Ancient Adaptative Evolution of *ACE2* in East Asians

Wenqiang Wang and Guan-Zhu Han*

College of Life Sciences, Nanjing Normal University, Jiangsu, China

*Corresponding author: E-mail: guanzhu@njnu.edu.cn.

Accepted: 21 July 2021

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been posing an unprecedented challenge to global public health. SARS-CoV-2 and several other coronaviruses utilize angiotensin-converting enzyme 2 (ACE2) as their entry receptors. The *ACE2* gene has been found to experience episodic positive selection across mammals. However, much remains unknown about how the *ACE2* gene evolved in human populations. Here, we use population genetics approaches to investigate the evolution of the *ACE2* gene in 26 human populations sampled globally. We find the *ACE2* gene exhibits an extremely low nucleotide diversity in the East Asian populations. Strong signals of selective sweep are detected in the East Asian populations, but not in the other human populations. The selective sweep in *ACE2* is estimated to begin in East Asian populations ~23,600 years ago. Our study provides novel insights into the evolution of the *ACE2* gene within human populations.

Key words: *ACE2*, coronavirus, selective sweep.

Significance

SARS-CoV-2 utilizes angiotensin-converting enzyme 2 (ACE2) as its entry receptor, but little is known about the evolution of the *ACE2* gene in human populations. We detect strong signals of selective sweep in the East Asian populations, but not in other human populations. Our study provides novel insights into the evolution of *ACE2* within the human populations.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the pathogen that causes Coronavirus Disease 2019 (COVID-19), has been posing an unprecedented threat to global public health, resulting in more than 179 million infections and 3.90 million deaths (https://www.who.int/emergencies/diseases/novel-coronavirus-2019; as to June 26, 2021). To date, at least seven coronaviruses have been known to infect humans, including human coronavirus (HCoV)-229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1, Middle East respiratory syndrome coronavirus (MERS-CoV), SARS-CoV, and SARS-CoV-2 (Drosten et al. 2003; Pene et al. 2003; Vabret et al. 2003; van der Hoek et al. 2005; Woo et al. 2005; Zaki et al. 2012; Zhu et al. 2020). Among them, four coronaviruses, namely HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1, are globally endemic and usually cause mild to moderate respiratory infections (Gaunt et al. 2010). Three novel coronaviruses, namely SARS-CoV, MERS-CoV, and SARS-CoV-2, with high case fatality rate have emerged in human populations in the past two decades (Petrosillo et al. 2020).

Binding of cellular surface protein known as viral receptor by spike (S) protein is the initial step of coronavirus infection. To date, the receptors of seven HCoVs have been identified: HCoV-OC43 and HCoV-HKU1 interact with 9-O-acetylsialic acid to invade host cells (Vlasak et al. 1988; Huang et al. 2015); HCoV-229E and MERS-CoV utilize amino-peptidase N (ANPEP) and dipeptidyl-peptidase 4 (DPP4) as their receptors, respectively (Yeager et al. 1992; Widagdo et al. 2016); SARS-CoV, HCoV-NL63, and SARS-CoV-2 utilize angiotensin-converting enzyme 2 (ACE2) as their receptors to initialize viral infection (Li et al. 2003; Hofmann et al. 2005; Zhou et al. 2020). ACE2 and its homolog ACE are the main enzymes of renin–angiotensin system (RAS), which is a key regulator of maintaining blood pressure homeostasis and fluid and salt balance (Tipnis et al. 2000; Bosso et al. 2020). The octapeptide angiotensin (ANG) II, an intermediate produced by ACE in RAS, promotes vasoconstriction, inflammatory, and fibrotic properties (Gaddam et al. 2014). ACE2 can convert Ang II to Ang (1–7), which mediates vasodilatation (Warner et al. 2004; Gaddam et al. 2014). Thus ACE2 acts as a negative regulator in the RAS to counterbalance the effects of ACE (Kuba et al. 2006). The human *ACE2* gene encodes an 805 amino acid protein with a highly conserved active zinc-binding motif His-Glu-X-X-His (HEXXH motif) and is located on chromosome X (Hamming et al. 2007).

Viral receptors, essentially "normal" host cellular proteins hijacked by viruses, have been thought to be subject to two conflicting directional forces, namely negative selection due to functional constraints to maintain their own cellular function and positive selection due to everchanging host-virus arms race (Wang et al. 2020). The *ACE2* gene has been found to experience episodic positive selection across mammals, including bats and primates (Demogines et al. 2012; Damas et al. 2020; Wang et al. 2020). Many positively selected sites overlap the interaction interface between ACE2 and coronaviruses, implying that ancient recurrent evolutionary arms races occurred between mammals and coronaviruses (Demogines et al. 2012; Damas et al. 2020; Wang et al. 2020). Our previous study indicates that the *ACE2* gene is likely to have experienced local adaptation in the Chinese Han population (Wang et al. 2020). However, much remains unknown about the evolution of the *ACE2* gene in human populations. In this study, we used population genetics approaches to explore the evolution of the *ACE2* gene in a total of 26 human populations sample globally, and detected strong signals of selective sweep in East Asian populations.

Results

Here, we employed a series of population genetics approaches to analyze the evolution of the *ACE2* gene in 26 human populations sampled globally, including seven African populations, four American populations, five East Asian populations, five European populations, and five South Asian populations (Sudmant et al. 2015). Initially, we found that the five East Asian populations, namely CDX (Chinese Dai in Xishuangbanna, China), CHS (Han Chinese South), CHB (Han Chinese), JPT (Japanese in Tokyo, Japan), and KHV (Kinh in Ho Chi Minh City, Vietnam), exhibit much lower nucleotide diversity than the other human populations (fig. 1*A*), raising the possibility that the *ACE2* gene underwent natural selection in East Asian populations. To investigate the possibility of natural selection shaping the evolution of *ACE2* in human populations, we first employed the neutrality test to detect selection signals. We found that the *ACE2* gene displays a *D* value of significantly less than 0 in all the five East Asian populations, CDX ($D = -1.75$; $P = 0.009$ and $P = 0.028$ for coalescence simulations under consistent population size and population growth, respectively), CHB ($D = -1.93$; $P = 0.006$ and $P = 0.015$), CHS ($D = -1.74$; $P = 0.013$ and $P = 0.031$), JPT ($D = -1.66$; $P = 0.022$ and $P = 0.040$), and KHV ($D = -1.82$; $P = 0.009$ and $P = 0.021$). In contrast, no significant signal of natural selection was detected in other human populations (fig. 1*B*). Moreover, we also merged the population genetic data based on the continents (AFR, African; AMR, Ad Mixed American; EAS, East Asian; EUR, European; SAS, South Asian). The continent-level analyses show the *ACE* gene of the East Asian super-population exhibits the lowest nucleotide diversity, and signal of selective sweep was only detected in the East Asian super-population ($D = -2.233$; $P < 0.001$) (supplementary table S1, Supplementary Material online). These results suggest that the East Asian populations might have experienced a common selective sweep in the past.

We also performed integer neighbor joining network analyses to investigate haplotype structures of *ACE2* among different human populations. The haplotype networks for all the five East Asian populations display a star-like structure, and many haplotypes of relatively low frequency are connected with the main haplotypes with short branches (supplementary fig. S1, Supplementary Material online), supporting selective sweep occurring in East Asian populations. Unlike East Asian populations, the *ACE2* haplotypes in non-East Asians appear to be more scattered with long step length (supplementary fig. S1, Supplementary Material online). Statistical analyses based on haplotypes also show that five East Asian populations exhibit the lowest nucleotide diversity and *D* values of significantly less than 0 (supplementary table S2, Supplementary Material online). Moreover, we merged the population genetic data according to the continent of populations for selection analysis. We found that the pattern of the main haplotype surrounding excess low-frequency haplotypes became more pronounced in the East Asian populations (supplementary fig. S1 and table S1, Supplementary Material online). These results further confirm that the *ACE2* gene of East Asian populations is likely to have underwent a selective sweep.

We also employed the integrated haplotype score (iHS) statistic to detect evidence of selective sweep in the *ACE2* gene in human populations. We found that four East Asian populations possess a high proportion of SNPs with extremely high |iHS| values within the *ACE2* gene (80% in CDX, 69% in CHB, 50% in JPT, 80% in KHV; fig. 2). Most of the non-East Asian populations do not possess SNPs with extremely high |iHS| values in the *ACE2* gene, except three populations with sporadic outlying SNPs (1.4% in LWK [Luhya in Webuye, Kenya], 2.4% in ITU [Indian Telugu in UK], 7.3% in STU [Sri Lankan Tamil in UK]; fig. 2). Moreover, we also merged the population genetic data based on the continents. The continent-level analyses show that 60% of SNPs within the *ACE2* gene display extremely high |iHS| values in East Asian populations, whereas no SNP with extremely high |iHS| values

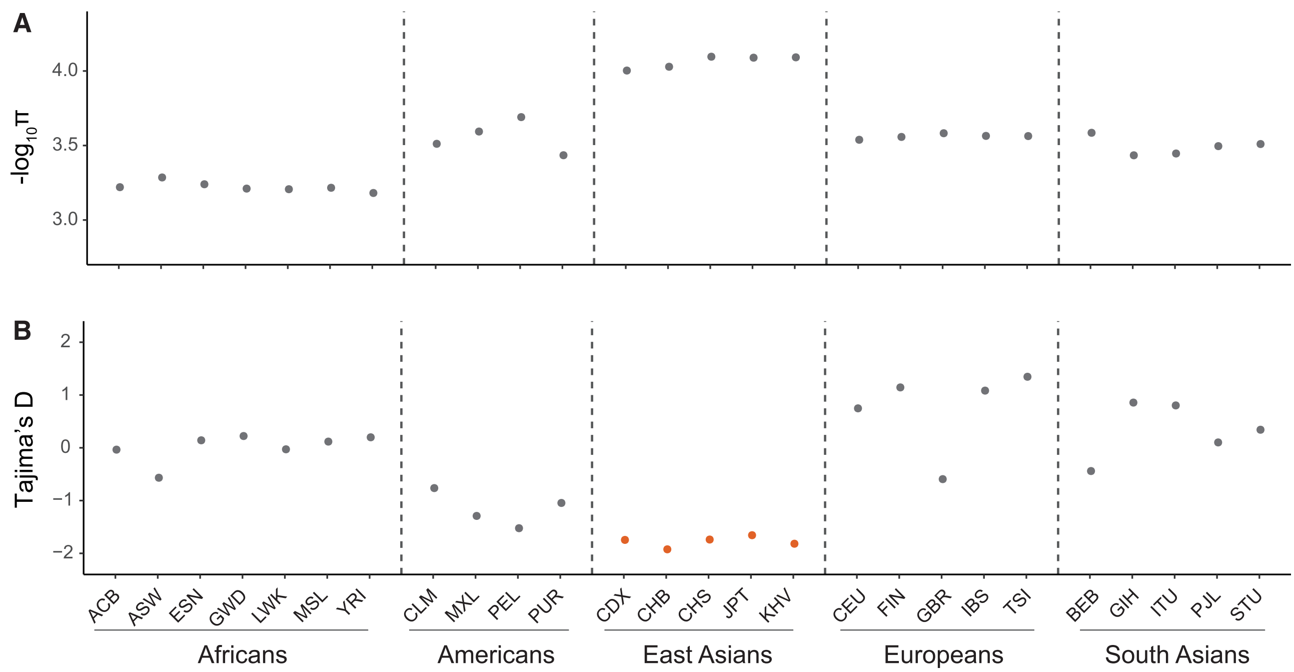

Fig. 1.—Signals of natural selection in the *ACE2* in human populations. (*A*) Nucleotide diversity (π) of the *ACE2* gene in 26 human populations (see supplementary methods, Supplementary Material online, for human population abbreviations). Populations from different continents are separated by dashed lines. (*B*) Tajima's *D* values of the *ACE2* gene in 26 human populations. Significant *D* values ($P < 0.05$) are highlighted in orange.

was identified in the human populations from the other continents (supplementary fig. S2, Supplementary Material online). Taken together, our results further support that the *ACE2* gene might have experienced a unique selective sweep in the East Asian populations.

To date the selective sweep occurring in the *ACE2* gene in the East Asian populations, we used a method based on ancestral recombination graph to estimate the starting time of selection for each SNPs of the *ACE2* gene. The selective sweep in *ACE2* was estimated to begin in East Asian populations approximately 844 generations ago (~23,630 years, with generation time of humans to be 28 years) (fig. 3).

Discussion

In this study, we found strong signal of selective sweep in the *ACE2* gene in East Asian populations. The selective sweep was estimated to begin near 25,000 years ago. What drove the adaptive evolution of *ACE2* in East Asian populations? ACE2 is the receptor for at least three human coronaviruses, including SARS-CoV, SARS-CoV-2, and HCoV-NL63 (Li et al. 2003; Hofmann et al. 2005; Zhou et al. 2020). The arms race between viruses and hosts has been found to drive the accelerated adaptive evolution in viral receptors (Wang et al. 2020). Previous studies show positively selected sites detected in bat *ACE2* genes overlap almost perfectly to its interaction interface with SARS-CoV and HCoV-NL63, indicating ACE2 may have been utilized by coronaviruses for millions of years (Demogines et al. 2012). Therefore, we propose that an *ACE2*-utilizing virus epidemic is likely to occur in East Asia around 25,000 years ago and might have been circulating for some time, which might have driven the adaptive evolution of *ACE2* in the common ancestor of East Asians. Consistent with this possibility, unique signals of selective sweep in East Asians were also detected in 42 other proteins that interact with coronaviruses, and the selective sweep was estimated to start around 25,000 year ago (Souilmi et al. 2021). However, it is hard to find direct evidence for the occurrence of the ancient virus epidemic. Moreover, the possibility that nonviral factors have driven the adaptive evolution of the *ACE2* gene in East Asians cannot be unambiguously excluded. Nevertheless, our study provides novel insights into the evolution of *ACE2* in human populations.

Materials and Methods

Neutrality Test

The genetic variation data of the *ACE2* gene from 26 human populations were retrieved from the pilot 3 phase of the 1,000 Genomes Project (Sudmant et al. 2015). Considering that *ACE2* is located on the X chromosome, only female population genetic data were utilized in this analysis to avoid the conflict caused by the coexistence of different ploidies. The nucleotide diversity (π) and Tajima's *D* values of the *ACE2* gene were estimated using the DNASP v6, and the statistical significance of the neutrality test is evaluated based on coalescent simulations with constant population size and no

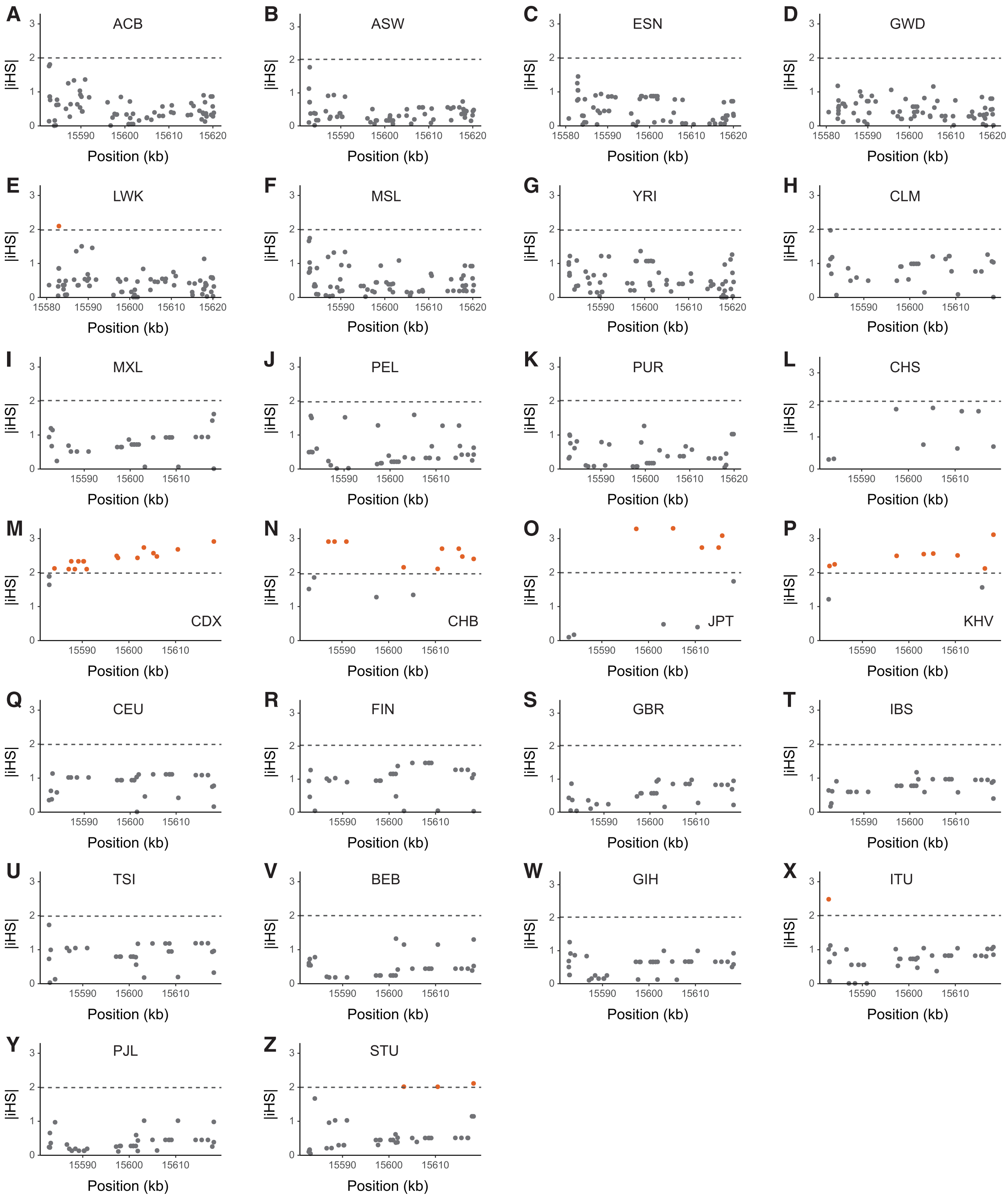

Fig. 2.—The |iHS| scores of SNPs within the *ACE2* gene in human populations. Dotted lines represent the top 5% of |iHS| scores at chromosome X, and outlier SNPs are labeled in orange. (*A*–*G*) belong to Africans. (*H*–*K*) belong to Americans. (*L*–*P*) belong to East Asians. (*Q*–*U*) belong to Europeans. (*V*–*Z*) belong to Southeast Asians.

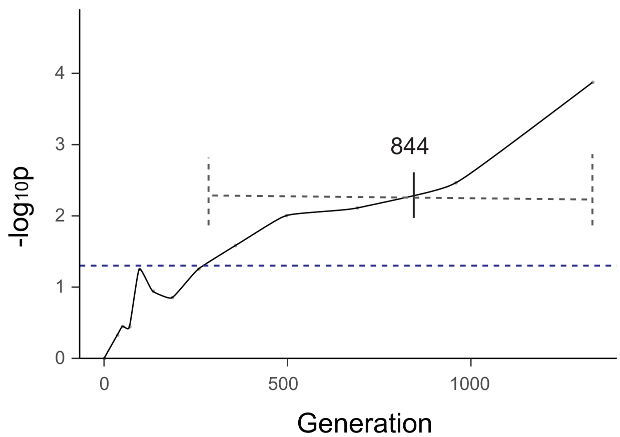

Fig. 3.—Timing of selective sweep occurring in the *ACE2* gene in East Asian populations. Generations of the selection for the SNP with lowest *P* value, rs2106809, are shown. The significance of $P = 0.05$ is labeled by a blue dotted line. The two vertical dotted lines represent the upper generation and lower generation of selection, and the average value is labeled using a vertical solid line.

recombination (Rozas et al. 2017). For those populations with significant D values, we further conducted 1,000 coalescence simulations using the msms software (Ewing and Hermisson 2010) under the demographic model of population growth (Gravel et al. 2011) and with the recombination rate inferred from AAmap (Hinch et al. 2011). The command used for simulation is: msms -N 10,000 -ms 50 1,000 -s 50 -r 23 41,000 -G 38.698 -eN 0.0315 0.2955 -oTPi. Tajima's *D* is considered to be different from 0, only when *P* values are significant in both simulations.

We also merged the population genetic data based on the five continents (AFR, African; AMR, Ad Mixed American; EAS, East Asian; EUR, European; SAS, South Asian) and performed neutrality test. AFR includes ESN (Esan in Nigeria), GWD (Gambian in Western Division, Gambia), LWK (Luhya in Webuye, Kenya), MSL (Mende in Sierra Leone), and YRI (Yoruba in Ibadan, Nigeria); AMR includes CLM (Colombian in Medellin, Colombia), MXL (Mexican ancestry in Los Angeles, California), PEL (Peruvian in Lima, Peru), and PUR (Puerto Rican in Puerto Rico); EAS includes CDX (Chinese Dai in Xishuangbanna, China), CHB (Han Chinese), CHS (Han Chinese South), JPT (Japanese in Tokyo, Japan), and KHV (Kinh in Ho Chi Minh City, Vietnam); EUR includes CEU (Utah residents with Northern and Western European ancestry), FIN (Finnish in Finland), GBR (British in England and Scotland), IBS (Iberian populations in Spain), and TSI (Toscani in Italy); SAS includes BEB (Bengali in Bangladesh), GIH (Gujarati Indians in Houston), ITU (Indian Telugu in UK), PJL (Punjabi in Lahore, Pakistan), and STU (Sri Lankan Tamil in UK). Note that AFR only includes five indigenous African populations, but does not include two African-descended populations, ACB (African Caribbean in Barbados) and ASW (African Ancestry in Southwest United States), with unique recent history.

Haplotype Network Construction

Haplotype data were transformed from VCF files in the 1,000 Genomes Project. We reconstructed the haplotype network of human populations based on the integer neighbor joining method implemented in PopART v1.7 (Leigh and Bryant 2015). Statistic based on haplotypes was also performed using PopART v1.7 (Leigh and Bryant 2015).

iHS Analysis

Selscan was performed to evaluate iHS by scanning all the SNPs of chromosome X, and only biallelic SNPs with the lowest frequency greater than 0.05 were used in the analysis (Szpiech and Hernandez 2014). After normalizing the unstandardized iHS scores separately in each population with 10 equally sized allele frequency bins by norm v1.3.0, we calculated the absolute value of these data, given that both the significantly high positive and negative values represent long haplotype homozygosity, and their difference is limited to whether ancestral alleles or derived alleles maintain long haplotype homozygosity. We ranked all the SNPs on chromosome X according to the |iHS| scores to delimit the top 5% region, and inferred whether SNPs of the *ACE2* gene are enriched in the plateau area. We also merged the population genetic data based on the five continents and performed iHS test.

Timing Selection

An ancestral recombination graph approach implemented in Relate was utilized to estimate the starting time of the selective sweep (Speidel et al. 2019). The recombination rate data used for the analysis were extracted from Hinch's research (Hinch et al. 2011). Relate calculates the *P* value based on the frequency of the derived alleles and the number of lineages in the general coalescent tree at specific generation to quantify the selection of each SNP. We identified the SNP with the lowest *P* value in the *ACE2* gene, and the lowest *P* value must be less than 10^{-3} to ensure the reliability of the selection. After identified the SNP with lowest *P* value, rs2106809, we calculated the selection time of the *ACE2* gene based on the formula Time = (upper generation + lower generation)/ 2 × generation time, where the upper generation represents the generation with the most significant *P* value, and the lower generation represents the generation with the least significant *P* value but less than 0.05.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (31922001).

Data Availability

No new data were generated in the course of the study.

Literature Cited

Bosso M, et al. 2020. The two faces of ACE2: the role of ACE2 receptor and its polymorphisms in hypertension and COVID-19. Mol Ther Methods Clin Dev. 18:321–327.

Damas J, et al. 2020. Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. Proc Natl Acad Sci U S A. 117(36):22311–22322.

Demogines A, Farzan M, Sawyer SL. 2012. Evidence for ACE2-utilizing coronaviruses (CoVs) related to severe acute respiratory syndrome CoV in bats. J Virol. 86(11):6350–6353.

Drosten C, et al. 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med. 348(20):1967–1976.

Ewing G, Hermisson J. 2010. MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. Bioinformatics 26(16):2064–2065.

Gaddam RR, Chambers S, Bhatia M. 2014. ACE and ACE2 in inflammation: a tale of two enzymes. Inflamm Allergy Drug Targets. 13(4):224–234.

Gaunt ER, Hardie A, Claas ECJ, Simmonds P, Templeton KE. 2010. Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method. J Clin Microbiol. 48(8):2940–2947.

Gravel S, et al. 2011. Demographic history and rare allele sharing among human populations. Proc Natl Acad Sci U S A. 108(29):11983–11988.

Hamming I, et al. 2007. The emerging role of ACE2 in physiology and disease. J Pathol. 212(1):1–11.

Hinch AG, et al. 2011. The landscape of recombination in African Americans. Nature 476:170–175.

Hofmann H, et al. 2005. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. Proc Natl Acad Sci U S A. 102(22):7988–7993.

Huang X, et al. 2015. Human coronavirus HKU1 spike protein uses O-acetylated sialic acid as an attachment receptor determinant and employs hemagglutinin-esterase protein as a receptor-destroying enzyme. J Virol. 89(14):7202–7213.

Kuba K, Imai Y, Penninger JM. 2006. Angiotensin-converting enzyme 2 in lung diseases. Curr Opin Pharmacol. 6(3):271–276.

Leigh JW, Bryant D. 2015. PopART: full-feature software for haplotype network construction. Methods Ecol Evol. 6(9):1110–1116.

Li W, et al. 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 426(6965):450–454.

Pene F, et al. 2003. Coronavirus 229E related pneumonia in immunocompromised patients. Clin Infect Dis. 37(7):929–932.

Petrosillo N, Viceconte G, Ergonul O, Ippolito G, Petersen E. 2020. COVID-19, SARS and MERS: are they closely related? Clin Microbiol Infect. 26(6):729–734.

Rozas J, et al. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 34(12):3299–3302.

Souilmi Y, et al. 2021. An ancient coronavirus-like epidemic drove adaptation in East Asians from 25,000 to 5,000 years ago. Curr Biol. doi: 10.1016/j.cub.2021.05.067.

Speidel L, Forest M, Shi S, Myers SR. 2019. A method for genome-wide genealogy estimation for thousands of samples. Nat Genet. 51(9):1321–1329.

Sudmant PH, et al. 2015. An integrated map of structural variation in 2,504 human genomes. Nature 526(7571):75–81.

Szpiech ZA, Hernandez RD. 2014. Selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol Biol Evol. 31(10):2824–2827.

Tipnis SR, et al. 2000. A human homolog of angiotensin-converting enzyme: cloning and functional expression as a captopril-insensitive carboxypeptidase. J Biol Chem. 275(43):33238–33243.

Vabret A, Mourez T, Gouarin S, Petitjean J, Freymuth F. 2003. An outbreak of coronavirus OC43 respiratory infection in Normandy, France. Clin Infect Dis. 36(8):985–989.

van der Hoek L, et al. 2005. Croup is associated with the novel coronavirus NL63. PLoS Med. 2(8):e240.

Vlasak R, Luytjes W, Spaan W, Palese P. 1988. Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. Proc Natl Acad Sci U S A. 85(12):4526–4529.

Wang W, Zhao H, Han GZ. 2020. Host-virus arms races drive elevated adaptive evolution in viral receptors. J Virol. 94(16):e00684–20.

Warner FJ, Smith AI, Hooper NM, Turner AJ. 2004. Angiotensin-converting enzyme-2: a molecular and cellular perspective. Cell Mol Life Sci. 61(21):2704–2713.

Widagdo W, et al. 2016. Differential expression of the Middle East respiratory syndrome coronavirus receptor in the upper respiratory tracts of humans and dromedary camels. J Virol. 90(9):4838–4842.

Woo PC, et al. 2005. Clinical and molecular epidemiological features of coronavirus HKU1 associated community acquired pneumonia. J Infect Dis. 192(11):1898–1907.

Yeager CL, et al. 1992. Human aminopeptidase N is a receptor for human coronavirus 229E. Nature 357(6377):420–422.

Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med. 367(19):1814–1820.

Zhou P, et al. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579(7798):270–273.

Zhu N, et al. 2020. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 382(8):727–733.

Associate editor: Wen-Hsiung Li