First detection of SARS-CoV-2 proteins in wastewater samples by mass spectrometry

Linda Lara-Jacobo

Ontario Tech University

Golam Islam

Ontario Tech University

Jean-Paul Desaulniers

Ontario Tech University

Andrea Kirkwood

Ontario Tech University

Denina Simmons (✉ denina.simmons@ontariotechu.ca)

Ontario Tech University

Letter

Keywords: SARS-CoV-2, COVID-19, liquid-chromatography mass spectrometry, wastewater, proteomics

Posted Date: June 4th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-237512/v4>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Environmental Science & Technology on March 29th, 2022. See the published version at <https://doi.org/10.1021/acs.est.1c04705>.

FIRST DETECTION OF SARS-COV-2 PROTEINS IN WASTEWATER SAMPLES BY MASS SPECTROMETRY

Lara-Jacobo, Linda R.; Islam, Golam; Desaulniers, Jean-Paul; Kirkwood, Andrea E; and Simmons, Denina B.D.*

*Corresponding author

KEYWORDS: SARS-CoV-2, COVID-19, liquid-chromatography mass spectrometry, wastewater, proteomics

INTRODUCTION

On March 12, 2020, the World Health Organization (WHO) declared COVID-19 as a global pandemic. COVID-19 is produced by a novel β -coronavirus known as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) [1]. Several studies have detected SARS-CoV-2 RNA in urine, feces, and other biofluids from both symptomatic and asymptomatic people with COVID-19 [2], suggesting that SARS-CoV-2 RNA could be detected in human wastewater [3]. Thus, wastewater-based epidemiology (WBE) is now used as an approach to monitor COVID-19 prevalence in many different places around the world [4-10]. Reverse transcription quantitative polymerase chain reaction (RT-qPCR) is the most common SARS-CoV-2 detection method in WBE, but there are other methods for viral biomolecule detection that could work as well. The aim of this study was to evaluate the presence of SARS-CoV-2 proteins in untreated wastewater (WW) influents collected from six wastewater treatment plants (WWTPs), from Durham Region, Ontario, Canada, using a LC-MS/MS-based proteomics approach. We identified many SARS-CoV-2 proteins in these wastewater samples, with peptides from pp1ab being the most consistently detected and with consistent abundance.

MAIN TEXT

SARS-CoV-2 contains a single RNA genomic strand (+) that encodes at least 29 proteins including four structural proteins: spike glycoprotein (S), membrane (M), the envelope (E) and nucleocapsid (N) [11]. The remaining proteins are known as nonstructural proteins (nsp) [12] which are produced by the host cell after viral infection [13]. The S protein, with a size of 180-200 kDa, is a transmembrane domain protein anchored in the viral membrane, whose main function is to allow the virus to fuse with the host cell. To avoid detection by the host immune system, these spike proteins are coated with polysaccharides (hence the term glycoprotein) that allow the virus to camouflage [14, 15]. The M protein is the most abundant structural protein of this virus, and while it is thought to also be a glycoprotein, its function is not fully understood [16].

There are three additional transmembrane proteins that are similarly not as well studied as the S protein, and are named after their open reading frame location: ORF3a, ORF7a, and ORF8. The ORF3a proteins regulates the subcellular environment in the host and plays an important role in the defense against infection by inducing apoptosis [17]. Both ORF7a and ORF8 are accessory proteins whose functions are not determined [18, 19].

Finally, the E protein forms a protein-lipid ion-transporting pore, and is involved in virus morphogenesis, assembly, and the induction of apoptosis [20, 21]. Replicase polyprotein 1ab (ppb1a) and replicase polyprotein 1a (ppb1a) are non-structural proteins (NSPs) involved in the transcription and replication of the virus and are produced after host cell infection by the virus. These proteins inhibit host translation and maintain optimal cellular conditions for the replication of SARS-CoV-2 [22].

In wastewater-based studies of COVID-19, the main detection technique of SARS-CoV-2 has been RT-qPCR, which has been used to measure SARS-CoV-2 in untreated wastewaters in Australia [23], Canada [5, 6], Germany [8], Netherlands [10], Spain [9] and USA [7]. In these studies, scientists detected the concentration of the virus (number of copies/L) using an average of the genes N1 and N2 (encoding the Nucleocapsid protein(s)) [4-6, 8-10]. Although this technique has produced satisfactory results for monitoring population-level infection, measuring viral proteins using liquid chromatography tandem mass spectrometry (LC-MS/MS) offers great advantages, including the considerably shorter run time [24], potentially improved sensitivity [25], the ability to multiplex and perform simultaneous non-targeted measurements, and a significantly reduced cost per-sample compared to RT-qPCR [24]. In the present study, our goal was to determine if we could detect the presence of SARS-CoV-2 proteins in untreated wastewater using a top-down proteomics approach with LC-MS/MS, and then to evaluate the potential use of this methodology as a routine method for the detection of SARS-CoV-2 in wastewaters.

We analyzed 115 influent samples from 2 WWTPs (see Methods section for full details). In total, we identified 245 unique peptides (962 total peptides) from eight SARS-CoV-2 proteins. 1. Structural proteins: M (6 peptides), S (15 peptides), and N (20 peptides), 2. Transmembrane proteins: ORF3a (8 peptides), ORF7a (16 peptides) and ORF8 (10 peptides), and 3. the NSPs, post-infection protein: pp1ab (170 peptides) (See Table S1 for complete protein identification details). The peptides were identified firstly with our untargeted method, and then we incorporated the m/z and RT into a target list (see Table S3 for target list). Figure 1 displays the number of peptides for each protein in influent

samples over the course of the sampling period as a heatmap for each WTP. The sequence coverage for each protein was: 13.33% (M), 38.02% (S), 13.33% (N), 30.26% (ORF3a), 40.50% (ORF7a), 57.02% (ORF8) and the non-structural protein 45.67% (pp1ab) (Figure 2). It is immediately striking that we frequently detected the most peptides for pp1ab compared to the other 6 SARS-CoV-2 proteins.

Thus, we determined that the pp1ab protein was the most consistently present and abundant protein in the Durham Region WWs. Considering that the structural proteins will degrade as the SARS-CoV-2 virus travels through the human gut and the wastewater system, it makes sense that we did not detect structural proteins with as much consistency or intensity as the pp1ab protein. In contrast, after LC-MS/MS analyses of nasal swabs and gargle samples collected from positive COVID-19 patients, the most abundant peptides were from structurally abundant proteins (N most often, then the E, M, and S proteins) [25-27]. Those structural proteins are all part of the viral envelope, which makes them the easiest to detect in a freshly collected clinical sample, but they are also the proteins that are most-exposed to environmental factors, and will likely degrade first in wastewaters known to contain many chemicals with the potential to degrade SARS-CoV-2 [28].

What is more interesting to consider is the fate of the pp1ab protein. As we currently understand SARS-CoV-2 pathology, the pp1ab protein would not be found inside the viral envelope because it is made by human cellular machinery *after* SARS-CoV-2 infection, and so we wonder, how it is possible that this protein is the most abundant in wastewaters? First, we considered the possibility that we were obtaining false positives; however, our most abundant and consistently detected peptide sequence

(KAIKCVUPQADVEWKFY) had a local FDR < 0.1%, and only matched to SARS-CoV-2 viruses after a BLAST search against the entire Uniprot database (1000 Hits, default settings). Additionally, our results demonstrated > 45% sequence coverage of the pp1ab protein with other high quality peptide matches to pp1ab in our combined dataset from 23 weeks of wastewater samples across 2 different wastewater catchments. Therefore, we are confident that we are not observing a false positive result.

We posit that there are two main factors which could affect the persistence of the pp1ab protein in the environment: (1) the abundance of the protein in the feces and urine in SARS-CoV-2-infected people within the community, and (2) the size and structure of the pp1ab protein. After extensive literature search, we cannot find any reports about the abundance of the pp1ab protein in urine or feces, with the majority of papers that include pp1ab in the text describing only its general function in SARS-CoV-2 pathogenesis [29-31]. Thus, it is difficult to discuss if this viral protein is shed in high levels through urine and feces, because we did not measure this directly. However, based on our results, we think that it is highly probable that this protein is abundant in human excrement. This same protein has been found in high abundance in nasopharyngeal and oropharyngeal samples from a large clinical sample cohort using LC-MS/MS [32], and as more evidence emerges, we may learn more about this protein in other body fluids.

Additionally, the pp1ab protein is very large (7096 amino acid residues in length and 794,058 Da in mass) [33] which could be a second factor contributing to the abundance of pp1ab in wastewater. In infected cells, pp1ab is cleaved into multiple functional proteins by proteolytic auto-processing. If the protein is released from cells before complete auto-processing, its sheer size could make the pp1ab protein more resistant to complete

degradation while travelling through the human gut and municipal wastewater system. Additionally, the peptide that we detected with the greatest frequency falls into the nsp14 region (amino acid positions 6261-6276), which is cleaved into the proofreading exoribonuclease (ExoN). We could not find any available information on the structure of ExoN, and so we cannot determine if our most abundant peptide falls into a protected pocket or region of ExoN, but we did attempt to model the peptide's likely structure using PEP-FOLD3, an online, freely available modelling tool (RPBS, Parisian Resource in Structural Bioinformatics). The model suggests that our peptide would form a hairpin-like structure which could be stabilized by hydrogen bonding (Figure S1). It is difficult to make any assumptions about the stability of the ExoN protein based upon this information, but we are certain that what we have observed in our study warrants further investigation into the 3D structure of this particular SARS-CoV-2 NSP. We also observed much of the pp1ab sequence coverage in the nsp 2 and nsp 3 (3C-like proteinase) regions, which may indicate that these regions (or proteins) are more resistant to degradation than other proteins in the pp1ab polyprotein sequence.

We employed a cross correlation analysis comparing the abundance of pp1ab in the wastewater from each treatment plant to the number of new reported (by onset date) cases of COVID-19 by onset date in the corresponding municipalities (Figures 3 and 4). The normalized (using daily flow) pp1ab signal from WWTP1 had a lag value of 6 days and a low positive, statistically significant correlation with the cumulative active cases ($r = 0.4527$, $p = <0.0001$ in 155 pairs of data); and normalized pp1ab signal from WWTP2 had a lag of 5 days, with a low positive, statistically significant correlation ($r = 0.4945$, $p = <0.0001$ in 156 pairs of data). Even though the pearson correlation coefficient is low, the

correlation remains very significant. As well, the lag time suggests that we can detect the presence of COVID-19 proteins from infected patients up to one week in advance of clinical case results. There are many reasons why the correlations would be low. The presence of pp1ab in the feces of asymptomatic patients/people who have not yet presented with symptoms, and variations in wastewater composition due to industrial influents, storm events, and degradation of sample material during residence time would all contribute to a high level of variability. As well, we are correlating two polynomial curves, would not produce a linear correlation, and thus a Pearson r coefficient that is closer to 0.0 could still be generated from a strong relationship. The differences in lag between each WWTP could be due to the population level using the wastewater system, and the residence time of the proteins within the wastewater catchment (from toilet to WWTP).

To compare, similar significant but weak correlations were obtained when we performed the same cross correlation analysis between the normalized N1 RNA by qPCR and the number of new reported cases of COVID-19 by onset date (Figures 3 and 4). There was a moderate correlation ($r = 0.7556$) with statistical significance ($p < 0.0001$) in 157 pairs of data from WWTP1; and there was a weak correlation ($r = 0.4500$) with statistical significance with ($p < 0.0001$) in 151 pairs of data from WWTP2, which also likely reflects many unknowns that contribute to high variability in the wastewater data (i.e.: storm events, population usage, degradation). Of greater interest was the fact that the proteins were more predictive of the cases by onset date in terms of the lag time, where the protein signals were 6 and 5 days ahead, compared to 4 and 10 days behind the cases by onset date for the RNA measured by qPCR. This could mean that when human cells are

infected, pp1ab protein production increases immediately, which is then followed by a more intense viral RNA, observed a few days later - which makes sense as the pp1ab proteins are the means by which the SARS-CoV-2 virus replicates, as part of the viral life cycle [34].

To our knowledge, we are the first to use LC-MS/MS to measure SARS-CoV-2 proteins in wastewater. Neault et al (2020) detected SARS-CoV-2 structural proteins with immunoblotting and then measured their abundance with an immune-linked PCR method. They found that proteins were detected more frequently than RNA and concluded that proteins were present in higher abundance. As well, they found that protein and RNA were visually correlated, but no statistical correlation analyses were performed. We attempted to correlated our RNA and protein data, but there was no significant correlation to be found, even with cross correlation (data not shown). Thus, we think that a protein-based LC-MS/MS approach may have more utility than immunoblot and RNA methods as an early warning signal because it correlates earlier with case data, and because it is possible to measure many other human-protein biomarkers of SARS-CoV-2 infection and susceptibility simultaneously in the same sample using our shotgun approach, which is a topic we are exploring for future work.

Here, we have presented a MS-based method for WWs samples that specifically detects SARS-CoV-2 proteins, 6 days ahead of the case data. We were able to identify unique peptides of at least eight proteins related to the SARS-CoV-2 virus and COVID-19 infection. We noticed a consistent presence in all the samples of the NSP protein pp1ab, which is only transcribed after host cell infection. Given the results from the present study, we suspect that the pp1ab protein is present in high abundance in the urine and (more

likely) feces of active COVID-19 cases, and that this protein might make an excellent alternative biomarker for testing people for infection in a less invasive manner compared to the common nasal swab qPCR test. Our next steps are (1) to quantify our strongest signal peptide of pp1ab using a stable isotope labelled peptide standard, and (2) to conduct a multivariate analysis of all human proteins detected within the samples to identify proteins we can use for normalization, as well as other protein biomarkers of infection and susceptibility for COVID-19. Our ultimate goal is to use this methodology for sensitive, accurate, and routine detection of SARS-CoV-2 population-level infection in WW samples, thereby providing robust monitoring data and proof of principle that establishes the value of wastewater-based epidemiology for future public health studies.

METHODS

Protein Precipitation and Digestion

Twenty-four, 24-hour composite influent samples from each of the three wastewater treatment plants (WWTP) from Durham Region (Ontario, Canada) were obtained over the course of 15 weeks – for a total of 171 samples across all sampling times and locations. Wastewaters from 4 municipalities are treated by these three WWTPs, although the boundaries of the wastewater catchments do not necessarily correspond to a single municipality. The WWTP and municipalities were related as follows: WWTP1 – Whitby (population = 135,556); and WWTP2 – Oshawa (population = 170,071. In the present study, all the samples were collected as “fresh” (unpasteurized) 24-hr composite samples. We added 50 mL of acetone (4 °C) to 50 mL of the wastewater and stored overnight at -20 °C to precipitate proteins from the liquid phase. The samples were then centrifuged

for 15 min at 4000 rpm. After centrifugation, we carefully removed and discarded the supernatant, and the pellet was transferred to a low-retention 1.5 mL micro-centrifuge tube, and then left in the fume hood for 60 minutes to allow any remaining acetone to evaporate. We then re-suspended the pellet in 50 μ L of a 100mM Triethylammonium bicarbonate buffer (TEAB) (CAS number 15715-58-9 1.0M solution, Sigma-Aldrich) and continued to reduce, alkylate, and digest the proteins in the sample into peptides using a previously described method [35].

LC-MS/MS Peptide Analysis

Peptides were identified with an Agilent 1260 Infinity Binary Pump HPLC coupled to a quadrupole Time-of-Flight mass spectrometer (Agilent 6545 LC-QTOF). The samples were loaded onto the separation column (Agilent ZORBAX 300SB-C18, Micro Bore Rapid Resolution 1.0 x 50 mm 3.5-Micron) and eluted with a 50-minute gradient (Solvent A = 0.1% formic acid in 95:5 water: acetonitrile; Solvent B = 0.1% formic acid in 95:5 acetonitrile: water; 0-2 mins 2% Solvent B; 2-22 mins 40% Solvent B; 22-27 mins 60% Solvent B; 27-32 mins 85% Solvent B; 32-50 mins 2% Solvent B.). The column was kept at 40 °C, with a flow rate set at 0.100 μ L/min, and the autosampler compartment was kept at 6 °C. Auto MSMS analysis was performed and after peptides were identified by database search, the m/z and retention time of each was added to an inclusion list for targeted MS/MS analysis. Specific instrumental settings and acquisition method details are provided in the supplemental materials.

SARS CoV-2 detection by qPCR

Total RNA was extracted from raw influent samples using a commercially available QIAGEN Powermicrobiome kit (QIAGEN, Canada). One-step RT-qPCR was performed

on the RNA samples using the Reliance One-Step Multiplex RT-PCR supermix 4x (BioRad Laboratories Inc, Canada). The 20µL reaction was composed of 5µL of RNA template input of each sample, 500nM Forward (GAC CCC AAA ATC AGC GAA AT) and Reverse (TCT GGT TAC TGC CAG TTG AAT CTG) primers, 125nM of custom TaqMan probe (FAM-ACC CCG CAT TAC GTT TGG TGG ACC-MGBNFQ) (Thermo-Fisher Scientific, Massachusetts, USA) targeting the N1 gene, 8µL of pcr grade H2O and 5µL of mastermix. The qPCR reactions were all ran in triplicates and according to MIQE guidelines using a CFX Connect qPCR thermocycler (Bio-Rad, Hercules, CA), RT was performed at 50°C, 10 minutes, followed by polymerase activation at 95°C for 10 minutes, and 45 cycles of denaturation, annealing/extension at 95°C/10 s, then 60°C/30 s, respectively. Serial dilutions of the viral RNA standard were run on every 96-well PCR plate to produce standard curves used to quantify the copies of SARS-CoV-2 genes. Additionally, RT-qPCR runs were validated with the use of non-template-controls (NTCs), positive controls, negative controls.

Normalization

The normalization of both data (Protein pp1ab and qPCR) was determined by the daily flow and the population served by WWTP (see the equations below).

$$N_1 = \frac{Vs_1 F}{P}$$

With N_1 normalization for Protein pp1ab (Total intensity counts (TIC) of protein per inhabitant per day), Vs_1 viral signal (TIC/m³), F flow (m³/day) and P population served (inhabitant).

$$N_2 = \frac{Vs_2 F}{P}$$

With N_2 normalization for qPCR (Num. of N1 copies per inhabitant per day), V_{s2} viral signal corrected with the estimation of recovery (~40%) (Num. of copies/m³), F flow (m³/day) and P population served (inhabitant).

Database search and analysis of identified proteins

For the viral protein identification, peptide spectral data was extracted, filtered, and searched against UniProtKB/Swiss-Prot *SARS-CoV-2* database (downloaded September 2020). The database contained fourteen SARS-CoV-2 proteins. The search was set in Spectrum Mill MS Proteomics Workbench (Agilent G2721AA/G2733AA) using an automated workflow to search for peptide sequence matches. Protein hits were initially filtered using peptide score > 5, and %SPI > 60%. The complete data base (Table S1) was initially validated using a less stringent filter, but we validated those results with a stronger one with a FDR <1% (Table S2).

Statistical analyses

For the statistical analyses, a cross correlation was performed first with all the 160 pairs of data (x: Protein pp1ab 7-day midpoint mean viral signal, y: 7-day midpoint mean in reported new cases of COVID-19 by onset date) and with lags between -15 to 15 as the maximum (cross correlation in time series). The lag with the highest value of r was selected and was used to perform a Pearson correlation analysis (one-tailed with a confident interval of 95%) using Prism software (v.9. GraphPad). In the first group of cross-correlation analysis, abundance of normalized pp1ab in the wastewater was the X variable and the number of reported new cases of COVID-19 by onset date was the Y variable. In the second group of cross-correlation analysis, the X variable was 7-day normalized midpoint mean viral signal (N1 RNA normalized by qPCR) and the Y variable

275 was 7-day midpoint mean reported new cases of COVID-19 by onset date. The number
276 of reported new cases of COVID-19 by onset date in each municipality was obtained
277 directly from the web site: Durham Region COVID-19 Data Tracker [36], where only the
278 new reported cases were selected daily.

279

280 **ACKNOWLEDGMENTS**

281 Durham Region, Ontario Tech University, MITACS, Ontario Clean Water Agency.

282

283

284 REFERENCES

- 285 1. Li, Q., et al., *Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected*
286 *Pneumonia*. N Engl J Med, 2020. **382**(13): p. 1199-1207.
- 287 2. Wang, W., et al., *Detection of SARS-CoV-2 in Different Types of Clinical Specimens*. JAMA, 2020.
288 **323**(18): p. 1843-1844.
- 289 3. Kitajima, M., et al., *SARS-CoV-2 in wastewater: State of the knowledge and research needs*. Sci
290 Total Environ, 2020. **739**: p. 139076.
- 291 4. Ahmed, W., et al., *First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia:*
292 *A proof of concept for the wastewater surveillance of COVID-19 in the community*. Sci Total
293 Environ, 2020. **728**: p. 138764.
- 294 5. D'Aoust, P.M., et al., *Catching a resurgence: Increase in SARS-CoV-2 viral RNA identified in*
295 *wastewater 48 hours before COVID-19 clinical tests and 96 hours before hospitalizations*. 2020.
- 296 6. D'Aoust, P.M., et al., *Quantitative analysis of SARS-CoV-2 RNA from wastewater solids in*
297 *communities with low COVID-19 incidence and prevalence*. 2020.
- 298 7. Sherchan, S.P., et al., *First detection of SARS-CoV-2 RNA in wastewater in North America: A study*
299 *in Louisiana, USA*. Sci Total Environ, 2020. **743**: p. 140621.
- 300 8. Westhaus, S., et al., *Detection of SARS-CoV-2 in raw and treated wastewater in Germany -*
301 *Suitability for COVID-19 surveillance and potential transmission risks*. Sci Total Environ, 2021.
302 **751**: p. 141750.
- 303 9. Randazzo, W., et al., *SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low*
304 *prevalence area*. Water Res, 2020. **181**: p. 115942.
- 305 10. Lodder, W. and A.M. de Roda Husman, *SARS-CoV-2 in wastewater: potential health risk, but also*
306 *data source*. The Lancet Gastroenterology & Hepatology, 2020. **5**(6): p. 533-534.
- 307 11. Siu, Y.L., et al., *The M, E, and N structural proteins of the severe acute respiratory syndrome*
308 *coronavirus are required for efficient assembly, trafficking, and release of virus-like particles*. J
309 Virol, 2008. **82**(22): p. 11318-30.
- 310 12. Kim, D., et al., *The Architecture of SARS-CoV-2 Transcriptome*. Cell, 2020. **181**(4): p. 914-921 e10.
- 311 13. *Post-translational modifications of coronavirus proteins: roles and function*.
- 312 14. Huang, Y., et al., *Structural and functional properties of SARS-CoV-2 spike protein: potential*
313 *antivirus drug development for COVID-19*. Acta Pharmacol Sin, 2020. **41**(9): p. 1141-1149.
- 314 15. *Site-specific glycan analysis of the SARS-CoV-2 spike*.
- 315 16. Thomas, S., *The Structure of the Membrane Protein of SARS-CoV-2 Resembles the Sugar*
316 *Transporter SemiSWEET*. Pathog Immun, 2020. **5**(1): p. 342-363.
- 317 17. Ren, Y., et al., *The ORF3a protein of SARS-CoV-2 induces apoptosis in cells*. Cell Mol Immunol,
318 2020. **17**(8): p. 881-883.
- 319 18. Nelson, C.A., et al., *Structure and intracellular targeting of the SARS-coronavirus Orf7a accessory*
320 *protein*. Structure, 2005. **13**(1): p. 75-85.
- 321 19. Pereira, F., *Evolutionary dynamics of the SARS-CoV-2 ORF8 accessory gene*. Infect Genet Evol,
322 2020. **85**: p. 104525.
- 323 20. Nieto-Torres, J.L., et al., *Severe acute respiratory syndrome coronavirus envelope protein ion*
324 *channel activity promotes virus fitness and pathogenesis*. PLoS Pathog, 2014. **10**(5): p. e1004077.
- 325 21. Nieto-Torres, J.L., et al., *Severe acute respiratory syndrome coronavirus E protein transports*
326 *calcium ions and activates the NLRP3 inflammasome*. Virology, 2015. **485**: p. 330-9.
- 327 22. Prentice, E., et al., *Identification and characterization of severe acute respiratory syndrome*
328 *coronavirus replicase proteins*. J Virol, 2004. **78**(18): p. 9977-86.

23. Ahmed, W., et al., *Surveillance of SARS-CoV-2 RNA in wastewater: Methods optimisation and quality control are crucial for generating reliable public health information*. Curr Opin Environ Sci Health, 2020.
24. Singh, P., et al., *A rapid and sensitive method to detect SARS-CoV-2 virus using targeted-mass spectrometry*. J Proteins Proteom, 2020: p. 1-7.
25. *Fast and low-cost detection of SARS-CoV-2 peptides by tandem mass spectrometry in clinical samples*.
26. Nikolaev, E.N., et al., *Mass-Spectrometric Detection of SARS-CoV-2 Virus in Scrapings of the Epithelium of the Nasopharynx of Infected Patients via Nucleocapsid N Protein*. J Proteome Res, 2020. **19**(11): p. 4393-4397.
27. Ihling, C., et al., *Mass Spectrometric Identification of SARS-CoV-2 Proteins from Gargle Solution Samples of COVID-19 Patients*. J Proteome Res, 2020. **19**(11): p. 4389-4392.
28. Whitney, O.N., et al., *Sewage, Salt, Silica and SARS-CoV-2 (4S): An economical kit-free method for direct capture of SARS-CoV-2 RNA from wastewater*. medRxiv, 2020.
29. Yan, S. and G. Wu, *Spatial and temporal roles of SARS-CoV PL(pro) -A snapshot*. FASEB J, 2021. **35**(1): p. e21197.
30. Naqvi, A.A.T., et al., *Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach*. Biochim Biophys Acta Mol Basis Dis, 2020. **1866**(10): p. 165878.
31. V'Kovski, P., et al., *Coronavirus biology and replication: implications for SARS-CoV-2*. Nat Rev Microbiol, 2020.
32. Cardozo, K.H.M., et al., *Establishing a mass spectrometry-based system for rapid detection of SARS-CoV-2 in large clinical sample cohorts*. Nat Commun, 2020. **11**(1): p. 6201.
33. UniProt, *UniProt*. 2019.
34. Chen, Y., Q. Liu, and D. Guo, *Emerging coronaviruses: Genome structure, replication, and pathogenesis*. J Med Virol, 2020. **92**(4): p. 418-423.
35. Simmons, D.B., et al., *Proteomic profiles of white sucker (Catostomus commersonii) sampled from within the Thunder Bay Area of Concern reveal up-regulation of proteins associated with tumor formation and exposure to environmental estrogens*. Environ Sci Technol, 2012. **46**(3): p. 1886-94.
36. Region, D., *Status of COVID-19 Cases in Durham Region*. 2020.

Figure 1. The number of peptides found for each protein detected in wastewaters as a heatmap for each WWTP.

Figure 2. Sequence coverage of SARS-CoV-2 structural and non-structural proteins found in wastewater influents.

Figure 3. Cross-correlation function (CCF) between smoothed viral signal protein (7-day midpoint mean protein: pp1ab) in wastewater and smoothed new cases of COVID-19 by Onset Date (7-day midpoint mean in reported new cases (by reporting date) with different time lags from -15 to 15 days. (A) WWTP1, (B) WWTP2. CCF between smoothed N1 gene copies in wastewater and smoothed new cases of COVID-19 by Onset Date (C) WWTP1 and (D) WWTP2. The time at which there is the highest CCF was used as the Lag for the cross correlations (shown in Figure 4)

Figure 4. Cross correlation analyses of the total intensity of pp1ab to the total cumulative active cases of COVID-19 by onset date. The lag number represents the time delay (in days) between the best correlation between when the protein levels increase in the water and when the reported case numbers increased. The Pearson correlation coefficient is represented by “r”. (A) WWTP1, (B) WWTP2 for protein. And for RNA measurement by qPCR (C) WWTP1 D) WWTP2

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [MS1SI270521.pdf](#)