Increased PCR screening capacity using pool addresses

Adrian Viehweger [1,4] *, Felix Kühnl [2] *, Christian Brandt [3,4], Brigitte König [1]

* These authors contributed equally.

[1] Medical Microbiology, University Hospital Leipzig

[2] Bioinformatics, University Leipzig

[3] Infection Medicine, University Hospital Jena

[4] nanozoo GmbH, Leipzig

Abstract

Effective public health response to viral outbreaks such as SARS-CoV-2 is often informed by real-time PCR screening of large populations. Pooling samples can increase screening capacity. However, when a traditional pool is tested positive, all samples in the pool need individual retesting, which becomes ineffective at a higher proportion of positive samples. Here, we report a new pooling protocol that mitigates this problem by replicating samples across multiple pools. The resulting "pool address" allows the sample status to be resolved more often than with traditional pooling. At 2% prevalence and 20 samples per pool, our protocol increases screening capacity by factors of 5 and 2 compared to individual testing and traditional pooling, respectively. The corresponding software to layout and resolve samples is freely released under a BSD license (https://github.com/phiweger/clonepool).

Main text

Group testing has long been used to screen larger collections of samples, most of which are expected to test negative.[1] In pathogen outbreaks such as the current SARS-CoV-2 pandemic, this "pooling" allows screening of large populations to guide public health response and restrict spreading. However, in most laboratories, the screening capacity is limited by the number of PCR reactions that can be performed in a day. It is, therefore, desirable to maximize the number of samples that can be tested per reaction.

Various approaches have been proposed to do this in the context of SARS-CoV-2 RT-PCR testing.[2,3] One problem with the traditional pooling approach, where several samples are collected and tested collectively, is that the number of positive pools that require individual retesting increases rapidly with the number of positive samples in the overall population, henceforth called "prevalence". A high prevalence renders pooling ineffective. To mitigate this, we propose to test samples in replicates and distribute them across multiple pools. The resulting "pool address" can then be used to resolve samples in one pool, given information from other pools that contain a replicate. While some previous studies have taken a similar approach implicitly,[2] it has neither been investigated systematically for more replicates than two, nor is there any software that would generate and resolve the corresponding pooling layout for laboratory use.

We therefore introduce "clonepool", a pooling framework to maximize the effective number of samples s_e per PCR reaction. "Effective" refers to the fact that samples in positive pools, whose status cannot be resolved in the pooled run, are assumed to be retested individually. The maximum number of samples for a given pool size p, number of pools n and number of replicates r is calculated as $s_m = \frac{pn}{r}$. The effective number of samples can then be calculated from the number of unresolved samples s as $s_e = \frac{s_m}{p+s}$.

The clonepool algorithm first distributes all sample replicates randomly across the available pools, with the limitation that a sample's replicates do not co-occur in the same pool. After the pools have been tested, the algorithm tries to resolve the samples' status in two iterations: In a first iteration, all samples that have at least one replicate in a negative pool are marked negative. In the second iteration, samples that only occur in positive pools and where at least one replicate is in a pool where all other samples are negative, are marked positive (red, orange). All other samples cannot be resolved, and need to be retested individually. The longer the set of pools a sample is distributed across, i.e., the longer its "pool address", the more samples can be resolved given a particular prevalence.

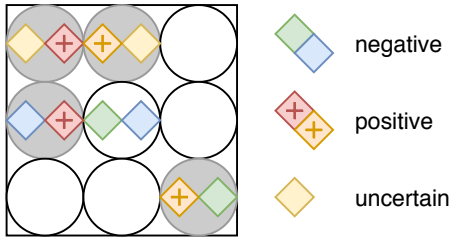

Figure 1: Illustration of the clonepool algorithm. Positive pools (circles) are marked grey, negative ones in white. Samples are depicted as squares, positive ones with a "+". In a first iteration, all samples that have at least one replicate in a negative pool are marked negative (blue, green). In the second iteration, samples that only occur in positive pools and where at least one replicate is in a pool where all other samples are negative, are marked positive (red, orange). All other samples cannot be resolved and have to be retested individually (yellow).

In our simulation test of the proposed clonepool algorithm, we assume no pipetting errors, which can be achieved e.g., through the use of a pipetting robot. We also assume that 94 pools are available, which corresponds to a 96-well plate with two wells reserved for a positive and a negative control. Furthermore, we assume that there are no false positive or false negative PCR reactions.

Two parameters determine which pooling scheme is most effective (Fig. 2). If both the prevalence and the number of samples per pool are low, traditional pooling without replicates yields the highest number of samples per reaction. However, if the prevalence increases or more samples are pooled, anyone pool is more likely positive. Sample replicates will then resolve more samples. In our testing experience, we observed a prevalence of about 5%, but this value is subject to variability e.g., depending on a population's pre-test probability.

The number of samples that can be pooled without affecting the PCR sensitivity is limited by the PCR cycle threshold (Ct) for the target, i.e., the cycle at which amplification becomes detectable over background noise (typically ten times the standard deviation, SD). Usually, Ct values above 35 are treated as unspecific amplification. SARS-CoV-2 amplifies at low Ct values due to high viral titers (Ct 18-25 depending on the material and number of days post-infection).[4,5] A 20-fold dilution, i.e., pooling 20 samples, would cause the Ct value to increase by about 4.3 cycles ($2^x = d$, where d is the dilution and x the shift in Ct), still comfortably above the detection limit.

At a prevalence of 5% SARS-CoV-2 positive samples, and for ten samples per pool and two replicates per sample, we simulate that 2.61 times the number of samples can be processed compared to testing samples individually (SD 0.13). This result is in line with previous estimates using a slightly different version of

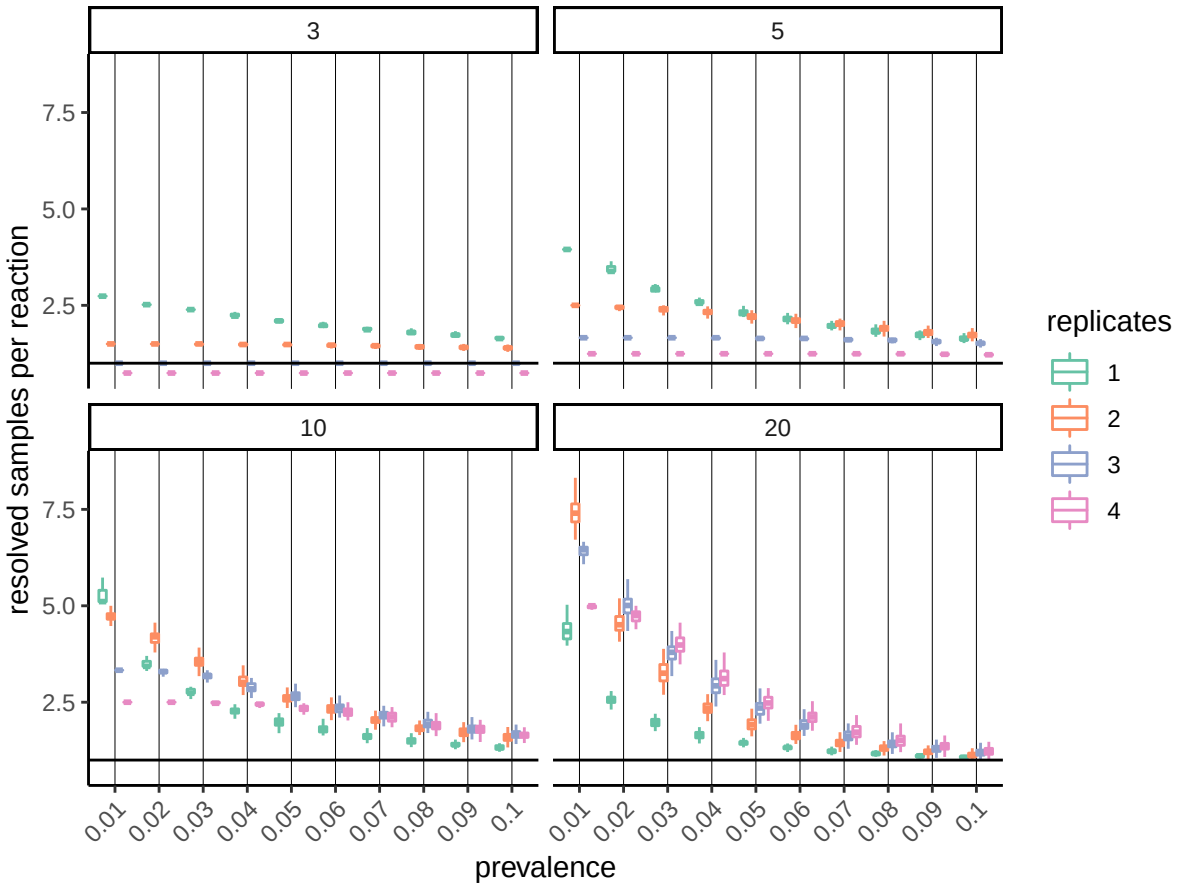

Figure 2: Simulation results for different percentages of positive samples (x-axis), replicates (colors), and pool sizes (panels). The target metric is the effective number of samples per PCR reaction, which includes the individual retesting of samples that cannot be resolved in the first pooling run.

the 2-replicate scheme.[2] Using two replicates increases the effective number of samples per reaction by 31% compared to pooling without replicates. At 2% prevalence and 20 samples per pool – a scenario more akin to screening large populations – 5.01 times the number of samples can be screened compared to individual testing (SD 0.28), and the increase over no replicate pooling is 193%. These presented values correspond to *in silico* simulations, and require further validation in the laboratory.

In conclusion, our pooling protocol based on sample replicates can substantially increase the number of samples per PCR reaction when screening large populations during pathogen outbreaks, such as SARS-CoV-2. The protocol can be tuned to local laboratory conditions such as pool size and proportion of positive samples. The accompanying software supports the protocol's implementation and routine use.

References

1. Dorfman, R. The detection of defective members of large populations. *Ann. Math. Stat.* **14**, 436–440 (1943).

2. Sinnott-Armstrong, N., Klein, D. & Hickey, B. Evaluation of group testing for SARS-CoV-2 RNA. *Infectious Diseases (except HIV/AIDS)* (2020).

3. Eis-Huebinger, A. M. *et al.* Ad hoc laboratory-based surveillance of SARS-CoV-2 by real-time RT-PCR using minipools of RNA prepared from routine respiratory samples. *Infectious Diseases (except HIV/AIDS)* (2020).

4. Wölfel, R. *et al.* Virological assessment of hospitalized patients with COVID-2019. *Nature* (2020).

5. Holshue, M. L. *et al.* First case of 2019 novel coronavirus in the united states. *N. Engl. J. Med.* **382**, 929–936 (2020).