

1 Increased PCR screening capacity using pool addresses

2 Adrian Viehweger ^{1,4} *, Felix Kühnl ² *, Christian Brandt ^{3,4}, Brigitte König ¹

3 * These authors contributed equally.

4 ¹ Medical Microbiology, University Hospital Leipzig

5 ² Bioinformatics, University Leipzig

6 ³ Infection Medicine, University Hospital Jena

7 ⁴ nanozoo GmbH, Leipzig

8 Abstract

9 We report a pooling protocol that increases the number of samples that can be processed with PCR
10 by multiples compared to per-sample testing and naive pooling. This is achieved through the newly
11 introduced concept of a *pool address*: A sample is replicated accross multiple pools, which in many
12 cases allows it to be resolved as positive even though it has not been tested individually. Our protocol
13 allows pooling to be effective even at a high target prevalence where common approaches become inef-
14 fective. The corresponding software to layout and resolve samples is freely released under a BSD license
15 (<https://github.com/phiweger/clonepool>).

16 Main text

17 Group testing has long been used to screen larger collections of samples, most of which are expected to
18 be negative.¹ With the current SARS-CoV-2 pandemic, screening large sample collections for the virus
19 enables effective public health response to restrict its spread. However, in most laboratories the capacity
20 is limited by the number of PCR reactions that can be performed in a day. Thus, it is desirable to
21 maximize the number of samples that can be tested, while minimizing the number of reactions required
22 to do so.

23 Various approaches have been proposed to do this in the context of SARS-CoV-2 RT-PCR testing.^{2,3}
24 One problem with the traditional pooling approach, where a number of samples is collected and tested
25 collectively, is that the number of positive pools that require individual retesting increases rapidly with the
26 number of positive samples in the overall population. This renders pooling ineffective. One approach to
27 mitigate this problem is to test samples in replicates and distribute those replicates accross multiple pools.
28 This “pool address” can then be used to resolve samples in one pool given information from the other
29 pools their replicates occur in. While previous studies have taken this approach implicitly,² it has neither
30 been investigated systematically for a larger number of replicates than two, nor is there any software that
31 would generate and resolve the corresponding pooling layout for laboratory use.

32 We introduce “clonepool”, a pooling framework to maximize the effective number of samples s_{eff} per
33 PCR reaction. “Effective” refers to the fact that samples in positive pools whose status cannot be resolved
34 in the pooled run are assumed to be retested individually.

35 The maximum number of samples for a given pool size p , number of pools n and number of replicates r
36 is calculated as

$$s_{max} = \frac{pn}{r}$$

37 The effective number of samples can then be calculated from the number of unresolved samples s as:

$$s_{eff} = \frac{s_{max}}{p + s}$$

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

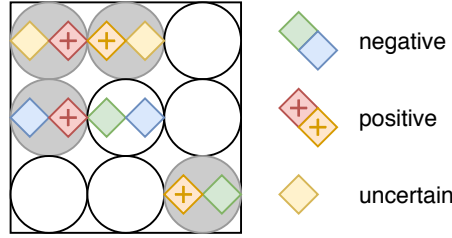


Figure 1: Illustration of the clonepool algorithm. Positive pools (circles) are marked grey, negative ones in white. Samples are depicted as squares, positive ones with a "+". In a first iteration, all samples that have at least one replicate in a negative pool are marked negative (blue, green). In the 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).

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

50 Two parameters determine which pooling scheme is most effective (Fig. 2). If the prevalence and the
51 number of samples per pool are low, using only one "replicate" yields the highest number of samples per
52 reaction. However, if the prevalence increases or more samples are pooled, there is an increased probability
53 for any one pool to become positive, in which case the use of sample replicates helps resolve more samples.
54 In our testing experience, we observe a prevalence of about 5%, but this value is subject to variability
55 e.g. depending on a population's pre-test probability. The number of samples that can be pooled without
56 affecting the PCR sensitivity is limited by the PCR cycle threshold (Ct) for the target, i.e. the cycle at
57 which amplification becomes detectable over background noise. Usually, Ct values above 35 are treated
58 as unspecific amplification. For example, 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.

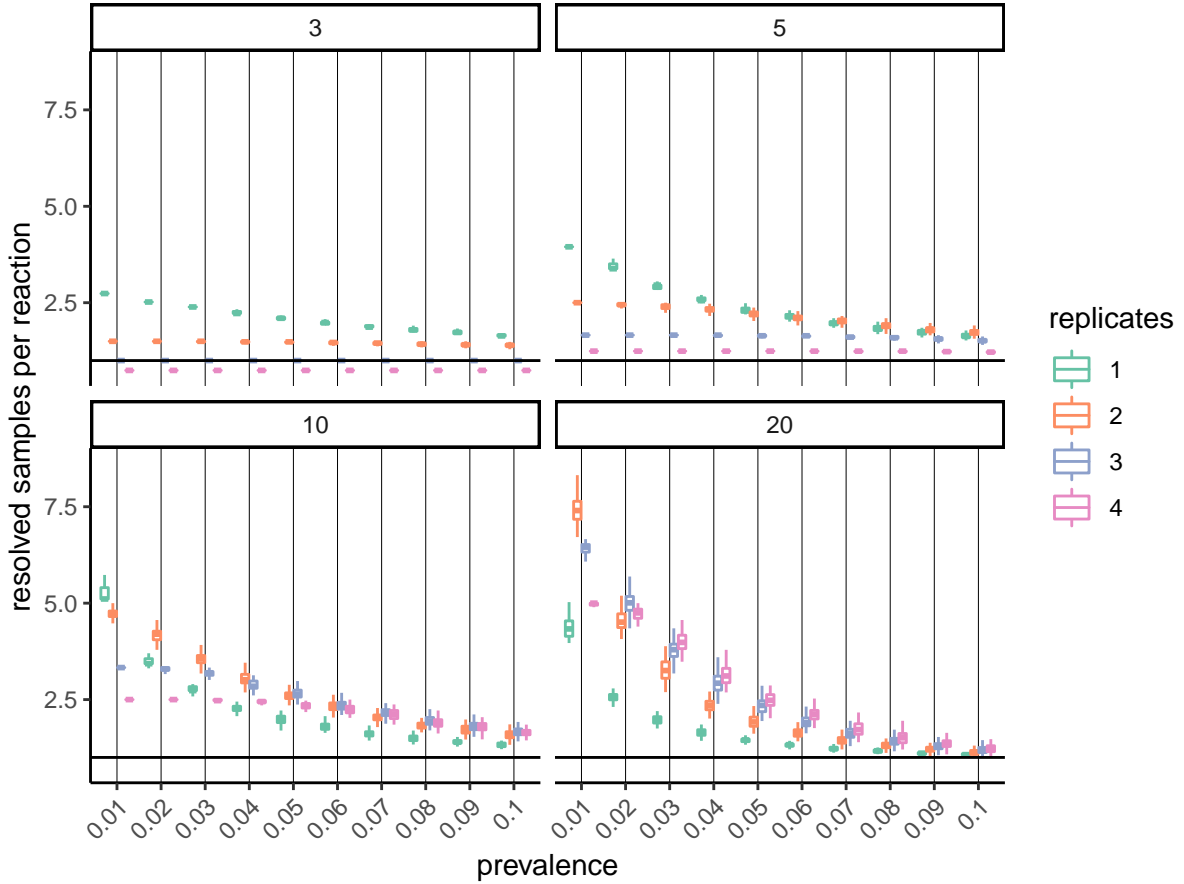


Figure 2: Simulation results for different percentages of positive samples (x-axis), replicates (colors) and pool sizes (panels). The target metric is the effective number of samples per PCR reaction, which includes the individual retesting of samples that cannot be resolved in the first pooling run.

At a prevalence of 5% SARS-CoV-2 positive samples, and for 10 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). Using two replicates increases the effective number of samples per reaction by 31% compared to pooling without replicates. This is in line with previous estimates using a slightly different version of the 2-replicate scheme.² 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 testing in the laboratory.

70 In conclusion, our pooling protocol based on sample replicates can substantially increase the number of
71 samples per PCR reaction when screening large populations for a target such as SARS-CoV-2. The proto-
72 col can be tuned to local conditions such as pool size and positive sample prevalence. The accompanying
73 software supports the protocol's implementation and routine use.

74 References

- 75 1. Dorfman, R. The detection of defective members of large populations. *Ann. Math. Stat.* **14**, 436–440
76 (1943).
- 77 2. Sinnott-Armstrong, N., Klein, D. & Hickey, B. Evaluation of group testing for SARS-CoV-2 RNA.
78 *Infectious Diseases (except HIV/AIDS)* (2020).
- 79 3. Eis-Huebinger, A. M. *et al.* Ad hoc laboratory-based surveillance of SARS-CoV-2 by real-time RT-
80 PCR using minipools of RNA prepared from routine respiratory samples. *Infectious Diseases (except*
81 *HIV/AIDS)* (2020).
- 82 4. Wölfel, R. *et al.* Virological assessment of hospitalized patients with COVID-2019. *Nature* (2020).
- 83 5. Holshue, M. L. *et al.* First case of 2019 novel coronavirus in the united states. *N. Engl. J. Med.* **382**,
84 929–936 (2020).