SARS-CoV-2 detection in multi-sample pools in a real pandemic scenario: a screening strategy of choice for active surveillance

Running title: SARS-CoV-2 detection by pooling samples

Andrés Marcos Castellaro1,2*, Pablo Velez3, Guillermo Giaj Merlera3, Juan Rondan Dueñas3, Félix Condat3, Jessica Gallardo3, Aylen Makhoul3, Camila Cinalli3, Lorenzo Rosales Cavagliera3, Guadalupe Di Cola4,5, Paola Sicilia6, Laura López7, Facultad de Ciencias Químicas UNC Group1, José Luis Bocco1,2, María Gabriela Barbás8, María Belén Pisano4,5, Viviana Ré4,5, Andrea Belaus3, Gonzalo Castro6.

1- Dpto. de Bioquímica Clínica, Facultad de Ciencias Químicas, Universidad Nacional de Córdoba, Córdoba, Argentina.  
2- Centro de Investigaciones en Bioquímica Clínica e Inmunología (CIBICI), CONICET, Universidad Nacional de Córdoba, Córdoba, Argentina.  
3- Unidad de Biología Molecular, Centro de Excelencia en Productos y Procesos de Córdoba (CEPROCOR), Córdoba, Argentina.  
5- Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Córdoba, Argentina.  
6- Laboratorio Central, Ministerio de Salud de la Provincia de Córdoba.  
7- Dirección de epidemiología, Ministerio de Salud de la Provincia de Córdoba.  
8- Secretaría de prevención y promoción de la salud, Ministerio de Salud de la Provincia de Córdoba.  

*These authors contributed equally to this work.

*Corresponding author: Andrés Marcos Castellaro, PhD.  
Dpto. de Bioquímica Clínica  
CIBICI-CONICET-UNC  
Facultad de Ciencias Químicas  
Universidad Nacional de Córdoba  
Address: Haya de la Torre y M. Allende. Ciudad Universitaria.  
Zip code: X5000HUA Córdoba. ARGENTINA  
Mail: andres.castellaro@unc.edu.ar / castellaroandresm@gmail.com
Abstract

Background The current COVID-19 pandemic has overloaded the diagnostic capacity of laboratories by the gold standard method rRT-PCR. This disease has a high spread rate and almost a quarter of infected individuals never develop symptoms. In this scenario, active surveillance is crucial to stop the virus propagation.

Methods Between July 2020 and April 2021, 11,580 oropharyngeal swab samples collected in closed and semi-closed institutions were processed for SARS-CoV-2 detection in pools, implementing this strategy for the first time in Córdoba, Argentina. Five-sample pools were constituted before nucleic acid extraction and amplification by rRT-PCR. Comparative analysis of cycle threshold (Ct) values from positive pools and individual samples along with a cost-benefit report of the whole performance of the results was performed.

Results From 2,314 5-sample pools tested, 158 were classified as positive (6.8%), 2,024 as negative (87.5%), and 132 were categorized as indeterminate (5.7%). The Ct value shift due to sample dilution showed an increase in Ct of 2.6±1.53 cycles for N gene and 2.6±1.78 for ORF1ab gene. Overall, 290 pools were disassembled and 1,450 swabs were analyzed individually. This strategy allowed correctly identifying 99.8% of the samples as positive (7.6%) or negative (92.2%), avoiding the execution of 7,806 rRT-PCR reactions which represents a cost saving of 67.5%.

Conclusion This study demonstrates the feasibility of pooling samples to increase the number of tests performed, helping to maximize molecular diagnostic resources and reducing the work overload of specialized personnel during active surveillance of the COVID-19 pandemic.

Keywords: SARS-CoV-2, pools, surveillance, pooling, COVID19

1. Introduction

The emergence of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) led humanity to an unprecedented pandemic with severe health consequences for the population worldwide. Diagnostic tests are essential due to their ability to detect and provide answers for pandemic management. However, they may be hampered because of the high demand that overwhelmed the healthcare system and the limited supply of reagents required for the setup of these tests. This aspect has been particularly worse in some geographic regions of the world, in low and middle-income countries as well. Thus, the context of SARS-CoV-2 pandemics motivated the design of alternative diagnostic strategies.

SARS-CoV-2 infection is characterized by the great diversity of signs and symptoms affecting patients. The most frequently associated with COVID-19 are fever, dry cough and generalized weakness though symptoms such as nausea, diarrhea, loss of smell, pharyngitis and enlarged tonsils have also been reported.[1]. However, about a quarter of infected people never develop symptoms (asymptomatic) and about half of the infected individuals do not manifest any symptom at the testing time (presymptomatic),[2, 3]. Both groups are of great concern as they contribute to the spread of the virus because they are not aware that they are infected. These groups are undetectable through passive surveillance but require instead active surveillance strategy based on massive testing methods. Hence, the need to increase testing capacity entails developing alternative strategies to optimize resources,
save time and reduce labor demand, thereby enhancing COVID-19 diagnosis which, in turn, is essential for evaluating the disease spread and for tracing the contacts of infected individuals,[4, 5]. Testing pooled samples by real-time reverse-transcriptase-polymerase chain reaction (rRT-PCR) could be a plausible way to deal with the huge demand for SARS-CoV-2 detection as long as it demonstrates optimum diagnostic performance. This strategy consists of mixing samples in a pool and then performing a single RNA extraction followed by a rRT-PCR assay. If the result is positive, then it must be identified which of the individual samples makes up the pool as positive by performing the RNA extraction and rRT-PCR test for each one of them. On the other side, if the rRT-PCR test is negative it is assumed that all individual samples composing the pool are also negative. Thus, according to this layout it is expected that fewer rRT-PCR tests and extraction reagents will be required instead of analyzing individual samples, leading to save reagents, time and labor demand. This pooling approach was previously developed for the analysis of several infectious diseases, e.g., malaria and HIV,[6, 7] and is currently used as a screening method in blood banks prior to transfusion,[8, 9]. Pooling nasopharyngeal swab samples for RNA virus detection, such as influenza has already been evaluated,[10]. In the case of COVID-19, the method was considered and analyzed as strategy by countries like Germany, Israel and the United States,[11-13] after the Food and Drug Administration (FDA) approved the emergency use of Quest SARS-CoV-2 rRT-PCR test for pooled samples on July, 2020,[14].

The aim of this work was the implementation of a multi-sample pool strategy during a specific period of COVID-19 pandemic, in order to decrease costs and response times, while increasing efficiency and testing capacity, thereby contributing to early patient assistance and control of disease spread.

2. Methods

2.1 Sample pooling

In July 2020, when the background SARS-CoV-2 community prevalence was between 3% and 5% in Córdoba, Argentina,[15] oropharyngeal swab samples in viral transport media (VTM) from testing centers were processed by pooling strategy following the protocol described by Ambrosi et al with minor modifications.[16]. Briefly, from five individual oropharyngeal swab samples, an aliquot of 60 µL of each was taken to create a pool, with a final volume of 300µL. Each pool of samples was processed for RNA extraction and subsequent rRT-PCR analysis. From August 2020 to April 2021, when the community viral prevalence exceeded 5%, oropharyngeal swab samples were collected in 234 closed and semi-closed institutions. Overall, 2314 pools were prepared for testing containing a mix of 5 individual samples each from initially 11580 patient samples.

2.2 Nucleic acid extraction and SARS-CoV-2 detection

RNA extraction of the pool was performed using Bioer MagaBio plus Virus DNA/RNA Purification Kit in addition to Bioer GenePure Pro fully automatic Nucleic Acid purification System (Bioer, Hangzhou, China) and EasyPure Viral DNA/RNA Kit (TransGene, Beijin, China), according to the manufacturer's instructions. A multiplex single step real time RT-PCR was carried out for amplification, using DisCoVery SARS-CoV-2 rRT-PCR Detection Kit, which is designed to detect two SARS-CoV-2 target genes: Open Reading Frame 1ab (ORF1ab) and nucleocapsid (N), along with the human ribonuclease P (RNAseP) gene as endogenous control (Safecare Biotech Hangzhou, China).

2.3 Classification criteria
Considering the cycle threshold (Ct) value for ORF1ab and N genes, the pools were classified as:

- Positive: pools with amplification of both genes with Ct ≤ 38.
- Negative: those without amplification of both genes, or with fluorescence signal only in one of them with Ct > 40.
- Suspected positive or indeterminate: pools that showed amplification in only one gene with Ct ≤ 40 and those that amplified both genes with Ct > 38.

To validate the previous results, all tests must amplify the endogenous RNAseP control gene.

2.4 Data processing

Results obtained by Real Time PCR (sample code and their respective Ct) were exported to plain text files from the AriaMx and Linegene software. From these files, the results were organized and interpreted with Perl scripts and conditional rules in Microsoft Excel and then converted into comma-delimited files. These data were incorporated into a database based on APACHE, MYSQL and PHPMYADMIN. A web interface was created to monitor the tests, evaluate coherence through PHP scripts, to visualize the data pools and their disassembly, and to carry out searches.

2.5 Statistical analysis

Differences in Ct values, mean, standard deviation and reduced number of tests were calculated using R version 4.0.5, 2021. The delta Ct value (ΔCt) was defined as the absolute change in Ct value when the pooled sample was tested compared to a positive sample that composed the pool, when it was tested individually. Therefore, a positive ΔCt value (i.e., an increase in Ct value of the pooled sample) represents the loss of rRT-PCR sensitivity as a consequence of individual sample dilution within a pool composed of 5 samples,[17]. ANOVA tests were done to compare groups, and a p-value of 0.05 or less was considered statistically significant.

3. Results

The study included 11570 samples analyzed under 2314 pools format, containing 5 samples each. We found 158 positive (6.8%) and 2024 negative pools (87.5%); 132 were classified as indeterminate (5.7%). Altogether, 290 pools were disassembled (158 classified as positive and 132 suspected to be positive) to analyze each one of the 1450 samples that composed them. As a result, the pooling strategy allowed to save 7806 tests, that is, 67.5% fewer tests were required for the screening, leading to a 3 times reduction in costs.

A hundred and five (105) of the positive pools contained a single positive sample, and 53 contained more than one positive sample. Results of Ct variation values comparing each pool and individual samples are shown in Figure 1. We observed that 5-sample pools containing one positive and four negative individual swabs samples, yielded higher Ct values than individual sample testing, exceeded by 2.6 cycles on average for both, N and ORF1ab genes (2.6±1.53 cycles for N gene and, 2.6±1.78 for ORF1ab gene). The differences between the Ct value of pooled and individual samples (ΔCt) are illustrated in Figure 2, showing the same mean value and comparable variability for both target genes. For most pools composed by more than one positive individual sample, the Ct values obtained were closer to the average of the individual Ct values, showing a pattern distribution in the middle area of the plot, and decreasing as the number of positive individual swabs increases within the same pool (Figure 3).
Regarding the 132 indeterminate pools, 19 contained only one positive sample (Table 1), 108 were constituted by all negative individual samples and five contained one suspected positive sample (inconclusive result according to the criteria provided in user manual of the rRT-PCR kit, so a new sample was requested). Analysis of the Ct values distribution of the indeterminate group, indicates that there are no significant differences (p<0.05) for both, N and ORF1ab genes, regardless of the 5-sample pools containing or not positive samples (Figure 4).

4. Conclusions and Discussion

Given the high transmission rates of SARS-CoV-2 and that almost a quarter of infected people are asymptomatic,[2, 3] one of the best strategies to control the virus spread is the early detection of those infected and their subsequent isolation. To achieve early detection of the infection, it is necessary to carry out active surveillance and massive testing. In this sense, different experimental or computational modeling tests have been carried out at the laboratory level to study the sensitivity and accuracy of the SARS-CoV2 detection strategies using different grouping techniques. However, at present there are very few studies reporting results of some of these strategies applied in a real setting, demonstrating their applicability on a large scale with patient samples.

This study demonstrates that the strategy based on the analysis of samples in a format of pools, composed of 5 samples each, is a reliable approach for an efficient screening of SARS-CoV-2 testing in a context of pandemics. In addition, it contributes to save costs of molecular diagnosis and labor demand of the health personnel in areas or institutions with low viral circulation. It can be useful to monitor the infection rate in closed or semi-closed establishments such as residential homes, police and military headquarters, prisons or hospitals, due to these places being populated by persons who are generally in close contact and/or with risk factors.[16].

It is important to highlight that optimum size pool must be determined according to the prevalence in the area under study.[4, 18]. Previous studies have analyzed different pool sizes, from 2- to 64-samples,[16, 19] and agreed that the number of samples in the pool is inversely proportional to the test sensitivity,[20, 21]. It has been reported that SARS-CoV-2 prevalence is a suitable criterion to define the most efficient pool size, defining 5-sample groups as appropriate when the prevalence rates are about 5%[4, 16].

According to our findings, as well as previous reports, pooling samples entails the loss of sensitivity and an increase of the Ct value of pools compared with the individual specimen due to the sample dilution, affecting especially detection of samples with low viral loads.[20-22]. The increase in the Ct value observed in our study (2.6 cycles on average) for both target genes was comparable to those reported in other studies of SARS-CoV-2 testing in pools constituted by 5 individual samples, where ΔCt ranges were between 2 and 3.4,[16, 20, 23, 24]. This variation results in more indeterminate tests than would be obtained in a single sample approach, mainly in those groups that contained individual positive samples with a Ct value close to 38. In both positive and indeterminate pools, the Ct value distribution for the N and the ORF1ab genes shows the same trend observed in individual samples: lower Ct values were detected for the N gene than for the ORF1ab target sequence.

Almost every suspected positive or indeterminate pool was classified in that way because only the N gene showed amplification, which could be due to a more sensitivity of this gene, contrasting with ORF1ab gene.[25]. Although in this indeterminate group we found only 19 pools (14.4%) that contained positive samples, the mean and range of the Ct values of both, N and ORF1ab genes, were similar to
those found for the remaining pools. Hence, it was not possible to establish a more rigorous cut-off Ct value to avoid disarming groups composed of samples that will ultimately be all negative. Therefore, it was mandatory to disassemble all indeterminate groups to avoid false-negative results. However, this approach allowed the correct identification of 99.8% of samples as positive (7.6%) or negative (92.2%), without the necessity to perform 7,806 tests, thus, saving 67.5% of costs and decreasing personnel labor demanding.

The pooling approach allows to deal with a variety of difficulties associated with a context of pandemics, most notably the limited availability of reagents, supplies, equipment and high labor demand of laboratory workers, as well as the high cost. However, the decision to implement this strategy must consider the total testing capacity and the disease prevalence in a specific geographic area. The limitation of this strategy to test the internal control of each sample, which is required to control the specimen quality, must also be underlined. So, false-negative results may occur if samples are improperly collected, transported, or handled. Hence, negative results obtained by pooled sampling do not preclude SARS-CoV-2 infection and should not be used as the only criteria for treatment or for other social management decisions.

The high sensitivity of rRT-PCR assays makes pool testing an efficient system that can be applied for resource optimization when positivity rate is low (e.g., 5% or lower), improving laboratory testing capacities without additional requirements in terms of equipment availability or qualified personnel,[19, 21, 26].

Our results show that the implementation of the strategy of pooling was able to save 67.5% of rRT-PCR reactions in a low viral circulation scenario. The approach of testing in pools was a positive experience that allow expanding sample processing capabilities allowing massive testing and early outbreak detections. Our findings could be taken into account as a strategy to be implemented in a new post-pandemic scenario, with an expected decrease of viral circulation (due to vaccine programs) to carry out large-scale testing in hospitals, care homes, schools, and other closed and semi-closed institutions.

**Ethics**

The Government of the Province of Córdoba through the Ministry of Health determined that the ethical review, approval and written informed consent are not required for the study of oropharyngeal swab samples obtained from human participants in the study “SARS-CoV-2 detection in multi-sample pools in a real pandemic scenario: a screening strategy of choice for active surveillance” in accordance with local legislation and institutional requirements. The Government of the Province of Córdoba wive the ethical review of the SARS-CoV-2 detection in multi-sample pools during 2020 and 2021 under strict confidentiality rules, based on the need for rapid surveillance and availability of methodologies for diagnosis of COVID-19.

**Acknowledgments**

The authors would like to acknowledge health care workers who take and processed the samples, all the staff of CEPROCOR laboratory and the team from Facultad de Ciencias Químicas, Universidad Nacional de Córdoba for their invaluable work and compromised attitude in favor of the community throughout the SARS-CoV-2 pandemic.

**Conflict of interests**

The authors declare no conflict of interest.
Author contributions

Conceptualization: AMC, MBP, VR and GC.
Methodology: AMC, PV, GGM, JRD, FC, JG, AM, CC, LRC, GDC, MBP, VR and AB.
Investigation: AMC, PV, GGM, JRD, FC, JG, AM, CC, LRC, GDC, FCQG, PS, MBP, VR, AB and GC.
Resources: MGB, LL and GC
Formal analysis: AMC, GDC, PV, JLB, MBP, VR and AB.
Writing - Original Draft: AMC and GDC.
Writing - Review & Editing: AMC, GDC, PV, AB, MBP, JLB and VR.
Visualization: AMC, MBP and VR.
Data Curation: AMC, PV, GDC, AB, LL, JLB, MGB and GC
Supervision: AMC, MBP, JLB and VR.
Funding acquisition: LL and MGB.

All authors contributed on the article and approved the submitted version.

Funding

This study was part of the active surveillance for SARS-CoV-2 dissemination carried out by the Ministry of Health of the Province of Córdoba, Argentina. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.


### Table 1. Ct values of positive individual samples from 19 indeterminate pools.

<table>
<thead>
<tr>
<th>N gene</th>
<th>ORF1ab gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled (Ct)</td>
<td>Individual Sample (Ct)</td>
</tr>
<tr>
<td>37.4</td>
<td>33.7</td>
</tr>
<tr>
<td>35.7</td>
<td>34.6</td>
</tr>
<tr>
<td>NA</td>
<td>37.9</td>
</tr>
<tr>
<td>32.9</td>
<td>29.6</td>
</tr>
<tr>
<td>35.2</td>
<td>33.0</td>
</tr>
<tr>
<td>37.0</td>
<td>33.5</td>
</tr>
<tr>
<td>38.9</td>
<td>34.4</td>
</tr>
<tr>
<td>36.3</td>
<td>32.7</td>
</tr>
<tr>
<td>35.5</td>
<td>31.2</td>
</tr>
<tr>
<td>37.3</td>
<td>32.1</td>
</tr>
<tr>
<td>36.9</td>
<td>34.7</td>
</tr>
<tr>
<td>34.6</td>
<td>31.1</td>
</tr>
<tr>
<td>NA</td>
<td>36.0</td>
</tr>
<tr>
<td>37.2</td>
<td>35.4</td>
</tr>
<tr>
<td>39.9</td>
<td>36.4</td>
</tr>
<tr>
<td>38.0</td>
<td>31.9</td>
</tr>
<tr>
<td>35.8</td>
<td>32.7</td>
</tr>
<tr>
<td>38.6</td>
<td>31.0</td>
</tr>
<tr>
<td>NA</td>
<td>36.5</td>
</tr>
</tbody>
</table>

NA: no amplification
Figure 1. **Pools with a single positive.** Scatter plots showing cycle threshold (Ct) values and Ct shifts detected by rRT-PCR in five-samples pools (squares), composed of four negative and one positive samples, with respect to the individual positive samples (circles) for N (top graph) and ORF1ab (bottom graph) genes. (2 column image)
Figure 2. **Ct value shift due to sample pooling.** Box plot showing the difference between the Ct values (ΔCt) of five-samples pools (composed of 1 positive and 4 negative samples) and the respective individual positive samples for N and ORF1ab target genes. (1.5 column image)
Figure 3. **Pools with two or more positives.** Scatter plots showing Ct values detected by rRT-PCR in five-samples pools (squares) and the individual positive samples (circles) from pools containing 2 (Pool # 1-27), 3 (Pool # 28-37), 4 (Pool # 38-47) or 5 (Pool # 39-53) positive samples. The analysis was performed for N (top graph) and ORF1ab (bottom graph) genes. (2 column image)

Figure 4. **Pools classified as indeterminate.** Box plot showing Ct values of N and ORF1ab genes from indeterminate pools which were finally classified as pools constituted by all negative samples (left), pools containing one positive sample (middle) and the individual positive samples after deconvolution (right). (2 column image)