



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A pooled testing strategy for identifying SARS-CoV-2 at low prevalence

Citation for published version:

Mutesa, L, Ndishimye, P, Butera, Y, Souopgui, J, Uwineza, A, Rutayisire, R, Ndoricimpaye, EL, Musoni, E, Rujeni, N, Nyatanyi, T, Ntagwabira, E, Semakula, M, Musanabaganwa, C, Nyamwasa, D, Ndashimye, M, Ujeneza, E, Mwikarago, IE, Mambo Muvunyi, C, Mazarati, JB, Nsanzimana, S, Turok, N & Ndifon, W 2020, 'A pooled testing strategy for identifying SARS-CoV-2 at low prevalence', *Nature*, vol. 589, pp. 276-280. <https://doi.org/10.1038/s41586-020-2885-5>

Digital Object Identifier (DOI):

[10.1038/s41586-020-2885-5](https://doi.org/10.1038/s41586-020-2885-5)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Nature

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



A pooled testing strategy for identifying SARS-CoV-2 at low prevalence

Leon Mutesa,^{1,2} Pacifique Ndishimye,^{3,2} Yvan Butera,² Jacob Souopgui,^{1,2,4} Annette Uwineza,² Robert Rutayisire,² Ella Larissa Nduricimpaye,² Emile Musoni,² Nadine Rujeni,² Thierry Nyatanyi,² Edouard Ntagwabira,² Muhammed Semakula,² Clarisse Musanabaganwa,² Daniel Nyamwasa,² Maurice Ndashimye,^{3,2} Eva Ujeneza,³ Ivan Emile Mwikarago,² Claude Mambo Muvunyi,² Jean Baptiste Mazarati,² Sabin Nsanzimana,² Neil Turok,^{3,5,6,*} and Wilfred Ndifon^{3,†}

¹*Centre for Human Genetics, College of Medicine and Health Sciences, University of Rwanda, Kigali, Rwanda*

²*Rwanda Joint Task Force COVID-19, Rwanda Biomedical Centre, Ministry of Health, Kigali, Rwanda*

³*African Institute for Mathematical Sciences, Kigali, Rwanda*

⁴*Department of Molecular Biology, Institute of Biology and Molecular Medicine, IBMM, Université Libre de Bruxelles, Gosselies, Belgium*

⁵*James Clerk Maxwell Building, University of Edinburgh, Edinburgh EH9 3FD, Scotland*

⁶*Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*

Suppressing SARS-CoV-2 will likely require the rapid identification and isolation of infected individuals on an ongoing basis. Reverse transcription polymerase chain reaction (RT-PCR) tests are accurate but costly, making regular testing of every individual expensive. The costs are a challenge for all countries and particularly for developing countries. Cost reductions can be achieved by pooling (or combining) subsamples and testing them in groups [1–7]. A balance must be struck between increasing the group size and retaining test sensitivity, since sample dilution increases the likelihood of false negatives for individuals with low viral load in the sampled region at the time of the test [8]. Likewise, minimising the number of tests to reduce costs must be balanced against minimising the time testing takes to reduce the spread of infection. Here we propose an algorithm for pooling subsamples based on the geometry of a hypercube that, at low prevalence, accurately identifies infected individuals in a small number of tests and rounds of testing. We discuss the optimal group size and explain why, given the highly infectious nature of the disease, largely parallel searches are preferred. We report proof of concept experiments in which a positive subsample was detected even when diluted 100-fold with negative subsamples (*cf.* 30-fold to 48-fold dilution in Refs. [9–11]). We quantify the loss of sensitivity due to dilution and discuss how it may be mitigated by frequent re-testing of groups, for example. With the use of these methods, the cost of mass testing could be reduced by a large factor which, furthermore, increases as the prevalence falls. Field trials of our approach are under way in Rwanda and South Africa. The use of group testing on a massive scale to closely and continually monitor infection in a population, along with rapid and effective isolation of infected people, provides a promising pathway to the longterm control of COVID-19.

* nturok@perimeterinstitute.ca

† wndifon@nexteinstein.org

I. INTRODUCTION

COVID-19 represents a major threat to global health. Rapidly identifying and isolating infected individuals is one of the most important available strategies for containing the virus. However, each diagnostic test for the SARS-CoV-2 virus costs 30-50 US dollars [12]. Therefore, testing every individual regularly, as may be required to eliminate the virus, is expensive. The costs are unaffordable for most low-income countries, which have limited available resources for massive SARS-CoV-2 testing. It is therefore important to ask: are there more efficient ways to find infected people?

The first step in testing, swab collection, is labour intensive but does not require expensive chemicals or equipment. It may therefore be feasible to collect swabs regularly from everyone. The next step involves RT-PCR machines [13]. These require expensive chemical reagents, currently in short supply, as well as skilled personnel. Reducing the cost requires that we minimise the total number of tests. The speed of testing is also a key concern because SARS-CoV-2 is so infectious. Each RT-PCR test takes several hours in the lab, time during which the virus can spread [14].

To find infected individuals, the naive approach is to test everyone, *i.e.*, to perform one test per person. However, at low prevalence it is far more efficient to pool (or combine) samples and test them together. This idea of group testing was proposed by Dorfman in 1943 [1]. At low viral prevalence p , Dorfman’s algorithm reduces the number of tests per person, required to find all infected individuals, to $\approx 2\sqrt{p}$ (see Appendix A). The algorithm we present is more efficient, requiring only $\approx ep \ln(1/p)$ tests per person at low p , where $e = 2.718\dots$ is Euler’s number. As an example, a survey of private residential households in England and Wales, released on 4 September, 2020 by the Office of National Statistics, estimated a prevalence of $p = 0.05\% (-.01\% + .02\%$ at 95%CI) [15–17]. For $p = 0.05\%$, Dorfman’s algorithm offers a 22-fold cost reduction while ours offers a 100-fold cost reduction. The main obstacle to achieving these large cost savings is the number of samples which can be pooled without compromising detection. In this paper, we present proof of concept experiments demonstrating that one positive sample in a pool of a hundred can still be reliably detected. We also discuss routes to increasing the pool size, as will be needed to gain the full benefits of group testing at low prevalence. *If larger pool sizes and the associated cost reductions can be achieved, group testing may provide an affordable pathway to the longterm control of SARS-CoV-2.*

In this paper, our focus is on population screening and not on protocols for use with at risk groups or in clinical settings. The prevalence is typically much higher among at risk groups or among individuals who present themselves for testing. For example, in the week ending 2nd September, 2020, 0.6% of the tests performed in hospitals in England were positive [18], suggesting a prevalence an order of magnitude greater than that in the wider population, quoted above. Second, in clinical settings the overriding concern should be to test the individual patient as quickly and accurately as possible. In most situations, that means performing an individual test. We are *not* suggesting group testing as a strategy for testing patients in clinics, specially those presenting with symptoms.

With this caveat, there are many potential applications. Sports teams where players and staff must be tested regularly are an example. A prominent rugby team in South Africa is trialling our method. Early results indicate cost savings of over an order of magnitude, with the successful detection of positives in groups of 81. Care homes provide a similar application. Pre-flight group tests can be used to clear passengers for commercial flights. Rwanda has adopted group testing as a national strategy. All air passengers are now required to undergo a test locally which has helped to revive tourism in the country. Group tests in University halls of residence, labs or departments could likewise enable safer in-person interaction. There is exciting potential for combining group testing with cheaper multiplex RT-PCR tests (see, *e.g.*, [19]) for the purposes of population

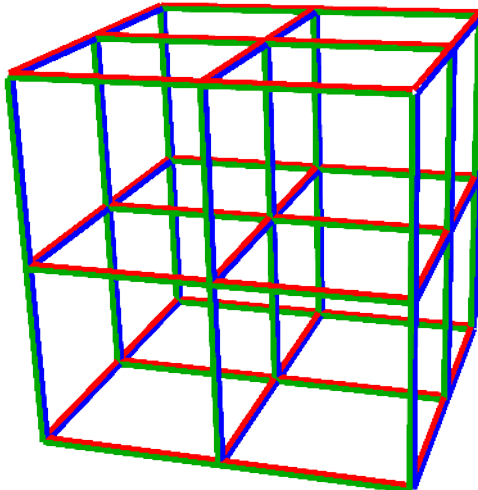


Figure 1: **Subsample pooling in the hypercube algorithm**, shown here for $D = L = 3$ and $N = 27 = 3^3$. Each vertex represents an individual. The hypercube is sliced into L slices, in each of the D principal directions. Samples from N/L individuals are pooled into a sample for each slice. For this example, the 3 sets of slices are shown in blue, red and green. If one infected individual is present, tests on each set of slices identify their coordinate in that direction. Hence only 9 tests uniquely identify them. As the viral prevalence falls, the optimal group size N and dimension D grow, and the efficiency gain rises.

screening. The combination may yield cost reductions of over two orders of magnitude.

Group testing algorithms generally require more than one round of testing. In Dorfman’s algorithm, a first round of group tests is followed by a second round in which each member of every positive group is tested. Our algorithm involves a similar first round of group tests, though with a larger group size. Positive groups proceed to a second round of “slice tests,” which usually suffices to identify all infected individuals without any need for individual tests. Occasionally one and, very rarely, more than one additional round of slice tests may be required. We compare our approach with other approaches in some detail (see Supplementary Information). There are adaptive algorithms requiring fewer tests but more rounds of testing, during which the viral prevalence can grow. We argue below that such searches are disfavoured at low prevalence. There are also non-adaptive algorithms requiring only one round of testing [4–7, 11]. While superficially attractive, they have some disadvantages compared to ours, for example a higher failure rate (see Supplementary Information, sections III and IX). In our approach, the first round of tests, performed on the groups, provides a valuable “sanity check” on the viral prevalence in the population being tested before the second round of more numerous “slice tests.” More generally, group tests will provide a highly efficient means of tracking the viral prevalence in a number of different populations in real time (see Supplementary Information, section VIII).

Group testing is most obviously effective when there are no infected individuals in the group: just one test suffices to show that no-one is infected. Our algorithm takes full advantage of this powerful result. In the first round of tests, subsamples from all group members are pooled and tested together. For our algorithm, the optimal group size is $N \approx 0.35/p$. The expected number of infected individuals in a group is 0.35 and a group will test negative over 70% of the time. Groups that test positive are passed on to the first round of slice tests, which we now describe.

II. WHEN ONE MEMBER OF A GROUP IS INFECTED

Consider the case where only one member of the group is infected. The idea behind our algorithm is geometrical: the group of individuals to be tested is represented by a set of N points on a cubic lattice in D dimensions, organised in the form of a hypercube with L points on a side (see Figure 1), so that

$$L^D = N. \quad (1)$$

Instead of directly testing the samples taken from every individual, we first divide each of them into D equal subsamples. These DN subsamples are recombined as follows. Slice the hypercube into L planar slices, perpendicular to one of the principal directions on the lattice. Form such a set of slices in each of the D principal directions and pool the L^{D-1} subsamples corresponding to each slice. Altogether, DL slices, each slice combining $N/L = L^{D-1}$ subsamples, are tested in parallel, in each round of slice tests. If there is one infected individual, then one slice out of the L slices, in each of the D directions, will yield a positive result. That slice indicates the coordinate of the infected individual, in the corresponding principal direction.

Therefore the number of tests required to uniquely identify the infected individual is

$$DL = DN^{1/D}, \quad (2)$$

where we used (1). Treating D as a continuous variable, the right hand side of (2) diverges at both small and large D , possessing a minimum at

$$D = \ln N, \quad (3)$$

corresponding to $L = e$ and a total of $e \ln N$ trials. In reality, D and L must be integers, but using $L = 3$ achieves almost the same efficiency (in the total number of trials, e is replaced with $3/\ln 3 \approx 2.73$, less than half a per cent larger, whereas using $L = 2$ or 4 gives $2/\ln 2 = 4/\ln 4 \approx 2.89$, more than 5 per cent larger). With no further constraint, finding one infected person in a population of a million, using $L = 3$, requires only 39 tests, performed in one round of testing. (To see this, note that $3^{13} > 10^6$, so a hypercube of side $L = 3$ in dimension $D = 13$ contains over a million points. A round of slice tests on this hypercube consists of $DL = 13 \times 3 = 39$ slice tests.)

III. PROOF OF CONCEPT

In real time PCR tests, the target RNA molecules are reverse transcribed into DNA which is replicated exponentially until it can be detected through fluorescence. In a perfectly efficient test, the number of DNA molecules doubles in every PCR cycle. The test is extremely sensitive: fewer than 10 molecules of viral RNA are sufficient [13]. A nasopharyngeal swab taken in the first 5 days of symptoms yields, on average, $\sim 2 \times 10^5$ viral RNA molecules per mL [20]. Asymptomatic patients appear to exhibit similar viral loads [21]. In the normal testing protocol, just $5\mu\text{L}$ of the solution, containing around 1000 RNA molecules, is included in the mix fed into the PCR machine. Samples taken earlier or later, or in younger patients whose antibodies have suppressed the virus, have less virus present. In practice, this is expected to be the most significant cause of testing error, taking the form of false negatives [17, 22, 23]. In pooled testing, positive subsamples are diluted with negative subsamples. Dilution by 100-fold, for example, means that only around 10 RNA molecules are likely to be present in the RT-PCR test. In principle, this should still be sufficient to yield a positive result.

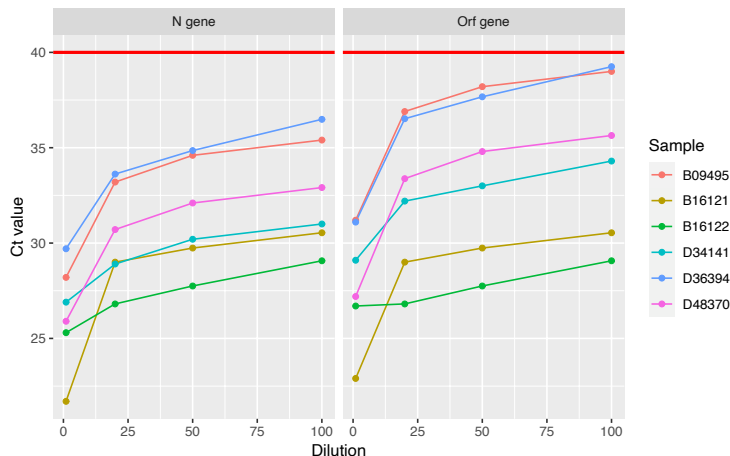


Figure 2: **Positive specimens are detected after a 100-fold dilution.** Each of six typical SARS-CoV-2-positive specimens was diluted through pooling with 19, 49, or 99 negative specimens. A Ct value (*i.e.*, the PCR cycle at which the fluorescence signal generated by a specimen exceeds the baseline signal) was determined for each pool through RT-PCR amplification of the N and $Orf1ab$ genes of SARS-CoV-2. For each gene, the Ct values are plotted against the dilution factor. The red horizontal lines indicate the Ct value (40) at or below which a specimen is considered positive. All Ct curves stay below the red lines even as the positive specimens are diluted 100-fold. (See Methods, Table 6)

As a proof of concept, using oropharyngeal swab specimens collected during COVID-19 surveillance in Rwanda, we investigated whether known positive specimens still test positive after they are diluted 20-, 50-, or 100-fold through pooling with negative specimens (see Methods). We used a RT-PCR test targeting the N and $Orf1ab$ genes of SARS-CoV-2, a combination used routinely for diagnostic screening for COVID-19 in Rwanda. The standard protocol is to consider a test positive if PCR amplification produces an above-background fluorescence signal for both target genes at a PCR cycle number, *i.e.*, a Ct value, ≤ 40 . Our key finding is that typical positive specimens can still be detected even after dilution by up to a 100-fold (see Figure 2). (For experiments demonstrating detection after 30-, 32-fold and 48-dilution, see Refs. [9–11]). As a consistency check, we determined the change in the Ct value (ΔCt) in going from a 50- to a 100-fold dilution. As noted above, a positive sample diluted 100-fold in principle requires one more cycle of PCR amplification than when diluted 50-fold, to achieve the same fluorescence signal, implying $\Delta Ct \approx 1.0$. Consistent with this expectation, we find $\Delta Ct \approx 1.0 \pm 0.15$ (1σ error) for the N gene and $\Delta Ct \approx 1.1 \pm 0.14$ for the $Orf1ab$ gene. The changes in Ct values for other dilutions are also consistent with this interpretation.

We estimated post-dilution sensitivities by combining (technically, convolving) the probability distribution for pre-dilution Ct values for positive samples (see Methods, Table 7) with the probability distribution for the increase ΔCt as inferred above. Treating both distributions as Gaussian, the distribution of post-dilution Ct values is also Gaussian, with mean given by the sum of the means and variance given by the sum of the variances. In this way, we estimated that a 40-cycle PCR test targeting the N -gene (resp. $Orf1ab$ gene) has post 20-, 50-, and 100-fold dilution sensitivities of 91% (resp. 85%), 88% (resp. 81%), and 85% (resp. 77%), respectively. We have confirmed these estimates using (i) an independent sample of Ct values for 107 positive specimens collected in Rwanda, in tests targeting the same two genes, and (ii) the published data set of 26 positive specimens identified in a recent study of pooled testing for SARS-CoV-2 in Israel [10], targeting a single gene (the E gene). All three data sets gave broadly consistent results.

The positive samples most likely to be missed due to dilution are those with the highest Ct

values before dilution, *i.e.*, those with the lowest viral load. The individuals concerned are likely to be the least infectious [24, 25]. Conversely the individuals producing samples with the *lowest* Ct values, least affected by sample dilution, whether symptomatic or asymptomatic, are the most important to detect since they are likely to be the most infectious. Nevertheless, it is important to consider ways in which the loss in sensitivity due to dilution might be mitigated. The most obvious is to re-test sufficiently often (say, every 3 days) to ensure a test occurs in the period of highest viral abundance, for any infected individual. Group tests involve the greatest degree of dilution in our method, but they are also the cheapest testing stage to repeat frequently and thereby to mitigate sensitivity loss. Second, the number of PCR cycles could be raised to 44, the maximum used in Ref. [10]. According to our analysis, this would raise the sensitivity after 100-fold dilution to 95% (resp. 90%) for the N (resp. $Orf1ab$) genes. Third, the volume of the sample used in the RT-PCR test can be increased from $5\mu\text{L}$ to $10\mu\text{L}$ (this is done in the lab we are working with in South Africa). Fourth, the viral concentration in the pooled sample might be raised by physical or chemical means such as ultra-centrifugation or precipitation. Finally, PCR machines might be re-engineered to allow larger sample volumes to be tested. All of these possibilities are worth exploring.

IV. WHEN MORE THAN ONE MEMBER OF A GROUP IS INFECTED

So far, we have assumed only one member of the group is infected. But what if 2, 3 or more members are infected? In normal circumstances, all we will have is an estimate of the prevalence p of the virus in the population from which the group has been drawn, *i.e.*, the probability that a person chosen at random is infected. A feature of group testing is that the first round of group tests, relatively few in number, allows us to conveniently update our knowledge of p , before any infected individuals have been identified (see Supplementary Information, section VIII).

Given p , the probability that k members of a group of size N are infected is described by a Poisson distribution with mean $\lambda = pN$. For λ well below unity, the probability falls rapidly with increasing k . At very low p , the optimal N is very large, so $D = \log_L N \gg 1$. The advantages of the hypercube algorithm are particularly clear in this limit. Therefore, we describe this limit first before discussing realistic values of D .

The first round of slice tests, as described in Section II, yield, for $L = 3$, a set of triples of zeros and ones, *i.e.*, $\{1, 0, 0\}$, $\{1, 1, 0\}$ or $\{1, 1, 1\}$ and permutations thereof, for every principal direction of the lattice. Let σ be the sum of the three values (so $\sigma = 1, 2$ or 3) and d_σ the number of directions in which the value σ occurs, so $d_1 + d_2 + d_3 = D$. For $D \gg 1$, the number of infected individuals k may be accurately inferred from the observed values of d_σ , even before any infected individuals are identified. Knowing k , we then find all infected individuals as follows: (i) if $k = 1$, then $d_1 = D$. Each positive slice indicates the coordinate of the infected individual in that direction. Hence, the infected individual is identified in one round of slice tests; (ii) if $k = 2$, $d_2 > 0$ but $d_3 = 0$. If $d_2 = 1$, the two infected individuals are immediately identified. If $d_2 > 1$, choose one of the directions with $\sigma = 2$, and treat the two positive slices as smaller hypercubes, each containing one infected individual. A further round of slice tests identifies one and the other is found by elimination; (iii) if $k = 3$ then, at large D , at least one direction has $\sigma = 3$. Choose one such direction and treat two of the positive slices as smaller hypercubes, each containing one infected individual. A slice test on each identifies two infected individuals and the third is found by elimination; if $k > 3$, the number of rounds of slice tests required to identify all infected individuals is slightly larger than k . However, for the optimal value of group size, the probability to have $k > 3$ infected members is negligibly small.

Hence, in the large D limit, to a good approximation k rounds of slice tests are required to

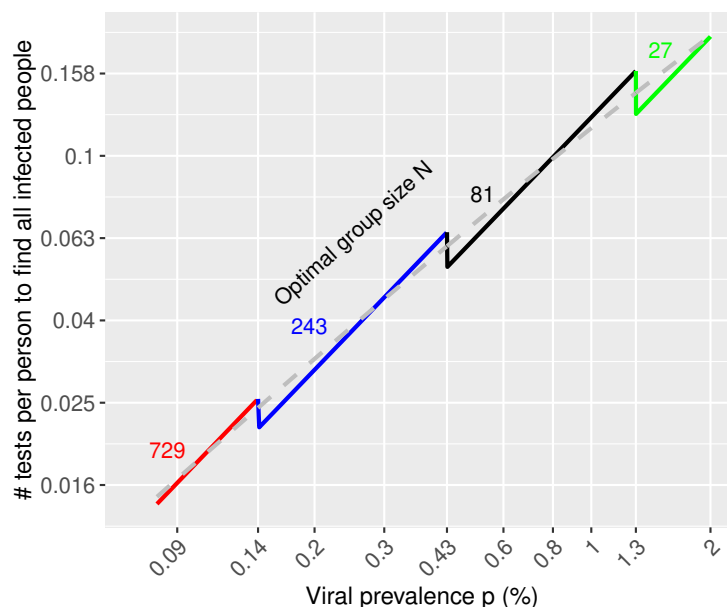


Figure 3: **Expected number of tests per person to find all infected individuals using the hypercube algorithm** (log-log plot). The dashed grey curve shows $ep \ln(0.734/p)$, the result obtained in the large D approximation, where the optimal group size $N \approx 0.35/p$. The coloured curves show the results obtained from a detailed analysis when the group size $N = 3^D$ with D an integer. Where $0.35/p$ is an exact power of 3, as at the left end of each coloured curve, optimal performance is attained. As p is increased, a growing fraction of sites in the 3^D hypercube are left empty, until the next exact power of 3 (see Supplementary Information)

identify k infected individuals. In the Supplementary Information, we show that, at low prevalence p , assuming Poisson statistics, the expected number of tests per person $\langle T \rangle / N$ required to identify all infected individuals is minimised for $N \approx 0.350/p$. At this optimal group size, $\langle T \rangle / N \approx ep \ln(0.734/p)$, shown as the dashed grey line in Figure 3. The reciprocal of this number is the efficiency gain relative to naive testing.

For practical applications, we are interested in the efficiency of the algorithm at modest values of D such as 3, 4 or 5. This requires a more intricate analysis, the details of which we relegate to the Supplementary Information. However, some simple and general statements may be made. First, when all directions yield $\sigma = 1$, only one individual is infected and they are immediately and uniquely identified. This is the most probable outcome of the first round of slice tests. Second, when $\sigma > 1$ in only one direction then two (or three) infected individuals are uniquely identified without further tests. If $\sigma > 1$ in more than one direction, a second round of slice tests is needed. We can eliminate any slice which tested negative in the first round and thus work with a smaller hypercube. We make only one approximation, namely we assume the infected samples are rare in the hypercube. They may then be treated as independent, randomly chosen points. Within this approximation, we compute the probabilities through to the second round of slice tests. Strikingly, we find that *the hypercube algorithm remains highly efficient at modest values of D* . For example, for $\lambda = 0.35$ and $D = 4$, in 93.3% of cases one round of slice tests suffices to identify all infected individuals. For the remaining 6.7% of cases, one more round suffices in all but 0.01% of cases, a very low theoretical failure rate (which does not include experimental errors). The expected total number of tests per person, for $D = 3, 4$ and 5, are plotted in Figure 3. Where $0.35/p$ is an exact power of 3, as at the left end of each coloured curve, optimal performance is attained. As p is

increased, an increasing fraction of sites in the 3^D hypercube are left empty until the next exact power is reached. Nevertheless, pooling still results in a high efficiency gain. As Figure 3 shows, the large D approximation provides a surprisingly good (and very convenient) fit to the low D results. (See Supplementary Information, sections IV-VII for further details).

V. LARGELY PARALLEL SEARCHES ARE PREFERRED

Some search methods require fewer tests but more rounds of testing. A binary search [2, 3], for example, finds one infected individual among N in $\sim \log_2 N$ tests, a factor of $e \ln 2 \approx 1.88$ less tests than needed by our hypercube algorithm at large D . However, the tests must be performed serially, requiring $\sim \log_2(1/p)$ rounds of testing. For $p = 0.4\%$ (or 0.15%) a binary search takes 8 (or over 9) rounds of testing whereas a hypercube search takes typically 2 and occasionally 3 (in both cases). For a highly infectious disease like COVID-19, saving time is crucial because infected individuals who are still at large can infect others. The doubling time for SARS-CoV-2 has been estimated at $\tau_2 \approx 2$ days [14, 26]. If each testing round takes τ days, the prevalence grows by $\sim (1/p)^{\tau/\tau_2}$ during a binary search. If this growth factor exceeds $e \ln 2$, a binary search will do worse than a hypercube search. Assuming $\tau \approx 1/3$ day, we find that for $p < 1\%$, the hypercube search is preferred. Another advantage of the hypercube search is that it includes many consistency checks. For example, finding $\sigma = 1$ in one direction and $\sigma = 0$ in the others indicates a testing error. In contrast, a binary search relies on repeated testing of the positive sample, so that a single false negative can prematurely terminate the search.

VI. CONCLUSIONS

The hypercube algorithm offers an attractive compromise between minimising the total number of tests to reduce costs and maximising the speed of testing to reduce the spread of the virus. We have demonstrated its viability for group sizes up to 100, showing that cost savings of a factor of nearly 20 can, in principle, already be achieved. We have quantified the loss of sensitivity due to dilution and discussed how it may be mitigated, for example through repeated group testing. These strategies could open the way to the use of larger pool sizes, bringing even greater cost savings at low prevalence. The most striking aspect of our approach is how rapidly the cost of testing a population can fall, pooled test sensitivity permitting, with decreasing prevalence. This should incentivise decision-makers to act firmly to drive the prevalence down through mass testing, contact tracing and isolation. Although mass testing is initially costly, it provides a clear route to reducing the prevalence. As the viral prevalence is reduced, keeping it low will, with the use of pooled testing, become progressively more affordable.

Acknowledgments: We thank the Rwanda Ministry of Health through RBC for stimulating discussions and correspondence. Kendrick Smith and Corinne Squire provided valuable encouragement and helpful references. We thank Andrew Jackson for discussions. Research at AIMS is supported in part by the Carnegie Corporation of New York and by the Government of Canada through the International Development Research Centre and Global Affairs Canada. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities.

Contributions: L.M., S.N. coordinated the experiments; P.N., T.N., E.N., M.S., C.M., D.N., M.N. contributed to patient recruitment and data collection from the community; P.N., Y.B., J.S., A.U., R.R., E.L.N., E.M., E.M., N.R., I.E.M., J.B.M., C.M.M. and E.U. contributed to laboratory

RT-PCR test validation, data analysis and interpretation. W.N. and N.T. contributed to the theory.

-
- [1] R. Dorfman, “The Detection of Defective Members of Large Populations,” *Ann. Math. Stat.* **14** (1943) 436.
- [2] F.K. Hwang, “A method for detecting all defective members in a population by group testing,” *J. Amer. Stat. Assoc.* **67** (1972) 605.
- [3] A. Allemann, “An efficient algorithm for combinatorial group testing,” in Aydinian H., Cicalese F., Deppe C. (eds), *Information Theory, Combinatorics, and Search Theory*, Lecture Notes in Computer Science, **7777**, Springer, 2013 .
- [4] M. Aldridge, O. Johnson, and J. Scarlett, “Group testing: an information theory perspective,” *Foundations and Trends in Communications and Information Theory*, to appear , [arXiv:1902.06002](https://arxiv.org/abs/1902.06002).
- [5] S. Ghosh, R. Agarwal, M. A. Rehan, S. Pathak, P. Agrawal, Y. Gupta, S. Consul, N. Gupta, R. Goyal, A. Rajwade, and M. Gopalkrishnan, “A compressed sensing approach to group-testing for covid-19 detection,” [arXiv:2005.07895](https://arxiv.org/abs/2005.07895) [q-bio.QM].
- [6] B. W. Heng and J. Scarlett, “Non-Adaptive Group Testing in the Linear Regime: Strong Converse and Approximate Recovery,” [arXiv:2006.01325](https://arxiv.org/abs/2006.01325) [cs.IT].
- [7] A. Z. Broder and R. Kumar, “A note on double pooling tests (preliminary version),” [arXiv:2004.01684](https://arxiv.org/abs/2004.01684) [cd.DM].
- [8] Ingrid Arevalo-Rodriguez *et al.*, “False negative results of initial RT-PCR assays for COVID-19: a systematic review,” *medRxiv preprint* (August 13, 2020) . <https://doi.org/10.1101/2020.04.16.20066787>.
- [9] S. Lohse *et al.*, “Pooling of Samples for testing for SARS-CoV-2 in asymptomatic people,” *Lancet Infectious Diseases* (April 28, 2020) . [https://doi.org/10.1016/S1473-3099\(20\)30362-5](https://doi.org/10.1016/S1473-3099(20)30362-5).
- [10] I. Yelin *et al.*, “Evaluation of COVID-19 RT-qPCR test in multi-sample pools,” *Clinical Infectious Diseases* (May 02, 2020) . <https://doi.org/10.1093/cid/ciaa531>.
- [11] Noam Shental *et al.* , “Efficient high throughput SARS-CoV-2 testing to detect asymptomatic carriers,” *Science Advances* **6** (2020) eabc5961.
- [12] Medicare, “Medicare administrative contractor (MAC) COVID-19 test pricing,” <https://www.cms.gov/files/document/mac-covid-19-test-pricing.pdf>.
- [13] V.M. Corman, O. Landt, M. Kaiser *et al.* , “Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR,” *Euro. Surveill.* **25** (2020) 2000045.
- [14] S. Sanche, Y.T. Lin, C. Xu, E. Romero-Severson, M. Hengartner, R. Ke, “High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus,” *Emerg. Infect. Dis.* **26** (July, 2020) . <https://doi.org/10.3201/eid2607.200282>.
- [15] Office of National Statistics, “COVID-19 Infection Survey (Pilot): England and Wales, 4 September (2020),” . <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/englandandwales4september2020>.
- [16] K. B. Pouwels, T. House, *et al.*, “Community prevalence of SARS-CoV-2 in England: Results from the ONS Coronavirus Infection Survey Pilot,” . <https://www.medrxiv.org/content/early/2020/07/07/2020.07.06.20147348.full.pdf>.
- [17] Office of National Statistics, “COVID-19 Infection Survey (Pilot): methods and further information,” . <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/covid19infectionsurveypilotmethodsandfurtherinformation#test-sensitivity-and-specificity>.
- [18] Office of National Statistics, “Weekly Statistics for NHS Test and Trace (England) and coronavirus testing (UK): 27 August to 2 September,” . <https://www.gov.uk/government/publications/nhs-test-and-trace-england-and-coronavirus-testing-uk-statistics-27-august-to-2-september-2020/weekly-statistics-for-nhs-test-and-trace-england-and-coronavirus-testing-uk-27-august-to-2-september-2020/annex-table-1>.

- [19] Martin A.M. Reijns *et al.*, “A sensitive and affordable multiplex RT-qPCR assay for SARS-CoV-2 detection,” *medRxiv preprint* (July 16, 2020) . <https://doi.org/10.1101/2020.07.14.20154005>.
- [20] R. Wölfel, V.M. Corman, W. Guggemos *et al.* , “Virological assessment of hospitalized patients with COVID-2019,” *Nature* . <https://doi.org/10.1038/s41586-020-2196-x>.
- [21] Q.X. Long *et al.* , “Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections,” *Nature Medicine* **26** (2020) 1200.
- [22] J. Watson, P. F. Whiting, and J. E. Brush, “Interpreting a covid-19 test result,” *BMJ* **369:m** (2020) 1808. <https://www.bmj.com/content/bmj/369/bmj.m1808.full.pdf>.
- [23] L. M. Kucirka, S. A. Lauer, O. Laeyendecker, and J. Lessler, “Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction-Based SARS-CoV-2 by Time Since Exposure,” *Ann. Intern. Med.* *173* (4) (2020) 262 (2020) . <https://pubmed.ncbi.nlm.nih.gov/32422057>.
- [24] Rita Jaafar *et al.*, “Correlation between 3790 qPCR positives samples and positive cell cultures including 1941 SARS-CoV-2 isolates,” *Clinical Infectious Diseases* (September 28, 2020) . <https://doi.org/10.1093/cid/ciaa1491>.
- [25] R. F. Service, “One number could help reveal how infectious a COVID-19 patient is. Should test results include it?,” . <https://www.sciencemag.org/news/2020/09/one-number-could-help-reveal-how-infectious-covid-19-patient-should-test-results>.
- [26] N. Afshordi, B. Holder, M. Bahrami, and D. Lichtblau, “Diverse local epidemics reveal the distinct effects of population density, demographics, climate, depletion of susceptibles, and intervention in the first wave of covid-19 in the united states,” [arXiv:2007.00159](https://arxiv.org/abs/2007.00159) [q-bio.PE].
- [27] DAAN Gene Company Ltd., “DAAN - RT - PCR Reagent Set for COVID-19,” . <https://prolabcorp.com/daan-rt-pcr-reagent-set-for-covid-19-real-time-detection-for-48-samples-research-use-only>.
- [28] World Health Organisation, “SARS-CoV-2 Nucleic Acid Tests,” . https://www.who.int/diagnostics_laboratory/200414_eul_covid19_ivd_update.pdf?ua=1.

METHODS

Observational study design: We conducted an experiment to evaluate the hypothesis that known SARS-CoV-2 positive oropharyngeal swab specimens collected during COVID-19 surveillance in Rwanda will test positive after they are combined with as many as 99 known SARS-CoV-2 negative specimens. This was followed by an observational study that aimed to apply our hypercube algorithm to increase the efficiency of community testing for COVID-19 in Rwanda. In the experiment, two different sets of sample pools were tested for SARS-CoV-2 using RT-PCR. Each set consisted of three sample pools containing one known SARS-CoV-2 positive sample diluted in ratios of 1:20, 1:50, and 1:100 by combining it with equivalent amounts of 19, 49, and 99 known SARS-CoV-2 negative samples, respectively (see Figure 2 in the paper and Table 6 here). In the observational study, 1280 individuals selected from the community were tested for SARS-CoV-2 using RT-PCR. One third of the individuals were participants in a screening for Severe Acute Respiratory Infections (SARI) and Influenza Like Illness (ILI) conducted in 30 per cent of the health facilities found across the 30 districts of Rwanda. The remaining two thirds were from COVID-19 screening of at-risk groups in the capital city of Kigali. The latter group is comprised mainly of people (market vendors, bank agents, and supermarket agents) who remained active during the lockdown imposed by the Government of Rwanda to contain COVID-19. Table 5 summarises the characteristics of the study participants.

The positive fraction of RT-PCR tests for SARS-CoV-2 conducted in Rwanda in March 2020 suggests an upper-bound of 2 per cent for the virus prevalence in the country. Using $p = 2$ per cent in the hypercube algorithm indicated an optimal sample group size of 17.5. For convenience, the 1280 individual samples were combined in 64 groups of 20 samples before testing for SARS-CoV-2 (see Figure 4).

We used two established experimental protocols for SARS-CoV-2 testing, namely 1) a protocol by DAAN Gene Co., Ltd., Sun Yat-sen University, which is available online [27], and is also under review by the WHO [28], and 2) another by Corman *et al.*, [13] which is widely used by the scientific community. The first protocol is used for routine screening for SARS-CoV-2, while the second protocol is used only if the first one produces a positive result and confirmation is thus required.

Sample collection and pool design: Oropharyngeal swabs were collected by wiping the tonsils and posterior pharynx wall with two swabs, and the swab heads were immersed in 3 mL Viral Transport Medium (VTM). Samples were transported in VTM to the Rwanda National Reference Laboratory (NRL) immediately after collection. Samples that had to be transported over a long distance were stored in dry ice. Each sample had a volume of 3 mL, of which 200 μL were used for pool testing, and the remainder was temporarily stored at -20°C until the result of the pool testing was known. 200 μL of each sample were mixed with the same volume of other samples of the same pool in a FalconTM 15 mL conical tube and, after vortexing for 5 seconds, 200 μL of the mixture were pipetted for downstream RNA extraction. 5 μL of the extracted RNA were added to 20 μL of master mix to make 25 μL of total solution to be amplified by RT-PCR. If a pool tested positive, stored samples from that pool were processed to identify the positive ones. Individual samples were bar coded, making it easy to trace individuals that tested positive and minimising the risk of confusion of samples. Pool design and subsequent experimental analysis (see **RT-PCR for SARS-CoV-2** below) were implemented with the aid of a robot to reduce human error.

RT-PCR for SARS-CoV-2: Total viral RNA was extracted from swab specimens using the QIAamp Viral RNA 91 Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer’s instructions. RNA samples were screened for SARS-CoV-2 using a 2019-nCoV RNA RT-PCR test targeting two genes respectively encoding an open reading frame (denoted *Orf1ab*) and nucleocapsid protein (denoted *N*) (DAAN Gene Co., Ltd. Of Sun Yat-sen University, 19, Xi-

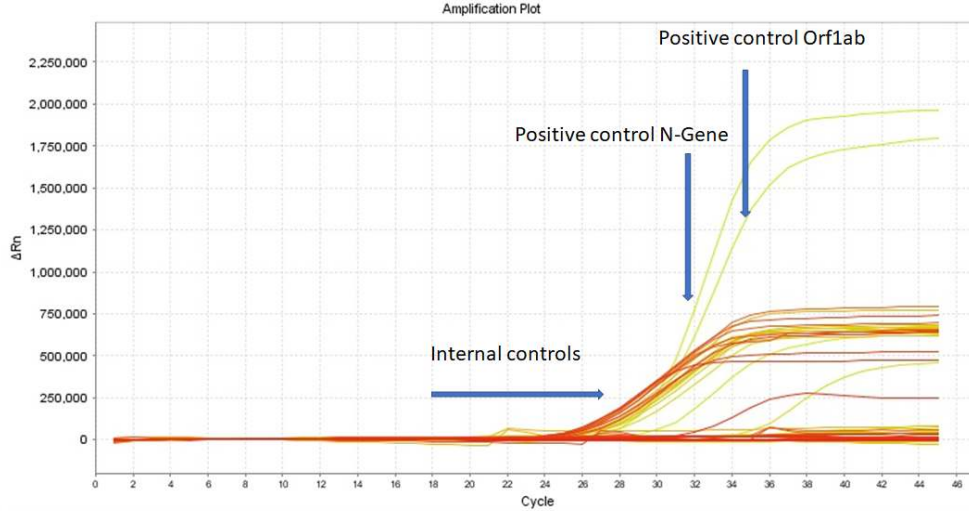


Figure 4: **Amplification plot for sample pools.** Each of 64 the sample pools described in the text tests negative for SARS-CoV-2: the RT-PCR fluorescence curves show below-threshold net fluorescence values. In contrast, for both target genes of the positive control, the fluorescence curves cross the threshold after 32 PCR cycles. ΔRn denotes the difference between the fluorescence signal generated by a sample and a baseline signal. The yellow curves reaching $\Delta Rn \sim 2,000,000$) and 1,750,000 represent the positive control for the *N* and *Orf1ab* genes, respectively. The other yellow and orange curves represent internal controls.

angshan Road, Guangzhou Hi-Tech Industrial Development Zone, China). For *Orf1ab*, CCCTGTGGGTTTTACTTAA and ACGATTGTGCATCAGCTGA were used as forward and reverse primers, respectively, together with a 5'-VIC CCGTCTGCGGTATGTGGAAAGTTATGG-BHQ1-3' probe. For *N*, GGGGAACCTTCCTGCTAGAAT and CAGACATTTTGCTCTCAAGCTG were used as forward and reverse primers, respectively, together with a 5'-FAM-TTGCTGCTGCTTGACAGATT-TAMRA-3' probe. The RT-PCR reaction was set up according to the manufacturer's protocol, with a total volume of 25 μ L. The reaction was run on the ABI Prism 7500 SDS Instrument (Applied Biosystems) at 50°C for 15 min for reverse transcription, denatured at 95°C for 15 min, followed by 45 PCR cycles of 94°C for 15 sec and 55°C for 45 sec. A threshold cycle (Ct value) <40 indicated a positive test, while Ct value >40 indicated a negative test. Positive controls for the reaction showed amplification as determined by curves for FAM and VIC detection channels, and a Ct value ≤ 32 . Positive tests were confirmed using LightMix SarbecoV E-gene and LightMix Modular SARS-CoV-2 RdRp RT-PCRs targeting the envelope (*E*) and RNA directed RNA Polymerase (*RdRp*) genes, respectively, as described by the manufacturer (TIB MOLBIOL Syntheselabor GmbH, Eresburgstr. 22-23, D-12103 Berlin, Germany). Both the primers used and the RT-PCR reaction conditions were previously described [13].

Statistical analysis: Ct values were tested for normality by using the Shapiro-Wilk test. A confidence bound for a sample of n Ct values was calculated as $\bar{C}_t \pm t_{df}^* \times s$, where \bar{C}_t is the sample mean, s is the sample standard error, and t_{df}^* is an appropriate quantile of the Student's t distribution with $n - 1$ degrees of freedom, df . A confidence bound for the sum of the means of two samples of Ct values of sizes n_1 and n_2 , respectively, was calculated using the same formula, with \bar{C}_t set to the sum of the individual sample means, s set to the sum of the individual sample standard errors, and df set to the smaller of $n_1 - 1$ and $n_2 - 1$. Statistical analysis was done using the R statistical computing environment (<https://www.r-project.org/>).

Loss of sensitivity due to dilution: To estimate the post-dilution sensitivities of PCR tests with different maximum numbers of PCR cycles, we combined two data sets. First, we used the mean and standard deviations of the number of additional PCR cycles required for a positive detection,

after a k -fold dilution of a positive specimen (Figure 2 in the paper, showing the data in Figure 6). Second, we used the mean and standard deviation of Ct values for positive specimens sampled from a target population. We combined (or, more accurately, convolved) the two probability distributions, represented as Gaussians to calculate the sensitivity of a $\leq x$ cycle PCR test as the probability that the Ct value of a k -fold diluted positive specimen sampled from the same population will be $\leq x$. Using a representative sample of 33 positive specimens identified during clinical screening for SARS-CoV-2 in Rwanda (see Table 7), we estimate that a ≤ 40 -cycle PCR test targeting the SARS-CoV-2 N gene (resp. *Orf1ab*) has post 20-, 50-, and 100-fold dilution sensitivities of 95% (resp. 86%), 92% (resp. 82%), and 89% (resp. 77%), respectively. For a ≤ 44 -cycle PCR test targeting the N gene (resp. *Orf1ab*), we obtain post 20-, 50-, and 100-fold dilution sensitivities of 99% (resp. 96%), 98% (resp. 94%), and 98% (resp. 92%), respectively. (As mentioned in the article, a maximum of 44 PCR cycles was used in the recent study of pooled testing for SARS-CoV-2 in Israel [10].)

As further checks, we applied the same analysis to (1) an independent sample of 107 positive specimens collected in Rwanda, and (2) the previously published data set consisting of the 26 positive specimens identified in Ref. [10]. From the Rwandan data set, we estimated that a 40-cycle PCR test targeting the N -gene (resp. *Orf1ab*) has post 20-, 50-, and 100-fold dilution sensitivities of 91% (resp. 85%), 88% (resp. 81%), and 85% (resp. 77%), respectively. For a 44-cycle PCR test targeting the N -gene (resp. *Orf1ab*), the predicted sensitivities are 97% (resp. 94%), 96% (resp. 92%), and 95% (resp. 90%). The data set of Ref. [10] contains Ct values for only one gene - the E gene of SARS-CoV-2. Based on the arguments above, we assume for simplicity that diluting a positive specimen by 20-, 50- and 100-fold adds approximately 5, 6, and 7, respectively, to the original Ct value. Applying these assumptions to the data in Ref. [10], we infer post 20-, 50-, and 100-fold dilution sensitivities of 94%, 92% and 89% respectively for a ≤ 40 -cycle PCR test, and 99%, 98% and 97%, for a ≤ 44 -cycle PCR test. These results are comparable to those reported from our experiments. Together, they confirm that diluting positive samples does result in a loss of sensitivity, but that much of the loss can be offset by increasing the number of PCR cycles. In particular, sensitivities above 90% can be achieved for 100-fold dilution by using 44 PCR cycles, only 10% more than the number routinely employed.

Ethics approval: Ethics approval was obtained from the Rwanda National Ethics Committee (Ref: FWA Assurance No. 00001973 IRB 00001497 of IORG0001100/20March2020) and written informed consents were obtained from the patients.

Data and code availability: All data and codes (both in R and in Mathematica) mentioned in the paper are available from the corresponding authors upon reasonable request.

Characteristics	Female		Male		n
	n	%	n	%	
Cluster					
At-risk	327	38.3	526	61.7	853
SARI/ILI*	190	44.5	237	55.5	427
Total	517	40.4	763	59.6	1,280
Age group					
0-14	62	55.9	49	44.1	111
15-34	263	40.2	392	59.8	655
35-49	147	38.7	233	61.3	380
50-64	30	28.0	77	72.0	107
65+	15	55.6	12	44.4	27
Total	517	40.4	763	59.6	1,280
Mean age of participants (sd)	31.97 (14.27)				

Figure 5: **Characteristics of participants in field trial of hypercube algorithm in Rwanda.** For more information, see Observational study design. *SARI=Severe acute respiratory infections, ILI=Influenza-like illness.

ID	Dilution factor	N-gene	ORF ab1	E-gene	RdRp
B16121	1	21.7	22.9	24.5	28
B16121-20	20	29	29		
B16121-50	50	29.74	29.74		
B16121-100	100	30.54	30.54		
B16122	1	25.3	26.7	28.3	32
B16122-20	20	26.81	26.81		
B16122-50	50	27.75	27.75		
B16122-100	100	29.07	29.07		
B09495	1	28.2	31.2	27	29.7
B09495-20	20	33.2	36.9		
B09495-50	50	34.6	38.2		
B09495-100	100	35.4	39		
D34141	1	26.9	29.1	24	25.6
D34141-20	20	28.9	32.2		
D34141-50	50	30.2	33		
D34141-100	100	31	34.3		
D36394	1	29.7	31.1	30.5	31.9
D36394-20	20	33.62	36.52		
D36394-50	50	34.85	37.67		
D36394-100	100	36.49	39.25		
D48370	1	25.9	27.2	26	26.4
D48370-20	20	30.71	33.38		
D48370-50	50	32.1	34.8		
D48370-100	100	32.91	35.64		

Figure 6: **Positive specimens are detected after massive dilution.** Six SARS-CoV-2 positive specimens detected during COVID-19 screening in Rwanda were analysed. The positive specimens were detected by using a screening RT-PCR test targeting the *N* and *Orf1ab* genes of SARS-CoV-2 (Ct values from this test are reported in columns 3 and 4), and confirmed by using another RT-PCR test targeting the *E* and *RdRp* genes (Ct values reported in columns 5 and 6). We determined whether the screening test would have detected the positive specimens if they had been combined with 19, 49 or 99 known SARS-CoV-2 negative specimens. Three pools were thus formed per sample, with dilution factors given in column 2. For all 18 pools, fluorescence exceeded background levels at Ct values ≤ 40 (columns 3 and 4), implying the positive samples would have been detected even if diluted 100-fold.

	ID	N-gene	CT	ORF ab1	CT	E-gene	CT	RdRp	CT	Diagnosis
1	D15943	Positive	17.2	Positive	18.5	Positive	18.1	Positive	19.2	Positive
2	B17418	Positive	19.33	Positive	22.26	Positive	16.87	Positive	18.38	Positive
3	B46255	Positive	20	Positive	21.1	Positive	20.1	Positive	20.8	Positive
4	B46255	Positive	20.1	Positive	21.1	Positive	20.1	Positive	20.8	Positive
5	16121	Positive	21.7	Positive	22.9	Positive	24.5	Positive	28	Positive
6	C45273	Positive	22.78	Positive	24.38	Positive	22.85	Positive	23.55	Positive
7	D18543	Positive	23.1	Positive	24.2	Positive	23.1	Positive	24.2	Positive
8	B79243	Positive	23.2	Positive	24.1	Positive	23.3	Positive	23.3	Positive
9	B79243	Positive	23.2	Positive	24.1	Positive	23.3	Positive	23.3	Positive
10	D40438	Positive	23.29	Positive	26.68	Positive	24.47	Positive	25.73	Positive
11	B86037	Positive	24.64	Positive	27.14	Positive	21.3	Positive	23.25	Positive
12	D04800	Positive	25.2	Positive	26.9	Positive	24.8	Positive	27.1	Positive
13	D04800	Positive	25.2	Positive	26.9	Positive	24.8	Positive	27.1	Positive
14	D15947	Positive	25.3	Positive	26.4	Positive	25.6	Positive	27.1	Positive
15	16122	Positive	25.3	Positive	26.7	Positive	28.3	Positive	32	Positive
16	D48370	Positive	25.9	Positive	27.2	Positive	26	Positive	26.4	Positive
17	B90541	Positive	26.3	Positive	29.6	Positive	27.2	Positive	27.8	Positive
18	CV000054088	Positive	26.6	Positive	27.6	Positive	24.9	Positive	27.8	Positive
19	D34141	Positive	26.9	Positive	29.1	Positive	24	Positive	25.6	Positive
20	CV000019058	Positive	27.7	Positive	28.7	Positive	26.2	Positive	28.8	Positive
21	B87522	Positive	27.82	Positive	30.64	Positive	25.75	Positive	26.63	Positive
22	B09495	Positive	28.2	Positive	31.2	Positive	27	Positive	29.7	Positive
23	D17033	Positive	28.8	Positive	31.8	Positive	28.9	Positive	29.7	Positive
24	D36394	Positive	29.7	Positive	31.1	Positive	30.5	Positive	31.9	Positive
25	B46304	Positive	31.6	Positive	36.2	Positive	36.3	Positive	36.7	Positive
26	D15317	Positive	31.7	Positive	34.6	Positive	33.1	Positive	33.7	Positive
27	B86032	Positive	32.57	Positive	35.12	Positive	30.79	Positive	31.57	Positive
28	B65096	Positive	33.3	Positive	35.5	Positive	32.1	Positive	34.9	Positive
29	D11510	Positive	33.5	Positive	36.3	Positive	33.7	Positive	34.7	Positive
30	B64423	Positive	33.54	Positive	36.69	Positive	34.15	Positive	36.19	Positive
31	B46275	Positive	34.1	Positive	35.9	Positive	34.1	Positive	35.4	Positive
32	B86031	Positive	35.03	Positive	37.47	Positive	33.5	Positive	34.29	Positive
33	B43708	Positive	38.12	Positive	39.7	Positive	34.94	Positive	35.81	Positive

Figure 7: **Representative sample of confirmed positive specimens detected during COVID-19 screening in Rwanda.** The specimens were detected by using a screening RT-PCR test targeting the *N* and *Orf1ab* genes of SARS-CoV-2 (Ct values from this test are reported in columns 3 and 5). They were subsequently confirmed as positive by using another RT-PCR test targeting the *E* and *RdRp* genes (Ct values are reported in columns 7 and 9).