DEVELOPING A BETTER SARS-COV-2 DIAGNOSTIC TOOL USING RT-LAMP TECHNOLOGY

by

Yongjun Kwon

A Thesis
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Master of Science
Bioinformatics

Committee:

_____	Dr. Ancha Baranova, Thesis Chair
_____	Dr. Cohava Gelber, Committee Member
_____	Dr. Donald Seto, Committee Member
_____	Dr. Iosif Vaisman, Director, School of Systems Biology
_____	Dr. Donna Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
_____	Dr. Fernando Miralles-Wilhelm, Dean, College of Science
Date: _____	Fall Semester 2021 George Mason University Fairfax, VA

Developing a Better SARS-CoV-2 Diagnostic Tool Using RT-LAMP Technology

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

by

Yongjun Kwon
Bachelor of Arts and Science
The Ohio State University, 2011

Director: Cohava Gelber, CEO, Associate Professor
Caerus Discovery

Fall Semester 2021
George Mason University
Fairfax, VA

DEDICATION

This is dedicated to my loving family and my great friends at work.

ACKNOWLEDGEMENTS

There are many people I would like to thank for their both direct and indirect contributions. First of all, I would like to thank my PI, Dr. Cohava Gelber at Caerus Discovery for giving me an opportunity to pursue a higher degree of education while working. It is because of Dr. Gelber that I chose to enroll in George Mason University's Bioinformatics and Computational Biology (BCB) Master's program.

Second of all, I would like to thank my colleagues Dr. Mark Spear and Ms. Lauren Pincus. Dr. Spear's expertise in molecular biology and initiative made this project happen from its conception. Furthermore, I would like to thank Ms. Pincus for her great knowledge in chemical components and for other valuable contributions.

Last but not least, I would like to thank my graduate thesis advisory committee Dr. Baranova and Dr. Seto, a great resource in regards to bioinformatics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF CODES (PERL, R, ETC.)

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

BLAST ..... Basic Local Alignment Search Tool
cDNA ..... complementary DNA
CE ..... Conformitè Europëenne
DNA ..... Deoxyribonucleic acid
dsDNA ..... Double stranded DNA
EUA ..... Emergency Use Authorization
FDA ..... The United States Food and Drug Administration
LCV ..... Leucocrystal Violet
LoD ..... Limit of Detection
MCMs ..... Medical Countermeasures
mRNA ..... Messenger RNA
NTC ..... No Template Control
ORF ..... Open reading frame
PCR ..... Polymerase Chain Reaction
pH ..... a unit of measure used to indicate the concentration of hydrogen ions in a solution
PoC ..... Point-of-Care
PREP Act ..... Public Readiness and Emergency Preparedness Act
qPCR ..... quantitative PCR
RNA ..... Ribonucleic acid
RT-LAMP ..... Reverse transcriptase Loop-mediated isothermal amplification
Rxn ..... Reaction
SARS-CoV2 ..... Severe acute respiratory syndrome coronavirus 2
SNPs ..... Single nucleotide polymorphisms
VTM ..... Viral Transport Media

ABSTRACT

DEVELOPING A BETTER SARS-COV-2 DIAGNOSTIC TOOL USING RT-LAMP TECHNOLOGY

Yongjun Kwon, M.S.

George Mason University, 2021

Thesis Director: Dr. Cohava Gelber

SARS-CoV-2 belongs to the betacoronavirus genus, and is closely related to severe acute respiratory syndrome coronavirus (SARS-CoV). Since its emergence in Wuhan province, China in December of 2019, from a suspected bat or pangolin origin, the pathogen has spread rapidly, with millions of cases reported on every continent.

Currently, the World Health Organization reports 236 million confirmed cases globally, with an estimated case fatality rate of approximately 2.0%. Significantly higher mortality rates are observed in elderly patients, immunocompromised patients, and patients with other preexisting conditions, such as cardiovascular disease, cancer and diabetes. Given the exceptional transmissibility and relatively high mortality rate, the development of simple, robust yet accessible diagnostics are of utmost importance to public health.

As such, both myself and my colleagues have developed a SARS-CoV-2 rapid diagnostic kit suitable for use in point of care settings. This kit is for the *in vitro* qualitative

detection of the SARS-CoV-2 RNA in nasopharyngeal swabs, oropharyngeal swabs, and saliva collected from individuals with or without ongoing symptoms. Here, I will describe computational approaches used for developing the components included in this kit, the workflow, and will share the discoveries both I and my colleagues made to develop the novel SARS-CoV2 diagnostic kit.

CHAPTER ONE – JUSTIFICATION BEHIND THE DEVELOPMENT OF THE KIT

It has been repeatedly demonstrated that rapid diagnostic tests can have a significant impact on point-of-care (PoC) settings, particularly in resource constrained environments.[1] Here, we propose a novel, affordable and automated SARS-CoV-2 diagnostic kit that can reliably detect infected individuals utilizing RT-LAMP technology, without capital-intensive investments that are associated with other diagnostic technologies (e.g., qRT-PCR). [2,3]

Loop-mediated isothermal amplification (LAMP) is a DNA amplification technology developed in the early 2000s which allows DNA amplification with high specificity, efficiency and rapidity under isothermal conditions. Four to six specifically designed primers recognize six to eight regions of target DNA, allowing for amplification by a high processivity isothermal DNA polymerase. Two of the primers, which are designated "inner primers," contain sequences that are complimentary to the sense and antisense strands of the target DNA, and initiate the LAMP reaction.[4,5] For RT-LAMP, a reverse transcriptase is added to the master mix. Because LAMP recognizes the target by six distinct sequences initially and by four distinct sequences afterwards, it is expected to amplify the target sequence with high selectivity.[5] See **Figure 1** for the graphical illustration of the mechanism.

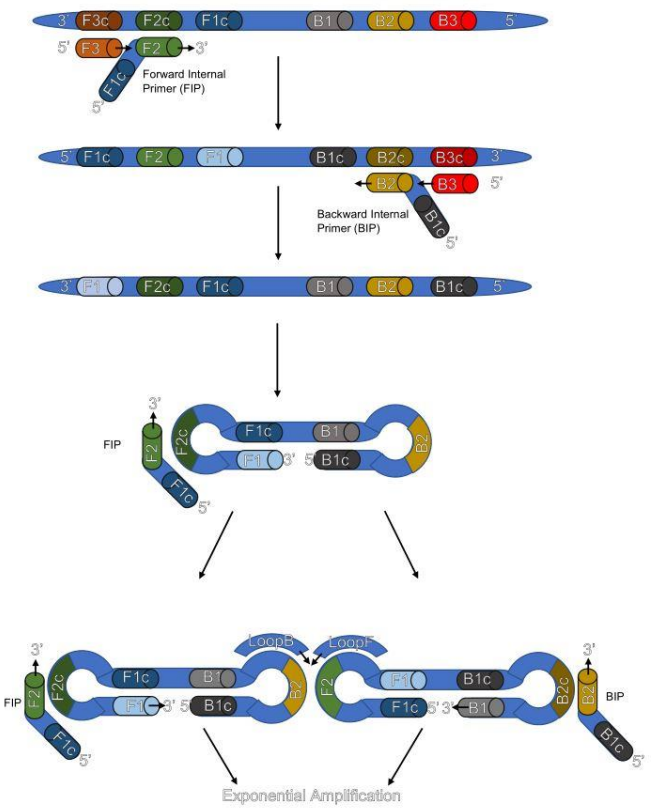

Figure 1. Graphical illustration of an RT-LAMP amplification

An additional advantage of RT-LAMP is that, while it is amenable to conventional fluorometric detection, non-fluorometric readouts are also viable due to prodigious product generation. This is made possible by the detection of reaction byproducts such as pyrophosphates and hydrogen ions, which are released during DNA polymerization.[6,7] For example, New England Biolabs' colorimetric LAMP master mix utilizes the pH change during the reaction to visualize their end result.[8–10] While this result can easily be interpreted with naked eye, the result can also be quantified by measuring the absorption ratio of 434nm (Yellow) and 558nm (Red). Alternatively, pyrophosphate salts accumulate to levels above their solubility limit, allowing for a turbidity-based readout of the reaction.

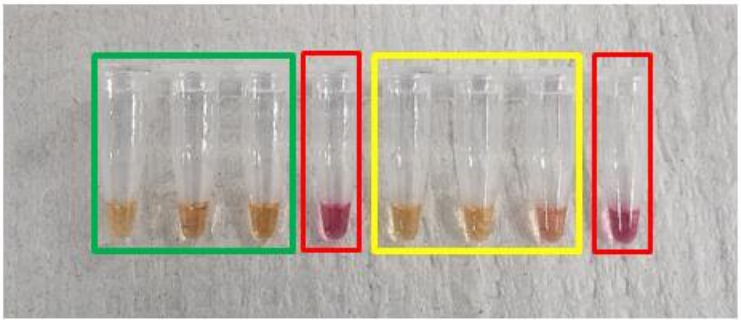

Figure 2. Example of the diagnostic end product using RT-LAMP technology. Green/Yellow: Positive Samples. Red: Negative Samples

As such, the development of a robust RT-LAMP primer set and protocol that enables highly sensitive colorimetric detection of the SARS-CoV-2 viral genome is of significant import.[11] Because of the strength of this color change, this technology allows point-of-care LAMP-based diagnostic that is accessible to the general public. Moreover, this product was envisioned for use in both PoC and non-PoC (e.g., airports, schools, etc.) settings given the relatively simplicity of the readout.

Current molecular diagnostics for SARS-CoV-2 rely on the amplification of target sequences in the viral genomic RNA or sub-genomic RNAs. Most of these assays use PCR to amplify the target sequences for probe-based detection and require a real-time thermocycler with fluorescent detection capabilities. As this equipment is expensive and dedicated for PCR-based assays, only few healthcare providers have direct testing capabilities. Additionally, inadequate healthcare infrastructure and limited financial resources exacerbate the SARS-CoV-2 pandemic in underdeveloped regions and countries,

leading to higher rates of morbidity and mortality. In part, this is attributable to limited or no access to facile diagnostic capabilities, which make communities and individuals incapable of correctly diagnose themselves from the disease and seek for right type of medical help. When a steady supply of electricity and clean water is not available or unreliable, maintaining access and distribution logistics for medical supplies, laboratory equipment, and trained medical staff may be impossible.

According to Miller et al., these kind of luxuries – vaccines, chemical prophylactics, well-educated medical consultants, and accurate diagnostic measures – are only available in centralized clinics.[1] Since the emergence of SARS-CoV-2 in 2019, these types of issues are more self-evident than ever, highlighting the need for cost-effective and rapid diagnostics that are viable in resource-constrained settings. This is why the low-cost, readily-available diagnostics at the PoC facilities are getting more important as the means of "closing the diagnostic gaps." This can be achieved by providing both patients and healthcare workers with rapid, comparable, and actionable diagnostic result which ultimately can be used to improve patient outcomes.

Herein, an alternative diagnostic technology is proposed for point-of-care testing of patient samples (i.e., saliva, nasal swabs, or feces). This technique provides numerous benefits in comparison to conventional quantitative PCR (qPCR) techniques as follows: (1) RT-LAMP is isothermal, meaning a single reaction temperature is needed, eliminating the need for a PCR thermocycler and allowing for detection at 55˚-65˚ in a simple water bath or heating block setup; (2) the assay described herein is colorimetric, with a visible readout of viral RNA amplification, eliminating the need for fluorescence detection

capabilities and allowing for simple user-based assessment of results without equipment; (3) the sensitivity of RT-LAMP is comparable to qPCR; and (4) with minor modification, the RT-LAMP assay can be measured by the naked eye, by spectrometer, by fluorometer, or can be multiplexed for detection of multiple viral RNAs from a single sample (with fluorescent detection). Additionally, while not intended as an at-home diagnostic, RT-LAMP is a sufficiently facile technique that at-home diagnostic kits are feasible, but would require a heating source (e.g., milk warmer, or sous vide machine).

Section One – Current Status of the Diagnostic Using Commercially Available RT-LAMP Reagent

All LAMP-based assays use a minimum of 4, and up to 6, primers for the generation of looped intermediates that undergo exponential amplification (**Figure 1**). The forward and backward inner primers (FIP and BIP) contain sequences complementary to internal viral RNA sequences, which results in looped intermediates that self-prime, ultimately forming long concatemers. The amplification is sufficiently prolific that the reaction pH drops due to the production of pyrophosphate and protons from DNA polymerization of dNTPs[6,7], allowing for colorimetric or fluorometric analysis with pH indicators. For the SARS-CoV-2 diagnostic, we designed 6 primers targeting the nucleocapsid, or N, RNA sequence (**Figure 3**, and **Table 14**: Sequence ID 001-006).

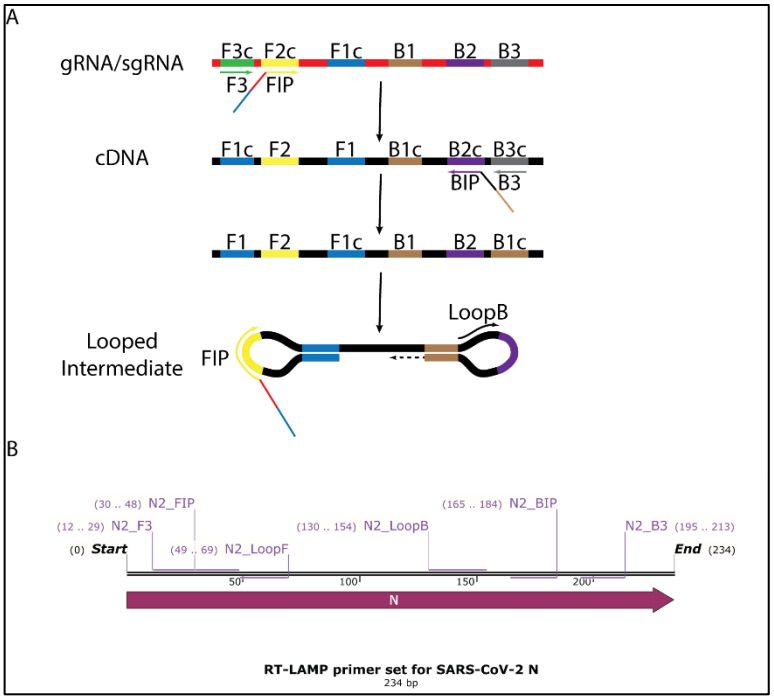

Figure 3. Schematic illustration of RT-LAMP Primers. (A) RT-LAMP utilizes 4-6 primers, which minimally include the B3 and F3 "bump" primers and the forward and backward inner primers (FIP and BIP). The FIP and BIP contain complementary sequences to internal residues, which result in looped intermediates (positive and negative-stranded) that undergoe exponential amplification. The optional LoopF and LoopB primers dramatically improve amplification of the looped intermediates. (B) A plasmid map showing the location of primers for the SARS-CoV-2 RT-LAMP diagnostic within the 5' end of the N sgRNA.

In proof-of-principle studies, the limits of detection were shown to be 10 copies per reaction or less (**Figure 4**), which is comparable to qPCR-based assays. Additionally, amplification of 50 copies of unextracted RNA was observed in one experiment (**Figure 4**), indicating that RNA extraction may not be necessary for some samples. While further testing in this regard is necessary, the potential of reliable, high-sensitivity detection of unprocessed patient samples markedly improves the ease of use for this diagnostic technology.

Currently, the SARS-CoV-2 colorimetric RT-LAMP uses a proprietary kit sold by New England Biolabs (catalog number: E1700S), which contains the Bst3.0 DNA

polymerase, reaction buffer, phenol red as a pH indicator, dNTPs, and a proprietary antibody inhibitor of the DNA polymerase that denatures at a temperature below the reaction temperature of 65˚C, preventing premature amplification when preparing the reaction at room temperature.

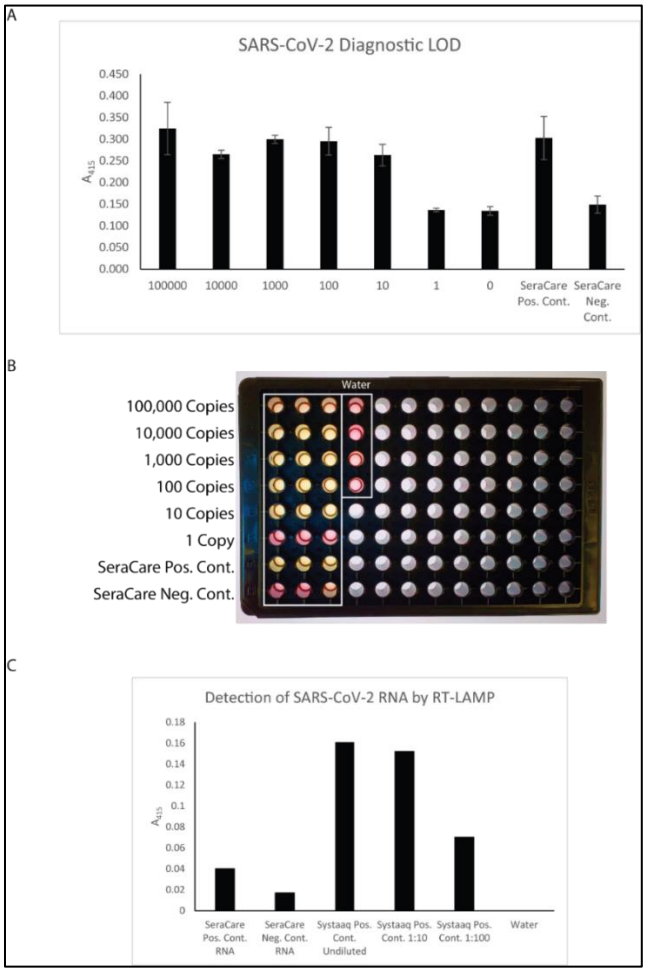

Figure 4. Proof-of-principle studies of the SARS-CoV-2 Diagnostic. (A) Extracted SARS-CoV-2 RNA from SeraCare's AccuPlex™ SARS-CoV-2 Reference Material Kit, or a positive control oligonucleotide from a commercial vendor, was admixed with primers and NEB's WarmStart® Colorimetric LAMP 2X Master Mix for 1 hour, after which the A_{415} was assessed. (B) A photo of the reactions, highlighting the pH-dependent change in phenol red color, with an increase in

A_{415} with decreasing pH. (C) Unextracted SeraCare controls were analyzed alongside dilutions of the commercial vendor's positive control. Background-subtraction was performed for (C), but not (A).

With this preliminary result established, further optimization and confirmation studies were conducted thereafter. In the following discussion, we confirmed that there are indeed many elements to be optimized regardless of the promising result in the proof-of-concept study: such as testing different set of primers, using different pH indicators, seeking "cleaner" samples by extracting the genetic materials from the virion and clinical matrix, and perfecting the amplification time to minimize the false positive/false negative rate while improving specificity and sensitivity of the diagnostic kit. The final LoD was determined at 1250~2500 viral copies range later in studies pursuant to an emergency use authorization (EUA) request to the FDA.

CHAPTER TWO – DEVELOPING A BETTER PRIMER SET FOR THE RT-LAMP SYSTEM

Given that this test is intended for diagnostic use, it is imperative to find the most optimized set of primers that will work in the RT-LAMP reaction. Different set of primers can perform differently, and cause suboptimal sensitivity, abnormal reaction dynamics due to primer dimers and self-annealing, and poor amplification kinetics, resulting in longer amplification times. During the optimization of the primers, we have observed different rates of false positives and false negatives depending on the primer sets. In the following section, I will talk about the endeavors involved in the optimization of the primer sets.

As a betacoronavirus, SARS-CoV-2 has a single stranded, plus-sense RNA genome of approximately 30kb in size, which encodes five open reading frames (ORFs). These ORFs are known to encode nonstructural proteins and structural proteins, which include the spike (S), envelope (E), membrane (M), and nucleocapsid (N). [12,13] Here, we decided to target the N gene because the N gene was deemed more likely to be highly conserved compared to the other structural protein candidates. Since the N protein is involved in forming the nucleocapsid core, requiring multiple protein-protein and protein-RNA interaction interfaces, few missense mutations are tolerated in the functional domains of the protein. All of the primers were generated from the reference SARS-CoV2 N gene using publicly available primer generator for the RT-LAMP reaction.[14]

Section One – Background Information and the Importance of the Detection Range

According to Sender et al.[15], using the **Equation 1**, they were able to conclude that during the peak infection, in the nasal mucosa and pharynx, the viral RNA copies can range from 10^6 to 10^8 (RNA copies per g). This provides a good standard where the kit can be evaluated. By using the **Equation 2**, I was able to confirm that our current LoD is well within the range.

Equation 1. Calculating Viral RNA Copies

$$\boldsymbol{C}^{\boldsymbol{lungs}}_{\boldsymbol{genome\ copies}} * \boldsymbol{M}_{\boldsymbol{lungs}} * \boldsymbol{F}_{\boldsymbol{virions\ to\ RNA\ copies}} = \boldsymbol{N}_{\boldsymbol{virions}}$$

Equation 2. Calculating the Baseline LoD Required

$$10^6\ Viral\ RNA\ copies/_g * \frac{150\ uL\ Clinical\ Sample\ Loading\ Volume}{80\ uL\ Kit\ Elution\ Volume} * \frac{10\ uL\ Transfer\ into\ the\ RTLAMP\ Rxn\ volume}{1000\ g/uL} = 18750\ Viral\ Copies$$

This means in both point-of-care and clinical settings, the diagnostic kit's LoD will most likely be within the range to detect SARS-CoV-2 RNA in infected individuals. According to the article by Li et al., where they compared the viral load of the Delta strain to that of original Wuhan 2019 strain, they observed up to 1000-fold increase in the viral titer in the Delta strain. In here, they observed viral peak at 6 days post-infection with 19A/19B strain and 4 days post-infection with Delta strain. They reported over 80% of the

Delta strain patients' samples contained >$6*10^5$ copies/mL which means that 4-6 days post infection, our diagnostic kit will effectively detect vast majority of the infected cases regardless of the variants. Though, to note, according to Wang et al., the viral load lowers to the undetectable levels 14 days post-infection even when using qPCR. So, it is fair to assess that our diagnostic kit can detect the majority of the patient samples between 4-14 days post-infection.

Section Two – *In silico* Studies Conducted

The FDA's EUA guidelines for developers of molecular diagnostics requires the following validation studies to be performed to support any EUA requests[16,17]: 1. Limit of Detection (LoD) Studies, 2. Inclusivity Studies, 3. Cross-reactivity Studies, 4. Microbial Interference Studies, 5. Endogenous/Exogenous Interference Substances Studies, and finally 6. Sample Stability Studies.

As of October 2021, all the required studies are done. Notably, the following items were accomplished by *in silico* studies in combination with wet lab studies to support the findings – inclusivity studies and cross-reactivity studies.

Inclusivity Study (Analytical Reactivity)

Background

According to FDA's guidelines, as mutations in the SARS-CoV-2 genome have been identified and updated, the genetic changes in the newer SARS-CoV-2 variants need to be compared with reference sequences such as Wuhan-Hu1 or USA-WA1/2020.

These new variants of SARS-CoV-2 can be identified by genomic sequences that contain one or more mutations in their genomic RNA. While the majority of these mutations are silent, they can result in amino acid insertions, deletions, and/or substitutions which can result in varying phenotypes (e.g., change in their antigenicity, virulence, transmissibility or pathogenicity). Additionally, such mutations in different variants can affect antibody epitopes relative to the Wuhan isolate, which potentially could impact the performance of *in vitro* tests based on testing patient antibodies or viral antigens.

Thus, the FDA suggests test developers should monitor new and emerging viral mutations and variants that could impact their molecular test performance on and ongoing basis. Database like GISAID make this happen by compiling the sequences worldwide and monitoring all the sequenced variants closely. See **Table 1.**

Table 1. Example of variant watchlist from GISAID database

Variant	#Genomes	#Top Location	#Top Clade	#Top Lineage	Co-occurring Changes List	#Co-occu	Δ#Loc(S)	#aachang	(SxC)
145H_452R_478K_68	24798	17142 England	24767 GK	23836 AY.4.2	Spike_A222V, Spike_T19R, Spike_E156G, Spike_D950N, Spike_F157del, Spike_T95I, Spike_R158del, Spike_D614(	28	44	4	176
452R_478K_681R_12	1563	457 England	1559 GK	400 AY.4	Spike_T19R, Spike_E156G, Spike_D950N, Spike_F157del, Spike_R158del, Spike_D614G, NSP4_T492I, M_I82T, NS	25	31	4	124
503X_504X	328	56 Indiana	318 GK	80 AY.43	Spike_G142D, Spike_T19R, Spike_E156G, Spike_L452R, Spike_F157del, Spike_P681R, Spike_R158del, Spike_T478	28	19	2	38
47X_503X_504X	136	20 California	136 GK	38 AY.43	Spike_G142D, Spike_T19R, Spike_E156G, Spike_L452R, Spike_P681R, Spike_F157del, Spike_R158del, Spike_T478	28	18	3	54
132A_452R_478K_68	60	22 Warminsko-Mazı	59 GK	50 AY.46	Spike_G142D, Spike_E156G, Spike_D950N, Spike_F157del, Spike_T19R, Spike_R158del, Spike_D614G, NSP4_T49	28	16	4	64
47X	90	16 Florida	88 GK	24 AY.43	Spike_G142D, Spike_T19R, Spike_E156G, Spike_L452R, Spike_F157del, Spike_P681R, Spike_R158del, Spike_T478	28	15	1	15
145H_152L_452R_47	33	10 Lower Saxony	33 GK	33 AY.4.2	Spike_A222V, Spike_E156G, Spike_D950N, Spike_F157del, Spike_T19R, Spike_T95I, Spike_R158del, Spike_D614(	29	13	5	65
356R_452R_478K_68	141	64 St Gall	141 GK	110 AY.43	Spike_E156G, Spike_D950N, Spike_F157del, Spike_T19R, Spike_R158del, Spike_D614G, NSP4_T492I, N_Q9L, NSF	29	12	4	48
143X	523	86 Grand-Est	490 GK	102 B.1.617.2	Spike_T19R, Spike_E156G, Spike_D950N, Spike_L452R, Spike_F157del, Spike_P681R, Spike_R158del, Spike_T478	28	11	1	11
148del_452R_478K_	20	7 West Virginia	20 GK	12 B.1.617.2	Spike_G142D, Spike_T19R, Spike_D950N, Spike_D614G, NSP4_T492I, M_I82T, NSP3_A488S, NSP3_P1228L, N_G21	23	10	4	40
478K_681R_1237I	10	1 England	8 GK	2 AY.4	Spike_G142D, Spike_T19R, Spike_E156G, Spike_D950N, Spike_F157del, Spike_R158del, Spike_D614G, M_I82T, N	16	10	3	30
434V_452R_478K_68	206	104 England	205 GK	123 AY.4	Spike_T19R, Spike_E156G, Spike_D950N, Spike_F157del, Spike_R158del, Spike_D614G, NSP4_T492I, M_I82T, NS	25	10	4	40
250I_346K_452R_47	29	9 Saxony	29 GK	29 AY.33	Spike_G142D, Spike_E156G, Spike_D950N, Spike_F157del, Spike_T19R, Spike_R158del, Spike_T29A, Spike_D614	31	10	5	50
143X_144X_145X_14	24	6 South Korea	24 GK	14 B.1.617.2	Spike_P681R, Spike_T478K, Spike_D614G, Spike_T19R, Spike_L452R, NS3_S26L, M_I82T, N_R203M, NSP13_P77L,	15	10	7	70
151N_452R_478K_68	9	1 England	9 GK	3 AY.4	Spike_T19R, Spike_E156G, Spike_D950N, Spike_F157del, Spike_R158del, Spike_D614G, NSP4_T492I, M_I82T, NS	25	9	4	36
378R_452R_478K_68	9	1 Turkey	9 GK	3 B.1.617.2	Spike_T19R, Spike_E156G, Spike_D950N, Spike_F157del, Spike_R158del, Spike_D614G, NSP4_T492I, M_I82T, NS	25	9	4	36
371T_452R_478K_68	22	7 Wales	22 GK	11 AY.44	Spike_G142D, Spike_T19R, Spike_E156G, Spike_D950N, Spike_F157del, Spike_R158del, Spike_D614G, NSP4_T49	25	9	4	36
5F_151I_452R_478K	12	2 New Jersey	12 GK	6 AY.44	Spike_T19R, Spike_E156G, Spike_D950N, Spike_F157del, Spike_R158del, Spike_D614G, NSP4_T492I, M_I82T, NS	25	9	5	45
452R_478K_484A_68	287	121 California	281 GK	138 AY.39.1	Spike_T19R, Spike_E156G, Spike_D950N, Spike_F157del, Spike_R158del, Spike_D614G, NSP4_T492I, M_I82T, NS	25	9	4	36
5F_452R_478K_522\	13	2 England	13 GK	5 B.1.617.2	Spike_T19R, Spike_E156G, Spike_D950N, Spike_F157del, Spike_R158del, Spike_D614G, NSP4_T492I, M_I82T, NS	25	9	5	45
502X_503X_504X	182	54 England	152 GK	60 B.1.617.2	Spike_T19R, Spike_L452R, Spike_D950N, Spike_P681R, Spike_T478K, Spike_D614G, NSP4_T492I, M_I82T, NSP3_F	25	9	3	27
250I_452R_478K_68	27	9 Hovedstaden	27 GK	27 AY.33	Spike_T19R, Spike_E156G, Spike_D950N, Spike_F157del, Spike_R158del, Spike_T29A, Spike_Q613H, Spike_D614	28	8	5	40
46X_47X	16	2 England	16 GK	8 AY.39	Spike_T19R, Spike_E156G, Spike_L452R, Spike_D950N, Spike_F157del, Spike_P681R, Spike_R158del, Spike_T478	25	8	2	16

This includes assessing and monitoring the prevalence of novel/notable variants in reputable databases such as GISAID.[18] These variants are prioritized based on frequency,

where the FDA currently considers a "significant frequency to be greater than 5% when considering at least 2000 sequences over a recent period of time."[17] These viral mutations and variants deemed "prevalent" and/or clinically significant would be the focus of studies identifying the possible effects of the mutations on assay performance.

This can be accomplished by *in silico* analysis of published SARS-CoV-2 sequences compared to the test's primers and probes. If mutations coincide with primer/probe binding sites, or are otherwise likely to affect assay, they should be tested with either clinical of appropriate contrived samples to assess the impact of the mutation or variant on the test's performance.[17] For example, a gene synthesis of the target amplicon can be performed to generate plasmid material harboring the relevant variant-specific mutation. These can be used for contrived samples in molecular diagnostics, wherein the plasmid is spiked into VTM or saliva.

Overview of the Study

Text-based Pairwise Sequence Alignment files were downloaded from the NCBI blastn search for each of the 6 primers in our established primer set. This was done to see if our "established" primer exhibits sequence identity among the vast majority of published SARS-CoV-2 variants. If sequence conservation in the primer binding sites is ubiquitous, or nearly so (e.g., 99% of published sequences), then the likelihood of false positives due to mismatches in the primer binding site is assumed to be exceedingly unlikely and irrelevant to assay performance.

These pairwise alignment text files were then parsed using a sequence identity algorithm written in Perl. This program processes the downloaded text dump files to

determine the degree of sequence identity between the primers and each of the database's SARS-CoV-2 RNA sequences. Here, blastn's nr/nt (nucleotide collection) database was used. Due to the prevalence of 100% sequence identity, it was deemed prudent to screen these results for instances in which less than 95% identity was observed.

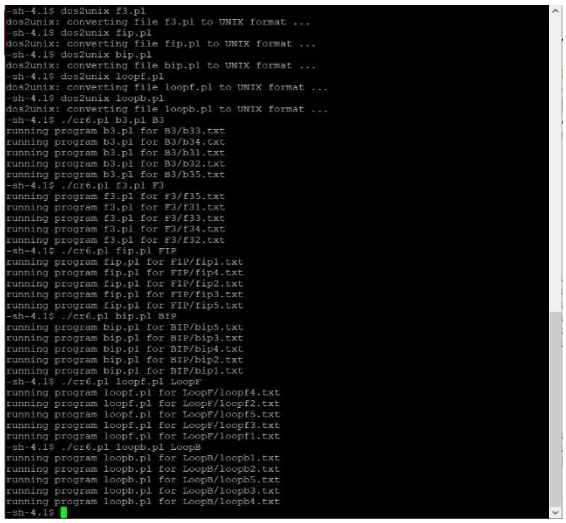

Figure 5. Blastn text dump files processed recursively on the Unix Secure Shell Environment using Perl

Results

Since the resultant text files exported from the BLAST search contained information that was not directly relevant to the inclusivity study, I decided to eliminate most of the text-heavy results and focused on showing the similarity in percentage to improve readability. Single mismatches were detected in primer binding sites in 16 of those sequences as shown in **Table 2**. Accession numbers are noted in the table.

As FIP/BIP primers are the most important for the formation of the looped amplification intermediate, mutations in said primers were prioritized in the analysis. As noted previously, the FIP and BIP primers are responsible for exponential amplification after generation of the first looped intermediates.[5] Therefore, mutations in the FIP/BIP binding sites are the most likely to affect assay performance. Mutations in the other primer binding sites, while still pertinent, may have more marginal impacts on amplification kinetics.

The inclusivity study was performed against a database of 5,000 SARS-CoV-2 sequences. Only 3 unique sequences contained mutations in the FIP and BIP portion. We additionally determined that the remaining mutations are less likely to affect amplification efficiency. Since sequence identity was retained among the vast majority of the 5,000 examined isolates, I chose to present those isolates with less than 95% sequence identity. Of the 5,000 isolates aligned, only 16 isolate sequences showed mismatches to the N8 primer set. This is especially impressive since 5000 sequences were aligned per primers, this effectively means, only a single hit was found for BIP and LoopB (0.02%), 4 hits for LoopF (0.08%), 8 hits for F3 (0.16%), 3 hits for FIP (0.06%), and 0 hits for B3 (0%). F3 exhibited the highest number of isolates with mismatched target sequence, perhaps due to lower sequence conservation in this region (e.g., a T cell epitope). Notably, in all instance mismatches were limited to one base, and no isolates were observed with less than 90% sequence identity.

Table 2. *In silico* inclusivity study results. A pairwise alignment of primer sequences against available SARS-CoV-2 sequences was performed. Where less than 100% identity was observed, the accession number and the corresponding primer were reported.

	Subject	% Identity	Alignment Length	Mismatches	Gap Opens	E Value
BIP	MW320777.1	95	20	1	0	0.07
LoopB	MW320747.1	95	20	1	0	0.07
LoopF	MW321330.1	95	20	1	0	0.07
	MW321246.1	95	20	1	0	0.07
	MW320935.1	95	20	1	0	0.07
	MW320774.1	95	20	1	0	0.07
F3	LR962981.1	95	20	1	0	0.07
	LR962923.1	95	20	1	0	0.07
	MW341912.1	95	20	1	0	0.28
	MW341904.1	95	20	1	0	0.28
	MW341848.1	95	20	1	0	0.28
	MW341842.1	95	20	1	0	0.28
	MW341841.1	95	20	1	0	0.28
	MW341837.1	95	20	1	0	0.28
FIP	MW321035.1	95	20	1	0	0.07
	MW320777.1	95	20	1	0	0.07
	MW341967.1	95	20	1	0	0.28
B3						

This is a very promising result since FDA determined 5% as a "significant frequency" out of 2000 sequences. In our *in silico* study, only 16 sequences out of total 5000 sequences showed less than 100% identity and even those showed higher than 90% identity. Additionally, 0.02% for BIP/LoopB, 0.08% for LoopF, 0.16% for F3, 0.06% for FIP, and 0% for B3 are far below FDA's set threshold. Thus, our finalized primer set should capture majority of the known SARS-CoV-2 strains and should have minuscule amount of false negative result that could be caused by escaping the primers' coverage.

Cross-reactivity Study (Analytical Specificity)/ Microbial Interference Studies

Background

The FDA also requires a cross-reactivity study for an EUA request. In this study, the risk of inadvertent detection of other microorganisms—those likely to be present in the relevant clinical matrices—is assessed *in silico* and in specified wet lab studies. If the primer is not sufficiently specific, the false positive rate will increase, which in turn renders the diagnostic kit inaccurate. Such a diagnostic, due to the potential damages incurred through generating inaccurate results, are not authorized by the FDA for distribution. Additionally, according to the FDA, cross-reactivity studies are required to demonstrate that the *in vitro* test does not react with related pathogens, high prevalence disease agents, and normal flora that are likely to be encountered in patient samples. As an example, *staphylococcus epidermidis* would be a common contaminant of NP swabs, and exhibiting no amplification from this genomic DNA would show that this common contaminant does not alter the specificity of the assay.

Here, *in silico* studies of target primers and probes was performed to determine the identity between the FDA's suggested list of microorganism and the test's primers/probes. The FDA recommends assessing the performance of the diagnostic in wet lab studies if there is ≥ 80% homology observed between the primers and the microorganism's genome: this entails purchasing or extracting nucleic acids from the microorganism and assess whether they result in false positives. Thus, a cross-reactivity study was performed to assess non-specific detection of other human pathogens likely present in respiratory samples.

Table 3. Recommended List of Organisms to be Analyzed *in silico* and by Wet Testing for All Respiratory Samples

High priority pathogens from the same genetic family	**High priority organisms likely present in respiratory samples**
Human coronavirus 229E	Adenovirus (e.g., C1 Ad. 71)
Human coronavirus OC43	Human Metapneumovirus (hMPV)
Human coronavirus HKU1	Parainfluenza virus 1-4
Human coronavirus NL63	Influenza A & B
SARS-CoV-1	Enterovirus (e.g., EV68)
MERS-coronavirus	Respiratory syncytial virus
	Rhinovirus
	Chlamydia pneumoniae
	Haemophilus influenzae
	Legionella pneumophila
	*Mycobacterium tuberculosis**
	Streptococcus pneumoniae
	Streptococcus pyogenes
	Bordetella pertussis
	Mycoplasma pneumoniae
	Pneumocystis jirovecii (PJP)*
	Pooled human nasal wash - *to represent diverse microbial flora in the human respiratory tract*
	Candida albicans
	Pseudomonas aeruginosa
	Staphylococcus epidermis
	Streptococcus salivarius

Overview of the Study

Utilizing BLAST, sequence similarity comparison between each primer and the whole genome of the microorganisms listed in **Table 4** was performed. Most of the primers

exhibited significant sequence similarity with SARS-CoV. However, this is not likely to be clinically relevant, as SARS-CoV is not prevalent.

This data was then exported to text dump files. Since the text files exported from BLAST are indecipherable without an algorithm, for all intents and purposes, these files are mined for relevant information that is exported into a more legible and analyzable format. In this case, percent identities were exported into Excel, as shown in the table below.

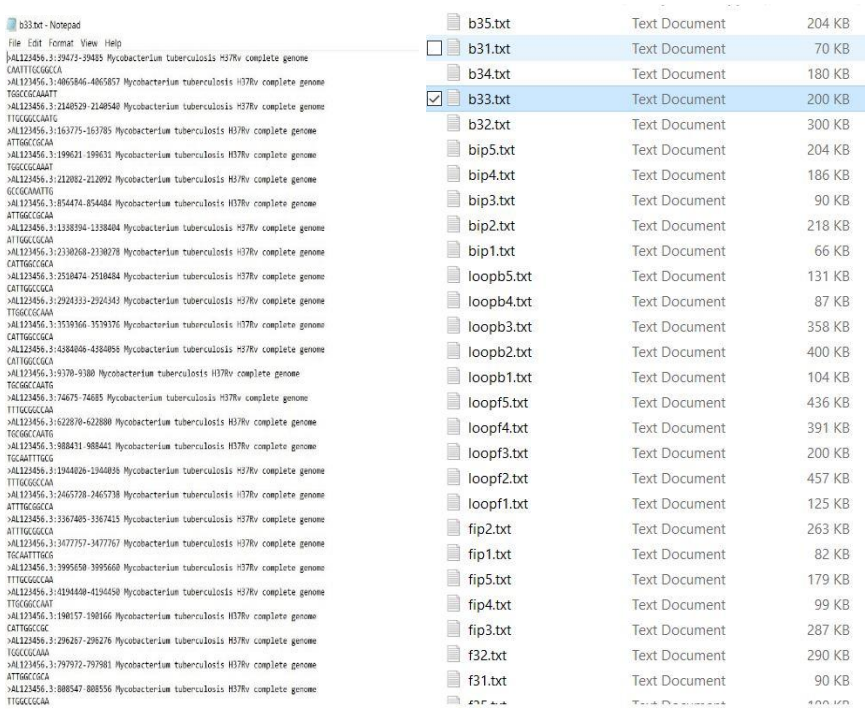

Figure 6. Example of the text dump files

Results

Primers B3, F3 and LoopF exhibited >80% sequence similarity as noted in the table below. However, as robust amplification, necessary for colorimetric detection, requires all primers to bind, it is unlikely that the low-efficiency primer binding mediated by 85% sequence identity would contribute to false-positives. Notably, in none of the assessed microorganisms, was >85% sequence identity observed for the FIP and BIP (except for 2003 SARS-CoV), which are the most pertinent for amplification and reaction specificity. B3 had multiple matches at 85% similarity cutoff. This is most likely due to the shorter primer length: at 18 mer, the probability of a perfect random match is 4^{18} (once every 68 billion base pairs). However, at the 85% identity cutoff, the probability is much more likely (e.g., once every 4 billion base pairs for 16 contiguous identical residues, and more probable still if including non-contiguous bases).

For the wet lab portion of the study, microorganism DNA or RNA was purchased from ATCC (more details can be found in **Table 12**). 1 µL of the purchased DNA or RNA was added to 9 µL of PCR-grade water, 2.5 µL of the primer mix, and 12.5 µL of RT-LAMP Master Mix. Reactions were incubated at 65°C for 20 minutes and the reactivity was assessed based on color development and is shown in **Table 4** and **Figure 7**.

Table 4. *In silico* and wet lab cross-reactivity study results. Non-specific detection of microorganism DNA or RNA was reported qualitatively based on the development of a yellow color in the reaction for the wet lab studies. For the *in silico*

portion of the cross-reactivity study, where pairwise alignment indicated >85% sequence identity, the corresponding primer is listed**.**

				In Silico Study Result		
	Average	STDEV	Result	85% Match	90% Match	95% Match
Human coronavirus 229E	0.968995	0.535748	-			
Human coronavirus OC43	1.137194	0.456966	-			
Human coronavirus HKU1	0.819208	0.009014	-			
Human coronavirus NL63	1.109229	0.358749	-			
SARS-CoV [2003]	0.818332	0.024444	-	LoopB/LoopF/BIP/FIP/F3/B3	LoopF/LoopB/FIP/BIP/B3	LoopB/LoopF/FIP/B3
Middle East Respiratory (MERS)syndrome coronavirus	0.846557	0.001148	-			
Human adenovirus 1	1.020236	0.244472	-			
Human adenovirus 2	1.209285	0.545541	-			
Human adenovirus 3	1.048658	0.202791	-			
Human adenovirus 4	0.897137	0.12893	-			
Human adenovirus 5	0.683029	0.002001	-			
Human adenovirus 6	0.814508	0.020517	-			
Human adenovirus 7	0.818297	0.017638	-			
Human metapneumovirus	0.806468	0.031888	-			
Human parainfluenzavirus 2	0.775679	0.004479	-			
Human parainfluenzavirus 3	1.280405	0.090777	-			
Influenza A virus (H3N2)	0.871993	0.012495	-			
Influenza B virus	0.805725	0.034184	-			
Enterovirus D68	0.773268	0.014118	-			
Human enterovirus 71	0.610825	0.022322	-			
Human respiratorysyncytial virus	0.776864	0.016	-			
Human rhinovirus 1A	0.760922	0.015447	-			
HRV Type 14			-			
Human rhinovirus 17	1.411243	0.026002	-			
Chlamydophila pneumoniae	0.573331	0.059018	-			
Haemophilus influenzae	1.105471	0.080173	-	B3		
Legionella pneumophila	0.806654	0.146312	-	B3		
Mycobacterium tuberculosis	1.407093	1.011345	-			
Streptococcus pneumoniae			-	LoopF		
Streptococcus pyogenes			-	B3		
Bordetella pertussis			-			
Mycoplasma pneumoniae	1.131624	0.515481	-			
Candida albicans	0.916783	0.357836	-	F3		
Human rhinovirus 1B	1.684091	0.048212	+			

Cross-reactivity was low among the vast majority of sample microorganisms. Some amplification below the threshold of 1.5 was detected for human rhinovirus 17, human parainfluenza virus 3, and mycobacterium tuberculosis. Only one sample exhibited amplification above the threshold—human rhinovirus 1B.

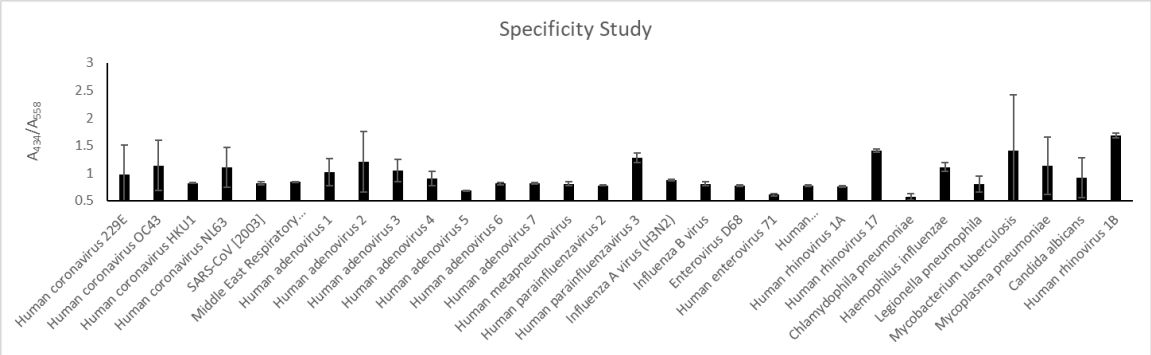

Figure 7. Result of the Wet-lab Confirmation Study Conducted Post *in silico* Study**.**

While some degree of sequence identity (e.g., 85%) was observed among the microorganisms for LoopF, F3, and—more prevalently—B3, the *in silico* study results did not broadly recapitulate the wet lab study results. Even though SARS-CoV (2003) bound to the FIP, LoopF, LoopB, and B3 with 90% identity, cross-reactivity was not observed in the wet lab study, which may highlight that sequence identity among all the primers may be required for non-specific amplification. It is currently believed that the amplification of rhinovirus sample was an artefact, given that no apparent sequence identity was observed. This may be due to contamination of the material purchased from ATCC or during preparation of the RT-LAMP plate.

Conclusion and Discussion

Through conducting the *in silico* inclusivity and cross-reactivity studies, we have demonstrated that the RT-LAMP kit is viable for further commercial development: the kit should not generate false positives, due to inadvertent detection of off-target microorganism genomic materials, or false negatives, due to frequent mutations in the primer binding sites in circulating SARS-CoV-2 variants. In the inclusivity study, by utilizing well-established and readily available genomic databases such as NCBI's nucleotides database, we were able to determine whether the primer sequences are conserved among the vast majority of available SARS-CoV-2 genomes. In the case of cross-reactivity study, by using publicly available databases, we were able to demonstrate that non-specific amplification of other pathogens/flora is highly unlikely. As noted

previously, the absence of binding to the FIP and BIP in the list of microorganisms largely precludes the possibility of non-target amplification, despite the occasional sequence identity observed in the other primers. This process allows molecular diagnostic kit developers to screen putative primer sets, assessing the relative risk of false positive or false negative results before purchasing the primers.

The resultant data can be exported into formats such as .txt, .csv, or .xls, that are more amenable to further processing and analysis by any user with sufficient data manipulation proficiency. In this case, a number of Perl codes written by myself were used to parse the data and to assess inclusivity and cross-reactivity.

These *in silico* studies can be used to screen prospective LAMP primer sets to exclude those sets most likely to generate false-positive results due to sequence similarity to off-target RNA/DNA from viruses, bacteria, other pathogens, or human cells that are present in samples. Similarly, false-negative results, due to poor conservation of the primer binding sites, can be avoided. Using these techniques, one can screen primers for SARS-CoV-2, other emerging infectious diseases, or specific genotypes in human patients.[19] In the latest case, one might include a clinically relevant single nucleotide polymorphism (SNP) as the 3' terminal residue of one of the inner primers: mismatch at this position in the BIP/FIP should severely impact amplification kinetics, resulting in a measurable delta in signal between the wild-type and SNP amplicon.

The development and widespread usage of *in silico* models are facilitated nowadays by rapidly advancing biological experimenting methods, and analytical tools that produce data-rich, high-throughput biological data. In the above studies, high-throughput Sanger

and next-generation sequencing identity thousands of variants whose sequences are uploaded into public databases—these databases made the inclusivity and cross-reactivity studies possible. Further, in our case, a more advanced primer selection workflow/pipeline could mine the RNA sequence from RNA viruses for secondary structures.

Such a pipeline may avoid regions of extended internal (i.e., intrastrand) binding, which, as an example, occurs in the stem portion of stem-loop secondary structures, that would compete with the primer, but include regions that are highly conserved.

CHAPTER THREE – INTERFERENCE STUDIES AND STABILITY STUDIES

Endogenous/Exogenous Interference Substances Studies

Additionally, we tested the effects of potentially interfering endogenous substances in both VTM and saliva extracted using a commercially available Nucleic Acid Purification Kit. The FDA notes common components likely to be present in NP swabs and OP swabs, which are noted in **Table 5**. These components can interfere with either the extraction process or, though residual contamination in the eluate, the detection of target nucleic acids. As such, the endogenous interfering substances study involves spiking these materials into positive and negative samples in a relevant clinical matrix (an NP swab in VTM and saliva, here) and performing the extraction and candidate diagnostic assay.

The results of this study are shown in **Table 6**. Menthol and benzocaine were dissolved into methanol at 100 and 300 mg/mL, respectively, before dilution to 3 mg/mL in saliva or and NP swab in VTM. Toothpaste was diluted to 5% in PCR-grade water and then diluted 10-fold into an NP swab in VTM or saliva. Nicotine was obtained as a 3 mg/mL solution from e-cigarette liquid (3 mg/mL nicotine in 50% v/v glycerol and polyethylene glycol, containing flavorings). Jurkat genomic DNA was used as the source of human genomic DNA. All other substances were directly diluted into the indicated clinical matrix. After preparation of each of the substances at their indicated concentrations in the two clinical matrices, chemically-inactivated SARS-CoV-2 (Microbiologics) was added directly to the sample, introducing a total of 60,000 copies. These solutions were then extracted and eluted into 80 µL of PCR-grade water, producing 750 copies per µL. 10

µL (i.e., 7,500 copies or 3-fold higher than the LoD in NP swabs) was then used for each RT-LAMP reaction. Samples were run in duplicate. As shown in **Table 6**, only toothpaste exhibited an effect, and this was limited to one of the NP swab samples.

Table 5. List of Potential Interfering Substances Recommended for Testing When the Candidate Test is Indicated for Respiratory Samples

Potential Interfering Substances	**Concentration**
Afrin Original nasal spray	15% v/v
Sore throat and cough lozenges such as Cepacol Lozenges (benzocaine/menthol)	3 mg/mL
Chloroseptic Sore Throat spray	5% v/v
Mouth Wash (Saliva)	5% v/v
Cough syrup (e.g., Robitussin)	5%
Mucin: bovine submaxillary gland, type I-S	2.5 mg/ml
Nicotine or Tobacco	0.03 mg/ml
Toothpaste (Saliva)	0.5% v/v

Table 6. Assessment of the effects of endogenous interfering substances.

		Saliva Standard	VTM Standard	Benzocaine + Menthol	Nasal Decongestant	Saline Spray	Nicotine	Toothpaste	Mouthwash	Sore Throat Spray	Human Genomic DNA	Whole human blood
Matrix		7,500 copies	7,500 copies	3 mg/mL (each)	15% v/v	1.25%	0.03 mg/mL	0.5% v/v	5% v/v	5% v/v	10 ng/mL	1% v/v
Saliva	Rep 1	1.94	1.79	1.94	1.67	1.98	2.41	1.66	1.63	1.96	1.84	1.79
	Rep 2	1.71	1.78	1.86	1.50	1.61	1.70	1.71	1.71	1.63	1.80	1.93
NP Swab	Rep 1	1.35	1.81	1.74	1.78	1.78	1.67	0.77	1.75	1.72	1.81	1.75
	Rep 2	1.17	1.63	1.85	1.67	1.75	1.66	1.84	1.83	1.78	1.78	1.97

Sample Stability Studies

Additionally, the FDA EUA guideline for molecular diagnostic includes stability studies. Here, stability is determined by storing the final kit components at recommended

temperatures (-20°C for the RT-LAMP primer/enzyme mix) and assayed periodically for performance. Stability can be defined thereby as the time point at which the LoD of the kit has declined below 2-3X the baseline LoD. RT-LAMP master-mix was mixed with the primer set in sufficient quantity for a baseline triplicate and triplicated samples through 6 months, with one triplicate per month. The resulting mixture was then aliquoted and stored at -20°C. After the initial freezing, a baseline reading, mimicking the first freeze-thaw cycle of material used after receipt by a customer, was performed. While this study is still ongoing, we have demonstrated that the mixture is stable for at least for 5 months post mixing at -20°C.

CHAPTER FOUR - DEVELOPMENT PROCESS OF THE SARS-COV2 DIAGNOSTIC KIT

This section will cover additional efforts to optimize and automate the diagnostic kit. Over the course of last year and half, from its conception, we expended a significant amount of time and resources to make this proof-of-concept kit into a functioning, automated and easy-to-use SARS-CoV-2 diagnostic, where it can be used in a PoC environment in developing countries. To achieve this, we had to resolve stochastic and non-stochastic false positive/false negative issues that can be associated with amplifying genetic materials in the RT-LAMP format. Additionally, we screened a number of different primer sets and automated the whole process such that the kit is user-friendly to those lacking laboratory technical skills.

The diagnostic kit was first developed in July of 2020. We established a prototype test using a primer set that was internally denoted N2. Early results indicated a high degree of sensitivity was possible with extended incubation times (i.e. 1 hour of amplification of RT-LAMP at 65°C) and with purified SARS-CoV-2 standard from an outside vendor. Later attempts to repeat this data using viral material in clinical matrix at more feasible incubation times (e.g., 20-30 minutes) demonstrated that this level of sensitivity is difficult to recapitulate. For instance, incubating for 1 hour introduces a high probability of generating false positives. To better explain this phenomenon, a 1-hour RT-LAMP incubation is comparable to performing qPCR over 80 cycles, or nested PCR, where one copy (or less) of contamination can result in a signal from nominally template-free control materials. As an example of this, in **Figure 8**, one of the negative control materials partially

developed at this time point. It was for these reasons that we opted to limit amplification incubations to 30 minutes or less, as false positives become far less prevalent. In a subsequent LoD study, it was we demonstrated an LoD of 625 copies per reaction (09-30-2020) (**Figure 9**).

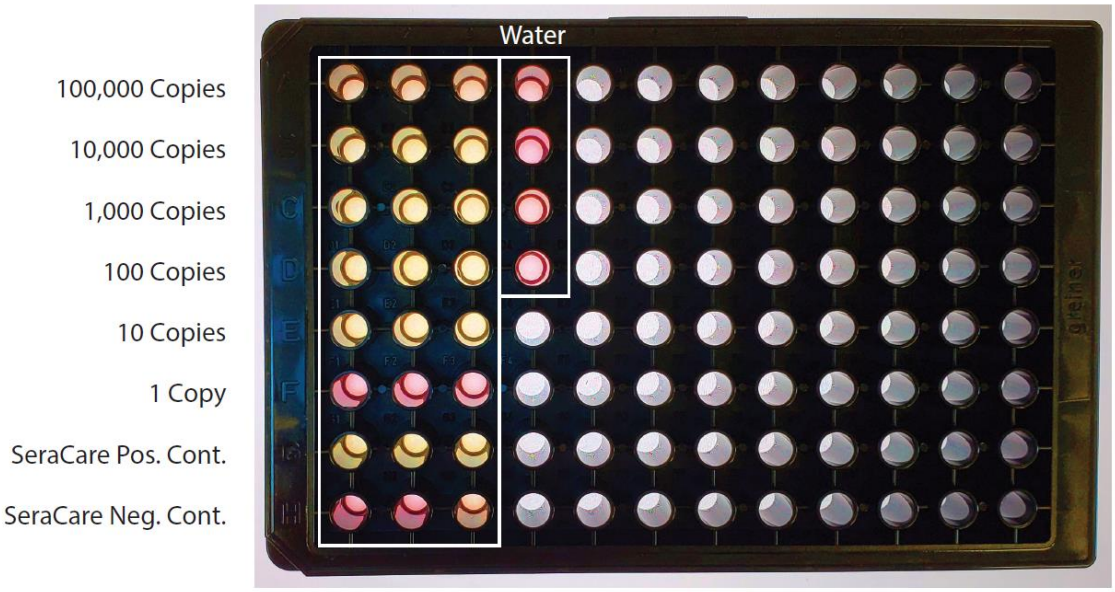

Figure 8. Small Scale LoD Study Using N2 Primer Sets

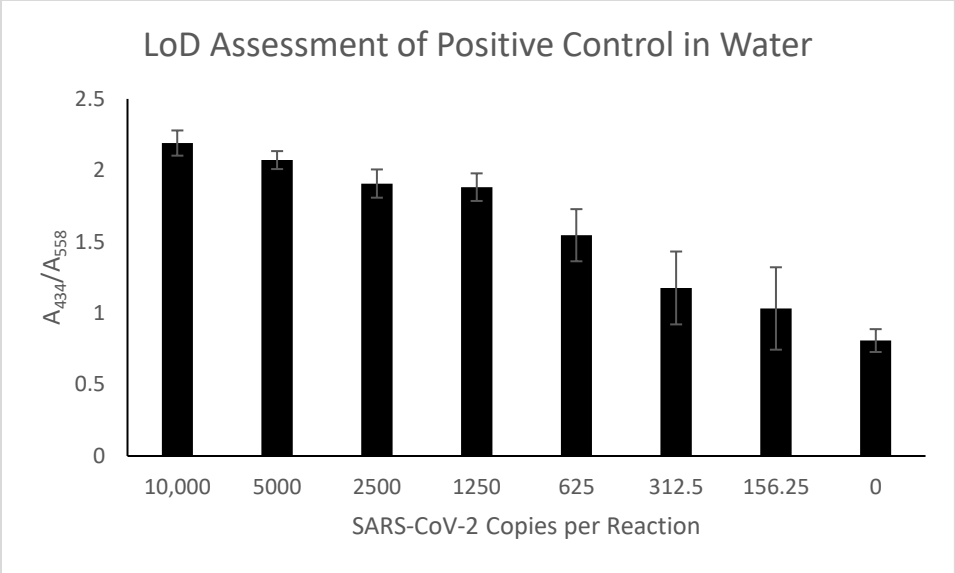

Figure 9 . LoD Assessment of Positive Control in Water

Figure 10. Larger Scale LoD Study Result Using N2 Primer Sets

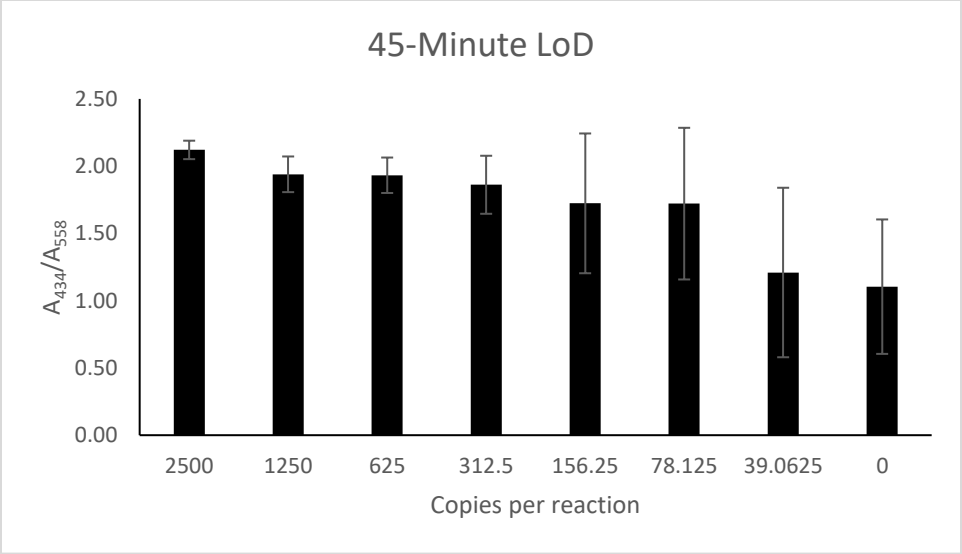

Figure 11. 45-Minute LoD Result Using N2 Primer Set

Based on the result in **Figure 8**, where marginal development of unextracted positive control material (see wells G1-G3) was observed, we hypothesized that it may be possible to perform the RT-LAMP on unextracted patient samples. For this, we utilized saliva and viral transport medium (VTM) as components that interfere with the colorimetric LAMP. It appeared that both the VTM and saliva prevent normal RT-LAMP colorimetric development (10-2-2020). It is likely the case that these clinical matrices do not prevent amplification, but merely buffer the reaction, preventing the pH reduction and color change. If this is true, it may be possible to use a lower volume of input material at the expense of a higher LoD, as the input sample volume reduction would mean fewer copies are introduced into the reaction. Consequently, it was deemed necessary to include an RNA extraction step to eliminate the interfering components in saliva and VTM, which are likely proteins and buffering salts.

This resulted in an internal discussion regarding optimization/minimization of the purification process. Commercial kits require many steps, including washes and elution. Our team sought a simplified purification procedure that was pursued later in October (see below). After, a repeat of the LoD study was done again with purified genetic material identified 1,250 viral copies per reaction as the LoD (10-3-2020).

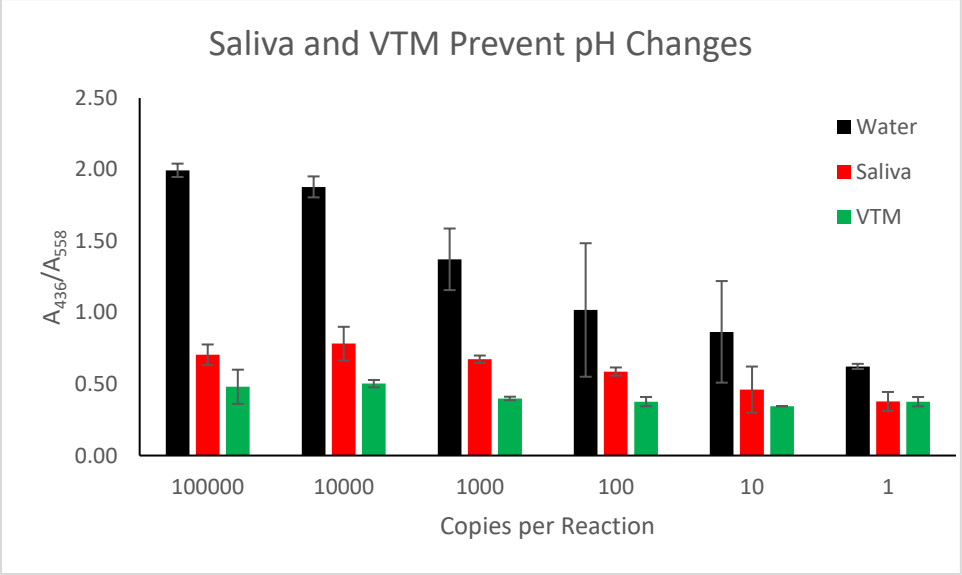

Figure 12. Attempt to Directly Using Different Types of Medium

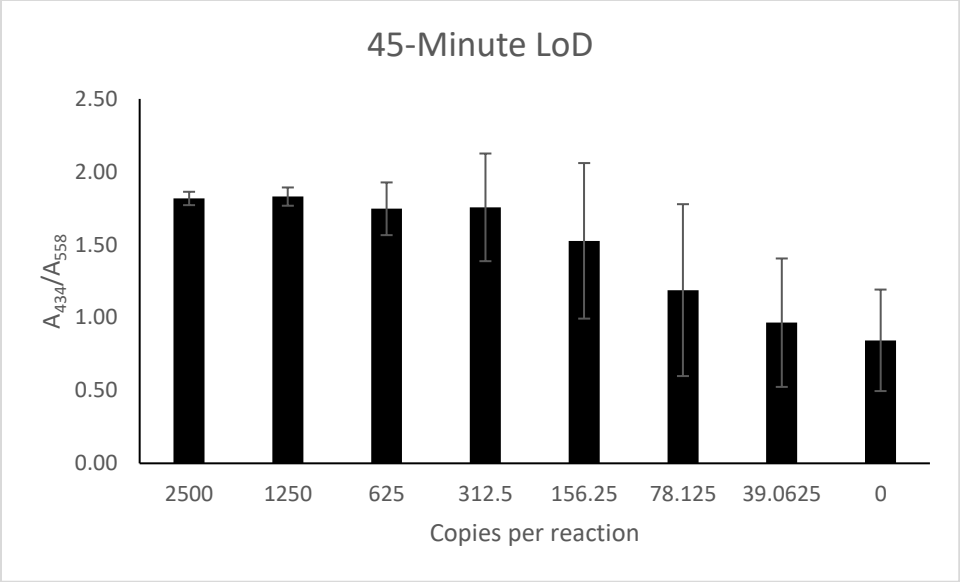

Figure 13. Repeat of LoD Study Using N2 Primer Set

We began to observe issues with background amplification (i.e., false positives) from saliva using N2 primer set (10-5-2020). This false positive issue continued through 10-20-2020, and we pursued alternative solutions. To resolve this issue, different magnetic affinity beads were assessed. It seemed plausible that the source of false positive may be due to the co-purification of other nucleic acids or enzymes that lower the reaction pH. In collaboration with Ceres Nanosciences, we sought to develop a purification process using their "NanoTrap" magnetic beads. These beads differ from conventional silica-based solid-phase extraction systems in that the beads bind to intact virions before washing, lysis and elution. A preliminary proof-of-principle extraction was performed on 10-16-2020 (**Figure 14**). However, despite developing the extraction protocol with Ceres NanoTrap particles, false positives still remained an issue at this point. In **Figure 15**, an attempted LoD study

using saliva spiked with SARS-CoV-2 and saliva only (NTC) extracted with NanoTrap particles was performed (10-19-2020). Note the large value for the background (NTC) sample.

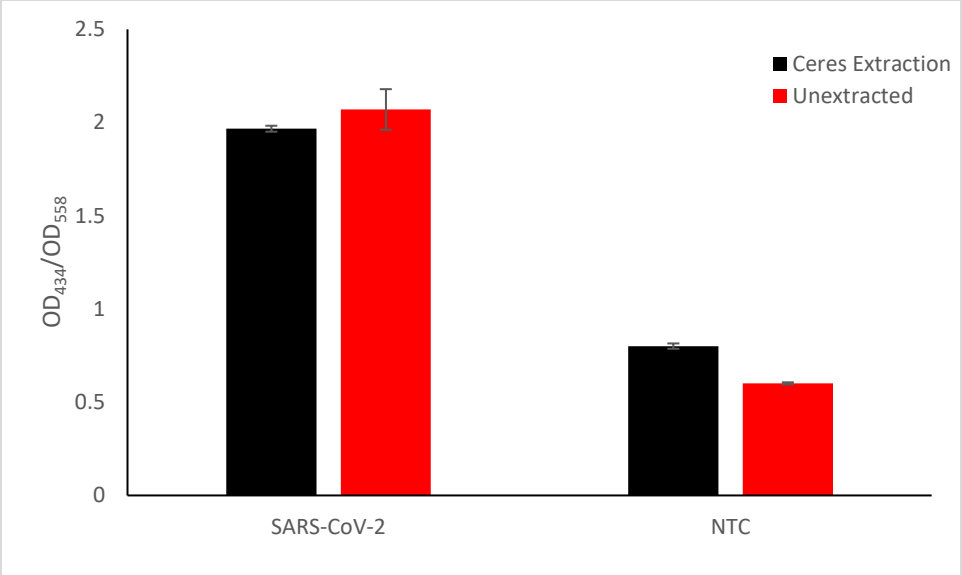

Figure 14. Utilizing an RNA Purification by Magnetic Beads

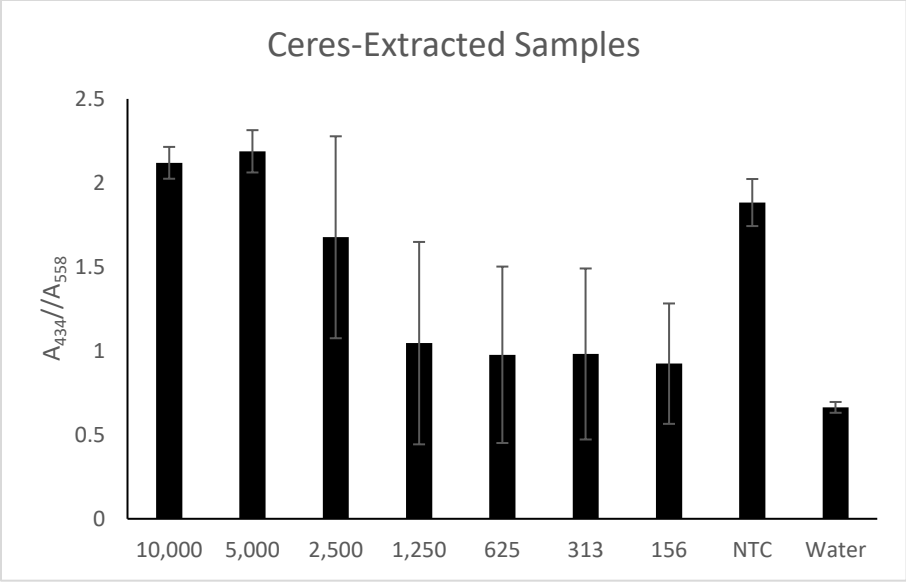

Figure 15. Third-Party Magnetic Bead Extraction Result

It was then postulated that different primer sets may exhibit different propensities for false positive development from saliva. Thus, primer sets N2-N6 were tested in parallel using a fluorescent variant of the RT-LAMP (10-21-2020). Note that N4 exhibits comparable sensitivity to N2, but with less background/false positive development from NTC.

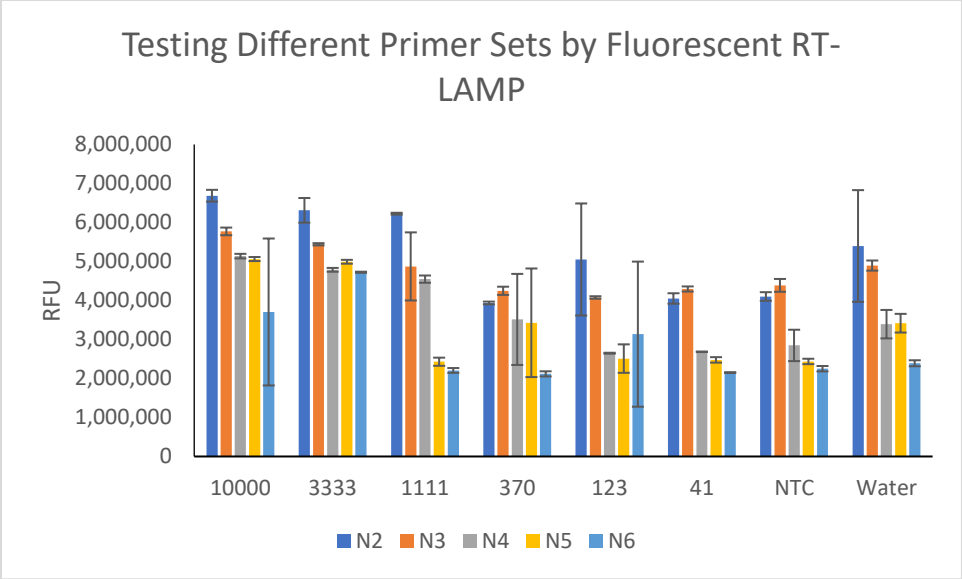

Figure 16. Fluorescent RT-LAMP Result with Different Set of Primers

Thereafter, we performed an LoD study using colorimetric RT-LAMP using Ceres NanoTrap particles and repeated the assessment of N2, N4, and N5. Note that N4, once again, performs better than N2 with lower NTC values (10-22-2020).

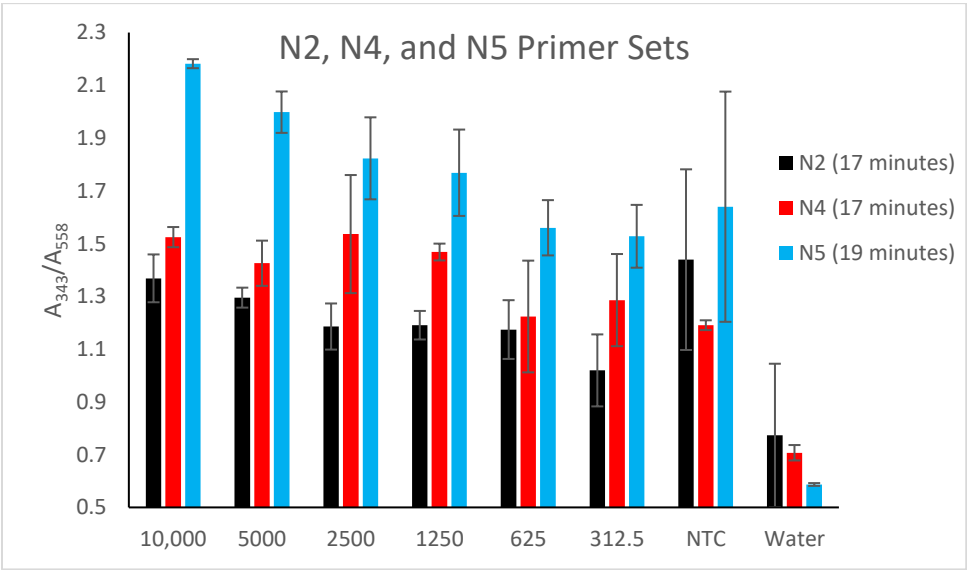

Figure 17. LoD Study with the Selected Primer Sets

Based on the observation earlier, a 20-replicate LoD study was performed using the new system (N4 + Ceres NanoTrap particle-based extraction). Here, a recurring phenomenon was noted—if the RT-LAMP was incubated for 17 minutes, the LoD was measured at ~2,500 copies per reaction with rare false positives: however, if the reaction proceeded longer than 18 minutes, the LoD was improved further at the expense of increased false positives (10-29-2020). In this case, as the reaction was terminated at 19 minutes, false positives were evident in the saliva only samples.

To pursue a protocol minimization for the extraction procedure, we tested the effects of eliminating the wash step and elution step (11-4-2020). The wash step could be eliminated (**Figure 18**), but the elution step was required (data not shown).

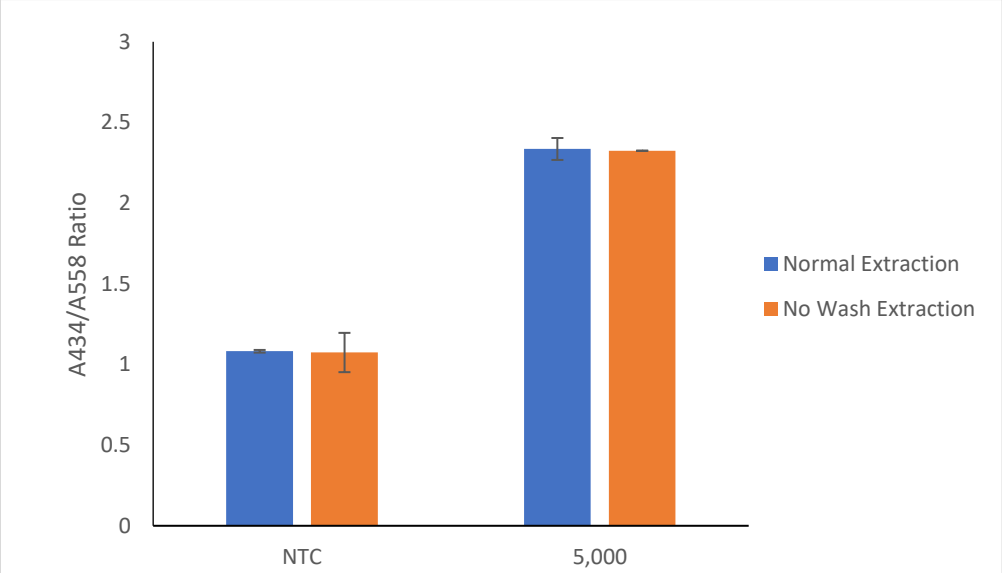

Figure 18. Attempt to Minimize the Extraction Procedure: No Washing Step

Additionally, we have tried many other detection options: a molecular beacon – a hairpin shaped oligonucleotide with an internally quenched fluorophores[20,21] – a fluorescent probe was tested with a primer set related to N2, denoted N2.1 (11-5-2020) The reaction failed to demonstrate fluorescent detection of target sequences. However, the development of a probe-based detection system could dramatically improve sensitivity and reduce false positives. This remains a goal for the RT-LAMP diagnostic. Furthermore, we assessed the potential of an alternative colorimetric detection reagent, denoted leucocrystal violet, LCV (11-12-2020)[22]. LCV is a colorless sulfite adduct that becomes blue after the nucleobase-mediated displacement of the sulfite. LCV was able to detect DNA in water, but in saliva, there was too much DNA from the Ceres NanoTrap-based extraction to allow for LCV-based detection (data not shown). As such, while LCV detection may be plausible, it may require dilution of the amplified material, which significantly complicates the assay.

We also attempted a real-time RT-LAMP operating at 45°C, with the goal of capturing data continuously to avoid false positive development due to over development of the reaction. This attempt failed, however. It was posited that lower temperatures increased non-specific binding of the primers—as such, no distinction between saliva only and spiked sample extracts could be observed.

Since this attempt, we decided to test different set of primers since we observed marginal improvement from the different set of primers.

Endeavors to Optimize the Primer Sets

At this juncture, we have tested and assessed many elements of a prospective SARS-CoV-2 RT-LAMP diagnostic; however, the primer optimization studies were partially inconclusive. Consequently, we assessed the performance of additional primer sets N7-N9 (11-13-2020). N4, N7, and N8 performed comparably. The hope was to identify primer sets that out-perform N4. However, it appeared that the non-specific amplification observed in extracted saliva samples was largely unrelated to the primer set. This spurred a more recent endeavor to reduce background by eliminating carryover proteins and DNA, which could additionally allow for LCV-based detection. Proteinase K and a dsDNAse were tested in following experiments.

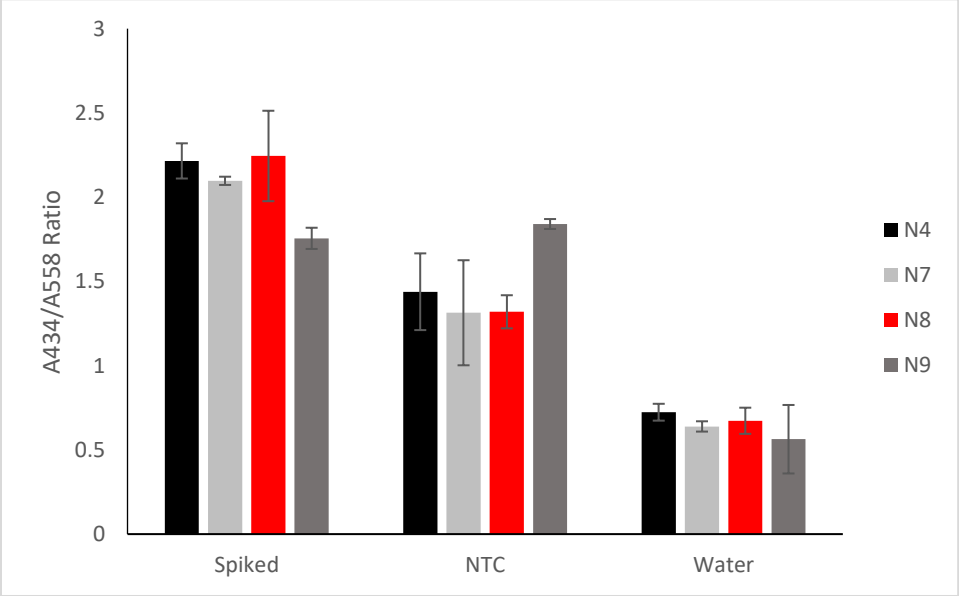

Figure 19. Assessing the Performance of Additional Sets of Primers

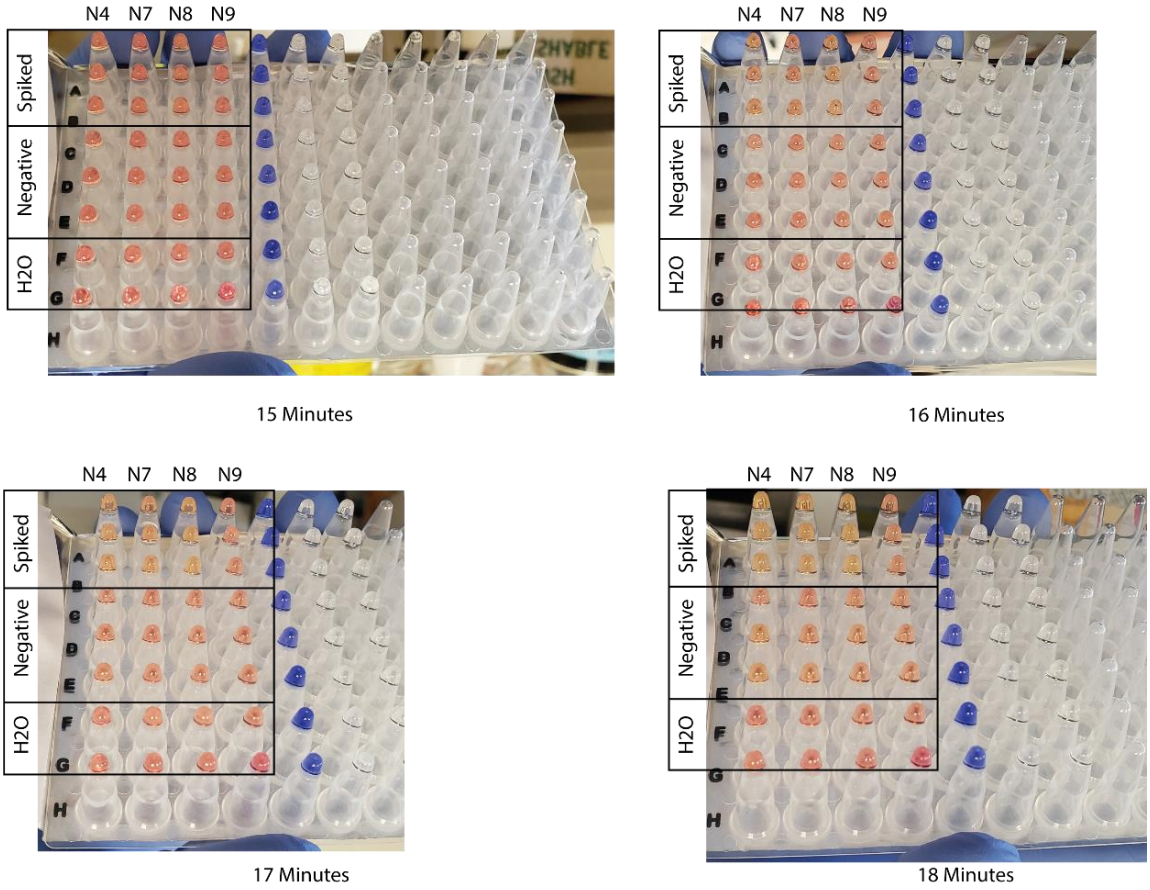

Figure 20. Assessing the Performance of Additional Sets of Primers Cont'd.

The dsDNAse was tested with N4 primers in the original colorimetric RT-LAMP and LCV-based colorimetric LAMP (11-19-2020). Unfortunately, the incubation with dsDNAse ablated the response to spiked samples, potentially indicating a sample degradation during the incubation period.

From an internal discussion, we decided to use viral transport media (VTM) rather than saliva samples. Focusing on spiked VTM became a new standard as it will mimic nasopharyngeal swab samples, which are resuspended into VTM. (11-20-2020) Notably,

VTM is a “clean” sample—there are no particulates, bacteria, or host cells that can interfere with the assay or result in false positives.

Additionally, it was noted that it would be advantageous since any potential “contaminants” would be present in genuine clinical samples at a lower titer than in saliva. Therefore, the assay should perform significantly better on this matrix than on saliva. An additional note is that the following assays were performed using the commercially available viral RNA extraction kit in lieu of the Ceres NanoTrap extraction protocol that we previously developed: it was deemed prudent to use a cleaner RNA prep for these studies to demonstrate the functionality of the diagnostic.

While the optimum primer sets for saliva were identified through previous primer screening – determining the N4 or N8 sets as better options – we had not established which sets perform well with spiked VTM. Thus, N2, N4, N7, and N8 were assessed in parallel during the duplicated primer screen on VTM (11-23-2020). Additionally, another test was done to assess the specificity of the primer set using only FIP/BIP portion: this was performed for N2 and N4 (denoted N2 and N4 FIP/BIP in **Figure 21**).

Notably, while N4 and N8 both performed well, N4 exhibited false negatives at 1,250 copies per reaction that were absent in N8. While this could be a stochastic sampling error, N8 appeared to out-perform N4. Also, while N7 performed comparably to N4, N2 performed worse than N4, N7, and N8, with a lower signal-to-noise/NTC ratio and a higher false positive rate in NTC samples.

The FIP/BIP pairs, comprising two of the six primers in the N2 or N4 set, could not discern spiked VTM from NTC (VTM only).

Table 7. Assessing the Performance of the New Primer Sets in Spiked VTM

	A434/A558											
	N2		N4		N7		N8		N2 FIP/BIP		N4 FIP/BIP	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	1.496	1.7	2.51087	2.520833	2.463918	2.382353	2.206897	2.144444	0.548736	0.55597	0.616601	0.604
B	1.479339	1.208633	0.666667	2.531646	2.45679	2.380952	1.936842	2.139535	0.601594	0.581673	0.808	0.839335
C	1.455285	1.409836	2.4375	1.82	0.686099	2.425	1.742574	2	0.639831	0.631356	0.7713	0.921053
D	0.937143	0.774359	0.751244	0.747368	0.719048	2.426667	1.873684	1.930233	0.694444	0.668122	0.767327	0.796296
E	0.666667	0.688679	0.659193	0.696682	0.659292	0.654378	0.727723	0.729064	0.621277	0.614407	1.608696	0.825
F	1.636364	0.918605	0.662281	0.662222	0.653333	0.641921	0.724138	0.75814	1.272059	0.605809	1.471074	1.023392
G	1.525862	0.698565	0.653333	0.654867	0.637931	0.646288	0.717822	0.727273	0.6	0.615721	1.191781	1.87619
H	1.163265	0.78125	0.729858	0.70852	0.700441	0.696581	0.747664	0.765	0.69469	0.704225	2.129032	1.913462

	A434/A558											
	N2		N4		N7		N8		N2 FIP/BIP		N4 FIP/BIP	
Copies	**Average**	**STDEV**	**Average**	**STDEV**	**Average**	**STDEV**	**Average**	**STDEV**	**Average**	**STDEV**	**Average**	**STDEV**
2500	1.598	0.14425	2.515851	0.007045	2.423135	0.057675	2.17567	0.04416	0.552353	0.005115	0.6103	0.00891
1250	1.343986	0.191418	1.599156	1.318739	2.418871	0.053625	2.038188	0.143325	0.591633	0.014086	0.823668	0.022157
625	1.43256	0.032137	2.12875	0.436638	1.555549	1.229589	1.871287	0.182027	0.635593	0.005992	0.846177	0.105891
312.5	0.855751	0.115106	0.749306	0.00274	1.572857	1.207469	1.901958	0.039986	0.681283	0.018613	0.781812	0.020485
	1.022457	0.443618	0.664763	0.016069	0.648857	0.008167	0.730693	0.014037	0.721545	0.269802	1.332689	0.390894
	1.022457	0.443618	0.664763	0.016069	0.648857	0.008167	0.730693	0.014037	0.721545	0.269802	1.332689	0.390894
NTC	1.022457	0.443618	0.664763	0.016069	0.648857	0.008167	0.730693	0.014037	0.721545	0.269802	1.332689	0.390894
Water	0.972258	0.270126	0.719189	0.015088	0.698511	0.002729	0.756332	0.012259	0.699458	0.006742	2.021247	0.152432

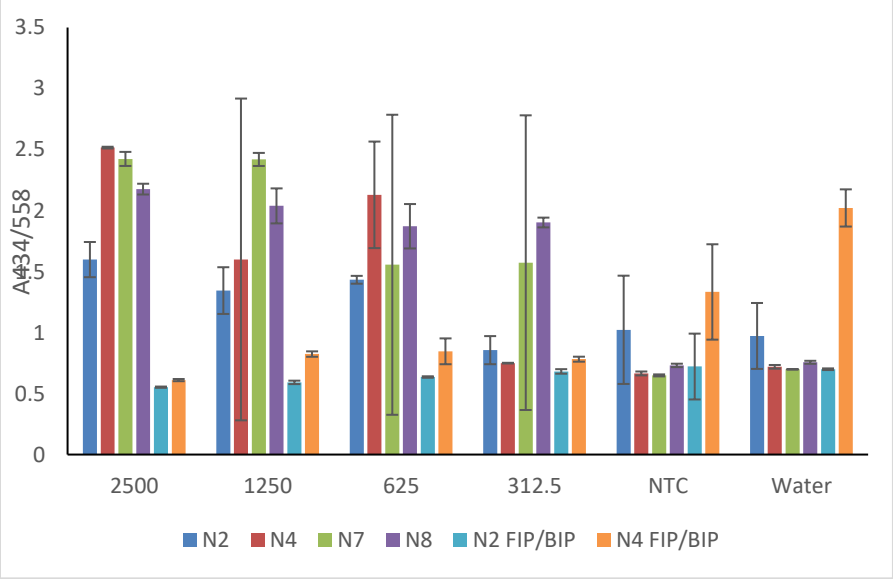

Figure 21. Assessing the Performance of the New Primer Sets in Spiked VTM Cont'd.

Thereafter, a triplicated repeat of the primer screen was performed, including N4, N7, and N8 primer sets (11-24-2020). See **Table 8** and **Figure 22.** The goal was to refine which primer set was superior from the previous screen, as the replicate number was too low to make a definitive judgement. Additionally, process control swabs from Microbiologics were included as separate samples for extraction, but due to the low copy number (1,000 RNA copies per swab), it was anticipated that detection would be intermittent.

Table 8. Assessing the Performance of New Sets of Primers, Triplicated

	A434/A558									Swab Samples			
	N4 Primer Set			N7 Primer Set			N8 Primer Set			N4	N7	N8	
	1	2	3	4	5	6	7	8	9	10	11	12	
A	2.06185567	2.23655914	2.195652174	2.170213	2.102041	2.083333333	2.488095	2.37931	2.344827586	1.896226	1.681034	0.659483	
B	0.701357466	0.659919028	2.104166667	2.139535	0.701357	0.695852535	2.227273	2.357143	2.168539326	0.692661	0.654167	0.643443	
C	2.244186047	0.719047619	2.30952381	0.735751	2.123596	0.754807692	0.741935	0.743316	2.075949367	0.694444	0.666667	0.683128	
D	0.75	0.753694581	0.768115942	0.743842	0.774194	0.743455497	0.785714	1.895522	1.826086957				
E	0.758454106	0.767567568	0.768421053	0.774011	0.774011	0.747368421	1.984375	2.277778	0.752475248				
F	0.696261682	0.713592233	0.726415094	0.701422	0.704433	0.694174757	0.708134	0.701493	0.703883495				
G	0.719806763	0.756097561	0.724637681	0.705882	0.701422	0.699074074	0.807229	0.695853	0.690909091				
H	0.747619048	0.753554502	0.766055046	0.739726	0.752252	0.735426009	0.733032	0.717949	0.734782609				

	A434/A558									Swab Samples			
Copies	N4 Primer Set			N7 Primer Set			N8 Primer Set			N4	N7	N8	
	Average	STDEV	TTEST	Average	STDEV	TTEST	Average	STDEV	TTEST				
2500	2.164688995	0.091374857	8.62634E-10	2.118529	0.045726	5.66356E-12	2.404078	0.074776	4.35841E-10	1.094444	1.000623	0.662018	Average
1250	1.15514772	0.822135637	0.103511025	1.178915	0.831926	0.086201278	2.250985	0.096512	2.44348E-09	0.694365	0.589287	0.019964	STDEV
625	1.757585825	0.899993589	0.009420848	1.204718	0.795828	0.068987747	1.187067	0.769795	0.076170774				
312.5	0.757270174	0.009572643	0.013072904	0.75383	0.017636	7.35863E-05	1.502441	0.621674	0.006400527				
156.25	0.764814242	0.005524546	0.004805734	0.76513	0.015382	9.79854E-06	1.671543	0.809342	0.008590122				
	0.722801836	0.019612232	0.219253033	0.701068	0.004153	0.374667057	0.717917	0.044171	0.313682266				
NTC	0.722801836	0.019612232	0.232859362	0.701068	0.004153	0.358611768	0.717917	0.044171	0.361193988				
Water	0.755742865	0.009410803	0.015556022	0.742468	0.008742	1.05701E-05	0.728588	0.009255	0.350279866				

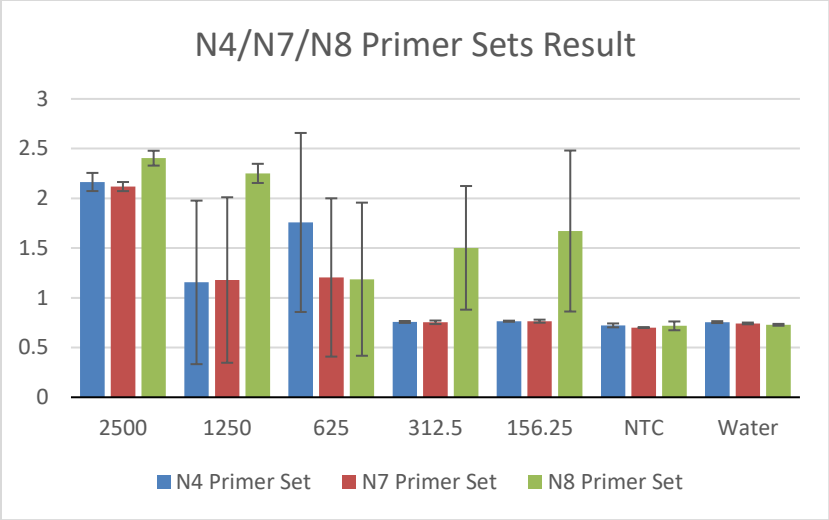

Figure 22. Assessing the Performance of New Sets of Primers, Triplicated

Here, N8 out-performed relative to N4 and N7, with a higher true positive rate at all tested concentrations. Tentatively, the LoD is estimated at 1,250-2,500 copies per reaction. As noted earlier, the LoD copies per reaction versus copies per mL is dependent on input sample volume, elution volume, and the portion of the eluate that is tested (See Equation 2). The copies per reaction can equal copies per mL if the input sample volume for extraction is 1 mL, the elution volume is 10-15 µL, and the entire elution is tested in the assay: alternatively, these numbers could be scaled equally and result in the same LoD. As such, the maximum LoD is 1,250-2,500 copies per mL, which is below the vast majority of positive sample titers. (11-24-2020)

As predicted, the process control swab results were spotty due to low input titers. As such, detection was random. It's possible that at arbitrarily long incubation times, detection rates would improve at the cost of potentially higher false positive rate.

To assess the LoD as per the CE and EUA guidelines, we performed a 22-replicated LoD study with the N8 primer set in spiked VTM. The LoD was determined as 2,500 copies per reaction at 20 minutes. Two false negatives appeared in the 1,250 copies per reaction concentration at 20 minutes. (11-25-2020) After this, we decided to pick N8 as our final primer candidate, and all following confirmation and optimizing experiments were done with the N8 primer set.

Up to this point, we have confirmed that the introduction of the purification process, whether it is a shortened or normal procedure, can reduce the false positive/false negative rate. Additionally, testing different sets of primers demonstrated that certain performance characteristics—amplification kinetics, false positive rate, and false negative rate—were highly dependent on the primer set used. In other words, specific sequences in each primer set and their binding affinity for target material versus non-target material (other primers, other nucleic acids, etc.) were driving these variations in performance. In the following section, I will cover our endeavor to automate the whole process after/while we were optimizing many aspects of our diagnostic tool for the SARS-CoV-2.

Table 9. LoD Study Result with the Best Candidate Primer Set

		A434/A5 58												
		N8 Primer Set - 22 Replicates												
Viral Copies		1	2	3	4	5	6	7	8	9	10	11	12	
2500	A	2.489362	2.563829787	2.628205128	2.57732	2.621053	2.604167	2.658228	2.516854	2.208791	2.642857	2.648352	0.608	
1250	B	2.552941	2.61038961	2.653333333	2.473684	2.594937	2.481013	0.693023	2.454545	2.532468	2.468354	2.255814	0.612245	NTC
625	C	0.639831	2.441558442	2.539473684	2.473684	0.721698	2.48	2.459459	2.368421	2.402439	2.345679	2.373494	0.64876	
312.5	D	0.652361	2.434210526	2.554054054	0.712195	2.12069	0.755102	2.308824	2.309859	2.447368	0.755102	2.183673	0.619433	
2500	E	2.43956	2.58974359	2.474358974	2.616438	2.4	2.573333	2.558442	2.035088	2.546667	2.6	2.53012	0.648069	
1250	F	2.417722	2.460526316	2.564102564	2.539474	2.545455	0.715686	2.608108	2.474359	2.4875	2.518987	1.439024	0.639831	Water
625	G	2.345679	2.46835443	2.46835443	2.425	0.714953	2.474359	2.4625	0.70283	2.519481	2.434211	2.45679	0.644068	
312.5	H	1.831683	0.660714286	2.426829268	0.656652	0.565385	2.536585	2.518987	0.679245	2.414634	2.5	1.772727	0.610039	

Table 10. LoD Study Result with the Best Candidate Primer Set Cont'd.

	N8 Primer Set				NTC			Water		
Viral Copies	Average	STDEV	TTEST	TP Rate	Average	STDEV	TTEST	Average	STDEV	TTEST
2500	2.523762	0.148991	5.00182E-19	100.00%	0.62211	0.018383		0.635501	0.017305	
1250	2.297338	0.569556	2.90165E-06	90.91%						
625	2.123557	0.691294	0.000132412	81.82%						
312.5	1.71804	0.828152	0.007806123	63.64%						

Table 11. LoD Study Result with the Best Candidate Primer Set Cont'd.

Viral Copies per reacton	N8 Primer Set 20-Replicate LoD																				NTC	Water
2500	2.49	2.56	2.63	2.58	2.62	2.60	2.66	2.52	2.21	2.64	2.65	2.44	2.59	2.47	2.62	2.40	2.57	2.56	2.04	2.55	2.60	2.53	0.61	0.61
1250	2.55	2.61	2.65	2.47	2.59	2.48	0.69	2.45	2.53	2.47	2.26	2.42	2.46	2.56	2.54	2.55	0.72	2.61	2.47	2.49	2.52	1.44	0.61	0.61
625	0.64	2.44	2.54	2.47	0.72	2.48	2.46	2.37	2.40	2.35	2.37	2.35	2.47	2.47	2.43	0.71	2.47	2.46	0.70	2.52	2.43	2.46	0.65	0.65
312.5	0.65	2.43	2.55	0.71	2.12	0.76	2.31	2.31	2.45	0.76	2.18	1.83	0.66	2.43	0.66	0.57	2.54	2.52	0.68	2.41	2.50	1.77	0.62	0.62

Endeavors to Automate the Diagnostic Kit

While optimization of the primer sets was ongoing, the complete automation of an RT-LAMP based rapid detection system for SARS CoV-2 was developed concurrently. This was achieved by utilizing an automated commercial nucleic acid extraction system. See **Figure 23.** This system will simultaneously extract viral RNA and perform colorimetric RT-LAMP for visual results from saliva and nasopharyngeal swab samples.

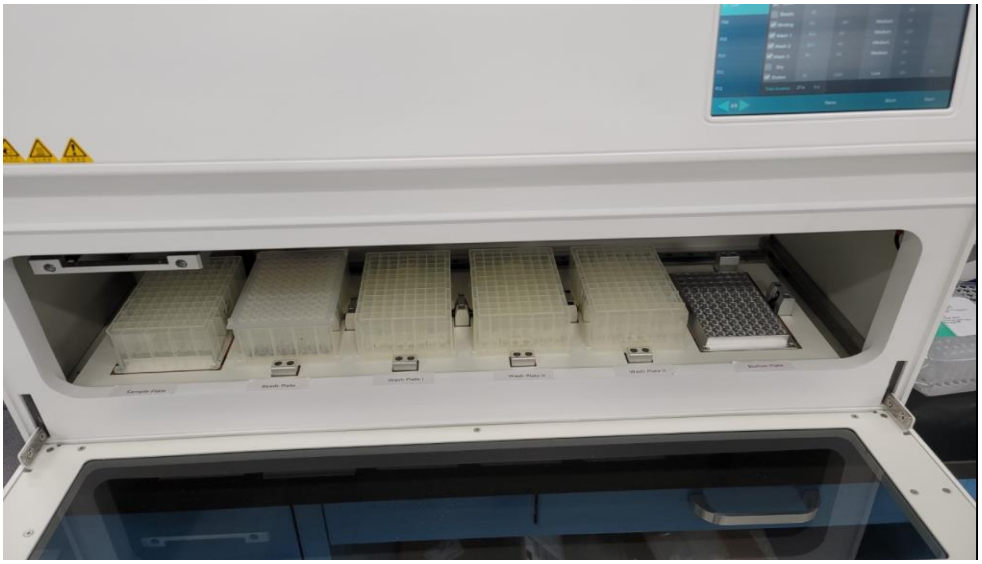

Figure 23. Picture of the automated SARS-CoV-2 Diagnostic Kit and machine

Once the N8 primer set was decided to be the final candidate for the primer, we ran multiple confirmation tests of which we were able to produce results successfully and reliably without any errors. See **Figure 24.** However, later it was later found that the slightest dimensionality change in the plasticwares could cause drastic change in the performance of the end-product.

Regardless, we were able to successfully reach the point where the kit could detect with excellent LoD (2500 viral copies) with reliable results. Of course, further optimization and confirmation may be required, but at the current state, the whole diagnostic can still provide tremendous relief to many places where reliable means of SARS-CoV-2 detection are not available.

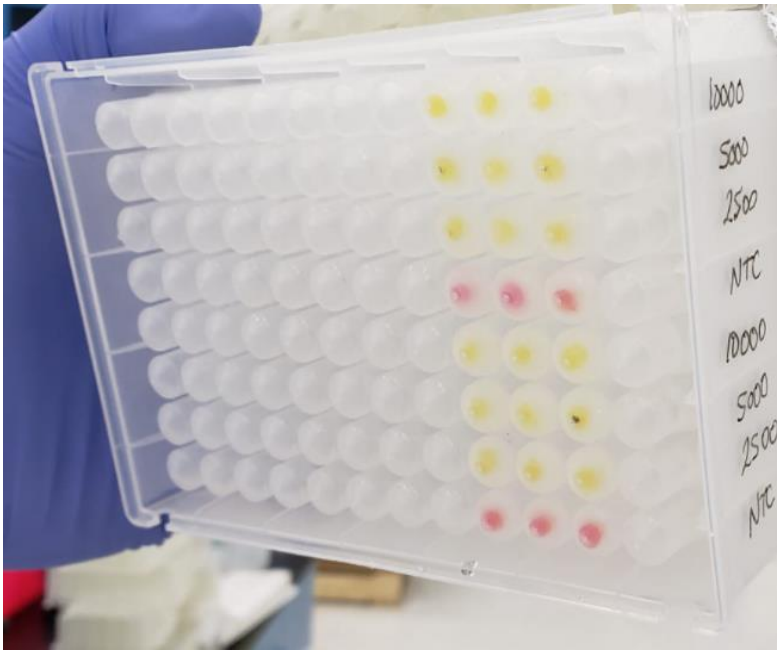

Figure 24. End result of the diagnostic using the automated kit

Conclusion and Discussion

Developing a diagnostic kit takes a lot of resources and commitment. Additionally, it requires a systematic approach to troubleshoot potential problems that can arise during the endeavor. Over the course of a year and a half, both myself and everyone who was involved in the project learned a lot about the process. I believe with this knowledge, we have developed a rigorous process for the development of other molecular diagnostics. Furthermore, this diagnostic kit design could be modified, with relative facility, for the detection of other pathogens. Such a system could systematically screen for dozens of respiratory pathogens from a single patient sample.

Here, we have developed a robust, facile, fully-automated diagnostic that can complete sample extraction, processing, and detection within as little as 35 minutes. This

system boasts the ability to generate results from ≤ 90 patient samples simultaneously from a single 96-deep-well plate, which are interpretable by technicians without additional instrumentation due to the colorimetric nature of the assay.

We believe this SARS-CoV-2 RT-LAMP diagnostic could provide an enormous benefit to the Global South, where qPCR capabilities may be restricted to major cities and hospital centers. Furthermore, we envision this diagnostic being used extensively in all countries in PoC and non-PoC facilities, with further testing as outlined by the FDA for such indications.

Chapter Five – Materials, Additional Figures, Tables, and Codes

In this section additional materials and figures used in the experiments will be covered.

References regarding 2019 SARS-CoV-2 Data[23–26]

Materials

Table 12. Organism List for the Cross Reactivity Study Confirmation

Organism	Catalog Number
Human coronavirus 229ERNAStrain: 229E	VR-740D
Betacoronavirus 1 RNAStrain: OC43Human coronavirus OC43	VR-1558D
Quantitative Synthetic Human coronavirus Strain HKU1 RNA	VR-3262SD
Quantitative Synthetic Human coronavirus Strain NL63 RNA	VR-3263SD
Quantitative Synthetic SARS-CoV [2003] RNA: nsp9, nsp11, N	VR-3280SD
Synthetic Middle Eastrespiratory syndromecoronavirus RNA	VR-3248SD
Human adenovirus 1Strain: Adenoid 71DNA	VR-1D
Human adenovirus 2Strain: Adenoid 6DNA	VR-846D
Human adenovirus 3Strain: GBDNA	VR-847D
Human adenovirus 4Strain: RI-67DNA	VR-1572D
Human adenovirus 5Strain: Adenoid 75DNA	VR-5D
Human adenovirus 6Strain: Tonsil 99DNA	VR-6D
Human adenovirus 7Strain: GomenDNA	VR-7D
Synthetic Humanmetapneumovirus RNA	VR-3250SD
Human parainfluenzavirus 2 RNAStrain: Greer	VR-92D
Human parainfluenzavirus 3 RNAStrain: C 243	VR-93D
Influenza A virus (H3N2)A/Aichi/2/68RNA	VR-1680D
Influenza B virus (BY)RNAB/Massachusetts/2/2012	VR-1813D
Enterovirus D68 RNAStrain: US/KY/14-18953	VR-1825D
Human enterovirus 71Strain: BrCrQuantitative RNA	VR-1775DQ

Human respiratorysyncytial virus RNAStrain: ATCC-2012-11	VR-1803D
Quantitative Genomic RNA from Human rhinovirus 1A	VR-1559DQ
HRV Type 14 ClonepWR3.26	VRMC-7
Quantitative Genomic RNA from Human rhinovirus 17	VR-1663DQ
Quantitative Genomic DNA from Chlamydophila pneumoniae	VR-1360DQ
Quantitative Genomic DNA from Haemophilus influenzae	51907DQ
Quantitative Genomic DNA from Legionella pneumophila	33152DQ
Quantitative Genomic DNA from Mycobacterium tuberculosis	25618DQ
Streptococcus pneumoniaespain23f-1genomic DNA	700669D-5
Streptococcus pyogenesT1genomic DNA	12344D-5
Bordetella pertussisgenomic DNA	BAA-589D-5
Mycoplasma pneumoniaegenomic DNA	29342D
Quantitative Genomic DNA from Candida albicans	MYA-2876DQ
Quantitative Genomic RNA from Human rhinovirus 1B	VR-1645DQ

Table 13. List of Items Used for the Optimization Process

Item	Vendor	Catalog Number
N8 Primer Set	IDT	N/A
WarmStart® Colorimetric LAMP 2X Master Mix (DNA & RNA)	NEB	M1800S
NEST Scientific 2.2mL 96-Well Deep Well Plate with V-Conical Bottom and Square Wells for KingFisher Extraction Platform, Sterile, 5/PK, 50/CS **OR** NEST Scientific NEST Scientific 0.5mL 96-Well Deep Well Plate with V-Conical Bottom and Square Wells for KingFisher Extraction Platform, Sterile, 5/PK, 50/CS	Stellar Scientfic	NST-503621 **OR** NST-500621
VWR Plate for PCR	VWR	89049-178
NEST Scientific 96 Tip Combs for KingFisher Flex Extraction Platform, Sterile, 5/PK, 50/CS	Stellar Scientific	NST-503361
Manual Extraction Kit	Commercial Vendor	TBD
Water for PCR nuclease-free	VWR	10220-396
15 mL Conical Tube	Multiple	Multiple
2.0 mL Skirted Tube with Tethered Cap, Graduated	VWR	490003-518

5mL, UniPlast Polystyrene Serological Pipettes, STERILE, Blue Striped, Individually Wrapped, 250/CS	Stellar Scientific	TC50-005
25mL Mini Serological Pipette, Sterile, Individually wrapped, 100/CS	Stellar Scientific	TC50-325
Powerpette® Pro Pipet Filler	VWR	75856-458
SILVERseal™ Sealer, Aluminium	VWR	82050-998
4-Inch Soft Roller	LI-COR	926-71000
AccuPlex™ SARS-CoV-2 Reference Material Kit	SeraCare	0505-0126

Codes

Code 1. Codes used for the inclusivity and cross reactivity studies

Codes

Recursive Program Execution

```
#!/usr/bin/perl

# Run the program of choice on all .fsa files in a directory tree


use strict;
use warnings;


die "usage: cr6.pl PROGRAM DIRECTORY\n" if scalar @ARGV < 2;
my ($prog, $dir) = @ARGV;

run_program_recursively($prog, $dir);


exit;


sub run_program_recursively {
    my($program, $directory) = @_;
    # Open the directory
    opendir(DIR, $directory) or die "Can't open directory $directory!";
    # Read the directory, ignoring special entries "." and ".."
    my @files = grep (!/^\.\.?$/, readdir(DIR));
    closedir(DIR);

    for my $file (@files) {
        # Get the full path to the file
        my $entry = "$directory/$file";

        # See if the directory entry is a regular file ending in .fsa
```

```
        if (-f $entry and $entry =~ /\.txt$/) {
            # give the outfile file the same name
            #  except change .fsa to .$program.out
            my $outfile = $entry;
            $outfile =~ s/\.txt$/.$program\.txt/;

            # run the program on the given file
            print "running program $program for $entry\n";
            system "./$program $entry";
        }

        # If the directory entry is a subdirectory
        elsif( -d $entry) {
            # Here is the recursive call to this subroutine
            run_program_recursively($program, $entry);
        }
    }
}
```

Code for Primer “FIP”

```
#!/usr/bin/perl

use strict;
use warnings;

# the argument list should contain the file name
die "usage: lorem_ipsum.pl filename\n" if scalar @ARGV < 1;

# get the filename from the argument list
my ($filename) = @ARGV;
# Open the file given as the first argument on the command line
open(INFILE, $filename) or die "Can't open $filename\n";


# variable declarations:
my @header = ();
my @sequence = ();
my $count=0;
my @outputsave=();
my $actseq = ' CCACTGCGTTCTCCATTCTGGTAAATGCACCCCGCATTACG';
my $actleng= length $actseq;
my $ratio=0;

read_fasta(\$filename, \@header, \@sequence, \$count);
#stat_fasta(\$filename, \@header, \@sequence, \$count);

my $outputfile = "FIP Match Result (80%+)";
      unless ( open(OUTPUTF, ">>$outputfile") ) {
            print "Cannot open file \"$outputfile\" to write to!!\n\n";
            exit;
      }
```

```
for (my $i = 0; $i < $count; $i++) {
		$ratio = ((length $sequence[$i])/$actleng);
		if ($ratio > 0.80){

		#print "This entry contains higher than 65% match!\n";
		#print "$header[$i]\n";
		print OUTPUTF "These entries contains higher than 80% match!
(FIP)\n";
		print OUTPUTF "$header[$i]\n";

		}

		}


exit;

#READ FASTA file
sub read_fasta{
	my ($filename, $header, $sequence, $count)= @_;
	my $n = -1;			# index of current sequence
	while (my $line = <INFILE>) {
		chomp $line;		# remove training \n from line
		if ($line =~ /^>/) {		# line starts with a ">"
			$n++;			# this starts a new header
			$$header[$n] = $line;		# save header line
			$$sequence[$n] = "";		# start a new (empty)
sequence
		}
		else {
		next if not @$header;		# ignore data before first header
		$$sequence[$n] .= $line	# append to end of current sequence
			}
	}
	$$count = $n+1;			# set count to the number of
sequences
	close INFILE;
	#print "Report for file $$filename\n\n";
	for (my $i = 0; $i < $$count; $i++) {
		#$$sequence[$i] =~ s/\s//g;
		#print "$$header[$i]\n";
		#print "$$sequence[$i]\n\n";
	#trimmer (\@sequence);
	}
	return $$count;
}

sub stat_fasta {

	my($filename, $header, $sequence, $count) = @_;
	for (my $i = 0; $i < $$count; $i++) {
```

```
                print "$$header[$i]\n";
                $$sequence[$i] =~ /\:(\d+)\-*?/;
                print "first number: $2\n\n";
                $$sequence[$i] =~ /\-(\d+)\s+/;
                print "second number: $2\n\n";

                }



}
```

Code for Primer “BIP”

```
#!/usr/bin/perl

use strict;
use warnings;

# the argument list should contain the file name
die "usage: lorem_ipsum.pl filename\n" if scalar @ARGV < 1;

# get the filename from the argument list
my ($filename) = @ARGV;
# Open the file given as the first argument on the command line
open(INFILE, $filename) or die "Can't open $filename\n";


# variable declarations:
my @header = ();
my @sequence = ();
my $count=0;
my @outputsave=();
my $actseq = ' CGCGATCAAAACAACGTCGGCCCTTGCCATGTTGAGTGAGA';
my $actleng= length $actseq;
my $ratio=0;

read_fasta(\$filename, \@header, \@sequence, \$count);
#stat_fasta(\$filename, \@header, \@sequence, \$count);

my $outputfile = "BIP Match Result (80%+)";
       unless ( open(OUTPUTF, ">>$outputfile") ) {
                print "Cannot open file \"$outputfile\" to write to!!\n\n";
                exit;
       }

for (my $i = 0; $i < $count; $i++) {
                $ratio = ((length $sequence[$i])/$actleng);
                if ($ratio > 0.80){

                #print "This entry contains higher than 65% match!\n";
                #print "$header[$i]\n";
```

```
                print OUTPUTF "These entries contains higher than 80% match!
(BIP)\n";
                print OUTPUTF "$header[$i]\n";

                }

                }


exit;

#READ FASTA file
sub read_fasta{
        my ($filename, $header, $sequence, $count)= @_;
        my $n = -1;                     # index of current sequence
        while (my $line = <INFILE>) {
                chomp $line;            # remove training \n from line
                if ($line =~ /^>/) {            # line starts with a ">"
                        $n++;                   # this starts a new header
                        $$header[$n] = $line;           # save header line
                        $$sequence[$n] = "";            # start a new (empty)
sequence
                }
                else {
                next if not @$header;           # ignore data before first header
                $$sequence[$n] .= $line      # append to end of current sequence
                        }
        }
        $$count = $n+1;                     # set count to the number of
sequences
        close INFILE;
        #print "Report for file $$filename\n\n";
        for (my $i = 0; $i < $$count; $i++) {
                #$$sequence[$i] =~ s/\s//g;
                #print "$$header[$i]\n";
                #print "$$sequence[$i]\n\n";
        #trimmer (\@sequence);
        }
        return $$count;
}

sub stat_fasta {

        my($filename, $header, $sequence, $count) = @_;
        for (my $i = 0; $i < $$count; $i++) {
                print "$$header[$i]\n";
                $$sequence[$i] =~ /\:(\d+)\-*?/;
                print "first number: $2\n\n";
                $$sequence[$i] =~ /\-(\d+)\s+/;
                print "second number: $2\n\n";

                }
```

```
}
```

Code for Primer “F3”

```
#!/usr/bin/perl

use strict;
use warnings;

# the argument list should contain the file name
die "usage: lorem_ipsum.pl filename\n" if scalar @ARGV < 1;

# get the filename from the argument list
my ($filename) = @ARGV;
# Open the file given as the first argument on the command line
open(INFILE, $filename) or die "Can't open $filename\n";


# variable declarations:
my @header = ();
my @sequence = ();
my $count=0;
my @outputsave=();
my $actseq = ' TGGACCCCAAAATCAGCG';
my $actleng= length $actseq;
my $ratio=0;

read_fasta(\$filename, \@header, \@sequence, \$count);
#stat_fasta(\$filename, \@header, \@sequence, \$count);

my $outputfile = "F3 Match Result (80%+)";
	unless ( open(OUTPUTF, ">>$outputfile") ) {
		print "Cannot open file \"$outputfile\" to write to!!\n\n";
		exit;
	}

for (my $i = 0; $i < $count; $i++) {
		$ratio = ((length $sequence[$i])/$actleng);
		if ($ratio > 0.80){

		#print "This entry contains higher than 65% match!\n";
		#print "$header[$i]\n";
		print OUTPUTF "These entries contains higher than 80% match!
(F3)\n";
		print OUTPUTF "$header[$i]\n";

		}

		}
```

```
exit;

#READ FASTA file
sub read_fasta{
	my ($filename, $header, $sequence, $count)= @_;
	my $n = -1;                    # index of current sequence
	while (my $line = <INFILE>) {
		chomp $line;            # remove training \n from line
		if ($line =~ /^>/) {          # line starts with a ">"
			$n++;                 # this starts a new header
			$$header[$n] = $line;         # save header line
			$$sequence[$n] = "";          # start a new (empty)
sequence
		}
		else {
		next if not @$header;         # ignore data before first header
		$$sequence[$n] .= $line     # append to end of current sequence
			}
	}
	$$count = $n+1;                   # set count to the number of
sequences
	close INFILE;
	#print "Report for file $$filename\n\n";
	for (my $i = 0; $i < $$count; $i++) {
		#$$sequence[$i] =~ s/\s//g;
		#print "$$header[$i]\n";
		#print "$$sequence[$i]\n\n";
	#trimmer (\@sequence);
	}
	return $$count;
}

sub stat_fasta {

	my($filename, $header, $sequence, $count) = @_;
	for (my $i = 0; $i < $$count; $i++) {
		print "$$header[$i]\n";
		$$sequence[$i] =~ /\:(\d+)\-*?/;
		print "first number: $2\n\n";
		$$sequence[$i] =~ /\-(\d+)\s+/;
		print "second number: $2\n\n";

		}



}
```

Code for Primer “B3”

```
#!/usr/bin/perl
```

```
use strict;
use warnings;

# the argument list should contain the file name
die "usage: lorem_ipsum.pl filename\n" if scalar @ARGV < 1;

# get the filename from the argument list
my ($filename) = @ARGV;
# Open the file given as the first argument on the command line
open(INFILE, $filename) or die "Can't open $filename\n";


# variable declarations:
my @header = ();
my @sequence = ();
my $count=0;
my @outputsave=();
my $actseq = ' GCCTTGTCCTCGAGGGAAT';
my $actleng= length $actseq;
my $ratio=0;

read_fasta(\$filename, \@header, \@sequence, \$count);
#stat_fasta(\$filename, \@header, \@sequence, \$count);

my $outputfile = "B3 Match Result (80%+)";
	unless ( open(OUTPUTF, ">>$outputfile") ) {
		print "Cannot open file \"$outputfile\" to write to!!\n\n";
		exit;
	}

for (my $i = 0; $i < $count; $i++) {
		$ratio = ((length $sequence[$i])/$actleng);
		if ($ratio > 0.80){

		#print "This entry contains higher than 65% match!\n";
		#print "$header[$i]\n";
		print OUTPUTF "These entries contains higher than 80% match!
(B3)\n";
		print OUTPUTF "$header[$i]\n";

		}

		}


exit;

#READ FASTA file
sub read_fasta{
	my ($filename, $header, $sequence, $count)= @_;
	my $n = -1;                    # index of current sequence
	while (my $line = <INFILE>) {
```

```
                chomp $line;            # remove training \n from line
                if ($line =~ /^>/) {          # line starts with a ">"
                        $n++;                 # this starts a new header
                        $$header[$n] = $line;         # save header line
                        $$sequence[$n] = "";          # start a new (empty)
sequence
                }
                else {
                next if not @$header;         # ignore data before first header
                $$sequence[$n] .= $line     # append to end of current sequence
                        }
        }
        $$count = $n+1;                     # set count to the number of
sequences
        close INFILE;
        #print "Report for file $$filename\n\n";
        for (my $i = 0; $i < $$count; $i++) {
                #$$sequence[$i] =~ s/\s//g;
                #print "$$header[$i]\n";
                #print "$$sequence[$i]\n\n";
        #trimmer (\@sequence);
        }
        return $$count;
}

sub stat_fasta {

        my($filename, $header, $sequence, $count) = @_;
        for (my $i = 0; $i < $$count; $i++) {
                print "$$header[$i]\n";
                $$sequence[$i] =~ /\:(\d+)\-*?/;
                print "first number: $2\n\n";
                $$sequence[$i] =~ /\-(\d+)\s+/;
                print "second number: $2\n\n";

                }



}
```

Code for Primer “LoopF”

```
#!/usr/bin/perl

use strict;
use warnings;

# the argument list should contain the file name
die "usage: lorem_ipsum.pl filename\n" if scalar @ARGV < 1;

# get the filename from the argument list
my ($filename) = @ARGV;
```

```
# Open the file given as the first argument on the command line
open(INFILE, $filename) or die "Can't open $filename\n";


# variable declarations:
my @header = ();
my @sequence = ();
my $count=0;
my @outputsave=();
my $actseq = ' TGAATCTGAGGGTCCACCAAA';
my $actleng= length $actseq;
my $ratio=0;

read_fasta(\$filename, \@header, \@sequence, \$count);
#stat_fasta(\$filename, \@header, \@sequence, \$count);

my $outputfile = "LoopF Match Result (80%+)";
	unless ( open(OUTPUTF, ">>$outputfile") ) {
		print "Cannot open file \"$outputfile\" to write to!!\n\n";
		exit;
	}

for (my $i = 0; $i < $count; $i++) {
		$ratio = ((length $sequence[$i])/$actleng);
		if ($ratio > 0.80){

		#print "This entry contains higher than 65% match!\n";
		#print "$header[$i]\n";
		print OUTPUTF "These entries contains higher than 80% match!
(LoopF)\n";
		print OUTPUTF "$header[$i]\n";

		}

		}


exit;

#READ FASTA file
sub read_fasta{
	my ($filename, $header, $sequence, $count)= @_;
	my $n = -1;			# index of current sequence
	while (my $line = <INFILE>) {
		chomp $line;		# remove training \n from line
		if ($line =~ /^>/) {		# line starts with a ">"
			$n++;			# this starts a new header
			$$header[$n] = $line;		# save header line
			$$sequence[$n] = "";		# start a new (empty)
sequence
		}
		else {
```

```
                next if not @$header;          # ignore data before first header
                $$sequence[$n] .= $line      # append to end of current sequence
                        }
        }
        $$count = $n+1;                      # set count to the number of
sequences
        close INFILE;
        #print "Report for file $$filename\n\n";
        for (my $i = 0; $i < $$count; $i++) {
                #$$sequence[$i] =~ s/\s//g;
                #print "$$header[$i]\n";
                #print "$$sequence[$i]\n\n";
        #trimmer (\@sequence);
        }
        return $$count;
}

sub stat_fasta {

        my($filename, $header, $sequence, $count) = @_;
        for (my $i = 0; $i < $$count; $i++) {
                print "$$header[$i]\n";
                $$sequence[$i] =~ /\:(\d+)\-*?/;
                print "first number: $2\n\n";
                $$sequence[$i] =~ /\-(\d+)\s+/;
                print "second number: $2\n\n";

                }



}
```

Code for Primer “LoopB”

```
#!/usr/bin/perl

use strict;
use warnings;

# the argument list should contain the file name
die "usage: lorem_ipsum.pl filename\n" if scalar @ARGV < 1;

# get the filename from the argument list
my ($filename) = @ARGV;
# Open the file given as the first argument on the command line
open(INFILE, $filename) or die "Can't open $filename\n";


# variable declarations:
my @header = ();
my @sequence = ();
my $count=0;
```

```
my @outputsave=();
my $actseq = ' GGTTTACCCAATAATACTGCGTCTT';
my $actleng= length $actseq;
my $ratio=0;

read_fasta(\$filename, \@header, \@sequence, \$count);
#stat_fasta(\$filename, \@header, \@sequence, \$count);

my $outputfile = "LoopB Match Result (80%+)";
	unless ( open(OUTPUTF, ">>$outputfile") ) {
		print "Cannot open file \"$outputfile\" to write to!!\n\n";
		exit;
	}

for (my $i = 0; $i < $count; $i++) {
		$ratio = ((length $sequence[$i])/$actleng);
		if ($ratio > 0.80){

		#print "This entry contains higher than 65% match!\n";
		#print "$header[$i]\n";
		print OUTPUTF "These entries contains higher than 80% match!
(LoopB)\n";
		print OUTPUTF "$header[$i]\n";

		}

		}


exit;

#READ FASTA file
sub read_fasta{
	my ($filename, $header, $sequence, $count)= @_;
	my $n = -1;			# index of current sequence
	while (my $line = <INFILE>) {
		chomp $line;		# remove training \n from line
		if ($line =~ /^>/) {		# line starts with a ">"
			$n++;			# this starts a new header
			$$header[$n] = $line;		# save header line
			$$sequence[$n] = "";		# start a new (empty)
sequence
		}
		else {
		next if not @$header;		# ignore data before first header
		$$sequence[$n] .= $line	# append to end of current sequence
			}
	}
	$$count = $n+1;			# set count to the number of
sequences
	close INFILE;
	#print "Report for file $$filename\n\n";
```

```
	for (my $i = 0; $i < $$count; $i++) {
		#$$sequence[$i] =~ s/\s//g;
		#print "$$header[$i]\n";
		#print "$$sequence[$i]\n\n";
	#trimmer (\@sequence);
	}
	return $$count;
}

sub stat_fasta {

	my($filename, $header, $sequence, $count) = @_;
	for (my $i = 0; $i < $$count; $i++) {
		print "$$header[$i]\n";
		$$sequence[$i] =~ /\:(\d+)\-*?/;
		print "first number: $2\n\n";
		$$sequence[$i] =~ /\-(\d+)\s+/;
		print "second number: $2\n\n";

		}



}
```

Additional Figures and Tables

Table 14. Primer Sequences

Sequences	
Sequence ID	5'-3' Nucleotide Sequence
Sequence ID 001: N2_F3	TGGACCCCAAAATCAGCG
Sequence ID 002: N2_FIP	CCACTGCGTTCTCCATTCTGGTAAATGCACCCCGCATTACG
Sequence ID 003: N2_LoopF	TGAATCTGAGGGTCCACCAAA
Sequence ID 004: N2_LoopB	GGTTTACCCAATAATACTGCGTCTT
Sequence ID 005: N2_BIP	CGCGATCAAAACAACGTCGGCCCTTGCCATGTTGAGTGAGA
Sequence ID 006: N2_B3	GCCTTGTCCTCGAGGGAAT
Sequence ID 007: N amplicon	TGGACCCCAAAATCAGCGAAATGCACCCCGCATTACGTTTGGTGGACCCTCAGATTCAACTGGCAGTAACCAGAATGGAGAACGCAGTGGGGCGCGATCAAAACAACGTCGGCCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACCG

	CTCTCACTCAACATGGCAAGGAAGACCTTAAATTCCCTCGAGGACAAGGC

REFERENCES

1. Miller, E. & Sikes, H. D. Addressing Barriers to the Development and Adoption of Rapid Diagnostic Tests in Global Health. *Nanobiomedicine (Rij)* **2**, 6 (2015).

2. Wang, Y. *et al.* Transmission, viral kinetics and clinical characteristics of the emergent SARS-CoV-2 Delta VOC in Guangzhou, China. *EClinicalMedicine* **40**, 101129 (2021).

3. Li, B. *et al. Viral infection and transmission in a large well-traced outbreak caused by the Delta SARS-CoV-2 variant*. 2021.07.07.21260122 https://www.medrxiv.org/content/10.1101/2021.07.07.21260122v1 (2021) doi:10.1101/2021.07.07.21260122.

4. Notomi, T. *et al.* Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res* **28**, e63 (2000).

5. Mori, Y. & Notomi, T. Loop-mediated isothermal amplification (LAMP): a rapid, accurate, and cost-effective diagnostic method for infectious diseases. *J Infect Chemother* **15**, 62–69 (2009).

6. Amaral, C. *et al.* A molecular test based on RT-LAMP for rapid, sensitive and inexpensive colorimetric detection of SARS-CoV-2 in clinical samples. *Sci Rep* **11**, 16430 (2021).

7. Tanner, N. A., Zhang, Y. & Evans, T. C. Visual detection of isothermal nucleic acid amplification using pH-sensitive dyes. *BioTechniques* **58**, 59–68 (2015).

8. Biolabs, N. E. SARS-CoV-2 Rapid Colorimetric LAMP Assay Kit | NEB. https://www.neb.com/products/e2019-sars-cov-2-rapid-colorimetric-lamp-assay-kit.

9. Biolabs, N. E. Loop-Mediated Isothermal Amplification | NEB. https://www.neb.com/applications/dna-amplification-pcr-and-qpcr/isothermal-amplification/loop-mediated-isothermal-amplification-lamp.

10. Thi, V. L. D. *et al.* A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples. *Science Translational Medicine* **12**, (2020).

11. Rapid point-of-care COVID-19 detection assay by RT-LAMP. *Harvard OTD* https://otd.harvard.edu/explore-innovation/technologies/rapid-point-of-care-covid-19-detection-assay-by-rt-lamp.

12. Dutta, N. K., Mazumdar, K. & Gordy, J. T. The Nucleocapsid Protein of SARS–CoV-2: a Target for Vaccine Development. *Journal of Virology* **94**, e00647-20.

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

14. LAMP primer designing software: PrimerExplorer. http://primerexplorer.jp/lampv5e/index.html.

15. Sender, R. *et al.* The total number and mass of SARS-CoV-2 virions. *PNAS* **118**, (2021).

16. Commissioner, O. of the. Emergency Use Authorization. *FDA* (2021).

17. Health, C. for D. and R. In Vitro Diagnostics EUAs. *FDA* (2021).

18. GISAID - Mission. https://www.gisaid.org/about-us/mission/.

19. Barh, D. *et al.* In Silico Models. *Animal Biotechnology* 385–404 (2014) doi:10.1016/B978-0-12-416002-6.00021-3.

20. Morozumi, M. *et al.* Simultaneous Detection of Pathogens in Clinical Samples from Patients with Community-Acquired Pneumonia by Real-Time PCR with Pathogen-Specific Molecular Beacon Probes. *J Clin Microbiol* **44**, 1440–1446 (2006).

21. Tyagi, S. & Kramer, F. R. Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol* **14**, 303–308 (1996).

22. Miyamoto, S., Sano, S., Takahashi, K. & Jikihara, T. Method for colorimetric detection of double-stranded nucleic acid using leuco triphenylmethane dyes. *Anal Biochem* **473**, 28–33 (2015).

23. Liu, P. *et al.* Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathog* **16**, (2020).

24. WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int.

25. Ritchie, H. *et al.* Coronavirus Pandemic (COVID-19). *Our World in Data* (2020).

26. Di Nardo, M. *et al.* A literature review of 2019 novel coronavirus (SARS-CoV2) infection in neonates and children. *Pediatric Research* 1–8 (2020) doi:10.1038/s41390-020-1065-5.

BIOGRAPHY

Yongjun Kwon received his Bachelor of Science from the Ohio State University in 2011. He was employed as a research associate in Caerus Discovery for three years and worked on multiple projects including the development of the point-of-care SARS-CoV 2 diagnostic from 2020 to 2021.