Designing Synthetic Positive Controls for Highly Multiplexed Amplicon Sequencing

JR Hensley[1,2], R Jin[1,3], GK Vestal[1,4], JL Rowell[1,5], ST Lucking[1,3], A Khan[1,3], AD Huang[1], HA Carleton[1], AJ Williams-Newkirk[1]

1. Enteric Diseases Laboratory Branch, Centers for Disease Control and Prevention, Atlanta, GA; 2. ASRT, Inc., Atlanta, GA; 3. WDS, Inc., Suwanee, GA; 4. Oak Ridge Institute for Science and Education, US Department of Energy, Oak Ridge, TN, USA

Synthetic positive controls are an efficient option for highly multiplexed amplicon sequencing panels.

Updated Abstract

Background: Fast characterization of the resistome of complex metagenomic samples has broad applications for public health, environmental monitoring, and other fields. Highly multiplexed amplicon sequencing (HMAS) is a new tool that enables rapid detection of thousands of antimicrobial resistance determinants (ARDs) directly from clinical samples without requiring isolation of host microbes. However, validation of every primer pair on a HMAS panel is more challenging than single-plex or traditional multiplex assays. Additionally, creating positive controls for ARDs using traditional plasmid-based methods creates substantial biosafety concerns. Therefore, we designed a pool of synthetic positive controls to validate our ARD HMAS panel.

Methods: The HMAS panel included 823 target amplicons covering 118 ARDs. Targets were selected from a published microarray for relevance to enteric bacteria. Primers were modified for compatibility with the Juno Targeted DNA Sequencing Library Preparation System (Fluidigm). Synthetic positive controls were designed based on public genome sequences for PhiX, *Streptomyces coelicolor*, and *Wolbachia pipientis* to match the length and GC content of the target amplicons in the ARD panel and were ordered as an oligo pool from Twist Bioscience. The Twist oligo pool was used as a template for panel amplification on the Juno. Amplicons (180-240 bp) were sequenced on the Illumina MiSeq using 2x250 bp chemistry. Mothur v.1.46.0 was used to perform sample demultiplexing, read assembly, quality filtering, and mapping. Mapping results were filtered using in-house R scripts.

Results: Amplicons were successfully detected for all synthetic positive control sequences. However, amplification efficiencies varied across targets.

Conclusions: The synthetic positive controls were an efficient way to validate a HMAS panel targeting sequences from many different genomes when using positive control plasmid constructs is unsafe. Furthermore, the synthetic control provides a commercially available custom positive control for use in production testing without fear of undetectable contamination. Future studies will include expanding the primer panel to include additional targets and testing stool specimens to evaluate the use of this panel for public health surveillance.

HMAS Panel Design

Antimicrobial Class	Number of Genes
Aminoglycosides	5
Beta-lactams	48
Macrolides	32
Phenicols	3
Polymyxins (colistin)	6
Quinolones	14
Tetracyclines	2
Trimethoprims	6
TOTAL	**118** (823 amps)

The current version of our highly multiplexed amplicon sequencing (HMAS) panel uses 823 primer pairs to amplify targets for 118 antimicrobial resistance determinants (ARDs) directly from stool samples. Future versions of the HMAS panel could potentially contain thousands of primer pairs.

Like all PCR-based methods, HMAS requires positive controls for initial assay validation as well as routine use to confirm test performance on each run. However, choosing appropriate and safe positive controls for hundreds of primer pairs on a HMAS panel is more challenging than single-plex or traditional multiplex assays.

Synthetic Control Design

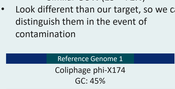
We need a positive control to verify function of each primer pair on every run, i.e. 823 primer pairs covering 118 AR genes. An optimal control would only require a single sample's space on the HMAS assay and be commercially available to support uniformity across surveillance labs.

Selecting or creating positive controls for ARDs using traditional methods is difficult and creates substantial biosafety concerns.

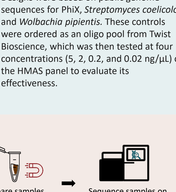
⃠ Mix multidrug resistant bacteria
- Select dozens of multidrug resistant (MDR) live bacteria, enough to cover 162 genes
- All network laboratories maintain stocks of these MDR bacteria
- Use a large number of wells on every run

⃠ Plasmid-based method
- Design and order a plasmid containing 118 AR gene sequences
- All network laboratories now have a super plasmid in their labs

✓ Synthetic Positive Controls
- Design targets for each primer pair
- Order as oligo pool
- Receive uniform, quality-controlled DNA
- No biosafety risk

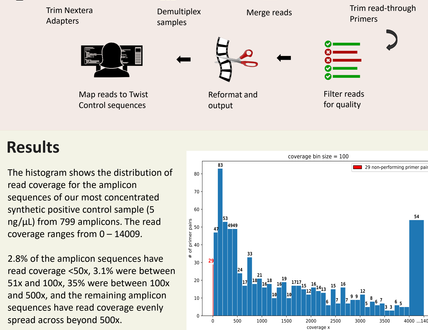


Synthetic positive controls were designed to:
- Behave like our targets in PCR reactions
  - Same length (180 – 240 bp)
  - Similar GC % (23 – 72%)
- Look different than our target, so we can distinguish them in the event of contamination



Designs were based on public genome sequences for PhiX, *Streptomyces coelicolor*, and *Wolbachia pipientis*. These controls were ordered as an oligo pool from Twist Bioscience, which was then tested at four concentrations (5, 2, 0.2, and 0.02 ng/µL) on the HMAS panel to evaluate its effectiveness.

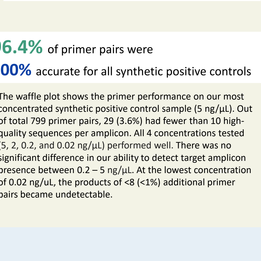
Method Workflows

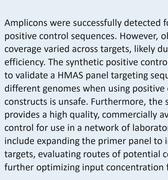


Results

The histogram shows the distribution of read coverage for the amplicon sequences of our most concentrated synthetic positive control sample (5 ng/µL) from 799 amplicons. The read coverage ranges from 0 – 14009.

2.8% of the amplicon sequences have read coverage <50x, 3.1% were between 51x and 100x, 35% were between 100x and 500x, and the remaining amplicon sequences have read coverage evenly spread across beyond 500x.





96.4% of primer pairs were
100% accurate for all synthetic positive controls

The waffle plot shows the primer performance on our most concentrated synthetic positive control sample (5 ng/µL). Out of total 799 primer pairs, 29 (3.6%) had fewer than 10 high-quality sequences per amplicon. All 4 concentrations tested (5, 2, 0.2, and 0.02 ng/µL) performed well. There was no significant difference in our ability to detect target amplicon presence between 0.2 – 5 ng/µL. At the lowest concentration of 0.02 ng/uL, the products of <8 (<1%) additional primer pairs became undetectable.

Conclusion

Amplicons were successfully detected for 96.4% of synthetic positive control sequences. However, observed depth of coverage varied across targets, likely due to varying primer pair efficiency. The synthetic positive controls were an efficient way to validate a HMAS panel targeting sequences from many different genomes when using positive control plasmid constructs is unsafe. Furthermore, the synthetic control provides a high quality, commercially available custom positive control for use in a network of laboratories. Future studies will include expanding the primer panel to include additional targets, evaluating routes of potential contamination, and further optimizing input concentration for cost efficiency.

Conc. (ng/µL)	Successfully detected %	Cost per run
5	96.4	$ 12.50
2	96.4	$ 5.00
0.2	96.2	$ 0.50
0.02	95.7	$ 0.05

Acknowledgments

GK Vestal is supported in part by a fellowship from the Oak Ridge Institute for Science and Education administered by the United States Department of Energy and funded by the Centers for Disease Control and Prevention (CDC). We thank Jason Folster for help in selecting AMR targets. We also thank Tom Slezak (Lawrence Livermore National Lab), Shea Gardner (LLNL), Nish Mulakken (LLNL), Jonathan Allen (LLNL), Ken Frey (Naval Medical Research Center), Gary Vora (U.S. Naval Research Laboratory), Gary An (University of Chicago), Tom Brettin (Argonne National Laboratory), Dion Antonopoulos (ANL), Sarah Owens (ANL), and Nick Greenfield (One Codex) for developing and sharing the AMR panel with us. Created using icons from The Noun Project. The Centers for Disease Control and Prevention does not endorse any commercial products used in this study

CONTACT INFO

Jasmine Hensley
yhe7@cdc.gov

A. Jo Williams-Newkirk
igy7@cdc.gov