Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Brief Communication

*Haemophilus influenzae* and SARS-CoV-2: Is there a role for investigation?

Biswaayoti Borkakotya,*, Nargis K. Bali

* Indian Council of Medical Research-Regional Medical Research Centre for NE Region, Bokel, Dibrugarh, 786010, Assam, India
* Department of Clinical Microbiology, Sher-i-Kashmir Institute of Medical Sciences, Bokel, Srinagar, Jammu & Kashmir, India

A R T I C L E   I N F O

Keywords:

*Haemophilus influenzae*
COVID-19
SARS-CoV-2
E-gene
qRT-PCR

A B S T R A C T

During the current pandemic of COVID-19, the authors observed that during screening test for SARS-CoV-2 targeting the E-gene by qRT-PCR, few nasopharyngeal/oropharyngeal samples showed amplification signals at late cycle threshold (C\textsubscript{T} value) > 35 despite being negative for other confirmatory target genes. Thirty such samples (taken as cases) showed detectable C\textsubscript{T} of > 35 cycle in E-gene which were negative for other target genes of SARS-CoV-2 and 30 samples with undetectable fluorescence in E-gene were taken as controls for investigation. An in-vitro diagnostic approved commercial qRT-PCR multiplex kit detecting 33 respiratory pathogens which can also detect *Haemophilus influenzae* was used for screening the samples. It was observed that out of the 30 samples showing detectable C\textsubscript{T} > 35 in E-gene, 11 samples were positive for *Haemophilus influenzae* whereas in the controls only three samples were positive for *H. influenzae* (p-value: 0.03) which was statistically significant. Further, the probes and primers were screened against *H. influenzae* for matches in the genome. It was observed that all primers and probes for the E-gene of SARS-CoV-2 had over 13 bp long sequences matching 100% with multiple sites across the *H. influenzae* genome. This qRT-PCR primer & probes are being used extensively across India, and laboratories using them should be aware of the cross-reactivity of primers & probes with the *H. influenzae* genome.

Further, the authors observed that 95.9% (5415/5642) of COVID-19 positive cases detected in their laboratory were asymptomatic at the time of collection of samples. This warrants further investigations.

1. Introduction

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) responsible for the current global pandemic of coronavirus disease 2019 (COVID-19), is a single-stranded positive sense RNA virus of 29.9 kbp genome with four major structural proteins (spike, envelope, membrane and nucleocapsid protein), few accessory proteins and 16 non-structural proteins [1]. The guanosine-cytosine (GC) content of SARS-CoV-2 is approximately 38% and its GC content is similar to many bacterial species [2]. *Haemophilus influenzae* spp. are Gram negative facultative anaerobic coccobacilli that form part of the normal microflora of the oral cavity in human beings [2]. *Haemophilus influenzae* the type species of the genus, with an approximate genome size of ~1.8Mbp, has six serotypes (a-f) with serotype b mainly responsible for causing bacterial meningitis, lower respiratory tract infections, otitis media, bacteraemia etc. in infants and young children [2,3]. Various studies across the globe have reported the detection of pre-existing T-cell immune response in 20–50% of samples collected between 2015-2018 to SARS-CoV-2 even before the novel coronavirus was introduced to the human population [4]. This implicates that even in persons unexposed to SARS-CoV-2, various other closely related pathogens including other betacoronavirus infecting humans i.e. HCoV-OC-43 or HCoV-HKU-1 or alphacoronavirus HCoV-229E & HCoV-NL63 may induce partial T-cell immune response to SARS-CoV-2 [4], which may either be deleterious or protective. Further, the role of other closely-related pathogens sharing short nucleotide sequences in the genome coding for up to 8–11 amino acids epitopes should be investigated for any role in inducing T-cell response. During the current pandemic of COVID-19, it was observed by the authors that during screening test for SARS-CoV-2 targeting the E-gene by qRT-PCR, some nasopharyngeal/oropharyngeal samples showed amplification signals at late cycle threshold (C\textsubscript{T} value) > 35 despite being negative for other confirmatory target genes i.e. ORF1ab and RdRP genes. Therefore, an investigation was carried out in May–August 2020, to know if there was any relationship of SARS-CoV-2 with other respiratory pathogens including sharing of short nucleotide sequences of the SARS-CoV-2 genome, the primers and probes used for diagnosis were investigated with other common respiratory pathogens including *Haemophilus influenzae*. 

* Corresponding author. ICMR-Regional Medical Research Centre, Dibrugarh, 786010, Assam, India.
E-mail address: biswaborkakoty@gmail.com (B. Borkakoty).

https://doi.org/10.1016/j.ijmmb.2021.03.008
Received 26 September 2020; Accepted 15 December 2020
Available online 26 March 2021
0255-0857/© 2021 Indian Association of Medical Microbiologists. Published by Elsevier B.V. All rights reserved.
2. Material and methods

The authors selected 30 nasopharyngeal samples which had shown late C_{T} value of > 35 for E-gene of SARS-CoV-2 and were additionally negative for other confirmatory test targeting ORF1ab and RdRP genes as cases(group 1) for the study. In addition, 30 samples with undetectable fluorescence in E-gene were taken as controls (group 2) for investigation. SARS-CoV-2 were declared positive only after confirmatory genes targets (RdRp or ORF1ab genes or both) were positive in qRT-PCR along with the SARS-CoV-2 E-gene. An in-vitro diagnostics (IVD) approved commercial qRT-PCR multiplex kit detecting 33 respiratory pathogens (FTD respiratory pathogen 33, Fast Track Diagnosis, Netherlands) which can also detect Haemophilus influenzae was used for screening the samples. Further, analysis of the forward and reverse primers (E_Sarbeco_F1 (5'-ACAGTGAGTTAAATGTAATTACG-3') and probe (E_Sarbeco_P1 (5'-FAM-ACACCTAGCGATCTTACTGCGCTTCG-BHQ1)) were performed for cross-reactivity with Haemophilus influenzae. These primers and probes sequences shared by the World Health Organization (WHO) to the National Institute of Virology, Pune, India for COVID-19 diagnosis are being used extensively across India for diagnosis by qRT-PCR. Both the forward, reverse primers and probe were screened in the basic local alignment search tool (BLAST) available online at the National Centre for Biotechnology Information (NCBI) database. The nucleotide BLAST search was adjusted automatically for short sequence search and targeted Haemophilus influenzae (taxid:727) and optimized for somewhat similar sequences (BLASTN). Further, to know if SARS-CoV-2 shares similar sequences of over 15 bp across the genome of H. influenzae and other Haemophilus spp., nucleotide BLAST search of reference strain of SARS-CoV-2 (NC_045512.2) with similar parameters as above was performed in NCBI BLAST tool.

3. Results and discussions

It was observed that out of the 30 samples showing detectable C_{T} value > 35 in E-gene of SARS-CoV-2, 11 samples were positive for Haemophilus influenzae whereas in controls only three samples were positive for H. influenzae (p-value: 0.03) which was statistically significant. Table 1 shows the details of the screening of 33 different respiratory pathogens performed by a commercial multiplex qRT-PCR assay in the two groups. Apart from H. influenzae, the next common respiratory pathogen detected was Klebsiella pneumonia (7/30) followed by Staphylococcus aureus (5/30) and Streptococcus pneumoniae (3/30). In a meta-analysis study conducted elsewhere early in the pandemic (data upto April 2020), it was reported that only 7% hospitalized cases of COVID-19 had a bacterial co-infection with a significantly higher prevalence of co-infection (14%) in ICU settings [3]. The common bacterial co-infection detected were due to Mycoplasma pneumoniae, Pseudomonas aeruginosa and Haemophilus influenzae. [5] Further, another study from China (pre-print) have reported that Haemophilus influenzae was the most common coinfection in SARS-CoV-2 subjects [6]. However, there are no reported study till now studying the cross-reactivity of Haemophilus influenzae with SARS-CoV-2 in qRT-PCR test.

As shown in Fig. 1, all the primers and probes of E-gene of SARS-CoV-2 had over 14 bp (100% matched, Expect value of 0.44–6.8) across multiple sites of the complete genome of Haemophilus influenzae. It is to be noted that out of the combined total of 74 bp length of primers and probes of E-gene, 44 bp (60%) had 100% match either with plus or minus strand of Haemophilus influenzae. When primers and probes of the E-gene of SARS-CoV-2 were analyzed against H. influenzae genome, it was seen that they shared over 13bp long (100% match with forward and reverse primers) and 14 bp (100% match with probe) across multiple sites of H. influenzae (Fig. 2). It was observed that H. influenzae strain accession no. CP044497.1, (PittGG strain) CP043770.1 (biotype aegyptius strain) etc; with overall Expect value of 2e-08 and query coverage of 4% (which corresponds to roughly 1200 bp of the 29.9 kbp SARS-CoV-2 genome) has 91 matches across the whole genome with 95–100% matches in upto 21 bp long sequences. Similarity matches in both plus and minus strand show multiple matches of SARS-CoV-2 genome including ORF1ab, spike protein, envelope protein, ORF8 at 100% similarity in over 16 bp long sequences. Further, to know if other bacterial genera had significant matches in 15bp or more long stretches of SARS-CoV-2, similar BLASTN search were performed. It was found that there were 100% matches upto 23bp for Escherichia fergusonii (single match SARS-CoV-2 RefSeq position: 3176–3198) and Salmonella enterica subs. houtena (a single match across the genome with SARS-CoV-2 RefSeq position: 15518–15540). Further it was seen that the query-coverage was less than 1% for all the different Enterobacteriaceae family compared to 2–4% for the Haemophilus group.

This preliminary investigation did not look into the T-cell or B-cell epitopes shared between the Haemophilus group and SARS-CoV-2 pathogens. There is a probability that peptide of 8–11 amino acid length (that can trigger T-cell immune response) might match with peptide sequences of H. influenzae. Though, it is still inconclusive if protective immunity is conferred with past infections with closely related human coronavirus (HCoVs) including SARS [7], but the possibility of H. influenzae or H. influenzae type b vaccination inducing T-cell immune response against SARS-CoV-2 needs to be investigated further. It has been observed that T-cell reactivity associated with CD4+ T-cells against SARS-CoV-2 was detected in 50% of donor blood samples obtained between 2015 to 2018 even before SARS-CoV-2 appeared in human population [8]. It is speculated that T-cell immune response against SARS-CoV-2 in unexposed individuals might originate from past exposure to the HCoVs or common cold coronaviruses (CCCs) [8]. As the seropositivity of CCs in human population is high (~90%) [9], the significance of pre-existing immune reactivity for SARS-CoV-2 requires further investigations. It has also been observed at the authors' laboratory in Assam, India, at the time of sample collection, 95.9% (5415/5642) COVID-19 positive cases were asymptomatic or had very mild symptoms after screening approximately 1,43, 000 samples subjected to SARS-CoV-2 RT-PCR test (authors unpublished data). Due to the protocol implemented by the state of government of Assam, test for SARS-CoV-2 irrespective of symptoms among travelers coming from outside the state in airports and railway stations including primary contacts of positive cases, a high number of asymptomatic

| Respiratory Pathogen | Group 1 (n = 30) | Group 2 (n = 30) | p-value (Fisher’s exact test) |
|----------------------|-----------------|-----------------|-----------------------------|
| 1. Haemophilus influenzae | 11 | 03 | 0.03 |
| Type B/H influenzae | 07 | 02 | 0.14 |
| 2. Klebsiella pneumonia | 05 | 02 | 0.42 |
| 3. Staphylococcus aureus | 03 | 01 | 0.61 |
| 4. Streptococcus pneumoniae | 02 | 00 | 0.49 |
| 5. Moraxella catarrhalis | 02 | 00 | 1.0 |
| 6. Human parainfluenza virus-1/2/3/4 | 00 | 00 | – |
| 7. Influenza A/Influenza B/ H1N1(2009 H1N1) /RSV A:4b/IBRV/Coronavirusa:NL63/OC43/229E/ HKU1/MHPV/Bocavirus/ Parchoevirus /Entrovirus /Adenovirus | 00 | 00 | – |
| 8. Chlamydia pneumoniae/ Mycoplasma pneumoniae/ Legionella pneumophila /L infetnaciae/ Pneumocystis jiroveci/ Bordetella pertussis/ Salmonella species | 00 | 00 | – |
SARS-CoV-2 infections were detected among asymptomatic subjects. The fact that majority of COVID-19 cases are subclinical or asymptomatic has been observed too from other regions of the world. One early study from China documented that, COVID-19 has resulted either in very mild or asymptomatic infection in four out of five cases in China [10]. This has been observed in studies from India too. [11]. As per the second nation-wide SARS-CoV-2 sero survey study in India conducted during August–Sept 2020 by ICMR with a sample size of 29,082, a prevalence of 7.1% was reported reflecting an overall 74.3 million infections across India. This represents a 26 to 32 infections for every COVID-19 detected case in India by August 2020 in India [11]. Thus, the bulk of the SARS-CoV-2 infections is estimated to be subclinical or asymptomatic not needing medical care. The role of past infection with CCCs or other pathogens sharing common epitopes in providing partial cell mediated immunity or less severe infection is still unknown. This is an area which requires further investigations, if past infection with CCCs or other pathogens sharing common epitopes provides a degree of protective immunity to COVID-19 disease. A hypothesis can be drawn if detecting antibodies against CCCs (which may act as a biomarker) can predict if a person is susceptible to severe COVID-19 or mild/subclinical COVID-19 disease. There is a gap in understanding on the role of pre-existing immunity which may either be beneficial or may be deleterious resulting in increased severity of disease by mechanism such as ‘original antigenic sin’ which can cause an inferior immune response or an antibody-mediated disease enhancement [4], that can mount a cytokine storm. Further, the reason for the high prevalence of asymptomatic or subclinical cases of COVID-19 should be investigated and role of any protective immunity or tolerance if any conferred by past infections/vaccinations with Haemophilus spp. or any other pathogens including the common cold viruses should be investigated. Due to the low sample size, a caution is required while interpreting the results of this study implicating the role of Haemophilus influenzae cross-reacting with primers and probes designed for SARS-CoV-2, especially the E-gene target. However,
it will be prudent to keep in mind while designing new PCR based assays for SARS-CoV-2, the issue of cross-reactivity and non-specific amplifications due to *Haemophilus spp.* or other pathogens though at a lower efficiency, which might show up as late $C_T$ in real time PCR assays.

**Source of support**

Intramural fund of ICMR-RMRC, Dibrugarh and Regional VRDL, Dibrugarh, Department of Health Research, New-Delhi.

**Ethical approval**

The institutional ethics committee of Indian Council of Medical Research, Regional Medical Research Centre for NE region, Dibrugarh Assam, India approved (Meeting held on May 31, 2018) the Department of Health Research, Government of India funded scheme “Establishment of a network of Laboratories for managing Epidemics and Natural Calamities (VRDL)”. The scheme allows the microbiological diagnosis free of cost for early diagnosis & better management of cases and investigations of outbreaks & epidemics. The study also allows the use of left-over samples for anonymous testing for pathogens. For this preliminary study, separate ethical clearance was not sought as institutional ethics clearance for use of left-over samples (anonymous) has been obtained for samples collected during outbreak/epidemics or pandemics.

**Declaration of competing interest**

None. On behalf of both the authors, the corresponding & the lead author declare that there is no conflict of interest related to the submitted manuscript.

**Acknowledgement for financial support & specific scientific contribution only**

The authors acknowledge the funding from the Regional VRDL, Dibrugarh, Department of Health Research, Government of India.
References

[1] Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. J Med Virol 2020;92(4):418–423.

[2] Murphy TF. Haemophilus species including H. influenzae and H. ducreyi (chancroid). In: Bennett JE, Dolin R, Blaser MJ, editors. Mandell, Douglas, and Bennett’s Principles and Practice of Infectious Diseases. ninth ed. Philadelphia, PA: Elsevier; 2020. chap 225.

[3] Iskander M, Hayden K, Van Domselaar G, Tsang RSW. First complete genome sequence of Haemophilus influenzae serotype a. Genome Announc. 2017;5. e01506-e01516.

[4] Sette A, Crotty S. Pre-existing immunity to SARS-CoV-2: the knowns and unknowns. Nat. Rev. Immunol. 2020;20:457–8. https://doi.org/10.1038/s41577-020-0389-z.

[5] Lansbury L, Lim B, Baskaran V, Lim WS. Co-infections in people with COVID-19: a systematic review and meta-analysis. J. Infect. 2020;81(2):266–75.

[6] Yan J, Hu I, Zou S, Liu G, Cao B. Haemophilus influenzae coinfection is common in COVID-19 patients. (Pre-print) Available at: https://doi.org/10.21203/rs.3.rs-41416/v1.

[7] Raoult D, Zumla A, Locatelli F, Ippolito G, Kroemer G. Coronavirus infections: epidemiological, clinical and immunological features and hypotheses. Cell Stress 2020;4(4):66–74. https://doi:10.15698/cs2020.04.216.

[8] Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Modertbacher CR, et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. Cell 2020;181:1489–501.

[9] Gorse GJ, Patel GB, Vitale JN, O’Connor T. Prevalence of antibodies to four human coronaviruses is lower in nasal secretions than in serum. Clin. Vaccine Immunol. 2010;17(12):1875–80.

[10] Day M. Covid-19: four fifths of cases are asymptomatic, China figures indicate. BMJ 2020;369:m1375. https://doi.org/10.1136/bmj.m1375.

[11] Murhekar M, Bhatnagar T, Selvaraju S, Kumar VS, Thangaraj JSW, Shah N, et al. SARS-CoV-2 antibody prevalence in India: findings from the second nationwide household serosurvey, August - September 2020. Available at: SSRN: https://ssrn.com/abstract=3715460 or https://doi.org/10.2119/sars.3715460.