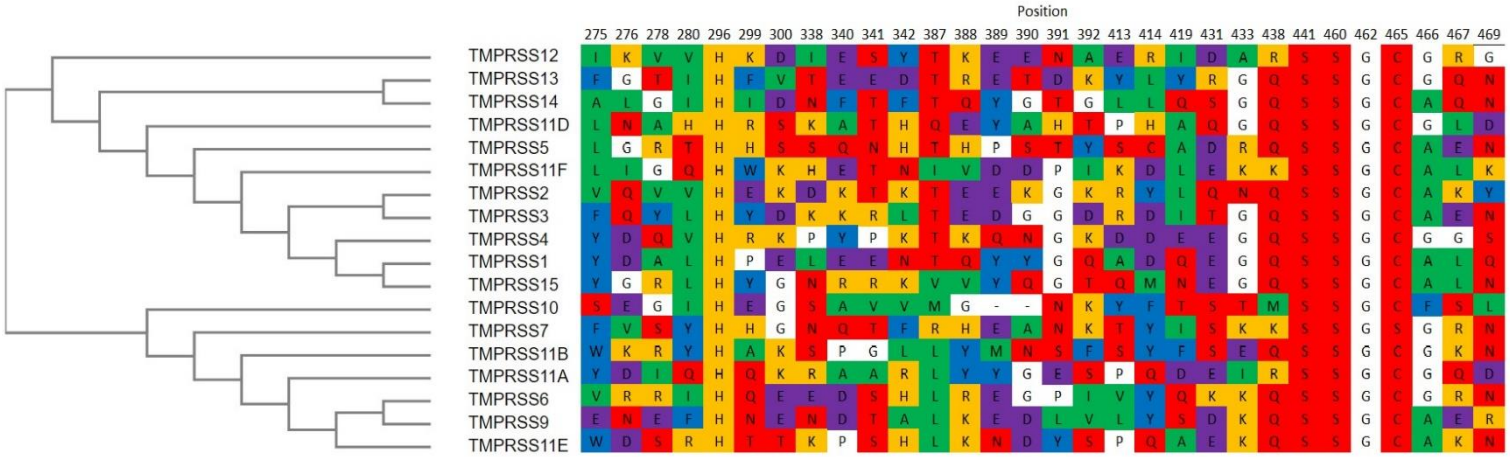


Figure S2. Alignment of the 30 sites and its phylogeny

4. Considerações finais

O presente trabalho contribuiu para atualizar a história das TTSPs, uma família de proteínas cujo último trabalho em relação a sua história de evolução molecular foi publicado em 2009. A oportunidade de um estudo dessa natureza se faz premente já que um ou mais membros dessa família (exemplo TMPRSS2) foi (foram) cooptado (s) evolutivamente pelo vírus SARS-CoV-2 para seu processamento e uma bem sucedida entrada na célula do hospedeiro *Homo sapiens*. SARS-CoV-2 é o agente causador da pandemia da COVID-19.

Dezoito parálogos TTSPs foram identificados no genoma do *Homo sapiens*, sendo que fomos capazes de alocar todos eles em quatro subfamílias, incluindo TMPRSS12 no clado das Matriptase.

Um alinhamento filogenético dessas 18 TTSP do *Homo sapiens* resgata o esperado considerando a história evolutiva das moléculas. Interessantemente quando a mesma análise é feita considerando somente os 30 sítios de TMPRSS2 que interagem com o vírus, a topologia da árvore se altera, sugerindo que há uma certa identidade em pontos chaves, não obstante a história filogenética das TTSPs, considerando suas sequências codificadoras completas. Em outras palavras, as proteínas dividiram-se um dois clados distintos, sendo que um deles continha todas as proteínas que já foram documentadas como capazes de realizar a clivagem da proteína Spike do SARS-CoV-2. No outro clado encontravam-se as TTSPs cuja interação com SARS-CoV-2 tem sido controversa ou descartada.

Finalmente, respondendo as questões principais formuladas:

1) O padrão de diversidade genética encontrado poderia ser resultado, mesmo que parcialmente, de pressão seletiva exercida por uma corrida armamentista biológica entre vírus-hopspedeiro?

Sim. Em relação a evolução da TMPRSS2 humana e seus ortólogos em outros mamíferos placentários, nosso trabalho identificou 33 sítios onde a variação encontrada entre pode ser explicada por seleção positiva, cinco deles em posições que interagem com o vírus, ou seja, somente 15%. Esse padrão de variação sugere que, talvez, a maior parte da variação entre espécies seja resultados de pressões seletivas que vêm moldando a função normal da TMPRSS2 e de seus ortólogos nas células das espécies correspondentes. Por outro lado, também indica que o vírus, através de sua proteína Spike, que necessita de clivagem, estaria sendo moldado evolutivamente para se ligar em posições nas proteases dos potenciais hospedeiros mamíferos com menos propensão a terem variação promovida por ação de seleção positiva devido a demandas do próprio hospedeiro. Isso daria vantagem ao vírus, pois ele teria, por exemplo, menos chance de perder afinidade com o hospedeiro ao mesmo tempo que daria mais chances para saltos zoonóticos. Nesse contexto, o padrão de variabilidade nos mamíferos placentários indica profundidade de tempo evolutivo, algo esperado já que TMPRSS2 e outras TTSPs vêm sendo cooptadas por outras famílias virais, tais como *Orthomyxoviridae*. Considerando que tanto as famílias de vírus como as de mamíferos placentários convivem a milhões de anos, esses achados indicam uma corrida armamentista evolucionária de longo tempo.

2) Porque TMPRSS2 foi preferencialmente cooptada evolutivamente pelo SARS-CoV-2 para o processamento da Spike, enquanto outras TTSPs não o foram?

Há um provável conjunto de cerca de 30 aminoácidos em sítios pouco variáveis devido à seleção positiva, o que garante uma identidade chave entre TMPRSS2 e outras TTSPs, incluindo de membros de diferentes subfamílias. É o conjunto de aminoácidos nesses sítios que potencialmente garante a reiterada cooptação evolutiva dessas proteases pelo SARS-CoV-2 e outros vírus para o processamento de suas Spikes. Por outro lado, TTSPs que não compartilham esse conjunto de aminoácidos não tem sido preferencialmente cooptada.

5. Referências

Abrams, E. M., Greenhawt, M., Shaker, M., Pinto, A. D., Sinha, I., & Singer, A. (2022). The COVID-19 Pandemic: Adverse Effects on the Social Determinants of Health in Children and Families. *Annals of Allergy, Asthma & Immunology*. https://doi.org/10.1016/j.anai.2021.10.022

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature Medicine*, *26*(450–452). https://doi.org/10.1038/s41591-020-0820-9

Azhar, E. I., El-Kafrawy, S. A., Farraj, S. A., Hassan, A. M., Al-Saeed, M. S., Hashem, A. M., & Madani, T. A. (2014). Evidence for Camel-to-Human Transmission of MERS Coronavirus. *New England Journal of Medicine*, *370*(26), 2499–2505. https://doi.org/10.1056/nejmoa1401505

Bertram, S., Dijkman, R., Habjan, M., Heurich, A., Gierer, S., Glowacka, I., Welsch, K., Winkler, M., Schneider, H., Hofmann-Winkler, H., Thiel, V., & Pohlmann, S. (2013). TMPRSS2 Activates the Human Coronavirus 229E for Cathepsin-Independent Host Cell Entry and Is Expressed in Viral Target Cells in the Respiratory Epithelium. *Journal of Virology*, *87*(11), 6150–6160. https://doi.org/10.1128/jvi.03372-12

Bieber, F. (2020). Global Nationalism in Times of the COVID-19 Pandemic. *Nationalities Papers*, *50*(1), 1–13. https://doi.org/10.1017/nps.2020.35

BöttcherFriebertshäuserE., Freuer, C., Sielaff, F., Schmidt, S., Eickmann, M., Uhlendorff, J., Steinmetzer, T., Klenk, H.-D., & Garten, W. (2010). Cleavage of Influenza Virus Hemagglutinin by Airway Proteases TMPRSS2 and HAT Differs in Subcellular Localization and Susceptibility to Protease Inhibitors. *Journal of Virology*, *84*(11), 5605–5614. https://doi.org/10.1128/jvi.00140-10

Bugge, T. H., Antalis, T. M., & Wu, Q. (2009). Type II Transmembrane Serine Proteases. *Journal of Biological Chemistry*, *284*(35), 23177–23181. https://doi.org/10.1074/jbc.r109.021006

Cao, Y., Li, L., Feng, Z., Wan, S., Huang, P., Sun, X., Wen, F., Huang, X., Ning, G., & Wang, W. (2020). Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. *Cell Discovery*, *6*(1). https://doi.org/10.1038/s41421-020-0147-1

Chen, W., Yuan, P., Yang, M., Yan, Z., Kong, S., Yan, J., Liu, X., Chen, Y., Qiao, J., & Yan, L. (2020). SARS-CoV-2 Entry Factors: ACE2 and TMPRSS2 Are Expressed in Peri-Implantation Embryos and the Maternal–Fetal Interface. *Engineering*, *6*(10), 1162–1169. https://doi.org/10.1016/j.eng.2020.07.013

Fam, B. S. O., Vargas-Pinilla, P., Amorim, C. E. G., Sortica, V. A., & Bortolini, M. C. (2020). ACE2 diversity in placental mammals reveals the evolutionary strategy of SARS-CoV-2. *Genetics and Molecular Biology*, *43*(2). https://doi.org/10.1590/1678-4685-gmb-2020-0104

Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., Tamir, H., Achdout, H., Stein, D., Israeli, O., Beth-Din, A., Melamed, S., Weiss, S., Israely, T., Paran, N., Schwartz, M., & Stern-Ginossar, N. (2020). The coding capacity of SARS-CoV-2. *Nature*, 1–6. https://doi.org/10.1038/s41586-020-2739-1

Gheware, A., Ray, A., Rana, D., Bajpai, P., Nambirajan, A., Arulselvi, S., Mathur, P., Trikha, A., Arava, S., Das, P., Mridha, A. R., Singh, G., Soneja, M., Nischal, N., Lalwani, S., Wig, N., Sarkar, C., & Jain, D. (2022). ACE2 protein expression in lung tissues of severe COVID-19 infection. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-07918-6

Glowacka, I., Bertram, S., Müller, M. A., Allen, P., Soilleux, E., Pfefferle, S., Steffen, I., Tsegaye, T. S., He, Y., Gnirss, K., Niemeyer, D., Schneider, H., Drosten, C., & Pöhlmann, S. (2011). Evidence that TMPRSS2 Activates the Severe Acute Respiratory Syndrome Coronavirus Spike Protein for Membrane Fusion and Reduces Viral Control by the Humoral Immune Response. *Journal of Virology*, *85*(9), 4122–4134. https://doi.org/10.1128/JVI.02232-10

Guaita Martínez, J. M., Carracedo, P., Gorgues Comas, D., & Siemens, C. H. (2022). An analysis of the blockchain and COVID-19 research landscape using a bibliometric study. *Sustainable Technology and Entrepreneurship*, *1*(1), 100006. https://doi.org/10.1016/j.stae.2022.100006

Hassan, Sk. S., Choudhury, P. P., & Roy, B. (2021). Rare mutations in the accessory proteins ORF6, ORF7b, and ORF10 of the SARS-CoV-2 genomes. *Meta Gene*, *28*, 100873. https://doi.org/10.1016/j.mgene.2021.100873

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N.-H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, *181*(2). https://doi.org/10.1016/j.cell.2020.02.052

Hofmann, H., & Pöhlmann, S. (2004). Cellular entry of the SARS coronavirus. *Trends in Microbiology*, *12*(10), 466–472. https://doi.org/10.1016/j.tim.2004.08.008

Hofmann, H., Pyrc, K., van der Hoek, L., Geier, M., Berkhout, B., & Pohlmann, S. (2005). Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proceedings of the National Academy of Sciences*, *102*(22), 7988–7993. https://doi.org/10.1073/pnas.0409465102

Hooper, J. D., Clements, J. A., Quigley, J. P., & Antalis, T. M. (2001). Type II Transmembrane Serine Proteases. *Journal of Biological Chemistry*, *276*(2), 857–860. https://doi.org/10.1074/jbc.r000020200

Horowitz, J. E., Kosmicki, J. A., Damask, A., Sharma, D., Roberts, G. H. L., Justice, A. E., Banerjee, N., Coignet, M. V., Yadav, A., Leader, J. B., Marcketta, A., Park, D. S., Lanche, R., Maxwell, E., Knight, S. C., Bai, X., Guturu, H., Sun, D., Baltzell, A., & Kury, F. S. P. (2022). Genome-wide analysis provides genetic evidence that ACE2 influences COVID-19 risk and yields risk scores associated with severe disease. *Nature Genetics*, *54*(4), 382–392. https://doi.org/10.1038/s41588-021-01006-7

Hussain, M., Jabeen, N., Amanullah, A., Ashraf Baig, A., Aziz, B., Shabbir, S., Raza, F., & Uddin, N. (2020). Molecular docking between human TMPRSS2 and SARS-CoV-2 spike protein: conformation and intermolecular interactions. *AIMS Microbiology*, *6*(3), 350–360. https://doi.org/10.3934/microbiol.2020021

Iser, B. P. M., Sliva, I., Raymundo, V. T., Poleto, M. B., Schuelter-Trevisol, F., & Bobinski, F. (2020). Definição de caso suspeito da COVID-19: uma revisão narrativa dos sinais e sintomas mais frequentes entre os casos confirmados. *Epidemiologia E Serviços de Saúde*, *29*(3). https://doi.org/10.5123/s1679-49742020000300018

Iwata-Yoshikawa, N., Okamura, T., Shimizu, Y., Hasegawa, H., Takeda, M., & Nagata, N. (2019). TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection. *Journal of Virology*, *93*(6). https://doi.org/10.1128/jvi.01815-18

Jackson, C. B., Farzan, M., Chen, B., & Choe, H. (2022). Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews Molecular Cell Biology*, *23*, 1–18. https://doi.org/10.1038/s41580-021-00418-x

Koch, J., Uckeley, Z. M., Doldan, P., Stanifer, M., Boulant, S., & Lozach, P. (2021). TMPRSS2 expression dictates the entry route used by SARS-CoV-2 to infect host cells. *The EMBO Journal*, *40*(16). https://doi.org/10.15252/embj.2021107821

Lamers, M. M., & Haagmans, B. L. (2022). SARS-CoV-2 pathogenesis. *Nature Reviews Microbiology*, 1–15. https://doi.org/10.1038/s41579-022-00713-0

Landau, L. J. B., Fam, B. S. de O., Yépez, Y., Caldas-Garcia, G. B., Pissinatti, A., Falótico, T., Reales, G., Schüler-Faccini, L., Sortica, V. A., & Bortolini, M. C. (2021). Evolutionary analysis of the anti-viral STAT2 gene of primates and rodents: Signature of different stages of an arms race. *Infection, Genetics and Evolution*, *95*, 105030. https://doi.org/10.1016/j.meegid.2021.105030

Li, X., Zai, J., Zhao, Q., Nie, Q., Li, Y., Foley, B. T., & Chaillon, A. (2020). Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. *Journal of Medical Virology*, *92*(6), 602–611. https://doi.org/10.1002/jmv.25731

Malone, B., Urakova, N., Snijder, E. J., & Campbell, E. A. (2022). Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nature Reviews Molecular Cell Biology*, *23*. https://doi.org/10.1038/s41580-021-00432-z

Matsuyama, S., Nao, N., Shirato, K., Kawase, M., Saito, S., Takayama, I., Nagata, N., Sekizuka, T., Katoh, H., Kato, F., Sakata, M., Tahara, M., Kutsuna, S., Ohmagari, N., Kuroda, M., Suzuki, T., Kageyama, T., & Takeda, M. (2020). Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. *Proceedings of the National Academy of Sciences*, *117*(13), 7001–7003. https://doi.org/10.1073/pnas.2002589117

Meyer, B. H., Prescott, B., & Sheng, X. S. (2022). The impact of the COVID-19 pandemic on business expectations. *International Journal of Forecasting*, *38*(2). https://doi.org/10.1016/j.ijforecast.2021.02.009

Mir, T., Almas, T., Kaur, J., Faisaluddin, M., Song, D., Ullah, W., Mamtani, S., Rauf, H., Yadav, S., Latchana, S., Michaelson, N. M., Connerney, M., & Sattar, Y. (2021). Coronavirus disease 2019 (COVID-19): Multisystem review of pathophysiology. *Annals of Medicine and Surgery*, *69*, 102745. https://doi.org/10.1016/j.amsu.2021.102745

Mousavizadeh, L., & Ghasemi, S. (2020). Genotype and phenotype of COVID-19: Their roles in pathogenesis. *Journal of Microbiology, Immunology and Infection*, *54*(2). https://doi.org/10.1016/j.jmii.2020.03.022

Nersisyan, S., Shkurnikov, M., Turchinovich, A., Knyazev, E., & Tonevitsky, A. (2020). Integrative analysis of miRNA and mRNA sequencing data reveals potential regulatory mechanisms of ACE2 and TMPRSS2. *PLOS ONE*, *15*(7), e0235987. https://doi.org/10.1371/journal.pone.0235987

Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, M., & Agha, R. (2020). The Socio-Economic Implications of the Coronavirus and COVID-19 Pandemic: A Review. *International Journal of Surgery*, *78*(78), 185–193. NCBI. https://doi.org/10.1016/j.ijsu.2020.04.018

Paoloni-Giacobino, A., Chen, H., Peitsch, M. C., Rossier, C., & Antonarakis, S. E. (1997). Cloning of the TMPRSS2 Gene, Which Encodes a Novel Serine Protease with Transmembrane, LDLRA, and SRCR Domains and Maps to 21q22.3. *Genomics*, *44*(3), 309–320. https://doi.org/10.1006/geno.1997.4845

Pummerer, L., Böhm, R., Lilleholt, L., Winter, K., Zettler, I., & Sassenberg, K. (2022). Conspiracy Theories and Their Societal Effects During the COVID-19 Pandemic.

Social Psychological and Personality Science, *13*(1), 194855062110002. https://doi.org/10.1177/19485506211000217

Robinson, E. L., Alkass, K., Bergmann, O., Maguire, J. J., Roderick, H. L., & Davenport, A. P. (2020). Genes encoding ACE2, TMPRSS2 and related proteins mediating SARS-CoV-2 viral entry are upregulated with age in human cardiomyocytes. *J Mol Cell Cardiol*, *147*, 88–91. https://doi.org/https://dx.doi.org/10.1016%2Fj.yjmcc.2020.08.009

Rozhkov, M., Ivanov, D., Blackhurst, J., & Nair, A. (2022). Adapting supply chain operations in anticipation of and during the COVID-19 Pandemic. *Omega*, 102635. https://doi.org/10.1016/j.omega.2022.102635

Sallard, E., Halloy, J., Casane, D., van Helden, J., & Decroly, É. (2020). Retrouver les origines du SARS-CoV-2 dans les phylogénies de coronavirus. *Médecine/Sciences*, *36*(8-9), 783–796. https://doi.org/10.1051/medsci/2020123

Singh, D., & Yi, S. V. (2021). On the origin and evolution of SARS-CoV-2. *Experimental & Molecular Medicine*, *53*. https://doi.org/10.1038/s12276-021-00604-z

Song, Z., Xu, Y., Bao, L., Zhang, L., Yu, P., Qu, Y., Zhu, H., Zhao, W., Han, Y., & Qin, C. (2019). From SARS to MERS, Thrusting Coronaviruses into the Spotlight. *Viruses*, *11*(1), 59. https://doi.org/10.3390/v11010059

Tang, T., Bidon, M., Jaimes, J. A., Whittaker, G. R., & Daniel, S. (2020). Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Research*, *178*, 104792. https://doi.org/10.1016/j.antiviral.2020.104792

Temgoua, M. N., Endomba, F. T., Nkeck, J. R., Kenfack, G. U., Tochie, J. N., & Essouma, M. (2020). Coronavirus Disease 2019 (COVID-19) as a Multi-Systemic Disease and its Impact in Low- and Middle-Income Countries (LMICs). *SN Comprehensive Clinical Medicine*, *2*(9), 1377–1387. https://doi.org/10.1007/s42399-020-00417-7

Temmam, S., Vongphayloth, K., Salazar, E. B., Munier, S., Bonomi, M., Regnault, B., Douangboubpha, B., Karami, Y., Chrétien, D., Sanamxay, D., Xayaphet, V., Paphaphanh, P., Lacoste, V., Somlor, S., Lakeomany, K., Phommavanh, N., Pérot, P., Dehan, O., Amara, F., & Donati, F. (2022). Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature*. https://doi.org/10.1038/s41586-022-04532-4

Thunders, M., & Delahunt, B. (2020). Gene of the month: TMPRSS2 (transmembrane serine protease 2). *Journal of Clinical Pathology*, jclinpath-2020-206987. https://doi.org/10.1136/jclinpath-2020-206987

Vaarala, M. H., Porvari, K. S., Kellokumpu, S., Kyllönen, A. P., & Vihko, P. T. (2001). Expression of transmembrane serine protease TMPRSS2 in mouse and human tissues. *The Journal of Pathology*, *193*(1), 134–140. https://doi.org/3.0.co;2-t">10.1002/1096-9896(2000)9999:9999<::aid-path743>3.0.co;2-t

Wirkner, J., Christiansen, H., Knaevelsrud, C., Lüken, U., Wurm, S., Schneider, S., & Brakemeier, E.-L. (2022). Mental Health in Times of the COVID-19 Pandemic. *European Psychologist*, *26*(4), 310–322. https://doi.org/10.1027/1016-9040/a000465

Yépez, Y., Marcano-Ruiz, M., Bezerra, R. S., Fam, B., Ximenez, J. P., Silva Jr, W. A., & Bortolini, M. C. (2022). Evolutionary history of the SARS-CoV-2 Gamma variant of concern (P.1): a perfect storm. *Genetics and Molecular Biology*, *45*(1). https://doi.org/10.1590/1678-4685-gmb-2021-0309

Zhang, C., Zhang, Y., Zhang, S., Wang, Z., Sun, S., Liu, M., Chen, Y., Dong, N., & Wu, Q. (2020). Intracellular autoactivation of TMPRSS11A, an airway epithelial transmembrane serine protease. *Journal of Biological Chemistry*, *295*(36), 12686–12696. https://doi.org/10.1074/jbc.ra120.014525

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., & Zheng, X.-S. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798). https://doi.org/10.1038/s41586-020-2012-7