A droplet digital PCR assay for quantification of bacteriophage viral vector titer and purity.

Peter J. Voorhees[1], Ronni M. Ponek[2], Jamie D. Liu[2], Samuel K. Lai[1,2*]

[1]Division of Pharmacoengineering and Molecular Pharmaceutics, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[2] Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

*Corresponding author:

Samuel K. Lai

Email: lai@unc.edu

Homepage: https://pharmacy.unc.edu/lai-research-group/

ABSTRACT: Bacteriophage (phage) vectors, bacterial viruses engineered to deliver specific transgenes into target bacteria, offer considerable promise as tools for tuning the microbiome. However, the field currently lacks standardized methods for characterizing phage vectors with the same rigor as those for viral vectors used in human gene therapy. To develop a robust and reproducible viral vector system, it is critical to measure both transduction efficiency and the true dose of vector particles used. Here, we use central composite design to develop a digital droplet PCR (ddPCR) assay for precise quantification of transgene and vector genome-packed phage vector titers following an established production process. This duplex assay uses probes specific to 2 unique DNA barcodes inserted into the vector genome and transgene cassette, enabling precise quantitation of each population with a dynamic range of nearly 3 orders of magnitude. We show biological activity assays can consistently underestimate the number of transgene-packed vectors, potentially greatly overestimating the true transduction efficiency, particularly when contamination by genome-packed vectors is high. This ddPCR approach can be broadly adopted to characterize phage particles for a variety of biomedical applications.

Introduction

Bacteriophages (phages), viruses that infect bacteria, have long been studied for their potential as antimicrobial agents(1–3). Recently, however, there is growing interest in their use as vehicles to deliver specific DNA that modulate the function of target bacteria, without killing(4–7). This approach has already been proven feasible *in vivo*, where phage vectors delivered transgenes that suppressed shiga toxin production in susceptible *Escherichia coli* (*E. coli*) in the mammalian gut, thereby ameliorating the pathogenic effects of these strains without disrupting an ecological niche that other opportunistic pathogens might colonize(8). The ability to precisely preform genetic manipulations on individual microbiome constituents *in situ* underscores phage vectors as an exciting new tool in microbiome engineering. Nonetheless, to support translational development, assays that offer robust and accurate characterization of phage vectors must be developed.

46 In human gene therapy, there are quality control and quantification assays for viral vectors
47 established to assess safety, potency, purity, and stability. While each of these is essential for
48 developing vector products for use in humans, it is critical to measure both potency and purity
49 even for *in vitro* assays, as well as *in vivo* preclinical studies. For instance, knowing the amount
50 (titer) of a viral vector is essential for consistent dosing, and can be quantified in terms of both
51 copies of transgenic DNA or biologically active particles in a given volume(9). Transgene copies
52 are often quantified by either qPCR or digital droplet PCR (ddPCR)(10,11), whereas transgene
53 expression assays in permissive cell lines are used to determine the number of biologically active
54 particles—those capable of mediating successful delivery of transgenic DNA (i.e.
55 transduction)(12,13). Purity must account for residual proteins and nucleic acids from the vector
56 production cell line, helper viruses or plasmids(14), as well as empty vectors and vector capsids
57 packed with non-target DNA, including viral genomic DNA.

58 Despite the importance of these quality control measures in developing effective and
59 reproducible viral vectors, analogous assays that determine potency and purity for phage vectors
60 have not yet been established. Similar to mammalian viral vectors, phage vector transduction
61 efficiencies can be affected by numerous factors, including the presence of cofactors, and the
62 size, complexity, and stability of the packed transgenic DNA, and thus can vary widely. As a result,
63 biological activity assays are unlikely to accurately reflect the number of transgene-packed phage
64 vectors. This is particularly important for phage vectors generated by modifying lytic phages, as
65 the bacterial cytotoxicity of genome-packed vectors can significantly impact measured
66 transduction levels due to target cell killing(15,16). We believe quantitative PCR assays are
67 therefore essential in not only understanding readouts of biological activity assays but also
68 informing the appropriate conditions with which to run these assays.

69 To meet this need, we report here a ddPCR-based assay for quantifying both the potency and
70 purity of phage vectors by measuring the quantity of capsids packed with transgenic DNA and

71 vector genome DNA. This assay was first optimized using central composite design (CCD), then
72 compared against two methods for measuring biologically active vector particles. As this assay is
73 based on small, standardized DNA barcodes, it is readily applicable to a wide variety of phage
74 vectors and we anticipate it will be crucial in the development of effective and reproducible phage
75 vector products.

76 **Materials and Methods**

77 ***Barcode generation***

78 A python script (DOI: 10.5281/zenodo.10869457) was written to computationally design the
79 DNA barcode sequences. Briefly, this script generates a specified number of random DNA
80 sequences of a set length, with a G/C content between 40-60%, without homopolymeric repeats
81 >4 bp, and without common restriction enzyme sites, common promoter and RBS consensus
82 sequences. The secondary structure of each sequence is then analyzed by RNAfold along a
83 sliding 50 bp window across each entire sequence. The maximum minimum free energy (MFE)
84 and average MFE are calculated for each sequence and sequences with a max MFE greater than
85 a specified threshold are removed. The remaining sequences are sorted by average MFE, and
86 bottom 90% are removed. Remaining sequences are then run through a BLAST search against
87 the NCBI Nucleotide database and the percent identity, alignment length, E-value, Bit Score, and
88 Query Coverage of the top BLAST match for each barcode is recorded. Next, the script finds the
89 best barcodes based on minimal similarity to existing sequences, as determined by the BLAST
90 search results, and minimal predicted secondary structure. Finally, the script finds the best pairs
91 of barcodes. Ranking each pair by various criteria including max MFE, average MFE, alignment
92 length, percent identity, and query coverage. The pairs are then sorted by weighted average rank
93 and intra pair orthogonality (Hamming distance), and the results are saved to a csv file.

94 ***Plasmid cloning***

95 8673 and 7898 barcode reference plasmids (pBc8673 and pBc7898) were cloned by Twist
96 Bioscience into pTwist Amp High Copy backbones. The transgene phagemid (pT7kan-mScarlet)
97 and gp17 helper plasmid (pT7gp17) were cloned through Gibson HiFi (New England Biolabs, cat.
98 no. E2621L) assembly of synthesized DNA fragments (IDT) (sequences in **Table S1**) according
99 to **Table S2**.

100 ***Phage genome assembly***

101 The barcoded T7 genome was generated through a previously described yeast-based assembly
102 and rebooting protocol(17). Briefly, the T7 genome was PCR amplified in fragments, excluding
103 the gp17 coding sequence. A yeast artificial chromosome was also amplified with regions of
104 homology to the termini of the T7 genome (primer sequences in **Table S3**). A double stranded
105 DNA fragment encoding the 7898 barcode flanked by 60 bp sequences homologous to the regions
106 up and downstream of gp17 was synthesized by Twist Bioscience. 1-4 μg of each fragment was
107 then transformed into EBY100 *S. cerevisiae* via electroporation. YAC-captured T7 genomic DNA
108 was purified from the yeast and transformed into *E. coli* NEB5α containing a T7 gp17 trans-
109 complementation plasmid. These *E. coli* were mixed with soft agar and plated on LB agar plates.
110 Plaques were then picked to isolate gp17-deficient barcoded T7 phage.

111 ***ddPCR central composite design***

112 Primers and Affinity Plus probes specific to the 8673 and 7898 barcodes were synthesized by
113 IDT (**Table S4**). To optimize the ddPCR assay a central composite design experiment was
114 designed using JMP 17.0.0. The factor setting ranges described in **Table 2** were used for the
115 annealing temperature, primer concentration, probe concentration, ramp rate, and cycle number.
116 The following four optimization goals were set: minimize 8673 relative error, minimize 7898
117 relative error, maximize 8673 droplet separation, and maximize 7898 droplet separation. Relative
118 error was defined as the relative difference between the measured barcode copies/well and the
119 actual barcode copies/well. Droplet separation was defined as the difference between the mean

120 amplitude of the positive and negative droplets. ddPCR was performed using each of the factor
121 combinations generated by JMP (**Table S5**) according to the manufacturers recommended
122 protocol (Bio Rad #1863024). Relative errors and droplet separation were then input into a JMP
123 Fit Least Squares model and the resulting desirability function was maximized.

124 ***ddPCR performance metric characterization***

125 5-fold serial dilutions of the barcode reference plasmids were made, ranging from 1.28×10^6 to
126 16 molecule/μL. Input plasmid concentrations were calculated using the barcode plasmid
127 molecular weight and the plasmid mass, as quantified by Qubit (Invitrogen #Q33230). ddPCR
128 was performed on these plasmid dilutions using the CCD-derived settings (T_a = 63°C, primer
129 concentration = 900 nM, probe concentration = 375 nM, ramp rate of 1°C/s, 50 cycles). The
130 relative error and coefficients of variance were calculated for both barcode reference plasmids at
131 each dilution point. Measured *vs.* actual values were plotted in log-log form and linear regression
132 was performed (GraphPad Prism 10.1.1) on the range of sample points that appeared linear. The
133 LLOD was calculated as the sum of the mean of the concentrations in the no template control
134 wells and 3 times the standard deviation of these wells.

135 ***Phage vector production***

136 *E. coli* NEB5α was transformed with pT7kan-mScarlet and pT7gp17. These cells were cultured
137 for 16 hrs overnight. Overnight cultures were passaged 1:100 into fresh media and cultures until
138 an OD_{600} of 0.6 was reached. Unless otherwise specified, these cultures were spiked with 1×10^8
139 gp17-deficient barcoded T7 helper phage. Infection was allowed to proceed for 4 hrs at 37°C with
140 shaking. Infections were then centrifuged at 10,000 x G for 10 min and supernatant was 0.2 μm
141 filtered to isolate phage vectors.

142 ***Phage DNA purification and ddPCR***

143 Phage DNA was purified using the Norgen Phage DNA Isolate Kit (Norgen Biotek Corp. #
144 46800) following the manufacturers recommended protocol. Purified DNA was digested with Sca-

145 I (NEB # r3122) and 3 5-fold dilutions were made of the completed digestion reaction. Each of
146 these dilutions was analyzed by the optimized ddPCR reaction and the highest concertation in
147 the dynamic range of the assay was used to calculate transgene and genome copies/mL.

148 ***Colony formation transduction assays***

149 5 5-fold dilutions of T7 phage vectors were made. *E. coli* NEB5α were cultured for 16 hrs
150 overnight. Overnight cultures were passaged 1:100 into fresh media and cultures until an OD_{600}
151 of 0.6 was reached. These cells were then mixed with 500 μL of each phage vector dilution to a
152 final concentration of $1x10^8$ cells/mL in a final volume of 1 mL. These transduction reactions were
153 incubated at 37°C for 90 min with shaking. Cells were then pelleted at 16,000 x G for 5 min and
154 resuspended in fresh LB. 4-fold dilutions of the transduced cells were made and plated on LB
155 plates supplemented with kanamycin (50 μg/mL). Plates were incubated for 16 hrs at 37°C and
156 colonies were counted to calculate the number of transduced cells (transductants) per mL.

157 ***Trans-complementation plaque formation assays***

158 *E. coli* NEB5α transformed with pT7gp17 were cultured for 16 hrs overnight. Overnight cultures
159 were passaged 1:100 into fresh media and cultures until an OD_{600} of 0.6 was reached. 800 μL of
160 these cells were mixed with 8 mL of molten soft agar and overlaid on prewarmed LB agar plates
161 supplemented with ampicillin (100 μg/mL). 12 10-fold diltions of each of the 5 5-fold phage vector
162 dilutions were made and spotted onto the pT7gp17-containing NEB5α lawn plates. Plates were
163 incubated incubated for 16 hrs at 37°C and plaques were counted to calculate the number of
164 plaque forming units (PFU) per mL. PFU/mL concentrations were plotted against ddPCR-
165 determined genome copies/mL and 2-way ANOVA with Šídák's multiple comparisons test was
166 used to evaluate differences between the two measures at each of the 5 phage vector dilutions.

167 **Results and Discussion**

168 *Development of orthogonal DNA barcodes*

ddPCR reactions commonly target amplification of a 60-200 bp DNA sequence(18). To maximize modularity and transferability of the assay, we inserted unique 300 bp DNA barcodes into both the transgenic DNA intended for phage vector packing as well as the genome of the helper phage used to generate the vectors. This allows us to, upon purification of encapsidated DNA from these vectors, use duplex ddPCR to probe these barcodes and quantify the concentration of transgene and genome-packed phages, in units of copies/mL **(Figure 1**).

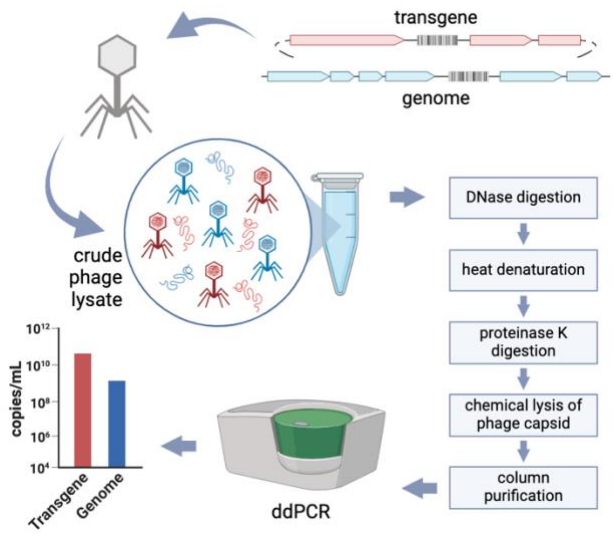


Figure 1. Phage vector characterization workflow using ddPCR.

Unique DNA barcodes are inserted into the transgene cassette and phage genome. Both DNA are then used to produce phage vectors, which typically consists of mixed population packed with either transgene or the genome DNA. Following a purification protocol that digests non-encapsidated DNA, phage packed-DNA is recovered and analyzed via ddPCR to determine the absolute quantity and proportions of transgene and genome-packed vector particles.

To choose barcodes with high compatibility, we focused on generating a pool of sequences with a G/C content between 40-60% and homopolymeric repeats <4 bps. We then filtered the barcode pool to remove those with common restriction endonuclease sites, with consensus promoter or ribosome binding site sequences, or with significant secondary structure. We then ranked the

filtered pool to minimize similarity to sequences in the NCBI nucleotide database. Of those with the lowest similarity to reported DNA sequences, pairs of remaining barcodes were ranked for orthogonality to each other. Of an initial pool of 10,000 candidates, we selected barcodes 8673 and 7898 (**Table 1**).

	Secondary Structure		Top BLAST Result			
Barcode	**Max MFE***	**Avg MFE***	**Alignment Length**	**Percent Identity**	**Query Coverage**	**Orthogonality†**
8673	-10.9	-5.090	28	96.4	9	250
7898	-11.8	-5.663	24	100	8	250

Table 1. Barcode metrics. *over a 50 bp sliding window. †defined by Hamming distance

ddPCR central composite design

Due to the sensitivity of ddPCR, reaction optimization is required to ensure accurate and reproducible readouts. The annealing temperature (T_a), primer concentration, probe concentration, ramp rate, and number of PCR cycles can all influence the readout of a ddPCR assay. We therefore decided to optimize

the duplex ddPCR assay using CCD (**Table 2**), with the goals of minimizing the relative error of the assay and maximizing the separation between positive and negative droplet populations for each barcode. To determine relative error, we determined the concentrations of two reference plasmids, each encoding a single barcode (either 8673 or 7898). We also calculated the theoretical concentration of barcode (copies/mL), which served as reference for comparison to the measured barcode concentrations.

Factor	Low	Central	High
T_a (°C)	57	60	63
[Primer] (nM)	450	900	1350
[Probe] (nM)	125	250	375
Ramp Rate (°C/s)	1	1.5	2
Cycles	30	40	50

Table 2. Central composite design factors.

Applying a least-squares model to the data, we found a moderate fit for both 8673 and 7898 relative errors (R^2 = 0.93 and 0.83, respectively) (**Figures 2A and 2B**), and a strong fit for both 8673 and 7898 droplet separation (R^2 = 0.96 and 0.98, respectively) (**Figures 2C and 2D**). In this model, the most significant main effects were the number of cycles ($p = 3.89 \times 10^{-6}$), followed by the annealing temperature ($p = 9.86 \times 10^{-4}$), the probe concentration ($p = 1.01 \times 10^{-3}$), and the ramp rate ($p = 1.30 \times 10^{-3}$). Interactions between the ramp rate and cycles ($p = 2.25 \times 10^{-3}$) and the probe concentration and cycles (p = 0.00995) were also significant (**Figure 2E**). The response surface graphs for relative error (**Figure 2F**) and separation (**Figure 2G**) as a function of ramp rate by cycles, and for relative error (**Figure 2F**) and separation (**Figure 2G**) as a function probe concentration by cycles appeared to indicate that the significance of these effects are primarily driven by the droplet separation. Notably, these graphs indicate a cycle number near 40 appears to minimize relative error for both 8673 and 7898.

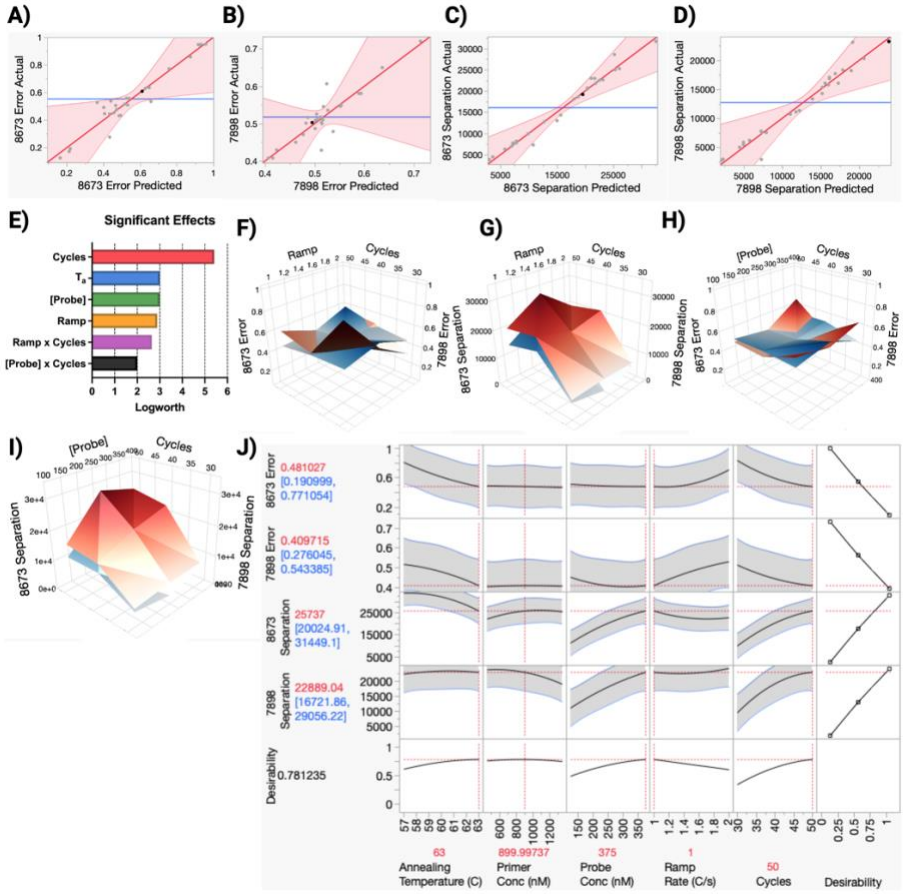


Figure 2. Central composite design modeling.

Actual *vs.* predicted model fit relative to the error of the assay for **(A)** transgene (8673) and **(B)** genome (7898) barcodes. Actual *vs.* predicted model fit to the separation of mean fluorescence intensity between positive and negative droplet populations of the assay for **(C)** transgene and **(D)** genome barcodes, respectively. **(E)** Plots of the log-transformation of the effect significance of design factors (T_a, annealing temperature;[Probe], probe concentration; Ramp, ramp rate). Response surface graphs of **(F)** relative error and **(G)** droplet separation as a function of the ramp rate x cycles; and **(H)** relative error and **(I)** droplet separation as a function of probe concentration x cycles (red, 8673; blue, 7898). **(J)** Plots of the maximized desirability function with respect to the 5 design factors (annealing temperature, primer concentration, probe concentration, ramp rate, and cycles) and 4 optimization goals (minimize relative error and maximize separation for the transgene and genome barcodes).

To determine factor settings that yield a balance of optimal relative error and droplet separation for both the 8673 and 7898 barcodes, we set a desirability function to minimize the relative errors and maximize the droplet separation for the 8673 and 7898 barcodes. A maximum desirability of 0.78 was achieved, corresponding to predicted relative errors of 48% and 41% and droplet separations of 25,737 and 22,889 a.u. for barcodes 8673 and 7898, respectively. The factor settings predicted to yield these values were a T_a = 63°C, a primer concentration = 900 nM, a probe concentration = 375 nM, a ramp rate of 1°C/s, and 50 PCR cycles (**Figure 2J**).

ddPCR assay performance characterization

Having characterized the effects of the PCR settings on assay performance, we next proceeded to assess the performance of the assay using the settings predicted to optimize the desirability function (herein, optimized ddPCR). To do so, we prepared eight dilutions of the reference plasmids in 5-fold intervals, ranging from a 6.44×10^5 to 8 copies/well. We then measured the concentration of these dilutions using optimized ddPCR and plotted against their predicted concentrations. In the dynamic range of the assay, we observed minimum relative errors of 51%

and 43%, closely mirroring the 48% and 41% relative errors predicted by the least-squares models (**Figure 3A and 3B**). This error was observed with high precision, indicating that it is systematic error in the assay and, thus, we next sought to correct for this using a calibration curve.

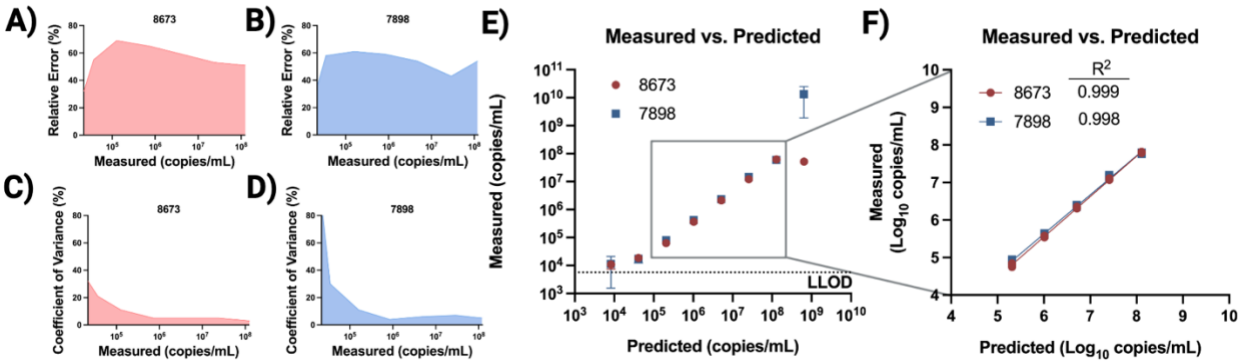


Figure 3. ddPCR assay performance characterization.

The relative error of the ddPCR measured concentrations, as compared to the predicted concentrations, for the **(A)** transgene and **(B)** genome barcodes. The coefficient of variance as a function of ddPCR measured **(C)** transgene and **(D)** genome barcode concentrations. **(E)** Concentrations of transgene (8673) and genome (7898) barcode DNA quantified by Qubit (predicted) compared to concentrations quantified by ddPCR (measured) with the lower limit of detection (LLOD) denoted. **(F)** Linear regression analysis of five of the concentrations in **(E)**.

We found a linear relationship between measured and predicted values at concentrations between 63 – 62,620 and 80 – 59,469 measured copies/well, with a precision (defined as the coefficient of variance) of 3-5% and 4-7% (**Figure 3C and 3D**), and a lower limit of detection (LLOD) of 5.66 and 6.66 copies/well, for 8673 and 7898 respectively (**Figure 3E**). These values correspond to a dynamic range of 6.34×10^4 – 6.26×10^7 and 7.98×10^4 – 5.95×10^7 measured copies/mL, with LLOD of 5,661 and 3,623 copies/mL, for 8673 and 7898, respectively **(Table 3)**. Performing linear regression on the data in the dynamic range of the assay revealed strong fits, with R^2 = 0.999 and 0.998 for 8673 and 7898, respectively (**Figure 3F**). With this strong fit, the following linear equations can be used to correct for the inherent error of the assay in the dynamic

range of measured template concentrations, where C_a is the true concentration and C_m is the measured concentration (**equations 1 and 2**).

$$C_a = 10^{(\log_{10} C_m + 0.9039)/1.075}$$

Equation 1. 8673 error correction.

$$C_a = 10^{(\log_{10} C_m + 0.6257)/1.042}$$

Equation 2. 7898 error correction.

	Transgene (8673) Barcode		Genome (7898) Barcode	
	copies/well	**copies/mL**	**copies/well**	**copies/mL**
Dynamic Range	63-62,620	6.34×10^4-6.26×10^7	80-59,469	7.98×10^4-5.95×10^7
LLOD	5.66	5661	6.66	3623
Precision*	3-5%		4-7%	

Table 3. ddPCR assay performance metrics. *in the dynamic range

Comparison of ddPCR to biological activity assays

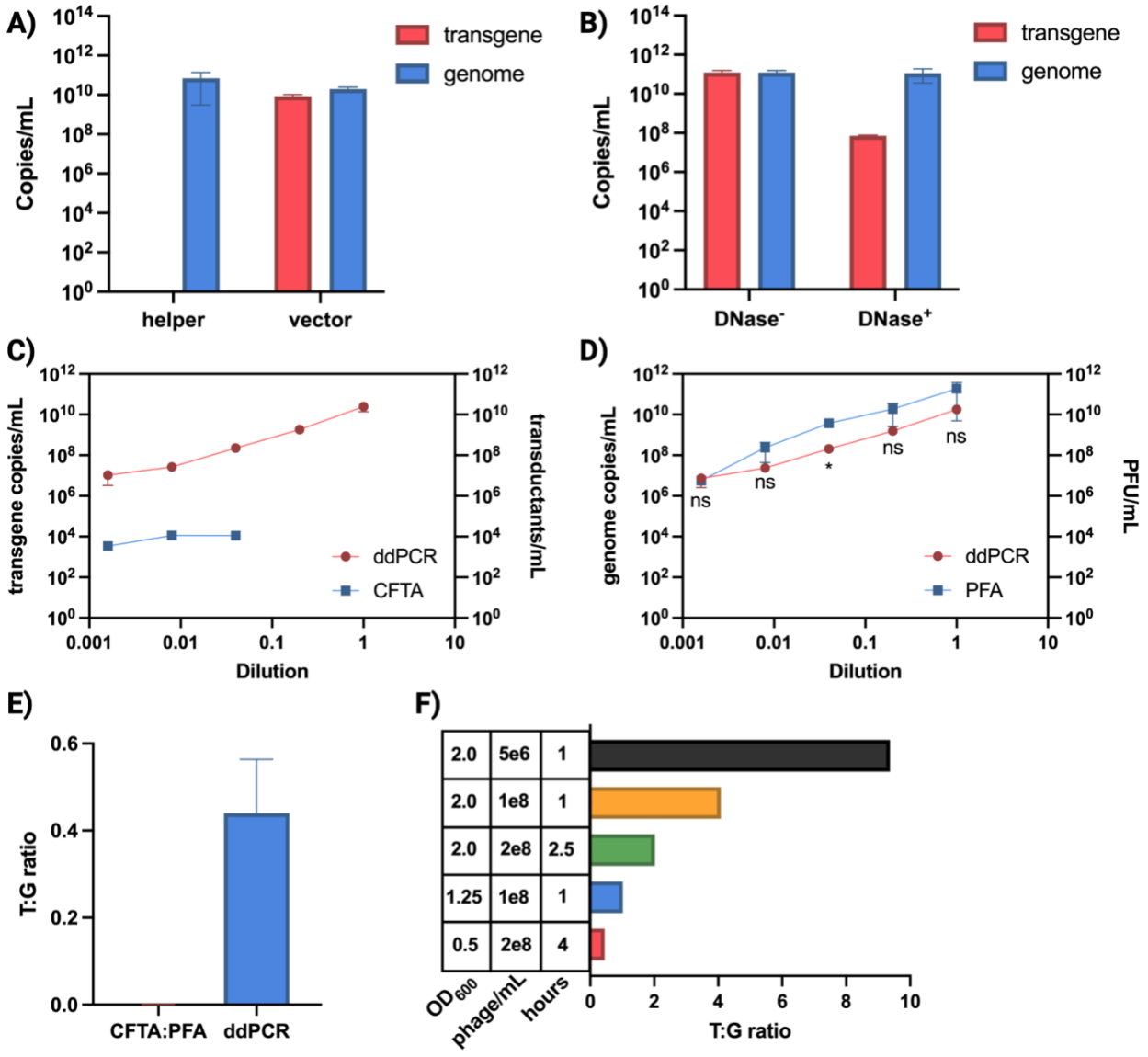
Assays that assess the biologically activity (i.e. cells transduced or killed) of a phage vector sample represent the *de facto* gold standard in phage characterization, but may not accurately reflect the true number of packed vector capsids present. To understand how the transgene/vector genome copies/mL titer compares to biological activity, we generated vectors based on the model phage T7. Briefly, the 8673 barcode was inserted into a plasmid containing a T7 packing signal and a kanamycin resistance cassette, whereas the 7898 barcode was inserted in the T7 genome in place of gene 17, which encodes the T7 tail fiber. Removal of this gene prevents any vector capsids packed with the genome from establishing a productive infection as tail-fiber deficient T7 particles lack the host-receptor binding domain(19). While tail-fiber competent vectors packed with the barcoded genomic DNA can still cause lysis if delivered to a target cell, the progeny from

284 this infection are tail fiber deficient, and thus unable to infect subsequent cells. *E. coli* containing
285 the barcoded transgene plasmid and a complementary plasmid expressing the T7 tail fiber were
286 then infected by barcoded T7 helper phage to generate the vectors.

287 Using the ddPCR assay optimized above, we first quantified the number of transgene and
288 genome DNA packed phages, with the ratio of these representing the purity of the phage vector
289 prep. The conventional process of generating T7 phage vectors involves infecting bacteria
290 possessing transgenic DNA of interest with a helper phage, which contains only genome-packed
291 capsids; the resulting T7 phage vector prep contains a mixture of both genome and transgene
292 packed phage particles. When samples of either T7 helper phage or T7 phage vectors were
293 analyzed, transgene-packed capsids were only detected in the vector sample, confirming high
294 assay specificity (**Figure 4A**). Next, to test the extent to which free DNA contributes to the ddPCR
295 signal, we spiked a helper phage sample with free transgene plasmid DNA. After DNA purification,
296 including a DNase treatment step to remove non-encapsidated free DNA, we observed a
297 reduction in transgene copies by >3 orders of magnitude, indicating that free DNA contributes to
298 <0.1% of the detected DNA copies (**Figure 4B**). This underscores the ability of our assay and
299 workflow to measure only transgene originating from phage particles.

300 We next assessed the vectors via biological activity assays, including colony formation
301 transduction assays (CFTA), and by a trans-complementation plaque formation assay (PFA) using
302 an *E. coli* strain with trans-complementary expression of the T7 tail fiber that allowed for plaque
303 visualization. For both, we tested the phage vector prep across 5 different dilutions at
304 5-fold intervals. The CFTA showed colony formation for only the 3 most dilute vector samples and,
305 notably, the colony counts at these 3 dilutions were nonlinear. In comparison, our optimized
306 ddPCR assay measured >3-orders of magnitude more transgene-packed vectors/mL than
307 transductants at each vector concentration and was able to quantify the 2 vector samples that
308 showed no transduction in the CFTA (**Figure 4C**). In contrast to the CFTA, there was no significant

309 difference observed between the PFA and the ddPCR assay in 4 out of 5 of the lysate dilutions
310 (**Figure 4D**). These ddPCR data were then used to determine the transgene to genome packed
311 vector ratio, which showed a transgene to genome packing ratio of 0.4 using these production
312 conditions (**Figure 4E**). When used to determine the packing ratio of phage vectors made using
313 a range of production conditions, the ddPCR assay revealed that packing ratios ranging from
314 ~0.5-9 transgene to genome- packed vector particles were achieved (**Figure 4F**). This validates
315 the ability of our assay to sensitively detect differences in ratios of transgene to genome-packed
316 vector particles among different preparations.



317

Figure 4. Characterization of phage vectors by ddPCR and biological activity assays.

(A) Concentrations of transgene and genome copies in DNA purified from bacteriophage lysates determined by ddPCR. **(B)** ddPCR quantified transgene and genome copies with and without DNase treatment to digest non-encapsidated DNA during DNA purification of a helper phage DNA spiked with free plasmid DNA. **(C)** Comparison of transgene-packed vector particle quantification by colony formation transduction assay (CFTA) and ddPCR. **(D)** Comparison of genome-packed vector particle quantification by plaque formation assay (PFA) and ddPCR. **(E)** The ratio of transgene-packed to genome-packed phage particles (T:G ratio) in the vector samples quantified in **(C)** and **(D)**. **(F)** The T:G ratios of phage vector samples made with varied production conditions: production strain cell density (OD_{600}), helper phage inoculum (phage/mL), and production infection time (hours).

Phage vectors represent a powerful new tool capable of genetically perturbing complex microbial communities with precision, while also offering the potential to facilitate entirely new therapeutic paradigms. To further their development, methods that can accurately characterize phage vectors are essential. For instance, clinical use of viral vectors for human gene therapy requires extensive characterization of safety, potency, purity and stability. At the core of each of these categories is accurate and reliable quantitative measurements of the number of vector capsids packed with the DNA of interest, as well as the abundance of any contaminating proteins and DNA. Such information is critical to ensure reproducible efficacy, safety and limited immunogenicity, to define the dose and/or dosing regimen, and to limit batch-to-batch variation in manufacturing. The same characterizations are almost certainly required by regulatory authorities to advance phage vectors into human studies.

qPCR and ddPCR are commonly used to quantify copies of vector-packed DNA when characterizing viral vectors used in human gene therapy. Although qPCR has traditionally been the more prevalent approach, ddPCR offers a number of advantages over qPCR. These include greater sensitivity of detection while simultaneously having less sensitivity to impurities that may disrupt PCR efficiency, and elimination of sources of error related to standard curves, which are

344 not needed for absolute quantitation(20). While ddPCR assays have been previously used to
345 quantify naturally occurring phages used as targeted antibiotics(21–23), there exist no established
346 ddPCR or qPCR assays for quantification of phage vectors and assessment of purity. We thus
347 sought to advance a ddPCR-based assay to address this gap.

348 Given the large number of input variables, we elected to utilize CCD to optimize the ddPCR
349 assay. Such response surface methodology was previously utilized in developing a ddPCR assay
350 for genetically modified crop certified reference materials23. We applied CCD to understand how
351 the assay sensitivity is impacted by changes in PCR annealing temperature, primer and probe
352 concentration, cycle ramp rate, and the number of PCR cycles, and to determine theoretically
353 optimal settings for these factors. We found, in comparison to droplet separation, assay accuracy
354 is relatively insensitive to changes in PCR conditions. Using the CCD-optimized settings, our
355 ddPCR assay offers a dynamic range of nearly 3 orders of magnitude for both the 8673 and 7898
356 barcodes (987-fold and 745-fold, respectively). Within the dynamic range, the high degree of
357 correlation between measured and known barcode concentrations allows calibration using the
358 corresponding equations of the regression lines (**equations 1 and 2**) to correct for the systematic
359 error of the assay, allowing for highly accurate quantitation of phage vector titers.

360 Phage vectors can be classified by both the underlying phage type (lytic or temperate)
361 and their replication competency. While a vector production system has been described for
362 temperate phages that yield vectors packed exclusively with the transgenic DNA of interest24,
363 many phage vectors to date, including all lytic phage-based systems, comprise of a mixed
364 population of both transgene and genome-packed vectors. Indeed, following the methods and
365 design of T7 phage vectors that are not capable of replication (and thus not at risk of horizontal
366 gene transfer)25,26, we show that there are substantial quantities of both transgene and genome-
367 packed vectors. This underscores the importance of an assay that can accurately enumerate both

368 populations, which we achieved here with high accuracy and sensitivity through barcodes
369 embedded in the two plasmids.

370 With this assay, there are a number of experimental procedures critical to ensuring accurate
371 quantitation. For instance, we found that DNase treatment during the DNA purification step
372 reduces free, non-encapsidated DNA by more than 3 orders of magnitude, making the latter
373 insignificant in calculating the final phage titers. To maximize assay precision and minimize the
374 LLOD, we found it critical to minimize freeze-thaw cycles of all reagents, and to prepare the assay
375 under a local fume extractor to reduce risk of contamination by aerosols.

376 We note important differences in titer estimates derived from conventional biological assays
377 such as CFTAs and PFAs *vs.* the ddPCR assay. For instance, ddPCR quantified transgene-
378 packed capsids at all 5 vector concentrations tested, whereas the the CFTA only yielded
379 transductants at the 3 most dilute concentrations, most likely due to overwhelming toxicity
380 associated with a higher number of genome-packed particles. Importantly, the number of
381 transductants measured was far lower than the actual number of transgene-packed phage vectors
382 determined by ddPCR at each of these concentrations. Conversly, the genome-packed titers
383 revealed by the ddPCR assay closely mirrored those given by the PFA—with any variation
384 between the two likely due to inefficiencies in the DNA purification process and errors inherent to
385 both the PFA and ddPCR assay—indicating that our optimized ddPCR assay is accurate with
386 purified vector DNA in addition to reference plasmids. Together with the genome-packed vector
387 data, the ddPCR assay indicated that genome-packed capsids outnumber transgene packed
388 capsids by ~2:1 in these particular vectors. This may explain the dramatic discrepancy between
389 the CFTA and ddPCR data, as well as the lack of observable transduction with the most
390 concentrated samples. Furthermore, when different production conditions are used, the ddPCR
391 assay revealed that a wide range of transgene to genome packing ratios could be achieved,
392 demonstrating the importance of such an assay in producing vectors with optimal efficacy.

Conclusions

The insights described here underscore why precise quantification of both transgene and vector genome-packed capsids in phage vector preparations is essential to a complete understanding of the vector system. In concert with data from biological activity assays, ddPCR readouts can help determine vector dosing, guide vector production, and ensure highly reproducible use both in preclinical and clinical studies. By carefully defining the error and precision of this assay, it can be used to quantify both the transgene and genome packed titers of phage vector samples with higher throughput and greater accuracy that previously existing biological activity assays. Furthermore, this work can be readily applied to any phage vector system of interest through the modular, highly-orthogonal barcodes developed here. With phage vectors poised to become versatile tools for microbiome engineering, we believe the approach reported here will likely be increasingly adopted and become a key assessment standard.

ASSOCIATED CONTENT

Supporting Information

Table S1. Synthesized DNA fragments used in this study (excel)

Table S2. Gibson assembly schemes (excel)

Table S3. Phage genome amplification primers (excel)

Table S4. ddPCR primer and probe sequences (excel)

Table S5. Central composite design conditions (excel)

Author Contributions

Conceptualization, P.J.V. and S.K.L.; Methodology, P.J.V.; Investigation, P.J.V., R.P., and J.L.; Formal Analysis, P.J.V.; Writing – Original Draft, P.J.V.; Writing – Review & Editing, P.J.V. and S.K.L.; Funding Acquisition, S.K.L.

ACKNOWLEDGMENTS

This work was supported by The David and Lucile Packard Foundation (2013-39274, SKL), and startup funds from UNC Eshelman School of Pharmacy. Figures were created with BioRender.com and GraphPad Prism.

REFERENCES

1. Strathdee SA, Hatfull GF, Mutalik VK, Schooley RT. Phage therapy: From biological mechanisms to future directions. Cell [Internet]. 2023 Jan 5 [cited 2023 Jan 5];186(1):17–31. Available from: http://www.cell.com/article/S0092867422014611/fulltext
2. Pires DP, Costa AR, Pinto G, Meneses L, Azeredo J. Current challenges and future opportunities of phage therapy. FEMS Microbiol Rev [Internet]. 2020 Nov 24 [cited 2023 Apr 23];44(6):684–700. Available from: https://academic.oup.com/femsre/article/44/6/684/5849005
3. Pires DP, Cleto S, Sillankorva S, Azeredo J, Lu TK. Genetically Engineered Phages: a Review of Advances over the Last Decade. Microbiology and Molecular Biology Reviews. 2016 Sep;80(3):523–43.
4. Voorhees PJ, Cruz-Teran C, Edelstein J, Lai SK. Challenges & opportunities for phage-based in situ microbiome engineering in the gut. Vol. 326, Journal of Controlled Release. Elsevier B.V.; 2020. p. 106–19.
5. Mimee M, Citorik RJ, Lu TK. Microbiome therapeutics — Advances and challenges. Vol. 105, Advanced Drug Delivery Reviews. Elsevier B.V.; 2016. p. 44–54.
6. Sheth RU, Cabral V, Chen SP, Wang HH. Manipulating Bacterial Communities by in situ Microbiome Engineering. Vol. 32, Trends in Genetics. Elsevier Ltd; 2016. p. 189–200.
7. Lee HL, Shen H, Hwang IY, Ling H, Yew WS, Lee YS, et al. Targeted Approaches for In Situ Gut Microbiome Manipulation. Genes (Basel) [Internet]. 2018 Jul 12 [cited 2022 Mar 2];9(7). Available from: /pmc/articles/PMC6071227/
8. Hsu BB, Way JC, Silver PA. Stable Neutralization of a Virulence Factor in Bacteria Using Temperate Phage in the Mammalian Gut. mSystems [Internet]. 2020 Feb 11 [cited 2022 Aug 30];5(1). Available from: https://journals.asm.org/doi/10.1128/mSystems.00013-20
9. Wright JF, Zelenaia O. Vector Characterization Methods for Quality Control Testing of Recombinant Adeno-Associated Viruses. Methods in Molecular Biology [Internet]. 2011 [cited 2024 Feb 11];737:247–78. Available from: https://link.springer.com/protocol/10.1007/978-1-61779-095-9_11
10. Martinez-Fernandez de la Camara C, McClements ME, Maclaren RE. Accurate Quantification of AAV Vector Genomes by Quantitative PCR. Genes (Basel) [Internet]. 2021 Apr 1 [cited 2024 Feb 24];12(4). Available from: /pmc/articles/PMC8074223/

11. Sanmiguel J, Gao G, Vandenberghe LH. Quantitative and Digital Droplet-Based AAV Genome Titration. Methods Mol Biol [Internet]. 2019 [cited 2024 Feb 11];1950:51. Available from: /pmc/articles/PMC7377043/
12. François A, Bouzelha M, Lecomte E, Broucque F, Penaud-Budloo M, Adjali O, et al. Accurate Titration of Infectious AAV Particles Requires Measurement of Biologically Active Vector Genomes and Suitable Controls. Mol Ther Methods Clin Dev [Internet]. 2018 Sep 9 [cited 2024 Feb 11];10:223. Available from: /pmc/articles/PMC6090651/
13. Aronson SJ, Bakker RS, Moenis S, van Dijk R, Bortolussi G, Collaud F, et al. A Quantitative In Vitro Potency Assay for Adeno-Associated Virus Vectors Encoding for the UGT1A1 Transgene. Mol Ther Methods Clin Dev [Internet]. 2020 Sep 11 [cited 2024 Feb 24];18:250–8. Available from: http://www.cell.com/article/S2329050120301248/fulltext
14. Wright JF. Manufacturing and characterizing AAV-based vectors for use in clinical studies. Gene Therapy 2008 15:11 [Internet]. 2008 Apr 17 [cited 2024 Feb 24];15(11):840–8. Available from: https://www.nature.com/articles/gt200865
15. Saragliadis A, Trunk T, Leo JC. Producing Gene Deletions in Escherichia coli by P1 Transduction with Excisable Antibiotic Resistance Cassettes. J Vis Exp [Internet]. 2018 Sep 1 [cited 2024 Feb 24];2018(139):58267. Available from: /pmc/articles/PMC6235078/
16. Tridgett M, Ababi M, Osgerby A, Ramirez Garcia R, Jaramillo A. Engineering Bacteria to Produce Pure Phage-like Particles for Gene Delivery. ACS Synth Biol [Internet]. 2021 Jan 15 [cited 2021 Mar 29];10(1):107–14. Available from: https://dx.doi.org/10.1021/acssynbio.0c00467
17. Ando H, Lemire S, Pires DP, Lu TK. Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing. Cell Syst. 2015 Sep 23;1(3):187–96.
18. Bio-Rad Laboratories. Droplet Digital ™ PCR Droplet Digital ™ PCR Applications Guide.
19. Dunne M, Prokhorov NS, Loessner MJ, Leiman PG. Reprogramming bacteriophage host range: design principles and strategies for engineering receptor binding proteins. Curr Opin Biotechnol. 2021 Apr 1;68:272–81.
20. Taylor SC, Laperriere G, Germain H. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. Scientific Reports 2017 7:1 [Internet]. 2017 May 25 [cited 2024 Mar 2];7(1):1–8. Available from: https://www.nature.com/articles/s41598-017-02217-x
21. Bichet MC, Patwa R, Barr JJ. Protocols for studying bacteriophage interactions with in vitro epithelial cell layers. STAR Protoc. 2021 Sep 17;2(3).
22. Morella NM, Yang SC, Hernandez CA, Koskella B. Rapid quantification of bacteriophages and their bacterial hosts in vitro and in vivo using droplet digital PCR. J Virol Methods. 2018 Sep 1;259:18–24.
23. Ács N, Gambino M, Brøndsted L. Bacteriophage Enumeration and Detection Methods. Front Microbiol. 2020 Oct 23;11:2662.
24. Dalmira FU, Melina PU, José-Benigno VT, Josefina LF, Raymundo GE, Abraham AS. Development, Optimization, and Evaluation of a Duplex Droplet Digital PCR Assay to

Quantify the T-nos/hmg Copy Number Ratio in Genetically Modified Maize. Anal Chem [Internet]. 2016 Jan 5 [cited 2024 Feb 11];88(1):812–9. Available from: https://pubs.acs.org/doi/full/10.1021/acs.analchem.5b03238

25. Lam CN, Mehta-Kolte MG, Martins-Sorenson N, Eckert B, Lin PH, Chu K, et al. A Tail Fiber Engineering Platform for Improved Bacterial Transduction-Based Diagnostic Reagents. ACS Synth Biol [Internet]. 2021 [cited 2021 Aug 1]; Available from: https://pubs.acs.org/doi/abs/10.1021/acssynbio.1c00036

26. Auster O, Globus R, Yosef I, Qimron U. Optimizing DNA transduction by selection of mutations that evade bacterial defense systems. RNA Biol [Internet]. 2019 Apr 3 [cited 2023 Feb 15];16(4):595. Available from: /pmc/articles/PMC6546373/

27. Yosef I, Goren MG, Globus R, Molshanski-Mor S, Qimron U. Extending the Host Range of Bacteriophage Particles for DNA Transduction. Mol Cell. 2017 Jun 1;66(5):721-728.e3.