AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE

Comparative Genomics and Integrated Network Approach Unveiled Undirected Phylogeny Patterns, Co-mutational Hot Spots, Functional Cross Talk, and Regulatory Interactions in SARS-CoV-2

Vipin Gupta,[a] Shazia Haider,[b] Mansi Verma,[c] Nirjara Singhvi,[d] Kalaisaran Ponnusamy,[e] Md. Zubbair Malik,[f] Helianthous Verma,[g] Roshan Kumar,[h] Utkarsh Sood,[i] Princy Hira,[a] Shiva Satija,[c] Yogendra Singh,[d] Rup Lal[i]

[a]PhiXGen Private Limited, Gurugram, Haryana, India
[b]Jaypee Institute of Information Technology, Noida, Uttar Pradesh, India
[c]Department of Zoology, Sri Venkateswara College, University of Delhi, New Delhi, India
[d]Department of Zoology, University of Delhi, Delhi, India
[e]School of Biotechnology, Jawaharlal Nehru University, New Delhi, India
[f]School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India
[g]Department of Zoology, Ramjas College, University of Delhi, Delhi, India
[h]P.G. Department of Zoology, Magadh University, Bodh Gaya, Bihar, India
[i]The Energy and Resources Institute, New Delhi, India

Vipin Gupta, Shazia Haider, and Mansi Verma contributed equally. Author order was determined by drawing straws.

ABSTRACT The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has resulted in 92 million cases in a span of 1 year. The study focuses on understanding population-specific variations attributing its high rate of infections in specific geographical regions particularly in the United States. Rigorous phylogenomic network analysis of complete SARS-CoV-2 genomes (245) inferred five central clades named a (ancestral), b, c, d, and e (subtypes e1 and e2). Clade d and subclade e2 were found exclusively comprised of U.S. strains. Clades were distinguished by 10 co-mutational combinations in Nsp3, ORF8, Nsp13, S, Nsp12, Nsp2, and Nsp6. Our analysis revealed that only 67.46% of single nucleotide polymorphism (SNP) mutations were at the amino acid level. T1103P mutation in Nsp3 was predicted to increase protein stability in 238 strains except for 6 strains which were marked as ancestral type, whereas co-mutation (P409L and Y446C) in Nsp13 were found in 64 genomes from the United States highlighting its 100% co-occurrence. Docking highlighted mutation (D614G) caused reduction in binding of spike proteins with angiotensin-converting enzyme 2 (ACE2), but it also showed better interaction with the TMPRSS2 receptor contributing to high transmissibility among U.S. strains. We also found host proteins, MYO5A, MYO5B, and MYO5C, that had maximum interaction with viral proteins (nucleocapsid [N], spike [S], and membrane [M] proteins). Thus, blocking the internalization pathway by inhibiting MYO5 proteins which could be an effective target for coronavirus disease 2019 (COVID-19) treatment. The functional annotations of the host-pathogen interaction (HPI) network were found to be closely associated with hypoxia and thrombotic conditions, confirming the vulnerability and severity of infection. We also screened CpG islands in Nsp1 and N conferring the ability of SARS-CoV-2 to enter and trigger zinc antiviral protein (ZAP) activity inside the host cell.

IMPORTANCE In the current study, we presented a global view of mutational pattern observed in SARS-CoV-2 virus transmission. This provided a who-infect-whom geographical model since the early pandemic. This is hitherto the most comprehensive

Citation Gupta V, Haider S, Verma M, Singhvi N, Ponnusamy K, Malik MZ, Verma H, Kumar R, Sood U, Hira P, Satija S, Singh Y, Lal R. 2021. Comparative genomics and integrated network approach unveiled undirected phylogeny patterns, co-mutational hot spots, functional cross talk, and regulatory interactions in SARS-CoV-2. mSystems 6:e00030-21. https://doi.org/10.1128/mSystems.00030-21.

Editor Paola Flórez de Sessions, Oxford Nanopore Technologies

Address correspondence to Rup Lal, ruplal@gmail.com.

SARS-CoV-2 mutations

Received 14 January 2021
Accepted 30 January 2021
Published 23 February 2021