

METHODOLOGY ARTICLE

Open Access

# BactQuant: An enhanced broad-coverage bacterial quantitative real-time PCR assay

Cindy M Liu<sup>1,2</sup>, Maliha Aziz<sup>1</sup>, Sergey Kachur<sup>1,5</sup>, Po-Ren Hsueh<sup>3</sup>, Yu-Tsung Huang<sup>4</sup>, Paul Keim<sup>1,2</sup> and Lance B Price<sup>1\*</sup>

## Abstract

**Background:** Bacterial load quantification is a critical component of bacterial community analysis, but a culture-independent method capable of detecting and quantifying diverse bacteria is needed. Based on our analysis of a diverse collection of 16 S rRNA gene sequences, we designed a broad-coverage quantitative real-time PCR (qPCR) assay—BactQuant—for quantifying 16 S rRNA gene copy number and estimating bacterial load. We further utilized *in silico* evaluation to complement laboratory-based qPCR characterization to validate BactQuant.

**Methods:** The aligned core set of 4,938 16 S rRNA gene sequences in the Greengenes database were analyzed for assay design. Cloned plasmid standards were generated and quantified using a qPCR-based approach. Coverage analysis was performed computationally using >670,000 sequences and further evaluated following the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines.

**Results:** A bacterial TaqMan<sup>®</sup> qPCR assay targeting a 466 bp region in V3-V4 was designed. Coverage analysis showed that 91% of the phyla, 96% of the genera, and >80% of the 89,537 species analyzed contained at least one perfect sequence match to the BactQuant assay. Of the 106 bacterial species evaluated, amplification efficiencies ranged from 81 to 120%, with  $r^2$ -value of >0.99, including species with sequence mismatches. Inter- and intra-run coefficient of variance was <3% and <16% for Ct and copy number, respectively.

**Conclusions:** The BactQuant assay offers significantly broader coverage than a previously reported universal bacterial quantification assay. BactQuant *in vitro* performance was better than the *in silico* predictions.

## Background

Scientists today are studying bacterial communities from diverse habitats, hosts, and health conditions based on the 16 S rRNA gene [1,2]. To date, most studies have focused on qualitative characterization based on the relative abundances of community bacterial groups [3-5]; however, quantitative characterization—i.e., measurement of the total bacterial load—provides valuable and complementary information when combined with these qualitative data [6]. Traditional culture-based approaches for quantifying bacterial load are inherently limited for assessing the complex bacterial communities that exist in many clinical and environmental samples. Likewise, standard culture-based methods are ineffective for quantifying many fastidious and uncultivable bacterial species [7].

Among culture-independent approaches, quantitative real-time PCR (qPCR) is currently best suited for measuring bacterial load, because of its intrinsic quantitative capability, ease of use, and flexibility in assay design [8,9]. Using the qPCR platform, we can design an assay capable of concurrently detecting and quantifying all unique bacteria that constitutes a complex community. Furthermore, by utilizing 16 S rRNA gene as the target of a broad coverage qPCR assay, results from the qPCR evaluation can be easily combined with 16 S rRNA gene-based qualitative characterization to fully describe the community of interest.

In the current paper, we present our design and validation of a broad-coverage quantitative real-time PCR assay—BactQuant—for quantifying 16 S rRNA gene copy number and estimating bacterial load. To accomplish this, we have employed a novel nucleotide distribution-based approach to effectively summarize a large 16 S rRNA gene sequence dataset for qPCR assay design. We further addressed a general limitation of the qPCR

\* Correspondence: lprice@tgen.org

<sup>1</sup>Division of Pathogen Genomics, Translational Genomics Research Institute, 3051 W. Shamrell Blvd., Suite 106, Flagstaff, AZ 86001 USA  
Full list of author information is available at the end of the article

platform—the normalization of in-run quantitative standards using fluorimetric or spectrometric methods—by developing an alternative qPCR-based method for quantifying plasmid standards. Lastly, we have complemented standard qPCR assay validation following MIQE guideline [10] with extensive *in silico* analysis using >670,000 16 S rRNA gene sequences from the Ribosomal Database Project [11].

## Methods

### Design of 16 S rRNA gene quantitative real-time PCR assay

Pre-aligned 16 S rRNA gene sequences (n = 4,938) were downloaded from the core set of the Greengenes database [12]. The alignment was analyzed to generate an output of nucleotide distribution—i.e., the summary of allele frequency at each nucleotide position in the 16 S rRNA gene multiple sequence alignment file—and diversity score using a 3% gap-filter setting and the Simpson's Diversity Index, respectively.

**Assay Design** The nucleotide distribution was examined to identify a conserved 500 bp region for assay design. In designing the assays, we applied the following rules: 1) primer sequences cannot have more than three degenerate bases and 2) the probe sequence cannot have any degenerate bases. The primer  $T_m$  was calculated using salt adjusted calculation from the online tool OligoCalc [13] and the probe  $T_m$  was calculated using the Primer Probe Test Tool for TaqMan<sup>®</sup> MGB quantification from the Primer Express<sup>®</sup> Software for Real-Time PCR version 3.0 (Applied Biosystems, Carlsbad, CA, USA) (Table 1).

### Computational analysis of assay specificity and coverage

- A. Specificity analysis. Specificity check was performed in GenBank using megablast against human, mouse, and fungal sequences from the nucleotide collection (nr/nt) [14].
- B. Collection and identification of bacterial 16 S rRNA gene sequence eligible for *in silico* coverage analysis. All 16 S rRNA gene sequence data used in the *in silico* coverage analysis were downloaded from the Ribosomal Database Project (RDP) Release 10 Update 20 [11]. Briefly, all bacterial 16 S rRNA gene sequences that were of “Good” quality and had a

length of 1200 bp or greater were extracted from RDP, along with taxonomic metadata and sequence IDs. Additionally, the *Escherichia coli* position data was kindly provided by staff at the RDP. The downloaded sequences were filtered based on *E. coli* position. Only sequences with data present in the qPCR assay amplicon of interest were considered to be eligible for sequence matching for the particular qPCR assay. Numerical and taxonomic coverage analysis was performed for the BactQuant assay and a published qPCR assay [15] by developing a web service for the RDP Probe Match Tool for sequence matching.

- C. Overview of sequence matching analysis for determining assay coverage. All sequence matching for the *in silico* coverage analysis was performed using two conditions: a) perfect match of full-length primer and probe sequences and b) perfect match of full-length probe sequence and the last 8 nucleotides of primer sequences at the 3' end. For each sequence matching condition, the *in silico* coverage analysis was performed at three taxonomic levels: phylum, genus, and species, as well as for all sequences eligible for sequence matching. The remaining taxonomic levels were omitted due to the large amounts of missing and inconsistent data. Details of *in silico* coverage analyses are as follows:
- D. Numerical coverage analysis. At each analysis level, unique operational taxonomic unit (OTU), i.e., each unique taxonomic group ranging from unique phyla to unique species, containing at least one sequence that is a sequence match (i.e., “match”) for all three components of the assay of interest were identified using the following requirement: [Forward Primer Perfect Match](union)[Reverse Primer Perfect Match](union)[Probe Perfect Match]. The *in silico* coverage analysis was performed in a stepwise fashion, beginning with all eligible sequences, then proceeding to analysis at the species-, genus-, and phylum-level. At each step, the taxonomic identification of each sequence was generated by concatenation of relevant taxonomic data (e.g., for species-level analysis, a unique taxonomic identification consisting of concatenated Phylum-Genus-species name was considered as one unique species). The sequence IDs were used in lieu of a

**Table 1 Primer and probe sequences of BactQuant, the new 16 S rRNA gene-based quantitative real-time PCR (bold letters denotes degenerate base)**

BactQuant		$T_m$ (°C)	<i>E. coli</i> region
Forward Primer	5'- CCTACGGG <b>D</b> GGC <b>W</b> GCA-3'	55.9–58.4	341–356
Reverse Primer	5'- GGACTAC <b>HV</b> GGGT <b>M</b> TCTAATC -3'	57.5–63.3	786–806
Probe	(6FAM) 5'-CAGCAGCCGCGTA-3' (MGBNFQ)	68.0	519–532

taxonomic identification for the first analysis step, which included all eligible sequences. The stepwise numerical coverage analysis was performed as follows: all eligible sequences underwent sequence matching with all three components of the assays of interest using a select matching condition (i.e., the stringent or the relaxed criterion). The sequence IDs of matched sequences were assigned and binned as Assay Perfect Match sequence IDs. For this first analysis step, the numerical coverage was calculated using the total number of sequences with Assay Perfect Match sequence IDs as the numerator and the total number of eligible sequences as the denominator. Next, at the species-level, all sequences assigned as Assay Perfect Match sequence IDs were dereplicated based on the concatenated Phylum-Genus-*species* taxonomic identifications. Species-level numerical coverage was then calculated using the total number of dereplicated taxonomic identifications as the numerator. Denominator was calculated using the dereplicated Phylum-Genus-*species* taxonomic identifications from all eligible sequences. As a result of the logic of this analysis pipeline, a species (i.e., a group of sequences sharing the same unique Phylum-Genus-*species* designation) was considered an assay sequence match and thus “covered”, when at least one Assay Perfect Match sequence ID was in the species group. The numerical coverage analysis was repeated on the genus-level using the dereplicated Phylum-Genus taxonomic identifications from the Assay Perfect Match sequence IDs bin (numerator) and from all eligible sequences (denominator), and lastly, on the phylum-level using Phylum taxonomic identifications. To facilitate calculation of assay coverage, two ambiguous phyla, “Bacteria Insertia Sedis” and “Unclassified Bacteria” were excluded from the phylum-level analysis. Sequences with genus, species, and strain names containing “unclassified” were included in the numerical coverage analyses due to their high abundance.

- E. Taxonomic coverage analysis. The *in silico* taxonomic coverage analysis was performed to generate a detailed output consisting of the taxonomic identifications that were covered or “uncovered” (i.e., no sequence match) at multiple taxonomic levels. A step-wise approach was again utilized for this analysis, beginning with all eligible sequences, performed as follows: First, the Assay Perfect Match sequence IDs were subtracted from the sequence IDs from all eligible sequences, with the resultant sequences assigned and binned as Assay Non-Perfect Match sequence IDs. Next, on the species-level, the Phylum-Genus-*species*

taxonomic identifications of all eligible sequences was first dereplicated, from which the “covered” species taxonomic identifications were subtracted. Species-level taxonomic coverage was then presented as a list of concatenated taxonomic identification of the covered and uncovered species. This was repeated with the genus- and phylum-level taxonomic identifications for genus- and phylum-level taxonomic coverage analyses. Output of taxonomic identifications from analysis using all eligible sequences was not presented in this manuscript due to its extensive size but is available in Additional file 1: Figure S1.

- F. Assay comparison using results from the *in silico* analyses. Results from the *in silico* analyses were summarized for assay comparison as follows: The numerical coverage for the BactQuantand published qPCR assays were calculated at three taxonomic levels, as well as for all eligible sequences using both sequence matching conditions and presented as both the numerator and denominator, and percent covered calculated as the numerator divided by the denominator. This was presented in Table 2. Additional comparison of the taxonomic coverage was performed by superimposing the genus-level numerical coverage of the BactQuant assay for each phylum onto a maximum parsimony phylogenetic tree. Construction of the phylum-level phylogenetic tree was performed using MEGA4 with representative full-length 16 S rRNA gene sequences from each of the 34 phyla analyzed [16]. In addition, each phylum was annotated as not covered or poorly covered by the published qPCR assay if the phylum was uncovered or if >50% of the genera within the phylum were uncovered, respectively. A list of the uncovered genera by phylum for the BactQuant assay was also generated. Comparison results using the stringent and relaxed criterion were presented in Figure 1 and Additional file 2: FigureS1, respectively.

#### Quantification and normalization of cloned plasmid standards

**Overview** To obtain accurately quantified plasmid standards for validation the BactQuant assay, a  $10^9$  copies/ $\mu$ l plasmid stock was quantified using a qPCR assay targeting portion of the vector using the second derivative maximum analysis algorithm on the LightCycler platform. The resultant crossing point value (i.e., Cp-value) is used in plasmid normalization. The details are as follows:

**Generation of normalized 16 S rRNA gene plasmid standards** Amplification of the full 16 S rRNA gene was performed using *E. coli* genomic DNA as the template and 16 S rRNA gene primers 27 F and 1492R as

**Table 2 Results from numerical coverage analysis performed by comparing primer and probe sequences from BactQuant and the published qPCR assays against >670,000 16 S rRNA gene sequences from RDP**

	BactQuant	Published qPCR Assay	Coverage Improvement
A. Perfect match using full length primers and probe			
Phyla	<b>91.2%</b> (31/34)	<b>61.8%</b> (21/34)	+ 29.4%
Genus	<b>96.2%</b> (1778/1849)	<b>80.3%</b> (1485/1849)	+15.8%
Species*	<b>83.5%</b> (74725/89537)	<b>66.3%</b> (59459/89646)	+17.2%
All Sequences*	<b>78.0%</b> (524118/671595)	<b>60.9%</b> (409584/672060)	+17.1%
B. Perfect match using 8-nt primers with full length probe			
Phyla	<b>91.2%</b> (31/34)	<b>67.7%</b> (23/34)	+23.5%
Genus	<b>97.7%</b> (1806/1849)	<b>82.1%</b> (1518/1849)	+15.6%
Species*	<b>89.1%</b> (79759/89537)	<b>70.9%</b> (63533/89646)	+18.2%
All Sequences*	<b>84.4%</b> (566685/671595)	<b>65.6%</b> (441017/672060)	+18.8%

The *in silico* analysis was performed using two sequence matching conditions.

\*The difference in number of sequences eligible for *in silico* evaluation is due to the difference in primer lengths and locations of the two assays.

previously described [17]. Visualization of PCR amplicon was performed using gel electrophoresis with SYBR 2% agarose gel. The resultant PCR amplicons were immediately used as the target gene insert with the TOPO<sup>®</sup> TA Cloning<sup>®</sup> Kit (with pCR<sup>®</sup>2.1 TOPO<sup>®</sup> vector) (Invitrogen Corp., Carlsbad, CA, USA) following the manufacturer's instructions. The resultant propagated cloned plasmids were purified using the QIAprep Spin Miniprep Kit (Qiagen Inc., Valencia, CA, USA). Sequence verification of the purified plasmids containing the 16 S rRNA gene insert was performed with capillary electrophoresis using BigDye<sup>®</sup> Terminator v3.1 Cycle Sequencing Kit on the 3130 Genetic Analyzer platform (Applied Biosystems, Carlsbad, CA, USA). Quantification of the cloned plasmids was performed by analyzing three 10-fold dilutions using the vector qPCR assay. Normalization was performed using the dilution factor  $2^{\Delta C_p}$ , where  $\Delta C_p = 10 - (C_p \text{ value of non-normalized cloned plasmids})$ .

#### Pan-bacterial qPCR assay optimization and initial specificity check

**Assay optimization** Using the normalized plasmid standards, different primer and probe titrations were tested on the on the 7900HT Real Time PCR System (Applied Biosystems) and evaluated based on reaction efficiency and assay dynamic range for 10  $\mu$ l and 5  $\mu$ l reaction volumes. For 10  $\mu$ l and 5  $\mu$ l reactions, the optimized conditions included 1  $\mu$ l of template into 9  $\mu$ l and 4  $\mu$ l of reaction mix, respectively, with the final reaction containing 1.8  $\mu$ M of each forward and reverse primer, 225 nM the TaqMan<sup>®</sup> probe, 1X Platinum<sup>®</sup> Quantitative PCR

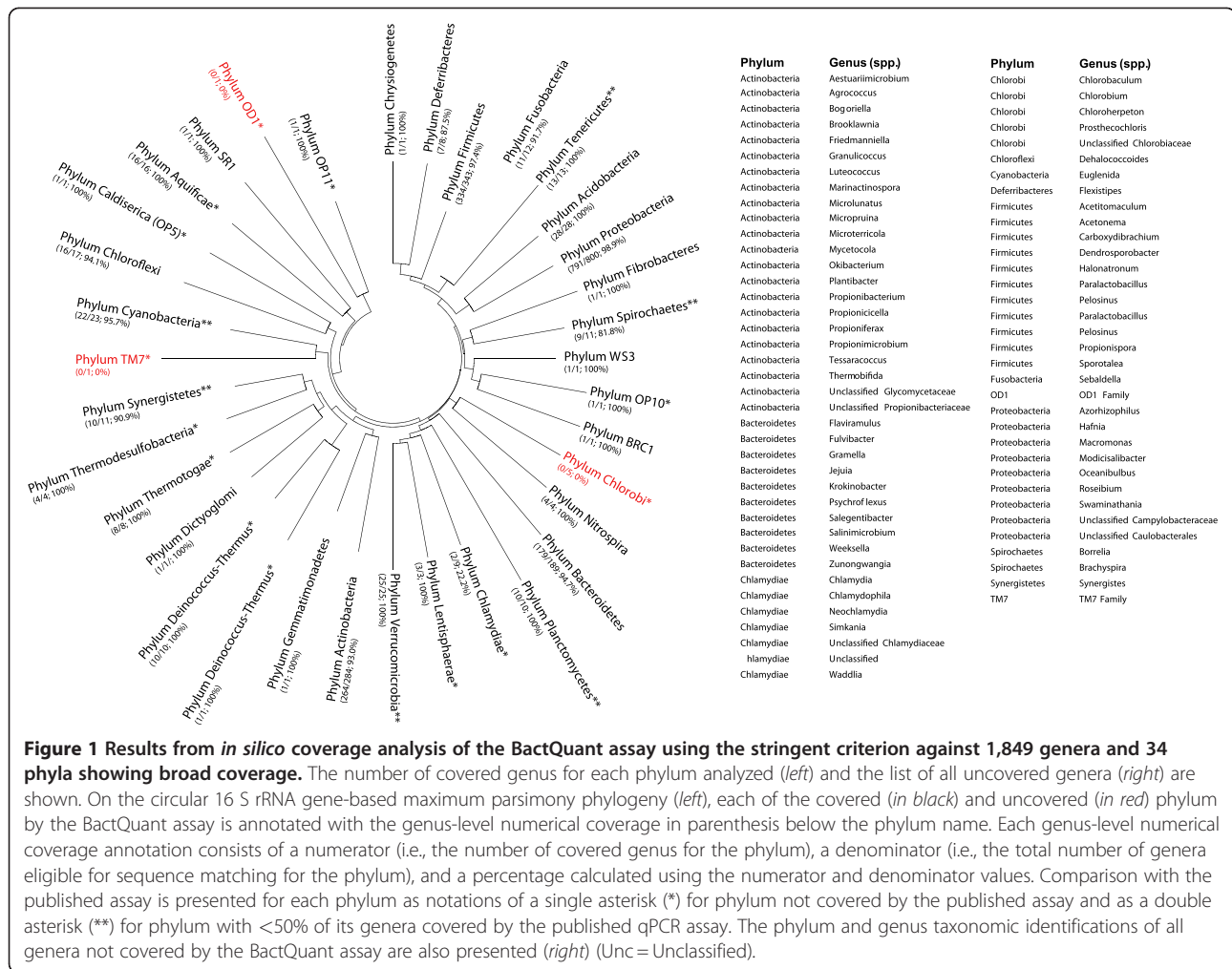
SuperMix-UDG w/ROX (Invitrogen Corp.) and molecular-grade water. Irrespective of reaction volume, each experiment included an in-run standard curve ( $10^2$ – $10^8$  in 10-fold serial dilutions) and no-template controls performed in triplicate. Amplification and real-time fluorescence detections were performed on the 7900HT Real Time PCR System (Applied Biosystems) using the following PCR conditions: 3 min at 50°C for UNG treatment, 10 min at 95°C for *Taq* activation, 15 s at 95°C for denaturation and 1 min at 60°C for annealing and extension x 40 cycles. Cycle threshold value (i.e., Ct value) for each 16 S qPCR reaction were obtained using a manual Ct threshold of 0.05 and automatic baseline in the Sequence Detection Systems v2.3 software (Applied Biosystems).

#### Initial specificity check against human and fungal genomic DNA

Using the optimized assay condition, the newly designed assay was tested against 1 ng, 100 pg, and 10 pg of human genomic DNA (Promega, Madison, WI, USA), *C. albicans* genomic DNA (American Type Culture Collection, Manassas, VA, USA), the normalized plasmid standards in triplicate reactions.

#### Laboratory analysis of assay performance using diverse bacterial genomic DNA

To assess our assay performance against diverse bacteria, we tested our assay against a diverse collection of bacterial genomic DNA to determine the assay efficiency and correlation coefficients. The details are as follows:



**Bacterial strains** *Arsenophonus nasoniae* ATCC 49151, *Budvicia aquatica* ATCC 51341, *Buttiauxella gaviniae* ATCC 51604, *Cedecea davisae* ATCC 33431, *Cellvibrio gilvus* ATCC13127, *Citrobacter freundii* ATCC 8090, *Clostridium difficile* ATCC 9689, *Cronobacter aerogenes* ATCC 13048, *Ewingella americana* ATCC 33852, *Edwardsiella tarda* ATCC 15947, *Escherichia vulneris* ATCC 33821, *Hafnia alvei* ATCC 29926, *Ewingella americana* ATCC 33852, *Klebsiella oxytoca* ATCC 49131, *Kluyvera ascorbata* ATCC 33433, *Leclercia adecarboxylata* ATCC 700325, *Leminorella richardii* ATCC 33998, *Moellerella wisconsensis* ATCC 35621, *Morganella morgani* ATCC 25830, *Obesumbacterium proteus* ATCC 12841, *Pantoea agglomerans* ATCC 27155, *Photorhabdus asymbiotica* ATCC 43950, *Plesiomonas shigelloides* ATCC 14029, *Pragia fontium* ATCC 49100, *Proteus mirabilis* ATCC 29906, *Providencia rustigianii* ATCC 33673, *Pseudomonas aeruginosa* ATCC 27853, *Pseudomonas andersonii* ATCC BAA-267, *Pseudomonas anguilliseptica* ATCC 33660, *Pseudomonas azotofixans* ATCC BAA-1049, *Pseudomonas fragi* ATCC 4973,

*Pseudomonas lundensis* ATCC 49968, *Pseudomonas luteola* ATCC 43273, *Pseudomonas mendocina* ATCC 25411, *Pseudomonas monteilii* ATCC 700476, *Pseudomonas mosselii* ATCC BAA-99, *Pseudomonas otitidis* ATCC BAA-1130, *Pseudomonas pseudoalcaligenes* ATCC 17440, *Pseudomonas putida* ATCC 12633, *Pseudomonas stutzeri* ATCC 17588, *Pseudomonas taetrolens* ATCC 4683, *Rahnella aquatilis* ATCC 33071, *Raoultella ornithinolytica* ATCC 31898, *Shigella dysenteriae* ATCC 13313, *Salmonella enterica* ATCC 13076, *Serratia liquefaciens* ATCC 27592, *Tatumella ptyseos* ATCC 33301, *Trabulsiella guamensis* ATCC 49492, *Yersinia enterocolitica* ATCC 9610, and *Yokenella regensburgi* ATCC 43001 were obtained from the American Type Culture Collection (Manassas, VA, USA). Bacterial propagation and enrichment were performed under the appropriate condition for each bacterial strain following ATCC recommendations.

**Extraction of bacterial genomic DNA** Extraction using the enriched broth was performed using ZR Fungal/

Bacterial DNA MiniPrep™ (Zymo Research, Irvine, CA, USA) following the manufacturer's instruction. Elution of the purified genomic DNA was performed using 100 µl of 1X TE buffer.

**Other sources of bacterial genomic DNA** Genomic DNA from *Bacteroides fragilis* ATCC 25285, *Bacteroides ureolyticus* GS-15 ATCC 43606, *Borrelia burgdorferi* strain B31 ATCC 35210, *Campylobacter jejuni* ATCC 700819, *Chlamydia trachomatis* ATCC VR-348B, *Chlamydophila pneumoniae* ATCC VR-1360D, *Fusobacterium nucleatum* ATCC 25586, *Pectobacterium atrosepticum* ATCC BAA-672, *Pseudomonas syringae* ATCC 11355, *Streptomyces violaceoruber* ATCC BAA-471, *Thermus thermophilus* ATCC BAA-163, *Treponema denticola* ATCC 35405, and *Vibrio cholera* ATCC 39315 were obtained from the American Type Culture Collection. DNA from *Mycobacterium avium*, subsp. *Avium*, *Mycobacterium abscessus*, *Mycobacterium bovis*, *Mycobacterium chelonae*, *Mycobacterium gastri*, *Mycobacterium goodii*, *Mycobacterium fortuitum*, *Mycobacterium kansasii*, *Mycobacterium marinum*, *Mycobacterium nonchromogenicum*, *Mycobacterium phlei*, *Mycobacterium smegmatis*, *Mycobacterium vaccae*, and *Mycobacterium xenopi* were kindly provided by National Taiwan University, Taipei, Taiwan. DNA from clinical isolates of *Acinetobacter baumannii*, *Klebsiella pneumoniae*, *Burkholderia pseudomallei*, *Coxiella burnetii*, *Enterobacter cloacae*, *Enterococcus faecium*, *Escherichia coli*, *Francisella tularensis*, *Haemophilus influenzae*, *Legionella pneumophila*, *Listeria monocytogenes*, *Moraxella catarrhalis*, *Neisseria gonorrhoeae*, *Pseudomonas aeruginosa*, *Salmonella enterica* subsp. *enterica* serovar *gallinarum*, *Staphylococcus arlettae*, *Staphylococcus capitis*, *Staphylococcus cohnii*, *Staphylococcus epidermidis*, *Staphylococcus equorum*, *Staphylococcus hominis*, *Staphylococcus haemolyticus*, *Staphylococcus kloosii*, *Staphylococcus lugdunensis*, *Staphylococcus saprophyticus*, *Staphylococcus xylosum*, *Streptococcus agalactiae*, *Streptococcus pneumoniae*, and *Viridans Streptococcus* and were kindly provided by a project supported by NIH/NIAID U01AI066581 at the Translational Genomics Research Institute, Flagstaff, AZ, USA.

**Experimental design** For sensitivity and efficiency analysis, bacterial genomic DNA from each species was analyzed in three 10-fold serial dilutions in triplicate reactions using the optimized 16 S qPCR conditions as described above.

**Data analysis** For each species tested, reaction efficiency and correlation coefficient were calculated using the data from tests against three 10-fold serial dilutions and presented in Table 3. Sequence comparison analysis was

performed by aligning the assay primer and probe sequences with 16 S rRNA gene sequences of the five uncovered species: *Borrelia burgdorferi* (Genbank Accession No. X98226), *Cellvibrio gilvus* (Genbank Accession No. GU827555.1), *Escherichia vulneris* (Genbank Accession No. AF530476), *Chlamydia trachomatis* (Genbank Accession No. NR025888), and *Chlamydophila pneumoniae* (Genbank Accession No. CPU68426) in SeqMan®. Amplification profile of the five uncovered species were annotated with results from the sequence comparison and presented in Additional file 3: Figure S3A-E.

#### Laboratory quantitative assay validation using pure plasmid standards and mixed templates

**Assay quantitative validation** For the assay quantitative validation, we followed the Minimum Information for publication of Quantitative real-time PCR Experiments, or the MIQE guidelines whenever applicable [10]. The MIQE guidelines were complemented with additional tests to determine assay performance in the presence of background fungal and human genomic DNA. In our experimental design, we included seven template conditions: plasmid standards alone and plasmid standards with 0.5 ng *C. albicans* genomic DNA (ATCC) and with 0.5 ng, 1 ng, 5 ng, and 10 ng of human genomic DNA per reaction in 10 µl reactions and plasmid standards alone in 5 µl reactions. For each condition assessed, three qPCR runs were performed to assess reproducibility, or inter-run variability. In each run, three replicate standard curves were tested across the 384-well plate to assess repeatability, or intra-run variability. All reactions were performed in triplicates.

**Data analysis** Using the data generated, the following assay parameters were calculated: 1) inter-run assay coefficient of variation (CoV) for copy number and Ct value, 2) average intra-run assay CoV for copy number and Ct value, 3) assay dynamic range, 4) average reaction efficiency, and 5) correlation coefficient ( $r^2$ -value). The limit of detection was not defined for the pure plasmid standards experiments due to variability in reagent contamination. At each plasmid standard concentration, the Ct standard deviation across all standard curves over three runs was divided by the mean Ct value across all standard curves over three runs to obtain the inter-run assay CoV. The CoV from each standard curve from each run (i.e., nine CoV were used in the calculation for each condition tested) were used to calculate the average and the standard deviation of the intra-run CoV. Linear regression of each standard curve across the full dynamic range was performed to obtain the slope and correlation coefficient values. The slope was used to calculate the reaction efficiency using  $\text{Efficiency} = 10^{(-1/\text{slope})} - 1$ . Of note, for each triplicate reaction with Ct standard deviation

**Table 3 The efficiency and  $r^2$ -value results from laboratory evaluation of the BactQuant assay using genomic DNA from ATCC strains and clinical isolates belonging to 106 unique bacterial species spanning eight bacterial phyla**

Species Name	Reaction efficiency	$r^2$ -value
<i>Streptomyces violaceoruber</i>	93%	>0.999
<i>Mycobacterium abscessus</i>	110%	>0.999
<i>Mycobacterium bovis</i>	106%	>0.996
<i>Mycobacterium chelonae</i>	101%	>0.999
<i>Mycobacterium gastri</i>	104%	>0.999
<i>Mycobacterium goodnae</i>	104%	>0.999
<i>Mycobacterium fortuitum</i>	93%	>0.999
<i>Mycobacterium kansasii</i>	107%	>0.999
<i>Mycobacterium marinum</i>	110%	>0.990
<i>Mycobacterium nonchromogenicum</i>	101%	>0.999
<i>Mycobacterium phlei</i>	104%	>0.999
<i>Mycobacterium smegmatis</i>	105%	>0.999
<i>Mycobacterium vaccae</i>	120%	>0.999
<i>Mycobacterium xenopi</i>	112%	>0.999
<i>Bacteroides ureolyticus</i>	92%	>0.999
<i>Bacteroides fragilis</i>	82%	>0.993
<i>Chlamydia trachomatis</i>	N/A	N/A
<i>Chlamydophila pneumoniae</i>	N/A	N/A
<i>Thermus thermophilus</i>	97%	>0.999
<i>Clostridium difficile</i>	88%	>0.987
<i>Listeria monocytogenes</i>	104%	>0.999
<i>Staphylococcus arlettae</i>	96%	>0.998
<i>Staphylococcus capitis</i>	95%	>0.993
<i>Staphylococcus cohnii</i>	104%	>0.999
<i>Staphylococcus epidermidis</i>	96%	>0.999
<i>Staphylococcus equorum</i>	85%	>0.997
<i>Staphylococcus hominis</i>	108%	>0.999
<i>Staphylococcus haemolyticus</i>	90–104%	>0.999
<i>Staphylococcus kloosii</i>	98%	>0.999
<i>Staphylococcus lugdunensis</i>	94%	>0.999
<i>Staphylococcus saprophyticus</i>	87–98%	>0.999
<i>Staphylococcus xylosus</i>	81–100%	>0.999
<i>Streptococcus agalactiae</i>	98%	>0.998
<i>Streptococcus pneumoniae</i>	98%	>0.999
<i>Streptococcus viridans</i>	103%	>0.999
<i>Enterococcus faecium</i>	91–111%	>0.999
<i>Enterococcus faecalis</i>	90–100%	>0.998
<i>Fusobacterium nucleatum</i>	90%	>0.999
<i>Burkholderia pseudomallei</i>	103%	>0.999
<i>Coxiella burnetii*</i>	100%	>0.998

**Table 3 (Continued)**

<i>Francisella tularensis</i>	100%	>0.999
<i>Legionella pneumophila</i>	98%	>0.999
<i>Neisseria gonorrhoeae</i>	95%	>0.997
<i>Pseudomonas aeruginosa</i>	90–100%	>0.999
<i>Pseudomonas mendocina</i>	93%	>0.999
<i>Pseudomonas andersonii</i>	90%	>0.999
<i>Pseudomonas otitidis</i>	93%	>0.999
<i>Pseudomonas stutzeri</i>	86%	>0.999
<i>Pseudomonas monteilli</i>	88%	>0.999
<i>Pseudomonas azotofixans</i>	84%	>0.999
<i>Pseudomonas mosselii</i>	92%	>0.999
<i>Pseudomonas luteola</i>	91%	>0.999
<i>Pseudomonas putida</i>	90%	>0.999
<i>Pseudomonas fluorescens</i>	96%	>0.999
<i>Pseudomonas taetrolens</i>	89%	>0.999
<i>Pseudomonas fragi</i>	93%	>0.999
<i>Pseudomonas syringae</i>	95%	>0.999
<i>Pseudomonas pseudoalcaligenes</i>	93%	>0.999
<i>Pseudomonas lundensis</i>	93%	>0.999
<i>Pseudomonas anguilliseptica</i>	93%	>0.999
<i>Cellvibrio gilvus</i>	92%	>0.999
<i>Acinetobacter baumannii</i>	100–105%	>0.999
<i>Arsenophonus nasoniae</i>	87%	>0.998
<i>Budvicia aquatica</i>	88%	>0.999
<i>Buttiauxella gaviniae</i>	107%	>0.999
<i>Cedecea davisae</i>	97%	>0.999
<i>Citrobacter freundii</i>	95%	>0.999
<i>Cronobacter sakazakii</i>	96%	>0.999
<i>Edwardsiella tarda</i>	106%	>0.999
<i>Enterobacter cloacae</i>	89–111%	>0.999
<i>Enterobacter aerogenes</i>	107%	>0.998
<i>Escherichia vulneris</i>	93%	>0.999
<i>Escherichia coli</i>	91–96%	>0.999
<i>Ewingella americana</i>	97%	>0.999
<i>Haemophilus influenzae</i>	91–110%	>0.999
<i>Hafnia alvei</i>	93%	>0.999
<i>Klebsiella oxytoca</i>	93%	>0.999
<i>Klebsiella pneumoniae</i>	95–100%	>0.999
<i>Kluyvera ascorbata</i>	100%	>0.999
<i>Leclercia adecarboxylata</i>	93%	>0.999
<i>Leminorella richardii</i>	94%	>0.999
<i>Moellerella wisconsensis</i>	93%	>0.999
<i>Moraxella catarrhalis</i>	91–106%	>0.999
<i>Morganella morganii</i>	95%	>0.999
<i>Obesumbacterium proteus</i>	114%	>0.994

**Table 3** (Continued)

<i>Pantoea agglomerans</i>	93%	>0.999
<i>Pectobacterium atrosepticum</i>	90%	>0.999
<i>Photorhabdus asymbiotica</i>	96%	>0.999
<i>Plesiomonas shigelloides</i>	93%	>0.999
<i>Pragia fontium</i>	100%	>0.998
<i>Proteus mirabilis</i>	98%	>0.999
<i>Providencia rustigianii</i>	93%	>0.999
<i>Rahnella aquatilis</i>	92%	>0.999
<i>Raoultella ornithinolytica</i>	94%	>0.999
<i>Salmonella enterica</i>	101%	>0.999
<i>Salmonella enterica subsp. enterica serovar gallinarum</i>	95%	>0.998
<i>Serratia liquefaciens</i>	94%	>0.999
<i>Shigella dysenteriae</i>	98%	>0.999
<i>Tatumella ptyseos</i>	101%	>0.999
<i>Trabulsiella guamensis</i>	95%	>0.999
<i>Yokenella regensburgei</i>	96%	>0.999
<i>Yersinia enterocolitica</i>	98%	>0.999
<i>Campylobacter jejuni</i>	89%	>0.999
<i>Vibrio cholerae</i>	85%	>0.996
<i>Borrelia burgdorferi</i>	90%	>0.999
<i>Treponema denticola</i>	82%	>0.999

\*No 16 S rRNA gene sequence available in the Ribosomal Database Project.

>0.3, the triplicates were compared and if a clear outlier was present ( $\Delta Ct > 0.3$  from other two replicates), the outlier well was excluded from analysis. Amplification profiles of the seven conditions tested were annotated and presented in Figure 2A-B and Additional file 4: Figure S4A-E. Results from laboratory quantitative validation using all conditions tested were summarized in Table 4. Detailed results of inter- and intra-run coefficient of variation for Ct value and copy number were presented for all conditions tested in Figure 3 and Additional file 5: Supplemental file 1A-C using scattered plots generated with the *vegan* package in R [18,19].

**Bacteria-to-human ratio calculations** Calculations were performed using the following copy number and genome size estimates: the average bacterial 16 S rRNA gene copy number per genome was estimated to be 3.94 copies as calculated by rrnDB [20] (accessed at <http://ribosome.mmg.msu.edu/rrnDB/index.php>) and the average human 18 S rRNA gene copy number per genome was estimated to be 400 copies [21]. The diploid human genome was estimated to be 5,758 Mb [22] or the mass equivalent of  $5,758 \text{ Mb} / (0.978 \times 10^3 \text{ Mb/pg}) = 5.887 \text{ pg}$  per diploid human genome [23].

## Results

### Assay design and initial specificity check

Using our 16 S rRNA gene nucleotide distribution output, we identified a conserved 500 bp region for assay design. Within this region, we selected three highly conserved sub-regions abutting V3-V4 for the design of a TaqMan<sup>®</sup> quantitative real-time PCR (qPCR) assay (Additional file 6: Supplemental file 2). Degenerate bases were incorporated strategically in the primer sequence to increase the unique 16 S rRNA gene sequence types matching the qPCR assay. No degeneracies were permitted in the TaqMan<sup>®</sup> probe sequence (Table 1). Initial *in silico* specificity analysis using megablast showed that the probe is a perfect match against human and *C. albicans* ribosomal DNA, due to its highly conserved nature, but the primers were specific and screening using human and *C. albicans* genomic DNA did not show non-specific amplification.

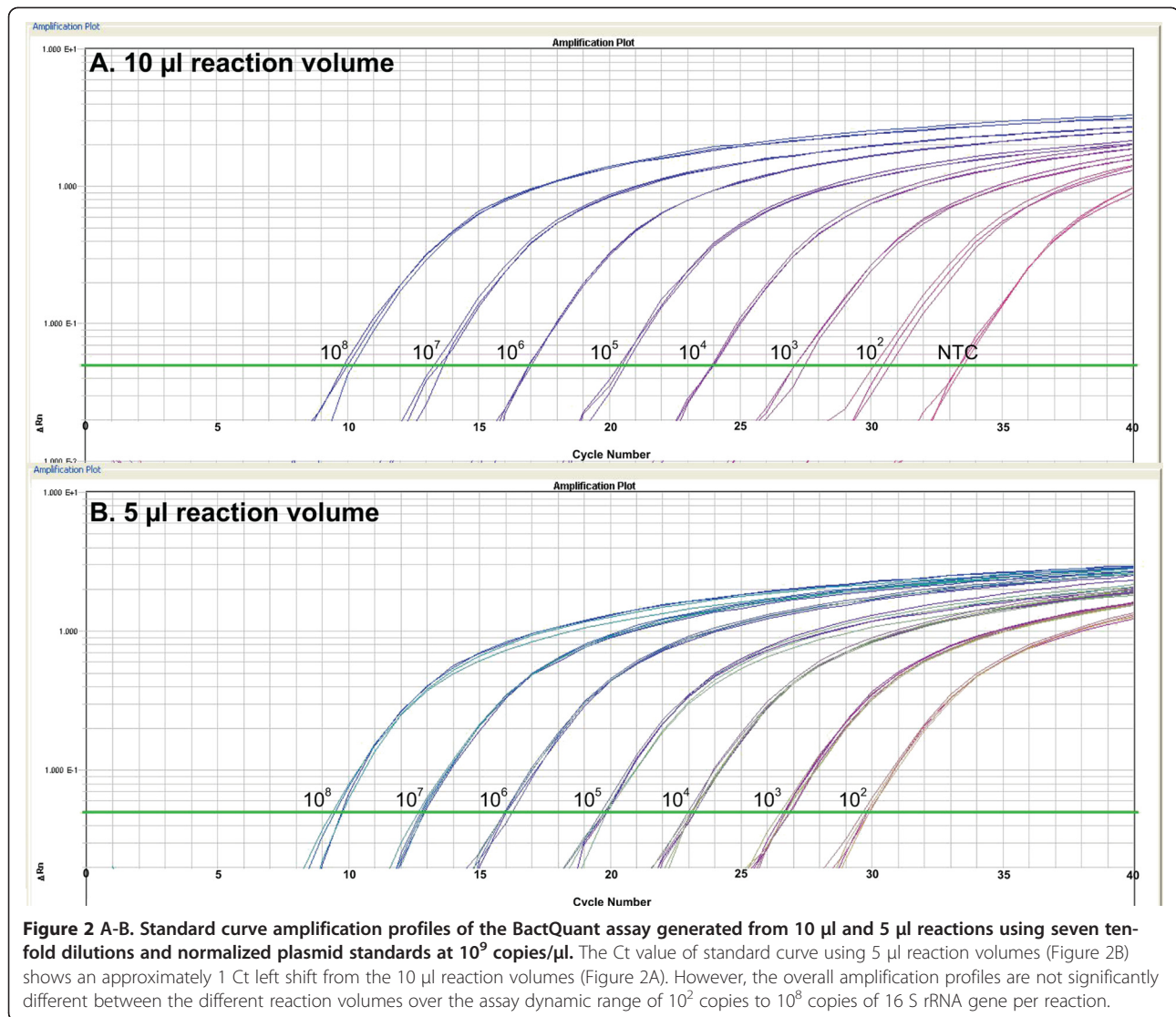
### *In silico* analysis of assay coverage using 16 S rRNA gene sequences from 34 bacterial phyla

Numerical and taxonomic *in silico* coverage analyses at the phylum, genus, and species levels were performed using 16 S rRNA gene sequences from the Ribosomal Database Project (RDP) as sequence matching targets. A total of 1,084,903 16 S rRNA gene sequences were downloaded from RDP. Of these, 671,595 sequences were determined to be eligible for sequence match comparison based on sequence availability in the *E. coli* region of the BactQuant assay amplicon. The *in silico* coverage analyses were performed based on perfect match of full-length primer and probe sequences (hereafter referred to as “stringent criterion”) and perfect match with full-length probe sequence and the last 8 nucleotides of primer sequences at the 3’ end (hereafter referred to as “relaxed criterion”).

Using the stringent criterion, *in silico* numerical coverage analysis showed that 31 of the 34 bacterial phyla evaluated were covered by the BactQuant assay. The three uncovered phyla being Candidate Phylum OD1, Candidate Phylum TM7, and Chlorobi (Figure 1). Among most of the 31 covered phyla, more than 90% of the genera in each phylum were covered by the BactQuant assay. The covered phyla included many that are common in the human microbiome, such as Tenericutes (13/13; 100%), Firmicutes (334/343; 97.4%), Proteobacteria (791/800; 98.9%), Bacteroidetes (179/189; 94.7%), Actinobacteria (264/284; 93.0%), and Fusobacteria (11/12; 91.7%). Only three covered phyla had lower than 90% genus-level coverage, which were Deferribacteres (7/8; 87.5%), Spirochaetes (9/11; 81.8%), and Chlamydiae (2/9; 22.2%) (Figure 1).

On the genus- and species-levels, 1,778 genera (96.2%) and 74,725 species (83.5%) had at least one perfect match using the stringent criterion. This improved to 1,803 genera (97.7%) and 79,759 species (89.1%) when the relaxed





criterion was applied (Table 2, Additional file 2: Figure S1). Using the same relaxed criterion, 566,685 or 84% of all eligible sequences were perfect matches with the BactQuant assay (Table 2). Detailed taxonomic information on the covered and uncovered OTUs for the BactQuant assay can be found in Additional file 5: Supplemental file 1. Additional file 6: Supplemental file 2.

During our *in silico* validation, a previously published qPCR assay was identified, which was used as a published reference for comparison [15]. The *in silico* comparison showed that the BactQuant assay covers more OTUs irrespective of the criterion applied (Table 2, Figure 1, Additional file 2: figure S1). Based on the stringent criterion, the published assay has 10 additional uncovered phyla in comparison to BactQuant; these were: Candidate Phylum OP11, Aquificae, Caldiseptica, Thermodesulfoacteria, Thermotogae, Dictyoglomi, Deinococcus-Thermus, Lentisphaerae,

Chlamydiae, and Candidate Phylum OP10 (Figure 1). Applying the relaxed criterion added two phyla, Aquificae and Lentisphaerae, to those covered by the published assay (Additional file 2: figure S1). The genus-level coverage of the published assay was also low, with fewer than 50% genus-level coverage in six of its covered phyla. For Cyanobacteria, Planctomycetes, Synergistetes, and Verrucomicrobia, only a single genus was covered by the published assay (Additional file 7: Supplemental file 3). In all, the BactQuant assay covered an additional 288 genera and 16,226 species than the published assay, or the equivalent of 15% more genera, species, and total unique sequences than the published assay (Table 2). Detailed taxonomic information on the covered and uncovered OTUs for the published qPCR assay can be found in Additional file 7: Supplemental files 3, Additional file 8: Supplemental files 4.

**Table 4 Laboratory quantitative validation results of the BactQuant assay performed using pure plasmid standards and different mixed templates**

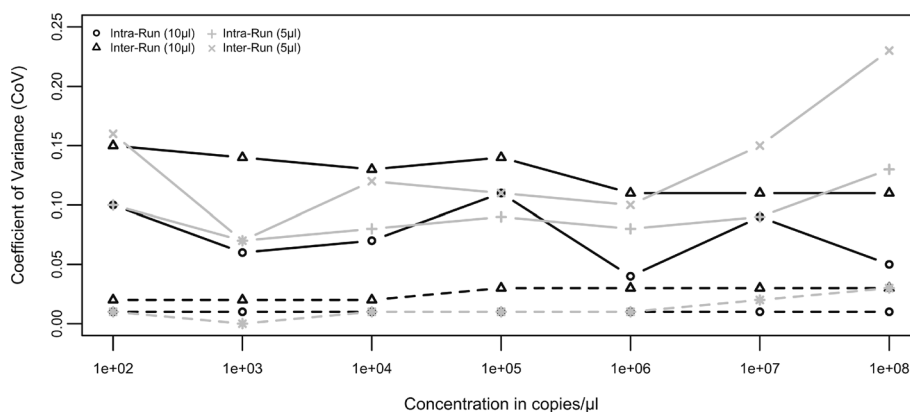
Templates used	Assay dynamic range	Average reaction efficiency (SD)	$r^2$ -value
Plasmid standards-only (10 $\mu$ l Rxn)	100–10 <sup>8</sup> copies	102% (2%)	>0.999
Plasmid standards-only (5 $\mu$ l Rxn)	100 – 10 <sup>8</sup> copies	95% (1%)	>0.999
Plasmid standards plus 0.5 ng human gDNA	100 – 10 <sup>8</sup> copies	99% (4%)	>0.994
Plasmid standards plus 1 ng human gDNA	100 – 10 <sup>8</sup> copies	101% (5%)	>0.994
Plasmid standards plus 5 ng human gDNA	500 – 10 <sup>8</sup> copies	96% (1%)	>0.999
Plasmid standards plus 10 ng human gDNA	1000 – 10 <sup>8</sup> copies	97% (2%)	>0.999
Plasmid standards plus 0.5 ng <i>C. albicans</i> gDNA	100 – 10 <sup>8</sup> copies	97% (1%)	>0.999

**Laboratory analysis of assay performance using diverse bacterial genomic DNA**

Laboratory evaluation of the BactQuant assay showed 100% sensitivity against 101 species identified as perfect matches from the *in silico* coverage analysis. The laboratory evaluation was performed using genomic DNA from 106 unique species encompassing eight phyla: Actinobacteria (n = 15), Bacteroidetes (n = 2), Deinococcus-Thermus (n = 1), Firmicutes (n = 18), Fusobacteria (n = 1), Proteobacteria (n = 66), Chlamydiae (n = 2), and Spirochaetes (n = 2). Overall, evaluation using genomic DNA from the 101 *in silico* perfect match species demonstrated  $r^2$ -value of >0.99 and amplification efficiencies of 81 to 120% (Table 3).

Laboratory evaluation against the five *in silico* uncovered species showed variable assay amplification profiles and efficiencies. Of these five species, *Chlamydia trachomatis*, *Chlamydophila pneumoniae*, and *Cellvibrio gilvus* were identified as uncovered irrespective of *in silico* analysis criterion. However, while *C. trachomatis* and *C. pneumoniae* showed strongly

inhibited amplification profile, *C. gilvus* amplified successfully with a  $r^2$ -value of >0.999 and an amplification efficiency of 92% (Additional file 3: Figure S3A-B & 3E). Two other species, *Borrelia burgdorferi* and *Escherichia vulneris*, which were uncovered only when using the stringent criterion, also showed successful amplification with a  $r^2$ -value of >0.999 and 90% and 93% reaction efficiency, respectively (Additional file 3: Figure S3C-D). Comparison of the assay and bacterial sequences showed that *C. trachomatis* and *C. pneumoniae* shared a single mismatch in the center of the probe sequence, whereas *C. gilvus* had a mismatch on the 3' end of the probe. The mismatch in *B. burgdorferi* and *E. vulneris* was a single base difference in 5' end of the reverse and the forward primer, respectively (Additional file 3: Figure S3A-E). These findings strongly suggest the location of the sequence mismatch is an important determinant of amplification outcome. Furthermore, it supports that the BactQuant assay's coverage in laboratory application is likely greater than determined by the *in silico* analyses.



**Figure 3 Inter- and intra-run coefficient of variation (CoV) for 10  $\mu$ l and 5  $\mu$ l reactions using seven ten-fold dilutions and normalized plasmid standards at 10<sup>9</sup> copies/ $\mu$ l calculated using data from multiple runs.** The data is presented for both copy number (solid line) and Ct value (dashed line). As would be expected, the CoV is higher for copy number than for Ct value and is also higher for inter-run than for intra-run. The CoV for copy number for both reaction volumes was consistently below 15% until at 10<sup>7</sup> copies for 5  $\mu$ l reactions. The CoV for Ct value was consistently below 5% for both reaction volumes.

### Laboratory quantitative assay validation using pure plasmid standards and mixed templates

To fully characterize the assay quantitative profile, the BactQuant assay was tested using different reaction volumes and against both pure and mixed templates containing bacterial and human rRNA gene targets. Laboratory evaluation using pure plasmid standards in 10  $\mu$ l and 5  $\mu$ l reaction volumes showed excellent amplification profiles, with an assay dynamic range of  $10^2$ – $10^8$  16 S rRNA gene copies per reaction (Figure 2A–B). For the 10  $\mu$ l reactions, the inter- and intra-run coefficients of variance (CoV) ranged from 1.58–2.94% and 0.64–1.25% for Ct values and from 10.60–15.36% and 4.02–10.51% for copy number, respectively (Figure 3). The inter- and intra-run CoV was comparable for the different reaction volumes, except for the higher CoV in 5  $\mu$ l reactions containing more than  $10^7$  plasmid copies (Figure 3). This suggests that the 5  $\mu$ l reaction volumes may be better suited for samples with low amounts of bacterial DNA. Establishment of the limit of detection (LOD) for the BactQuant assay using pure plasmid standards was not attempted because it was affected by the level of contaminants in reagents, as previously reported [15,24–28].

Further laboratory evaluations using mixed templates showed that the ratio of bacteria-to-human DNA ratio determined the assay dynamic range of the BactQuant assay (Table 4, Additional file 4: Figure S4A–E, Additional file 5: Additional file 9: Table S1A–C). Experiments using seven tenfold dilutions of plasmid standards with 0.5 ng and 1 ng human gDNA showed that the assay dynamic range was unchanged from pure plasmid standard. However, experiments using 5 ng and 10 ng of human gDNA showed narrower assay dynamic ranges of 500 -  $10^8$  and 1000 -  $10^8$  16 S rRNA gene copies per reaction, respectively. Based on this result, the LODs for 10  $\mu$ l reactions using templates containing 5 ng and 10 ng of human gDNA were estimated to be a bacteria-to-human ribosomal gene copy ratio of 500:339732 and 1000:679464, respectively. This could be further simplified to a bacteria-to-human ribosomal gene copy ratio of 1:679. From a genomic equivalent perspective, the LOD of the BactQuant assay was approximately at a bacteria-to-human ratio of 127:849.

### Discussion

We designed and evaluated a new expanded-coverage bacterial quantitative real-time PCR assay targeting the 16 S rRNA gene. To accomplish this, we curated a set of high-quality 16 S rRNA gene sequences for assay design and evaluated the coverage of our primers and as a union (rather than as separate entities). In

addition, we improved the quantitative capacity of our assay using a cloned plasmid standard. Our computational and laboratory analyses showed that BactQuant had superior *in silico* taxonomic coverage while retaining favorable *in vitro* performance. As would be expected, the diverse gene sequences targeted by BactQuant have resulted in variable reaction efficiencies. Nevertheless, laboratory evaluation showed 100% sensitivity against perfect match species from the *in silico* analysis.

To allow researchers to determine whether BactQuant covers key organisms in their target community, we provided additional detailed OTU coverage information in the Supplemental Files. We have applied the logic that an OTU was covered if it contained at least one perfect match sequence in the *in silico* analysis. 16 S rRNA gene sequences with ambiguous or degenerate bases at the primer and probe sites were considered non-perfect matches, thus making our coverage estimates more conservative. Lastly, although we prohibited the use of a degenerate probe to maximize our assay's quantitative ability, this approach may permit detection of specific taxa such as *Chlamydia* spp. and *Chlamydophila* spp.

For most studies, the desired measurement of bacterial load is the number of cells rather than 16 S rRNA gene copy number; however, the 16 S rRNA gene copy number varies among bacterial species and even among strains [29,30]. The range of copy number is estimated at one to 14, with most non-spore forming species having fewer than 10 copies per genome [20]. We use the average 16 S rRNA gene copy number per genome from rrnDB in our genomic equivalent estimation, but alternative approaches are possible. This, combined with logarithmic growth of bacteria, suggest that using estimated average copy number could be sufficient.

The *in silico* analysis was an important component of our validation of BactQuant against diverse bacterial sequence types, even though sequence matching is not a perfect predictor of laboratory performance [31]. Many factors are known to affect reaction efficiency, such as oligonucleotide thermodynamics, the type of PCR master mix used, and the template DNA extraction method. Concentration of background nontarget genomic DNA is another factor that can affect the quantitative parameters rRNA gene-based assays [32]. The interference of background human DNA with BactQuant dynamic range reported in this paper was most likely due to cross-reactivity of human DNA with the probe, which targets a region conserved even among eukaryotic organisms, including in the human 18 S rRNA gene. This may be overcome by using an intercalating reporter dye in place of a fluorescent probe as a qPCR reporter mechanism; however, the loss of tertiary-level of specificity is a

potential concern in direct application of an intercalating dye assay to specimens containing high amounts of non-target DNA.

Exogenous bacterial DNA, particularly from biologically synthesized reagents such as Taq DNA polymerase are a known limitation for analyzing samples with low bacterial load [28,33]. Recently, this issue has received renewed attention due to increased usage of next-generation sequencing and the frequent data contamination from exogenous bacterial DNA. Several methods have been evaluated for removing bacterial contaminants from Taq DNA polymerase, including UV irradiation [34,35], DNase I treatment, and ultrafiltration [36]. The level of *E. coli* contamination in Taq DNA polymerase has been estimated at  $10^2$  to  $10^5$  genome equivalents of bacterial DNA per unit of enzyme [28]. This is consistent with the lowest amount of contamination we have observed in our experiments, which were 5 and 10 copies of 16 S rRNA gene in 5  $\mu$ l and 10  $\mu$ l reactions, respectively. The ubiquity of bacterial DNA also makes the determination of assay specificity challenging.

Our use of qPCR-quantified plasmid standards addressed a major limitation in the preparation of qPCR quantification standards. The conventional approach of quantifying bacterial genomic DNA or plasmid standards necessitates converting mass (i.e., nanograms per  $\mu$ l) to copy number (i.e.,  $10^8$  copies per  $\mu$ l) and can introduce substantial error. Thus far, we have also successfully applied BactQuant to a diverse range of clinical specimens, including swab eluents, surgical specimens, and respiratory specimens, but we did not present these findings in this paper. To fully understand the likelihood of false negative results due to interference from human DNA or BactQuant's limit of detection will require additional evaluations.

## Conclusion

In summary, we have developed and evaluated a new broad-coverage qPCR assay—BactQuant—for bacterial detection and quantification that showed concurrently improved assay coverage and favorable quantitative parameters. Laboratory tests showed that *in vitro* performance was even better than predicted in the *in silico* analysis; however, our approach of evaluating assay coverage by considering the primer and probe sequences as a single unit is appropriate and necessary. In addition, when employing a copy number estimation method, such as qPCR, the quantification of standards is critical for accurate template quantification. Thus, our approach of quantifying plasmid standards using the intrinsic property of real-time PCR is another important step for any absolute quantification experiments using qPCR.

## Additional files

**Additional file 1: Figure S1.** Figure S1 containing the *in silico* coverage analysis using the relaxed criteria.

**Additional file 2: Figure S2A-E.** Standard curve amplification plots using mixed templates.

**Additional file 3: Figure S3A-E.** Amplification plots of the non-perfect match targets, including *C. trachomatis*, *C. pneumoniae*, *C. gilvus*, *B. burgdorferi*, and *E. vulneris*.

**Additional file 4: Figure S4A-E.** Coefficient of variance (CoV) distribution across assay dynamic range for mixed templates.

**Additional file 5: Supplemental File 1.** Detailed results for BactQuant using the stringent criteria.

**Additional file 6: Supplemental File 2.** Detailed results for BactQuant using the relaxed criteria.

**Additional file 7: Supplemental File 3.** Detailed results for published assay using the stringent criteria.

**Additional file 8: Supplemental File 4.** Detailed results from published assay using the relaxed criteria.

**Additional file 9: Table S1.** Base distribution output used in primer and probe design, with the bolded base signifying the selected base(s) and incorporation of more than one allele at a given nucleotide position was accomplished using degenerate bases. The alignment position information in the base distribution file contains many gaps as a result from the sequence alignment and differs from the *E. coli* region information from Table 1.

## Competing interests

The authors have declared that no competing interests exist.

## Authors' contributions

CML contributed to the overall study design, the acquisition, analysis, and interpretation of data, and drafting the manuscript, MA contributed to the bioinformatics portion of the study design and its implementation, SK participated in bioinformatics analysis and assay design, PRH and YTH both contributed to the acquisition and interpretation of laboratory data, PK conceived of the study and contributed to the overall study design, LBP contributed to the overall study design and helped to draft the manuscript. All authors read and approved the final manuscript.

## Funding

This work was supported by the National Institutes of Health (R01AI087409-01A1, R15DE021194-01), the Department of Defense (W81XWH1010870), and the TGen Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Acknowledgements

We thank Lindsey Watson (growth and extraction of ATCC isolates and performance of quantitative real-time PCR), Jordan Buchhagen (cloning and sequence verification of cloned standards), Jolene Bowers (generation of the 16 S rRNA gene plasmid clones), and Tania Contente-Cuomo (performance of quantitative real-time PCR) at Translational Genomics Research Institute for their assistance with different aspects of the laboratory work presented in this manuscript.

## Author details

<sup>1</sup>Division of Pathogen Genomics, Translational Genomics Research Institute, 3051 W. Shamrell Blvd., Suite 106, Flagstaff, AZ 86001 USA. <sup>2</sup>Center for Microbial Genetics and Genomics, Applied Research & Development Building, Northern Arizona University, 1298 S. Knoles Drive, Flagstaff, AZ 86011 USA. <sup>3</sup>Departments of Laboratory Medicine and Internal Medicine, National Taiwan University Hospital, National Taiwan University College of Medicine, No. 7, Chung-Shan South Road, Taipei, Taiwan. <sup>4</sup>Department of Internal Medicine, Far Eastern Memorial Hospital, No.21, Nanya S. Rd., New Taipei City, Taiwan. <sup>5</sup>Current Address: Ross University School of Medicine, 630 US Highway 1, North Brunswick, NJ 08902 USA.

Received: 28 October 2011 Accepted: 6 March 2012  
Published: 17 April 2012

## References

1. Tringe SG, Hugenholtz P: **A renaissance for the pioneering 16S rRNA gene.** *Curr Opin Microbiol* 2008, **11**(5):442–446.
2. Woo PC, Lau SK, Teng JL, Tse H, Yuen KY: **Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories.** *Clin Microbiol Infect* 2008, **14**(10):908–934.
3. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, *et al*: **Vaginal microbiome of reproductive-age women.** *Proc Natl Acad Sci U S A* 2011, **108**(Suppl 1):4680–4687.
4. Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, Bouffard GG, Blakesley RW, Murray PR, Green ED, *et al*: **Topographical and temporal diversity of the human skin microbiome.** *Science* 2009, **324**(5931):1190–1192.
5. Turnbaugh PJ, Hamady M, Yatsunencko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, *et al*: **A core gut microbiome in obese and lean twins.** *Nature* 2009, **457**(7228):480–484.
6. Zhang H, Parameswaran P, Badalamenti J, Rittmann BE, Krajmalnik-Brown R: **Integrating high-throughput pyrosequencing and quantitative real-time PCR to analyze complex microbial communities.** *Methods Mol Biol* 2011, **733**:107–128.
7. Zengler K, Toledo G, Rappe M, Elkins J, Mathur EJ, Short JM, Keller M: **Cultivating the uncultured.** *Proc Natl Acad Sci U S A* 2002, **99**(24):15681–15686.
8. Higuchi R, Dollinger G, Walsh PS, Griffith R: **Simultaneous amplification and detection of specific DNA sequences.** *Biotechnology (N Y)* 1992, **10**(4):413–417.
9. Bustin SA, Benes V, Nolan T, Pfaffl MW: **Quantitative real-time RT-PCR—a perspective.** *J Mol Endocrinol* 2005, **34**(3):597–601.
10. Bustin SA, Benes V, Garson JA, Hellems J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, *et al*: **The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments.** *Clin Chem* 2009, **55**(4):611–622.
11. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity G, *et al*: **The Ribosomal Database Project: improved alignments and new tools for rRNA analysis.** *Nucleic Acids Res* 2009, **37**(Database issue):D141–D145.
12. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL: **Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB.** *Appl Environ Microbiol* 2006, **72**(7):5069–5072.
13. Kibbe WA: **OligoCalc: an online oligonucleotide properties calculator.** *Nucleic Acids Res* 2007, **35**(Web Server issue):W43–W46.
14. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schaffer AA: **Database indexing for production MegaBLAST searches.** *Bioinformatics* 2008, **24**(16):1757–1764.
15. Nadkarni MA, Martin FE, Jacques NA, Hunter N: **Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set.** *Microbiology* 2002, **148**(Pt 1):257–266.
16. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24**(8):1596–1599.
17. D.J.Lane: **1991 Nucleic acid techniques in bacterial systematics.** John Wiley and Sons, New York, NY; Lane DJ: *Nucleic acid techniques in bacterial systematics.* New York, NY: John Wiley and Sons; 1991.
18. Jari Oksanen, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry, H. Stevens, Helene Wagner (2011). **vegan: Community Ecology Package.** R package version 2.0-2. Accessed at: [http://CRAN.R-project.org/package=vegan]
19. RDCTeam2008R: **A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna: R Foundation for Statistical Computing; 2008.
20. Lee ZM, Bussema C 3rd, Schmidt TM: **rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea.** *Nucleic Acids Res* 2009, **37**(Database issue):D489–D493.
21. Gaubatz J, Prashad N, Cutler RG: **Ribosomal RNA gene dosage as a function of tissue and age for mouse and human.** *Biochim Biophys Acta* 1976, **418**(3):358–375.
22. Consortium IHGS: **Finishing the euchromatic sequence of the human genome.** *Nature* 2004, **431**(7011):931–945.
23. Dolezel J, Bartos J, Voglmayr H, Greilhuber J: **Nuclear DNA content and genome size of trout and human.** *Cytometry A* 2003, **51**(2):127–128. author reply 129.
24. Bottger EC: **Frequent contamination of Taq polymerase with DNA.** *Clin Chem* 1990, **36**(6):1258–1259.
25. Dreier J, Stormer M, Kleesiek K: **Two novel real-time reverse transcriptase PCR assays for rapid detection of bacterial contamination in platelet concentrates.** *J Clin Microbiol* 2004, **42**(10):4759–4764.
26. Klaschik S, Lehmann LE, Raadt A, Hoefl A, Stuber F: **Comparison of different decontamination methods for reagents to detect low concentrations of bacterial 16S DNA by real-time-PCR.** *Mol Biotechnol* 2002, **22**(3):231–242.
27. Meier A, Persing DH, Finken M, Bottger EC: **Elimination of contaminating DNA within polymerase chain reaction reagents: implications for a general approach to detection of uncultured pathogens.** *J Clin Microbiol* 1993, **31**(3):646–652.
28. Rand KH, Houck H: **Taq polymerase contains bacterial DNA of unknown origin.** *Mol Cell Probes* 1990, **4**(6):445–450.
29. Klappenbach JA, Dunbar JM, Schmidt TM: **rRNA operon copy number reflects ecological strategies of bacteria.** *Appl Environ Microbiol* 2000, **66**(4):1328–1333.
30. Stevenson BS, Schmidt TM: **Life history implications of rRNA gene copy number in Escherichia coli.** *Appl Environ Microbiol* 2004, **70**(11):6670–6677.
31. Morales SE, Holben WE: **Empirical testing of 16S rRNA gene PCR primer pairs reveals variance in target specificity and efficacy not suggested by in silico analysis.** *Appl Environ Microbiol* 2009, **75**(9):2677–2683.
32. Ludwig W, Schleifer KH: **How quantitative is quantitative PCR with respect to cell counts?** *Syst Appl Microbiol* 2000, **23**(4):556–562.
33. Hughes MS, Beck LA, Skuce RA: **Identification and elimination of DNA sequences in Taq DNA polymerase.** *J Clin Microbiol* 1994, **32**(8):2007–2008.
34. Corless CE, Guiver M, Borrow R, Edwards-Jones V, Kaczmarek EB, Fox AJ: **Contamination and sensitivity issues with a real-time universal 16S rRNA PCR.** *J Clin Microbiol* 2000, **38**(5):1747–1752.
35. Sarkar G, Sommer SS: **Removal of DNA contamination in polymerase chain reaction reagents by ultraviolet irradiation.** *Methods Enzymol* 1993, **218**:381–388.
36. Silkie SS, Tolcher MP, Nelson KL: **Reagent decontamination to eliminate false-positives in Escherichia coli qPCR.** *J Microbiol Methods* 2008, **72**(3):275–282.

doi:10.1186/1471-2180-12-56

**Cite this article as:** Liu *et al*: **BactQuant: An enhanced broad-coverage bacterial quantitative real-time PCR assay.** *BMC Microbiology* 2012 **12**:56.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

