OXFORD GENETICS

GENETICS, 2021, 219(2), iyab117
DOI: 10.1093/genetics/iyab117
Advance Access Publication Date: 21 July 2021
Investigation

Precise measurement of the fitness effects of spontaneous mutations by droplet digital PCR in *Burkholderia cenocepacia*

Anita Rana [1], David Patton [1], Nathan T. Turner,[1] Marcus M. Dillon [2], Vaughn S. Cooper [3] and Way Sung [1,*]

[1]Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
[2]Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario M5S3B2, Canada
[3]Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA 15219, USA

*Corresponding author: Department of Bioinformatics and Genomics, University of North Carolina, Bioinformatics 263/335L, 9201 University City Blvd., Charlotte, NC 28223, USA. Email: wsung@uncc.edu

Abstract

Understanding how mutations affect survivability is a key component to knowing how organisms and complex traits evolve. However, most mutations have a minor effect on fitness and these effects are difficult to resolve using traditional molecular techniques. Therefore, there is a dire need for more accurate and precise fitness measurements methods. Here, we measured the fitness effects in *Burkholderia cenocepacia* HI2424 mutation accumulation (MA) lines using droplet-digital polymerase chain reaction (ddPCR). Overall, the fitness measurements from ddPCR-MA are correlated positively with fitness measurements derived from traditional phenotypic marker assays ($r = 0.297$, $P = 0.05$), but showed some differences. First, ddPCR had significantly lower measurement variance in fitness ($F = 3.78$, $P < 2.6 \times 10^{-13}$) in control experiments. Second, the mean fitness from ddPCR-MA measurements were significantly lower than phenotypic marker assays (-0.0041 *vs* -0.0071, $P = 0.006$). Consistent with phenotypic marker assays, ddPCR-MA measurements observed multiple (27/43) lineages that significantly deviated from mean fitness, suggesting that a majority of the mutations are neutral or slightly deleterious and intermixed with a few mutations that have extremely large effects. Of these mutations, we found a significant excess of mutations within DNA excinuclease and Lys R transcriptional regulators that have extreme deleterious and beneficial effects, indicating that modifications to transcription and replication may have a strong effect on organismal fitness. This study demonstrates the power of ddPCR as a ubiquitous method for high-throughput fitness measurements in both DNA- and RNA-based organisms regardless of cell type or physiology.

Keywords: Burkholderia cenocepacia; droplet digital polymerase chain reaction; fitness effects of mutations; mutation accumulation

Introduction

Mutations are the fundamental source of genetic variation and understanding the effects of mutations is of great practical importance for comprehending the nature of quantitative genetic variation and complex genetic diseases. Mutations provide the source material for natural selection to operate on, allowing organisms to evolve, adapt, and compete in changing environments. However, mutations can also be highly detrimental, disrupt gene function, and in humans have been linked to debilitating genetic diseases such as diabetes, cancer, and mental illness (Caballero and Keightley 1994; Zhang and Hill 2005; Eyre-Walker *et al.* 2006; Poledne and Zicha 2018).

Multiple evolutionary forces determine the fate of new mutations. One of these forces, natural selection, is expected to purge mutations that have strong deleterious effects, fix mutations that have strong beneficial effects, and ignore mutations that have weak effects relative to the power of random genetic drift (Barrett *et al.* 2006). The underlying fitness effects can provide us with a strong understanding of which mutations will be purged from a population, and which mutations may rise in frequency within a population, including those involved in antibiotic resistance and pathogenicity. In microbes that reproduce quickly, the effect of a mutation can be measured directly by comparing survivability traits (*e.g.*, growth rate or death rates) of the wild-type against the mutant in competition assays (Lenski *et al.* 2015) and these experiments have been a valuable tool to generate rough estimates of fitness (Eyre-Walker and Keightley 2007; Gordo *et al.* 2011). In the past, studies on fitness have shown that most spontaneous mutations have very small effects (Kibota and Lynch 1996a; Halligan and Keightley 2009; Trindade *et al.* 2010; Heilbron *et al.* 2014; Kraemer *et al.* 2016) and that deleterious mutations far outnumber beneficial mutations (Keightley and Lynch 2003; Eyre-Walker and Keightley 2007). On the contrary, recent studies suggest that the distribution of fitness effects of spontaneous mutations is multimodal, containing strong signatures of both deleterious and beneficial mutations (Zeyl and DeVisser 2001; Eyre-Walker and Keightley 2007; Loewe and Hill 2010). However, even though considerable progress has been made in this field,

Received: April 04, 2021. **Revised:** July 07, 2021. **Accepted:** July 09, 2021
© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America. All rights reserved.
For permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/genetics/article/219/2/iyab117/6325026 by UNC Charlotte user on 29 July 2022

the methods used to assay fitness are still extremely laborious and have inherent biases.

Spectrophotometry and colony counting, both of which measure cell density of the wild-type and mutant populations before and after competition, were some of the first approaches used to directly quantify fitness (Kibota and Lynch 1996b; Lynch *et al.* 2008). Known measurement biases in spectrophotometry include high levels of protein aggregation, external light sources, temperature and growth variation, and microbial biofilm formation, all which can interfere with spectrophotometry readings and generate inaccurate measurements of cell growth. These sources of variation in measurement can creates a large signal-to-noise imbalance that makes it difficult to accurately assay mutations that have a minor effect on fitness. Colony counting assays that rely on phenotypic expression of a character trait (*e.g.*, blue-white colonies) are subject to phenotypic lag (the delay in expression of a phenotype), poor plating efficiency, and selection against mutants, which may generate errors when used to measure cell growth.

More recently, fluorescence-activated cell sorting (FACS), real-time quantitative PCR (qPCR), and genomic barcoding methods have been applied to measure fitness with mixed results. FACS can quantify individual cells using fluorescent chemicals and has been shown to be highly accurate and consistent for larger eukaryotic cells (*Chlamydomonas reinhardtii*) (Kraemer *et al.* 2017). However, use of FACS is limited in bacteria as they are smaller and more difficult to separate, particularly in organisms that form biofilms (Gallet *et al.* 2012; Dillon *et al.* 2016). qPCR uses fluorescent amplification of DNA to identify and quantify a target and thus provides a cell-independent method of measuring fitness. However, qPCR is dependent on amplification efficiency of the target and requires the use of a standard curve, both of which may cause biases depending on the target locus. Furthermore, qPCR is plagued with large variation in signal, and because the majority of mutations have a small fitness effect ($s < 0.1$), limiting the use of qPCR to measure fitness (Eyre-Walker and Keightley 2007; Bataillon and Bailey 2014). Finally, genomic barcoding which, tags and tracks organisms using inserted DNA fragments, can have certain limitations, such as tagging effects dependent on the tagged loci, barcode contamination, and requires that the target can accept foreign DNA (Lea and Coulson 1949; Blundell and Levy 2014; Fasanello *et al.* 2020).

To improve the accuracy, repeatability, and the speed of fitness measurements, we employed Droplet Digital PCR (ddPCR) (Vogelstein and Kinzler 1999). ddPCR builds on the workflow of qPCR, whereby fluorescent amplification (or a fluorescent probe) indicates the presence of a target within a sample, and fluorescent signal indicates the quantity of that target within the sample (Figure 1). ddPCR improves upon this technology by partitioning a single sample into 20,000 individual qPCR reactions within microfluidic water-in-oil droplets (Vogelstein and Kinzler 1999; Morrison *et al.* 2006; Ottesen *et al.* 2006; Pinheiro *et al.* 2012). Amplification within each droplet will indicate the presence or absence of the nucleic acid within the sample, providing an accurate measurement of the absolute copies of the DNA or RNA target (Wloch *et al.* 2001; Pinheiro *et al.* 2012). With this technique, the relative abundance of a specific mutant can be tracked at high resolution (~1:20,000), and when applied to competition experiments can be used to generate highly accurate measurements of the relative fitness of competing populations over time.

Utilizing ddPCR for fitness measurements has many advantages over other methods. First, ddPCR targets nucleic acids, avoiding any biases or difficulties that may arise from colony counting, qPCR, and FACS methods described above. Second, unlike qPCR, which measures the number of cycles needed to reach a threshold concentration of PCR product, ddPCR provides absolute quantification of the individual droplets that contain the target product (Pinheiro *et al.* 2012) without the need for a standard curve. Third, ddPCR possesses a high signal-to-noise ratio that allows for detection of extremely rare targets and can be used to resolve minor differences in relative fitness or mutations with minor effects. Fourth, ddPCR can be performed using automated robotics (Bio-Rad QX200), reducing the chance of pipetting error or human biases. Finally, ddPCR can be expanded to a 96- or 384-well plate, providing reproducible, high-throughput measurements of fitness.

To measure the cumulative effects of spontaneous mutations on fitness, we apply ddPCR to mutation accumulation lines (ddPCR-MA). MA involves repeated bottlenecking of a population to minimize the influence of natural selection on new mutations. Under these conditions, the majority of mutations will accumulate in genomes through random genetic drift and the underlying mutation rates reflect errors that arise during DNA replication and repair. At the end of the MA experiment, whole-genome sequencing is applied to each MA line, providing the number, type, and genomic location of each mutation. Each MA line can then be competed against the ancestral genotype to determine the cumulative impact of the mutations on fitness. MA experiments have been conducted in many microbial species, including *Saccharomyces cerevisiae* (Wloch *et al.* 2001; Zeyl and DeVisser 2001; Dickinson 2008), *Escherichia coli* (Kibota and Lynch 1996b; Trindade *et al.* 2010), *Paramecium tetraurelia* (Sung *et al.* 2012), and other microbes (Kibota and Lynch 1996b; Halligan and Keightley 2009; Trindade *et al.* 2010; Heilbron *et al.* 2014; Kraemer *et al.* 2016), and ddPCR-MA provides a unique platform to perform reliable fitness measurements on spontaneous mutations across different species and taxa.

Here, we present a novel approach of measuring fitness using ddPCR on MA lines from *Burkholderia cenocepacia* HI2424. Each MA lineage of *B. cenocepacia* contains between 2 and 14 spontaneous mutations (Dillon and Cooper 2016) and combining ddPCR with MA (ddPCR-MA) allows us to measure the relative fitness effect of those mutations per generation against the wild-type progenitor. Importantly, the fitness effects of these *B. cenocepacia* MA lines has been previously measured using a traditional colony counting approach (Dillon and Cooper 2016), providing us with a point of comparison to benchmark our methods.

Materials and methods

Generation of bacterial strain and MA lines

Burkholderia cenocepacia HI2424 strain is a β-proteobacteria and was recovered from agricultural soil in upstate New York (LiPuma *et al.* 2002). It is a member of the *B. cepacia complex* (Bcc), a group of 17 closely related and phenotypically similar species (Vanlaere *et al.* 2009), most of which are soil saprophytes and phyto-pathogens that occupy a wide range of environmental niches. HI2424 is characterized as a representative of the *B. cenocepacia* PHDC clonal lineage (Chen *et al.* 2001; Vanlaere *et al.* 2009). This clone appears to be widely distributed as a human pathogen, having been recovered from cystic variance patients in 24 US states (Liu *et al.* 2003) and Europe (Coenye *et al.* 2004). Amongst *B. cepacia complex*, *B. cenocepacia*, and *B. multivorans* appear to form the most profuse and highly resistant biofilms (Mahenthiralingam and Vandamme 2005; Traverse *et al.* 2013) on both abiotic surfaces (*e.g.*, glass and plastics) as well as biotic

Downloaded from https://academic.oup.com/genetics/article/219/2/iyab117/6325026 by UNC Charlotte user on 29 July 2022

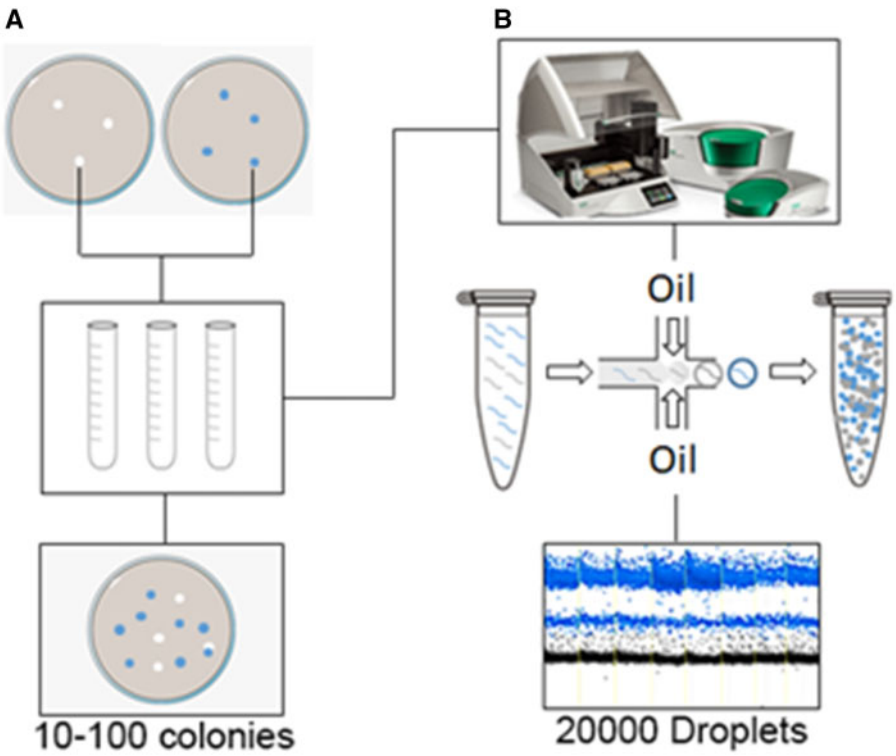

Figure 1 Droplet digital PCR (ddPCR) general workflow. (A) Traditional colony counting fitness measurements. Mutant and wild-type individuals undergo a competition assay within a test tube. Relative fitness is determined by the ratio of colonies exhibiting a detectable phenotype. (B) Fitness measurements using ddPCR (Biorad QX200). Mutant and wild-type individuals undergo the same competition assay within a test tube and DNA is extracted from the pool. Nucleic acids and PCR reagents are then subdivided into ~20,000 oil-emulsion droplets. Primers that target the wild-type, mutant, or both types can be used to amplify droplets that contain the DNA from the desired target(s). Droplets that do contain the target DNA are shown in blue, and those that do not amplify are shown in gray/black. Multiple loci can be targeted in a single reaction, bottom right panel shows example of a multiplex ddPCR assay with two targets. Optimization and amplification protocols can be found in Supplementary Materials.

surfaces such as epithelial cells (Coenye 2010). *Burkholderia* biofilm formation plays a key role in persistence can cause severe lung infections in immunocompromised patients or patients with cystic fibrosis (LiPuma 2005). The sticky and thick biofilms formed by many *Burkholderia* species make it difficult to measure cell growth and fitness using traditional methods (*e.g.*, spectrophotometer, FACS, colony counting).

The genome of *B. cenocepacia* HI2424 has been fully sequenced and is composed of three chromosomes (Dillon *et al.* 2015) (Chr1: 3.48 Mb, 3253 genes; Chr2: 3.00 Mb, 2709 genes; Chr3: 1.06 Mb, 929 genes) and a plasmid (0.164 Mb, 159 genes) (Agnoli *et al.* 2012). The genome is characterized by the presence of peh, which encodes a polygalacturonase involved in maceration of onion tissue (Gonzalez *et al.* 1997), and a type IV secretion system, similar in arrangement and homology to those of *Brucella suis* and *Sinorhizobium meliloti* (Engledow *et al.* 2004).

The mutation accumulation lines *B. cenocepacia* HI2424 used in this study were generated previously (Dillon and Cooper 2016). Seventy-five independent lineages were founded from a single colony of *B. cenocepacia* HI2424 and independently propagated every twenty-four hours onto fresh tryptic soy agar agar [30 g/l tryptic soy broth powder, 15 g/l agar] and incubated at 37°C for 217 days. Transfers between plates were performed by randomly choosing one colony and spreading it on a new plate, thus bottlenecking each line to a single cell at each transfer (Sung *et al.* 2016). Estimates of generation time were conducted monthly by serial dilution and plating of a single random colony from each lineage. This protocol was repeated until each MA line had undergone approximately 5554 generations of mutation accumulation. At the end of the MA experiment, one randomly chosen colony per MA line was stored at –80°C in 8% DMSO and sequenced using an Illumina Hi-Seq 2000 at the Hubbard Center for Genomic Studies at the University of New Hampshire. All reads were aligned to the *B. cenocepacia* HI2424 reference genome (LiPuma *et al.* 2002) with both the Burrows–Wheeler aligner (Li and Durbin 2009) and Novoalign (http://www.novocraft.com) (Dillon *et al.* 2015; Dillon and Cooper 2016) and the resulting alignments were used to identify mutations that had accumulated within each line. In sum, a total of between 2 and 14 mutations were identified in individual MA lines.

Measurement of the cumulative effect of spontaneous mutations on fitness

For fitness competition assays, we revived each of the 43 MA lineages of *B. cenocepacia* HI2424 and a wild-type *B. cenocepacia* HI2424 by growing them in tryptic soy broth (30 g/l) and agar (30 g/l, 1.5% agar). To quantify the cumulative effect of mutations on fitness (s_{cum}) of each of the 43 derived MA lineages, we conducted competitions between each MA lineage and *B. cenocepacia* HI2424 *lacZ* strain. Forty-four competition assays were performed (43 MA lineages, wild-type control) in which five colonies were randomly selected from each of the MA line and the control and thoroughly mixed in 3 ml of fresh tryptic soy broth in an 18 × 150 mm glass capped tube. DNA extraction was performed on 1 ml of this mixture and used to quantify the initial frequency of each competitor (N_{ia}; N_{ib}). The remaining 2 ml was then incubated in a 37°C and shaken at 30 rpm for 17 h, until early exponential phase. Four biological replicates were conducted for each MA assay for a total of

Downloaded from https://academic.oup.com/genetics/article/219/2/iyab117/6325026 by UNC Charlotte user on 29 July 2022

172 competitions. Genomic DNA was extracted before and after the competition assay for the absolute quantification of initial (N_{ia}; N_{ib}) and final (N_{fa}, N_{fb}) wild type and mutant populations to calculate relative fitness after competition. To measure the relative fitness of MA lines using ddPCR, we did the absolute quantification of the MA progenitor that had an inserted *lacZ* allele (a: Ancestor) and compared them to the absolute copies of genomic DNA that lack the *lacZ* allele (b: MA lines). ddPCR measurements were taken at the start and end of competition via appropriate dilutions where N_{ia} and N_{ib} were the initial DNA frequencies of the wild-type and mutant bacteria, respectively, and N_{fa} and N_{fb} were the final DNA frequencies of the wild-type and mutant bacteria, respectively (Chevin and Lande 2011; Perfeito *et al.* 2014). The difference in growth (Δr_{ab}) between the two strains was calculated as:

$$\Delta r_{ab} = \ln (N_{fa}/N_{ia}) - \ln (N_{fb}/N_{ib})$$

s_{cum} was then calculated, where G normalizes the measurement of s_{ab} to the number of generations that were elapsed:

$$s_{cum} = s_{ab} = \Delta r_{ab}/G$$

G is equal to:

$$G = \log 2(N_{fb}/N_{ib})$$

In ddPCR measurements, plating on X-Gal is not required as s_{cum} was instead measured by quantifying the absolute copies of genomic DNA. Nevertheless, in control experiments, both plate assays using X-gal (Dillon and Cooper 2016) and ddPCR assays (four biological replicates without X-gal) showed that the *lacZ* insertion had no significant effect on relative fitness in the tryptic soy media environment used in this study (Figure 2, A and B). Wild-type *B. cenocepacia* HI2424 strain contain a *lacZ* gene inserted at the attTn7 site. The insertion causes colonies to turn blue after exposure to X-gal (Choi *et al.* 2005). In traditional colony counting methods, overexpression of this phenotype is used to track the growth differences between the wild-type and each MA lineage but could consume cellular resources that might be deleterious and bias fitness measurements under certain conditions. The detailed molecular method for the insertion of *lacZ* into ancestral wild type *B. cenocepacia* HI2424 strain has been described in the Supplementary section.

Genomic DNA extraction and quantification

For ddPCR, genomic DNA was extracted from the reserved initial and final 1 ml of tryptic soy broth using the Wizard genomic DNA purification kit (Promega, Madison, WI, USA). The concentration of genomic DNA was quantified using a Nanodrop (ND-1000 Spectrophotometer, Thermo Fisher Scientific, Wilmington, DE, USA) and stored in DNA rehydration buffer at −20°C. For quality assessment of isolated nucleic acids, all samples had 260/280 ratios between 2.00 and 2.05 and 260/230 ratios between 1.05 and 2.01, with DNA concentrations estimated to be in the range of 100–500 ng/µl.

Sample preparation and optimization for ddPCR

To optimize multiplex ddPCR for this experiment, we diluted purified genomic DNA from the cultures at the end of the competition assay in 10 mM, nuclease-free Tris buffer (pH 7.5) and quantified each sample using the Qubit high sensitivity DNA broad range fluorometer assay (Thermo Fisher Scientific, Waltham, MA, US). DNA was digested using fast digest *EcoR1* restriction digestion enzyme (Fisher Scientific MA, USA) for 5 min at 37°C to allow for higher resolution of ddPCR without disrupting the target amplicon of interest. DNA concentrations from 0.020 ng/µl to 300 ng/µl were tested on ddPCR to determine the optimal DNA concentration for separation of wild-type and mutant populations (Supplementary Figure S1, S2 and Table S2), and was found to be optimized at 0.028 ng/µl. DNA for all samples were then diluted to 0.028 ng/µl for ddPCR. Calculations for final frequencies considered the dilutions that occurred at this step.

Primer sets were specifically designed to target the region of the *lacZ* insert in the wild-type *B. cenocepacia* HI2424 genome (Supplementary Table S7), to test the efficiency of the primers, we ran each primer individually for wild type and mutant lineages at 59°C (DNA template: 0.028 ng/µl, primers: 100–300 nM per reaction). For an optimal reaction, we expect two clearly separated bands of positive and negative droplets in 1D and 2D plots (Supplementary Figure S2). However, low binding affinity, inadequate annealing temperature, inaccurate primer or DNA

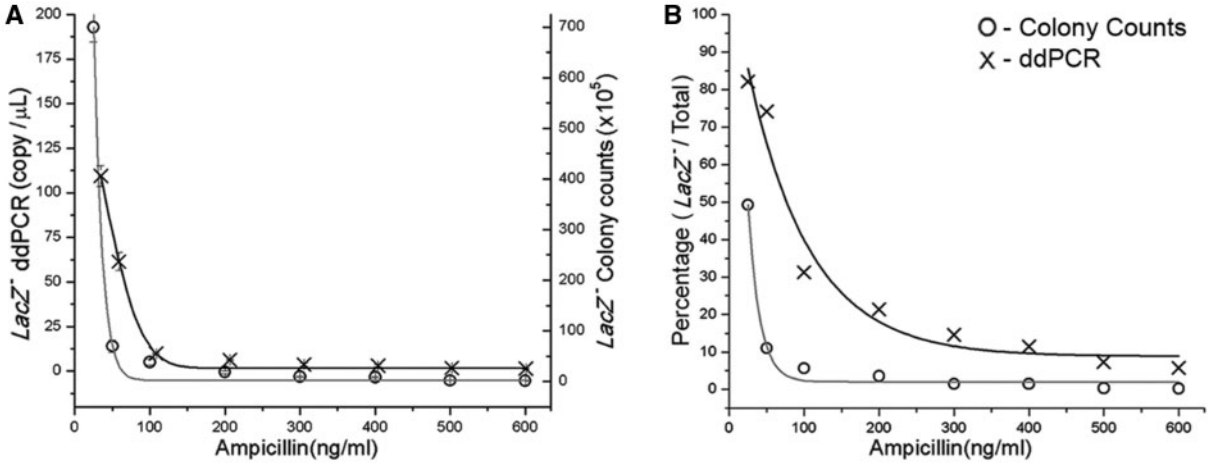

Figure 2 ddPCR quantification and colony counts of *B. cenocepacia* HI2424 with increasing ampicillin concentration. (A) ddPCR and colony counts of *lacZ*⁻ *B. cenocepacia* HI2424. Fitted regression equations for ddPCR [$y = 365.96 \times e^{(-x/20.18)} + 13.18 \times e^{(-x/598.04)} - 3.50$, $R^2 = 0.94$] and Colony Counts [$y = 20883.47 \times e^{(-x/7.17)} + 68.47 \times e^{(-x/157.82)} + 0.11$, $R^2 = 0.98$]. (B) Percentage of *lacZ*⁻ from ddPCR and colony counting at increasing concentrations. Fitted regression equations given as ddPCR: [$y = 103.78 \times e^{(-x/82.84)} + 8.79$, $R^2 = 0.96$]; Colony Counts: [$y = 236.19 \times e^{(-x/15.52)} + 2.01$, $R^2 = 0.99$]. All regression equations are significant at a $P < 0.001$ level.

Downloaded from https://academic.oup.com/genetics/article/219/2/iyab117/6325026 by UNC Charlotte user on 29 July 2022

template concentrations or buffer interference can result in indistinct positive and negative droplets (Supplementary Figure S2), generating "rain," or droplets that lie in between threshold points that cannot be resolved. To resolve ddPCR rain, we optimized PCR conditions for multiplex assays. Further details on DNA optimization and primer optimization can be found in Supplementary Materials (see Supplementary Information).

Eva green-based ddPCR WorkFlow

Eva Green-based ddPCR reaction mixtures contained 1× ddPCR Eva Green Super mix (Bio-Rad), 100 nM mutant, 300 nM wild type specific primers (Sigma Aldrich USA), and bacterial genomic DNA in a final volume of 22.5 µl. No template negative control wells were created by replacing input DNA solutions with TE buffer (pH 8.0). A 20 µl aliquot was taken from each of the assembled ddPCR mixtures and pipetted into each sample well of an eight-channel disposable droplet generator cartridge (Bio-Rad, Hercules, CA, USA). A 70 µl volume of Droplet Generation Oil for Eva Green (Bio-Rad) was then loaded into each of the eight oil wells. The cartridge was placed into the droplet generator (Bio-Rad) where a vacuum was applied to the outlet wells to simultaneously partition each 20 µl sample into nano-liter sized droplets. After ~1.5 minutes, the cartridge was removed from the generator, and the droplets that had collected in each of the independent outlet wells were transferred with a multichannel pipet to a 96-well polypropylene plate (Eppendorf, Hamburg, Germany). The plate was heat-sealed with foil using a PX1 PCR Plate Sealer (Bio-Rad) and placed in a conventional thermal cycler (C1000 Touch, Bio-Rad). Unless otherwise stated, thermal cycling conditions for all Eva Green assays consisted of an activation period (5 minutes at 95°C) followed by 40 cycles of a two-step thermal profile comprising of a denaturation step (30 seconds at 95°C) and a combined annealing extension step (60 seconds at 60°C). A dye-stabilization step was also included at the end of each Eva Green thermal cycling protocol (4°C for 5 minutes, 95°C for 5 min, and finally a 4°C indefinite hold). After PCR, the 96-well plate was loaded into the QX200 Droplet Reader (Bio-Rad), and the appropriate assay information was entered into the analysis software package (Quanta Soft, Bio-Rad). Droplets were automatically aspirated from each well and streamed in single-file past a two-color fluorescence detector and finally to waste. The quality of all droplets was analyzed and rare outliers (*e.g.*, doublets, triplets) were gated based on detector peak width. Analysis of the ddPCR data was performed with Quantasoft analysis software (Bio-Rad) that accompanied the QX200 Droplet Reader.

Data processing

Droplet digital PCR relies on the ability to distinguish between partitions that contain amplicons and those that do not. Partitions that contain amplicons can be identified by the presence of increased fluorescence using a variety of detection chemistries common to qPCR such as intercalating DNA dyes or fluorophore-labeled oligonucleotides. The target concentration is then calculated by taking the ratio of positive to negative droplets. The high-throughput and robotic digital PCR system is easier to use and more precise than other quantification methods (Hindson *et al.* 2011; Hayden *et al.* 2013).

The wild-type and the mutant strains from this study are nearly isogenic except for the small set of mutations that accumulated over the course of the MA experiment, so primers specific *B. cenocepacia* will amplify both strains. *LacZ* primers will only amplify the wild-type population. To ensure accurate quantification of multiplex targets, it is necessary to separate single (*B. cenocepacia* only) and double positive amplicons (*B. cenocepacia* and *lacZ*) (Whale *et al.* 2016). From our results, we will observe three visible clusters (i) negative partitions that contain no targets for either primer set, (ii) single-positive partitions (wild-type only partitions), and (iii) positive partitions that contain a signal from both wild type and mutant (combined cluster). The final single-positive cluster (mutant only partitions) is subsumed into the combined cluster since it always produces a signal in the presence of any amplicon. Following Whale *et al.* (2016), the number of mutants can be calculated using equations 1 and 2, with c_0 defined as the number of partitions in the double-negative cluster, c_{wt} as the wild type (wt) only cluster, and $c_{combined}$ as the combined cluster:

$$\text{Mutant type} = \ln(c_0 + c_{wt} + c_{combined}) - \ln(c_0 + c_{wt}) \quad (1)$$

$$\text{wild type} = \ln(c_0 + c_{wt}) - \ln(c_0) \quad (2)$$

These equations were used to separate and measure the mutant (equation 1) and wild-type frequencies (equation 2). These equations require that there are no interactions between the wild-type and mutant targets, which is the case, as each primer is unable to amplify the other target.

After PCR cycling, the fluorescence end-point signal associated with each partition is measured by the reader instrument. This signal was plotted on a one-dimensional (1D) scatter graph, with the event (partition) number along the *x*-axis and the fluorescent amplitude along the *y*-axis (Supplementary Figure S1 and S2). In a well-optimized assay, two visually distinct populations are observed; positive partitions that have high fluorescence, and negative partitions that have low (or background) fluorescence. An automatic threshold is set by the software to separate these two populations, and these measurements were taken at face value. Manual thresholds were used to separate c_{wt} and $c_{combined}$, and all measurements had clear separation between these two partitions (Supplementary Figure S1, S2, and Table S8).

In the duplex reactions, we discriminated between targets of different sizes based on their fluorescence intensity (McDermott *et al.* 2013), and each cluster was used to measure the wild type and mutant populations. This reduces technical errors, such as accumulated pipetting inaccuracy, thereby making it possible to measure smaller differences than the same comparison using parallel uniplex reactions (Whale *et al.* 2013), while also reducing reagent costs and runtime. We note that multiplexing or duplexing ddPCR comes at a slight cost to resolution of the overall precision of ddPCR measurements, which is dependent on the number of partitions used (Pinheiro *et al.* 2012). For studies that require higher resolution, probes can be designed that are specific to a mutation or uniplex ddPCR can be used (Dillon and Cooper 2016).

Statistical analysis

All statistical analyses were performed in Origin Pro 8.0 using the Stats analysis package. Shapiro–Wilk tests were performed to test whether the data violate the assumption of a normal distribution. Unless otherwise noted, the data did not violate a normal distribution and the mean and SEM were presented. If the data were not normally distributed, the results of the Shapiro–Wilk test were reported, and the mean and SD were presented. For independent two-tailed *t*-tests, all *P*-values were corrected for multiple comparisons using a Benjamini–Hochberg correction (Supplementary Table S2), ensuring that our false positive rate remains below 5% across all 43 lineages (Benjamini and Hochberg 1995). Corrected *P*-values that were below a threshold of 0.05 were considered significant. Nonparametric Pearson

Downloaded from https://academic.oup.com/genetics/article/219/2/iyab117/6325026 by UNC Charlotte user on 29 July 2022

product moment correlations were used to evaluate the correlation between the number of mutations in a lineage and its selection coefficient, as well as the correlation between the selection coefficients of lineages. Lastly, to test for effects of selection coefficients on the total number of mutations harbored in each lineage, we performed a regression analysis and ANOVA on the cumulative data set.

Data availability

GSA Figshare portal used to upload supplemental files containing Supporting Information, Figure S1–S5, and Table S1–S9: https://doi.org/10.25386/genetics.14920347. DNA sequences of *Burkholderia cenocepacia* HI2424 (Accession: NC_008542.1 GI: 116688024) and *lacZ* vector (Accession: M77789, M11662, Version: M77789.2 alleles) are available in GenBank. All data are available in the supplemental section and all strains and primers are available upon request.

Results

Droplet-digital PCR to measure microbial fitness

The application of ddPCR to measure fitness is a new use of the ddPCR platform (Boynton *et al.* 2017). Therefore, it was important that we first establish that ddPCR can accurately quantify microbial populations and that ddPCR measurements have low measurement variance. To this extent, we applied ddPCR to *B. cenocepacia* strains with and without a *lacZ* AmpR reporter construct (see Supplemental Materials) that was previously shown to have no effect on fitness in laboratory tryptic-soy liquid media culture (Dillon and Cooper 2016). The $lacZ^+$ and $lacZ^-$ strain of *B. cenocepacia* exhibited resistance and susceptibility to ampicillin respectively (Supplementary Table S1), so we grew both strains with increasing dosage of ampicillin and measured each bacterial population in triplicate following 17 h. of growth using both colony counting methods and ddPCR (see Materials and methods). We find that variance on the ddPCR measurement of the $lacZ^+$ population is significantly lower than that from colony counting methods ($F = 3.78$, $P < 2 \times 10^{-13}$, Figure 2, Supplementary Table S1), indicating that variation in fitness measurements can be minimized using ddPCR. The coefficient of variation (CV) is widely used to express the precision of an assay, providing a ratio of the standard deviation to the mean. We calculated the CV for ddPCR as 4.67% (less than 5% of the mean) and 20.44% for colony counting (Supplementary Table S1). This lower CV indicates higher precision in ddPCR as compared to colony counting in this control experiment.

Furthermore, we find that the fitted curve follows a more consistent decline as ampicillin concentration increases (Figure 2B), which suggests ddPCR can detect minor changes in the level of fitness that may not be detectable using traditional colony counting methods. We also find that ddPCR performed better at quantifying at high concentrations of ampicillin (≥50 ng/ml), while repeated plating was required to generate enough viable colonies to measure fitness using colony-counting methods (0–1 colonies). Taken together, we find that ddPCR can provide high-resolution fitness measurements with low measurement error.

To determine the optimal DNA concentration for this locus with ddPCR, we quantified $lacZ^+$ and $lacZ^-$ *B. cenocepacia* at known concentrations of DNA (280 ng/μl, 28 ng/μl, 2.8 ng/μl, 0.28 ng/μl, 0.028 ng/μl, 0.0028 ng/μl). We find that ddPCR has optimal separation at 0.028 ng/μl, with low variance in measurement and high precision (lowest measurement standard deviation at 0.028 ng/μl, ~5% CV, Supplementary Table S2), and the sample partitioning and removal of PCR amplification bias allows for accurate separation of $lacZ^+$ and $lacZ^-$ targets when using this locus (Supplementary Figure S1 and S2). The high accuracy observed is consistent with many different studies which have used ddPCR for nucleic acid quantification (Hindson *et al.* 2011; McDermott *et al.* 2013; Boynton *et al.* 2017).

Measuring the cumulative effect of mutations on fitness in *B. cenocepacia* HI2424 MA lines using ddPCR

Once we established the accuracy and resolution of ddPCR on the *lacZ* locus, we performed competition experiments with MA lineages of *B. cenocepacia* HI2424 transformed with $lacZ^+$ against the ancestral *B. cenocepacia* HI2424 strain ($lacZ^-$) in tryptic soy liquid media. DNA was extracted from the initial and final populations, quantified using ddPCR, and converted into selection coefficients as a measurement of relative fitness (see Data Processing, Figure 3A). The cumulative effect of mutations on fitness (s_{cum}) to the ancestor across all ddPCR-MA lineages in tryptic soy liquid media were -0.041 ± 0.038 (SD) per generation over the experiment and the data was normally distributed (Shapiro–Wilk's Test; $P = 0.092$) (Supplementary Tables S3 and S4). s_{cum} in these MA lines were also previously measured using colony-counting methods and remains one of the few studies of fitness effects in genotyped MA lines (Dillon and Cooper 2016). We compared these fitness effects derived from ddPCR-MA with those from colony counting methods, and found that ddPCR-MA has a broader distribution of fitness measurements and exhibited a significantly greater decline (1.68×) in mean fitness (t-test, $P = 0.005$, Figure 3A). The distribution from both methods exhibited signs of bimodality [Dip-test (Hartigan and Hartigan 1985), ddPCR $P = 0.048$, colony counting $P = 0.033$]. Both the colony counting, and ddPCR-MA experiments observed a small deleterious change in the cumulative effect of mutations on fitness in most lines (0 to −0.07 in 72.1% of MA lines from ddPCR and 76.7% of MA lines from colony counting).

When we compare the raw data from ddPCR-MA and colony counting methods, we find a weak but positive correlation between the cumulative effect of mutations on fitness $r = 0.297$, $P = 0.05$, Figure 4. To examine why the two methods deviated, we repeated colony counting methods in-house on a random subset of MA lineages (10) using the same procedures as described in Dillon and Cooper (2016). Our in-house measurements exhibited higher variance in these lines than previously reported (average colony counting variance in-house = 0.002; average colony count Dillon *et al.*, = 0.0009, Supplementary Table S6). Furthermore, the ten repeated in-house colony counting measurements showed a positive but non-significant correlation with ddPCR measurements ($\rho = 0.47$, $r^2 = 0.17$, Supplementary Figure S3 and Table S6). Taken together, this suggests that there are differences in the measurement variance and mutational effects for fitness measurements based on colony counting on solid media and those based on liquid growth in ddPCR.

Average effect of spontaneous mutations

To estimate the average fitness effect of individual mutations, we divided the cumulative effect of all mutations per line by the total number of mutations identified in the MA study (including base substitutions and indels). We contrasted the average selective effect (s) per mutation across all lines from ddPCR-MA [-0.0070 ± 0.0013 (SEM)], with those from colony counting methods with repeated measurements [-0.0040 ± 0.0007 (SEM)] (Dillon and Cooper 2016) and find a significant difference in the

Downloaded from https://academic.oup.com/genetics/article/219/2/iyab117/6325026 by UNC Charlotte user on 29 July 2022

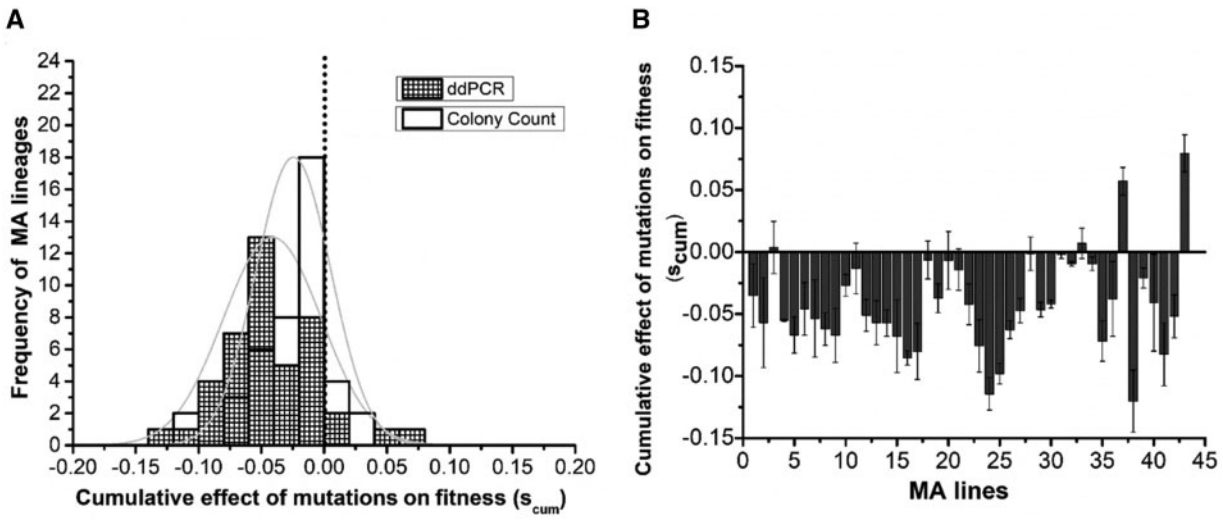

Figure 3 Cumulative effect of mutations on fitness in *B. cenocepacia* MA lineages. (A) Measurements for ddPCR and colony counting method for the cumulative effect of mutations on fitness overlaid with each other (s_{cum}) (Dillon and Cooper 2016). (B) Graph for $\mathbf{s_{cum}}$ relative to the ancestral strain shown for each MA line measured by ddPCR (gray) with the error bars indicating SEM. Nonlinear fitting equation for ddPCR, y = −0.022 + (−0.12/0.50sqrt(pi/2)) e(−2((x − 16.68)/w) ^2) and nonlinear fitting equation for colony count y = −0.044 + (2.06/5.44sqrt(pi/2)) e(−2((x − 95.55)/w)^2).

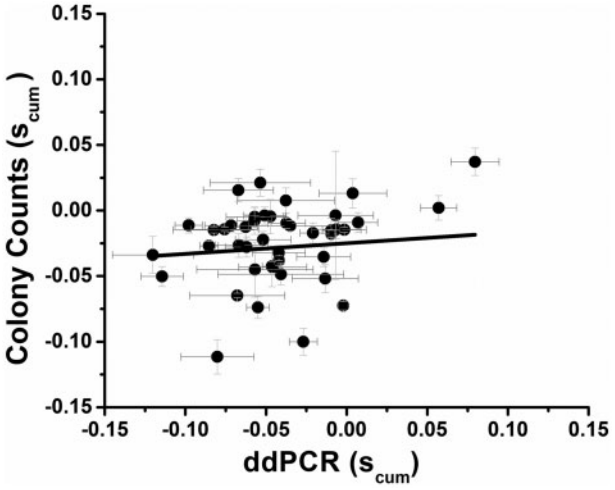

Figure 4 Correlation between the cumulative effect of mutations on fitness (s_{cum}) of *B. cenocepacia* MA lines measured using ddPCR and colony counting by Dillon and Cooper (2016) ($r = 0.297$, $P = 0.05$). Error bars indicate standard error of the mean. Linear regression equation y = 0.552x − 0.007, Chi^2 = 105.03 and R^2 = 0.794.

average effect of spontaneous mutations (*t*-test, $P = 0.006$, Supplementary Table S3). This order of magnitude difference is on par with fitness studies performed in different labs in *E. coli* K12 MG1655, which vary by one order of magnitude in colony counting (Kibota and Lynch 1996b) and microfluidic studies (Robert *et al.* 2018). Across different organisms, the average selection coefficient per mutation is on the order of 10^{-3}–10^{-4} for a majority of bacteria (Table 1). Eukaryotes show more variation in average selective coefficient per mutation ranging from 10^{-1} to 10^{-4}.

The deleterious effect (s_d) and the beneficial effect (s_b) of a single mutation were estimated, with $s_d = -0.0443 \pm 0.006$ (SEM) and sb = 0.04791 ± 0.0214 (SD) when measured by ddPCR and average sb = 0.013 ±.0.005 (SD); average $s_d = -0.048 \pm 0.007$(SD) when measured by cell counts (calculations described in Supplementary Text) (Dillon and Cooper 2016).

The average fitness effect of mutations in the MA lines were found to be lower than that of the ancestral wild type, consistent with the idea that most mutations are deleterious. However, the relationship between the total number of mutations and the decline in fitness was not significant ($r = -0.04$, $P = 0.77$, Supplementary Figure S4) supporting the idea that beneficial mutations were masking the effect of deleterious ones, that only a few mutations are driving large changes in relative fitness, or that antagonistic epistasis between deleterious mutations are muting the effect (Dillon and Cooper 2016; Böndel *et al.* 2019). Despite acquiring multiple mutations, we found that six MA lines did not differ significantly in fitness from the ancestral strain (Supplementary Table S4).

Mutations that lie in different regions can have different effects, and we expect that mutations that generate nonfunctional proteins like nonsense base-substitution mutations or coding insertion-deletion mutations (indels) to have the most deleterious effects, followed by missense base substitutions that generate modified proteins, then synonymous and non-coding base substitutions that do not alter protein sequences. Using colony counting methods, Dillon and Cooper (2016) had previously observed that there is a slight, but non-significant increase in nonsense and missense base substitutions, as well as coding indels in lineages with reduced fitness. ddPCR-MA showed a similar non-significant relationship (Supplementary Figure S5) between the same mutation classes and relative fitness. This suggests that either more mutations from these classes are required to detect a large fitness decline, that only a small number of mutations with protein sequences generate strong fitness effects, or that epistatic effects are obscuring these relationships.

Plasmids can sometimes contain important genes, and loss of plasmids can affect replication, transcription, and enzymatic activity, which in turn can affect survivability. The loss of 0.164 Mb plasmid in *B. cenocepacia* results in a significant loss in fitness in ddPCR and is consistent with what has been previously observed using colony counting methods ($s_d = -0.03$ by ddPCR-MA and $s_d = -0.05$ by colony counts, *t*-test, $P = 0.18$) (Dillon and Cooper 2016). While a majority of mutations appear to contribute a neutral or slightly deleterious effect, we did not find a significant relationship between the total number of mutations harbored within an MA line

Downloaded from https://academic.oup.com/genetics/article/219/2/iyab117/6325026 by UNC Charlotte user on 29 July 2022

Table 1 Cumulative fitness effect (s) scaled by generation time per mutation

Organism	Method used	MA lines	Average selective coefficient per mutation (s)	References
Escherichia coli MG1655	Microfluidics	1000[a]	−0.003	Robert *et al.* (2018)
Escherichia coli K12 MG1655	Colony Counts	50	−0.03	Kibota and Lynch (1996a)
Escherichia coli BWT strain	Spectrophotometry	50	−0.012	Gordo *et al.* (2011)
Burkholderia cenocepacia HI2424	Droplet digital PCR	43	−0.0070	This study
Burkholderia cenocepacia HI2424	Colony Counts	43	−0.0041	Dillon and Cooper (2016)
Salmonella typhimurium (S. enterica)	Spectrophotometry	12	−0.00145	Lind and Andersson (2008)
Saccharomyces cerevisiae	Phenotypic traits	151	−0.12–0.20	Hall and Joseph (2010)
Chlamydomonas reinhardtii CC-1373	Flow Cytometry	60	−0.0008	Kraemer *et al.* (2017)
Arabidopsis thaliana	Phenotypic traits	40	−0.09	Shaw *et al.* (2002)
Drosophila melanogaster	Viability traits	5	0–0.125	Sharp and Agrawal (2016)
Caenorhabditis elegans N2	Phenotypic traits	9	−0.12	Vassilieva *et al.* (2000)

[a]Estimated number of MA lines from microfluidic methodology (Robert *et al.* 2018).

and a change in fitness (Supplementary Figure S4). MA lineages containing numerous mutations did not show any deviation in relative fitness from the ancestral wild type (*t*-test, $P > 0.05$, MA4, MA24, MA29, MA31, MA43, and MA68, Supplementary Table S4). One possible explanation for this is that epistatic interactions between beneficial and deleterious mutations negate each other (Davies *et al.* 1999; Heilbron *et al.* 2014). Another possible explanation is that many mutations do not cause significant deviations in fitness.

Mutations of large effects

Twenty-seven out of 43 MA lines exhibited s_{cum} that significantly deviated from mean fitness (99% confidence interval, −0.057 to −0.027), suggesting that these 27 lines harbored mutations of large effect, or multiple mutations that combined to have a large effect. The lack of a correlation between the number of mutations and a decline in fitness suggests that these large effects are not simply additive and are likely to be caused by mutations of large effects or epistatic interactions between mutations of smaller effect. The mutations in these lines are known from high-throughput Illumina sequencing of *B. cenocepacia* MA (Dillon and Cooper 2016) and revealed multiple mutations within the repeat region of the excinuclease ABC subunit C (Locus Tag: BCEN2424_RS12080) and in LysR/M transcriptional regulators (Table 2, Supplementary Table S5).

Interestingly, MA62 and MA74 which had extreme increases in fitness and MA63 which had an extreme decline in fitness harbored mutations within the same excinuclease locus. Excinuclease is a *uvrC* homolog, and *uvrC* is known to be responsible for repairing UV damage (Crowley *et al.* 2006). Non-identical mutations, or mutations at different loci, were also found in LysR-like transcriptional regulators in MA lines with both significant increases and declines in fitness (Table 2). While these mutations may not necessarily be responsible for the changes in fitness, we do find that these types of mutations are overrepresented in the 27 MA lines that exhibited selection coefficients that significantly deviated from the mean (Table 2, Supplementary Table S9, χ^2-test, $P = 0.01$, df = 1). Repeats and paralogous loci are known to have higher mutation rates so these mutation types may simply arise more often than others. Nevertheless, transcriptional and replication loci are known to have significant effects that can contribute to these large changes in fitness (Whale *et al.* 2013; Sprouffske *et al.* 2018). These changes in LysR were not previously reported by Dillon and Cooper (2016) in their colony counting study, as they did not closely examine the types of mutations that were driving the large changes in observed fitness. Furthermore, higher precision estimates can make clear the contributions that different loci have to changes in fitness.

It should be noted that that the cumulative fitness effect of mutations can be measured by many different methods (ddPCR-MA, colony counting, etc., Table 1). The finding that overall mutations are slightly deleterious with some mutations of large effects is consistent with that found in previous studies (Dillon and Cooper 2016). Both ddPCR and colony counting methods also showed that a majority of the lineages significantly deviated from mean fitness, with both methods being positively correlated and showing a bimodal distribution.

Discussion

The effect that a mutation has on an organism can determine the fate of that mutation, and understanding the fitness effects of mutations is a critical part in determining the mutations that will rise to high frequency in different environments (Charlesworth *et al.* 1993; Charlesworth and Charlesworth 1998; Charlesworth *et al.* 2004; 2009), how complex traits that require multiple mutations evolve (Muller 1964; Kondrashov 1988), and the expected levels of genetic diversity and divergence that are generated over time. Studying fitness effects is extremely critical in understanding how mutations interact with each other (epistasis), the magnitude of interactive effects (multiplicative, additive, subtractive, or exponential effects), and how epistasis can drive the evolution of complex traits. Measuring fitness effects can also help us understand the evolution of mutator and antimutator alleles (Sniegowski *et al.* 1997; Tenaillon *et al.* 1999; Sniegowski *et al.* 2000), which determines not only the pace of cellular evolution, but also the development of genetic diseases and disorders.

Recent development of high-throughput genotyping using fluorescent based microfluidics has opened the door to highly accurate measurements of fitness. In this study, we leverage the exceptional properties of digital-droplet PCR (ddPCR) to quantify the relative fitness of mutant organisms after competition. Overall, we find that ddPCR-MA has high-resolution and precision, and that fitness measurements between ddPCR-MA and traditional colony counting methods are positively correlated (Figure 4). In comparison with traditional colony-counting methods, both methods observe a bimodal distribution for fitness effects and both methods can show that the fitness effects are slightly deleterious effect overall, with some mutations having large effects.

However, there are some significant differences that are shown in the ddPCR-MA experiment. ddPCR-MA reveals a

Downloaded from https://academic.oup.com/genetics/article/219/2/iyab117/6325026 by UNC Charlotte user on 29 July 2022

Table 2 Lys and excinuclease mutations found in MA lines with a large change in fitness measured using ddPCR

S_{cum}	Line_Chr_Site	Mut.	Type	Gene product
−0.1201	63_1_2680077	+12	COD	excinuclease ABC subunit C
−0.1201	63_1_2087196	C>T	S	LysM domain/BON superfamily protein
−0.0678	26_2_1996747	−8	COD	LysR family transcriptional regulator
−0.0572	24_3_598965	C>T	NS	LysR family transcriptional regulator
−0.0572	24_1_143249	G>C	S	LysR family transcriptional regulator
−0.0536	11_1_2680077	+6	COD	excinuclease ABC subunit C
−0.0269	14_1_743097	C>T	S	LysR family transcriptional regulator
−0.0209	65_1_46681	C>T	NS	LysR family transcriptional regulator
−0.0096	56_3_18850	G>C	NS	LysR family transcriptional regulator
0.0571	62_1_2680077	−6	COD	excinuclease ABC subunit C
0.0796	74_1_2680077	+6	COD	excinuclease ABC subunit C
0.0796	74_1_2750800	C>G	NS	LysR family transcriptional regulator

Cumulative effect of mutations on fitness (S_{cum}) shown for MA line where mutation was found. 99% confidence interval for S_{cum} is $-0.057 < X < -0.027$, significant selection coefficients are bolded. Line_Chr_Site indicates the MA line, chromosome, and the site of the mutation delimited by underscore. Mutation column (Mut.) is the size of the indel (#) or the base substitution mutation, with Type column denoting whether the mutation is synonymous (S), nonsynonymous (NS), intergenic (IG), or coding (COD). Mutations in excinuclease and Lys regulatory proteins are shown.

significantly lower total mean fitness and significantly larger average effect per mutation. There could be a few explanations for this. First, colony counting methods may be underestimating the overall effect. Although colony counting is a time-tested method for rough growth estimates, it is well documented that phenotypic lag or growth lag from solid media growth can lead to underestimation of cells counts and consequently underestimate fitness effects (Carballo-Pacheco *et al.* 2020).

Second, this difference may be due to variation in fitness measurements in different laboratory and environmental conditions. Growth time, handling, water, and growth temperature are all factors that can influence cell growth. It is important to note that mutations can have different effects in different conditions (Blount *et al.* 2012; Dillon and Cooper 2016). Growth on solid media is a very different condition than growth in liquid media, and the effect of mutations in each condition may vary considerably. For example, the number of context-specific mutations have been shown to vary under different growth media and conditions (Foster *et al.* 2015; Sung *et al.* 2015), and their effect may also vary significantly between liquid and plate growth. The MA experiment was evolved on solid media and thus the mutations are likely to have less effect on solid media competition experiments, leading to a small but consistently greater deleterious effect in liquid ddPCR assays.

Third, every technology has some biases and this includes ddPCR. Although ddPCR can quantify DNA at high resolution, it also measures DNA from dead cells. To minimize the effect of dead cells on fitness measurements, DNA was extracted from all competition experiments at the exponential phase (17 h) where live cells significantly outnumber dead cells. If higher precision is needed, propidium monoazide can be added to the extracted DNA prior to ddPCR to prevent amplification and quantification of dead cell DNA (Gobert *et al.* 2018).

Furthermore, despite the precision and accuracy in measurements, ddPCR may also have some issues such as false positives, poor signal thresholding, and molecular dropout. Proper assay design and validation are critical to minimize these issues. We took significant steps to optimize and test our marker, and investigators developing ddPCR fitness assays should take particular caution with marker design to ensure that the marker has limited fitness effects.

Thus, while it is nearly impossible to remove variation and biases in fitness assays across different experimental conditions and laboratory environments, high-throughput repeatable fitness measurements through ddPCR will allow us to rapidly measure fitness effects across different environmental conditions and identify environment-specific differences in fitness effects.

Across 43 MA lines, we found 27 MA lines that displayed significant changes in relative fitness. Upon further examination, we find that the ddPCR-MA lines that exhibited the largest beneficial and deleterious fitness contained locus-specific mutations within a repeat unit of the DNA repair enzyme excinuclease with both extreme positive and negative fitness. Furthermore, we observe a significant excess of mutations within transcriptional regulators in lines from both ends of the fitness spectrum (Table 2). While these mutations are overrepresented, they arise in a genetic background with other mutations, and may not be the cause of these large fitness differences. We propose that in the future, ddPCR-MA can be used on intermediate MA stocks that contain a subset of these mutations allowing us to tease out these interactions at high resolution.

For available data, intraspecies variation of the average fitness effect of a mutation does not differ more than two-fold between experiments regardless of method (Table 1). On the other hand, the average effect varies by three orders of magnitude in different microorganisms, with the average effect increasing with genome size. It would be expected that mutations would have a larger effect in smaller genomes with a higher density of coding sites, as the probability of a mutation landing within essential protein coding regions increases. However, the opposite is observed, with mutations in species with larger genomes having a larger average effect. It remains unclear why this is the case, so developing consistent, reliable methods of fitness measurements such as ddPCR-MA is critical to understand how these patterns evolve across species.

Although ddPCR assays are highly accurate, an automated robotic ddPCR machine has a large upfront cost (~100k USD). Thus, accuracy and repeatability have a high upfront cost. Each sample costs ~1.60$ to run, which is comparable to real-time PCR experiments that also use fluorescent labeling technology (Hayden *et al.* 2013; Taylor *et al.* 2017). Thus, the accuracy comes at high upfront cost. As more and more laboratories begin to use third-generation PCR technology, this cost should decrease. ddPCR assays provide high-throughput measurements and 96 samples take approximately 5 h to process once DNA is isolated, so when compared to other methods such as flow cytometry or microfluidics that require expensive equipment (Table 1), ddPCR can be completed more rapidly at a cheaper cost. Colony counting is considerably cheaper at a slightly lower resolution and accuracy,

Downloaded from https://academic.oup.com/genetics/article/219/2/iyab117/6325026 by UNC Charlotte user on 29 July 2022

but also requires the most working time. For ddPCR, the largest working time is the optimization of sample concentrations, PCR temperatures, and primers for the target loci. This study provides sample preparation steps, primers, and optimization protocols that can enable other labs to easily replicate and extend this work into other organisms.

To conclude, we have established that ddPCR-MA can be used to provide high-resolution measurements of fitness effects in microorganisms by tracking changes in their genotype frequency during competition. Our work shows a significant correlation between ddPCR and colony counting methods, and has identified the presence of specific mutation types associated with strong beneficial and deleterious effects. The ddPCR-MA methodology is agnostic with respect to cell type, can use fluorescent probes to target specific alleles, and this high-throughput, highly accurate method can be adapted to monitor genotype frequencies and fitness effects in any environment desired (even complex liquid environments such as blood). This platform will be predominantly useful when studying epistatic effects of spontaneous mutations in microorganisms and can be further leveraged to perform fitness assays in RNA viruses.

Acknowledgments

We thank Jaleh Abedini and Charanya Uppalapati for DNA isolation process and help in labeling competition assays tubes.

Funding

This work is supported by the National Science Foundation Division of Molecular and Cellular Biosciences grant number 1818125 to principal investigator W.S.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Author contributions

W.S. and A.R. conceived the ideas and designed methodology; W.S., A.R., M.M.D., D.P., and N.T.T. collected the data; A.R. and W.S. analyzed the data; A.R., W.S., M.M.D., and V.S.C. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Literature cited

Agnoli K, Schwager S, Uehlinger S, Vergunst A, Viteri DF, *et al.* 2012. Exposing the third chromosome of *Burkholderia cepacia* complex strains as a virulence plasmid. Mol Microbiol. 83:362–378.

Barrett RD, M'Gonigle LK, Otto SP. 2006. The distribution of beneficial mutant effects under strong selection. Genetics. 174:2071–2079. doi:10.1534/genetics.106.062406.

Bataillon T, Bailey SF. 2014. The year in evolutionary biology. Ann N Y Acad Sci. 1320:76–92.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. J R Stat Soc Series B. 57:289–300. doi:10.2307/2346101.

Blount ZD, Barrick JE, Davidson CJ, Lenski RE. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. Nature. 489:513–518.

Blundell JR, Levy SF. 2014. Beyond genome sequencing: lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer. Genomics. 104:417–430.

Böndel KB, Kraemer SA, Samuels T, McClean D, Lachapelle J, *et al.* 2019. Inferring the distribution of fitness effects of spontaneous mutations in *Chlamydomonas reinhardtii*. PLoS Biol. 17: e3000192.doi:10.1371/journal.pbio.3000192.

Boynton PJ, Stelkens R, Kowallik V, Greig D. 2017. Measuring microbial fitness in a field reciprocal transplant experiment. Mol Ecol Resour. 17:370–380. doi:10.1111/1755-0998.12562.

Caballero A, Keightley PD. 1994. A pleiotropic nonadditive model of variation in quantitative traits. Genetics. 138:883–900.

Carballo-Pacheco M, Nicholson MD, Lilja EE, Allen RJ, Waclaw B. 2020. Phenotypic delay in the evolution of bacterial antibiotic resistance: mechanistic models and their implications. PLoS Comput Biol. 16:e1007930.doi:10.1371/journal.pcbi.1007930.

Charlesworth B, Betancourt AJ, Kaiser VB, Gordo I. 2009. Genetic recombination and molecular evolution. Cold Spring Harb Symp Quant Biol. 74:177–186. doi:10.1101/sqb.2009.74.015.

Charlesworth B, Borthwick H, Bartolome C, Pignatelli P. 2004. Estimates of the genomic mutation rate for detrimental alleles in *Drosophila melanogaster*. Genetics. 167:815–826. doi:10.1534/genetics.103.025262.

Charlesworth B, Charlesworth D. 1998. Some evolutionary consequences of deleterious mutations. Genetica. 102/103:3–19.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics. 134: 1289–1303.

Chen JS, Witzmann KA, Spilker T, Fink RJ, LiPuma JJ. 2001. Endemicity and inter-city spread of *Burkholderia cepacia* genomovar III in cystic fibrosis. J Pediatr. 139:643–649. doi: 10.1067/mpd.2001.118430.

Chevin LM, Lande R. 2011. Adaptation to marginal habitats by evolution of increased phenotypic plasticity. J Evol Biol. 24:1462–1476. doi:10.1111/j.1420-9101.2011.02279.x.

Choi KH, Gaynor JB, White KG, Lopez C, Bosio CM, *et al.* 2005. A Tn7-based broad-range bacterial cloning and expression system. Nat Methods. 2:443–448. doi:10.1038/nmeth765.

Coenye T, Spilker T, Van Schoor A, LiPuma JJ, Vandamme P. 2004. Recovery of *Burkholderia cenocepacia* strain PHDC from cystic fibrosis patients in Europe. Thorax. 59:952–954.

Coenye T. 2010. Social interactions in the *Burkholderia cepacia* complex: biofilms and quorum sensing. Future Microbiol. 5:1087–1099.

Crowley DJ, Boubriak I, Berquist BR, Clark M, Richard E, *et al.* 2006. The uvrA, uvrB and uvrC genes are required for repair of ultraviolet light induced DNA photoproducts in Halobacterium sp. NRC-1. Saline Syst. 2:11–11. doi:10.1186/1746-1448-2-11.

Davies EK, Peters AD, Keightley PD. 1999. High frequency of cryptic deleterious mutations in *Caenorhabditis elegans*. Science. 285: 1748–1751.

Dickinson WJ. 2008. Synergistic fitness interactions and a high frequency of beneficial changes among mutations accumulated under relaxed selection in *Saccharomyces cerevisiae*. Genetics. 178: 1571–1578.

Dillon MM, Cooper VS. 2016. The fitness effects of spontaneous mutations nearly unseen by selection in a bacterium with multiple chromosomes. Genetics. 204:1225–1238. doi:10.1534/genetics.116.193060.

Dillon MM, Rouillard NP, Van Dam B, Gallet R, Cooper VS. 2016. Diverse phenotypic and genetic responses to short-term selection in evolving *Escherichia coli* populations. Evolution. 70:586–599. doi: 10.1111/evo.12868.

Downloaded from https://academic.oup.com/genetics/article/219/2/iyab117/6325026 by UNC Charlotte user on 29 July 2022

Dillon MM, Sung W, Lynch M, Cooper VS. 2015. The rate and molecular spectrum of spontaneous mutations in the GC-rich multichromosome genome of *Burkholderia cenocepacia*. Genetics. 200: 935–946. doi:10.1534/genetics.115.176834.

Engledow AS, Medrano EG, Mahenthiralingam E, LiPuma JJ, Gonzalez CF. 2004. Involvement of a plasmid-encoded type IV secretion system in the plant tissue watersoaking phenotype of *Burkholderia cenocepacia*. J Bacteriol. 186:6015–6024.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nat Rev Genet. 8:610–618.

Eyre-Walker A, Woolfit M, Phelps T. 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics. 173:891–900. doi:10.1534/genetics.106.057570.

Fasanello VJ, Liu P, Botero CA, Fay JC. 2020. High-throughput analysis of adaptation using barcoded strains of *Saccharomyces cerevisiae*. PeerJ. 8:e10118.

Foster PL, Lee H, Popodi E, Townes JP, Tang H. 2015. Determinants of spontaneous mutation in the bacterium *Escherichia coli* as revealed by whole-genome sequencing. Proc Natl Acad Sci U S A. 112:E5990–E5999. doi:10.1073/pnas.1512136112.

Gallet R, Cooper TF, Elena SF, Lenormand T. 2012. Measuring selection coefficients below 10(-3): method, questions, and prospects. Genetics. 190:175–186. doi:10.1534/genetics.111.133454.

Gobert G, Cotillard A, Fourmestraux C, Pruvost L, Miguet J, *et al.* 2018. Droplet digital PCR improves absolute quantification of viable lactic acid bacteria in faecal samples. J Microbiol Methods. 148: 64–73. doi:10.1016/j.mimet.2018.03.004.

Gonzalez CF, Pettit EA, Valadez VA, Provin EM. 1997. Mobilization, cloning, and sequence determination of a plasmid-encoded polygalacturonase from a phytopathogenic Burkholderia (Pseudomonas) cepacia. Mol Plant Microbe Interact. 10:840–851.

Gordo I, Perfeito L, Sousa A. 2011. Fitness effects of mutations in bacteria. J Mol Microbiol Biotechnol. 21:20–35.

Hall DW, Joseph SB. 2010. A high frequency of beneficial mutations across multiple fitness components in *Saccharomyces cerevisiae*. Genetics. 185:1397–1409. doi:10.1534/genetics.110.118307.

Halligan DL, Keightley PD. 2009. Spontaneous mutation accumulation studies in evolutionary genetics. Annu Rev Ecol Evol Syst. 40: 151–172. doi:10.1146/annurev.ecolsys.39.110707.173437.

Hartigan JA, Hartigan PM. 1985. The dip test of unimodality. Ann Statist. 13:70–84.

Hayden RT, Gu Z, Ingersoll J, Abdul-Ali D, Shi L, *et al.* 2013. Comparison of droplet digital PCR to real-time PCR for quantitative detection of cytomegalovirus. J Clin Microbiol. 51:540–546. doi:10.1128/JCM.02620-12.

Heilbron K, Toll-Riera M, Kojadinovic M, MacLean RC. 2014. Fitness is strongly influenced by rare mutations of large effect in a microbial mutation accumulation experiment. Genetics. 197:981–990. doi:10.1534/genetics.114.163147.

Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, *et al.* 2011. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Anal Chem. 83:8604–8610. doi:10.1021/ac202028g.

Keightley PD, Lynch M. 2003. Toward a realistic model of mutations affecting fitness. Evolution. 57:683–685; discussion 686–9.

Kibota TT, Lynch M. 1996a. Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. Nature. 381:694–696. doi: 10.1038/381694a0.

Kibota TT, Lynch M. 1996b. Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. Nature. 381:694–696.

Kondrashov AS. 1988. Deleterious mutations and the evolution of sexual reproduction. Nature. 336:435–440. doi:10.1038/336435a0.

Kraemer SA, Bondel KB, Ness RW, Keightley PD, Colegrave N. 2017. Fitness change in relation to mutation number in spontaneous mutation accumulation lines of *Chlamydomonas reinhardtii*. Evolution. 71:2918–2929. doi:10.1111/evo.13360.

Kraemer SA, Morgan AD, Ness RW, Keightley PD, Colegrave N. 2016. Fitness effects of new mutations in *Chlamydomonas reinhardtii* across two stress gradients. J Evol Biol. 29:583–593.

Lea DE, Coulson CA. 1949. The distribution of the numbers of mutants in bacterial populations. J Genet. 49:264–285.

Lenski RE, Wiser MJ, Ribeck N, Blount ZD, Nahum JR, *et al.* 2015. Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *Escherichia coli*. Proc Biol Sci. 282:20152292.doi:10.1098/rspb.2015.2292.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 25: 1754–1760.

Lind PA, Andersson DI. 2008. Whole-genome mutational biases in bacteria. Proc Natl Acad Sci USA. 105:17878–17883. doi: 10.1073/pnas.0804445105.

LiPuma JJ, Spilker T, Coenye T, Gonzalez CF. 2002. An epidemic *Burkholderia cepacia* complex strain identified in soil. Lancet. 359: 2002–2003.

LiPuma JJ. 2005. Update on the *Burkholderia cepacia* complex. Curr Opin Pulm Med. 11:528–533.

Liu L, Spilker T, Coenye T, LiPuma JJ. 2003. Identification by subtractive hybridization of a novel insertion element specific for two widespread *Burkholderia cepacia* genomovar III strains. J Clin Microbiol. 41:2471–2476.

Loewe L, Hill WG. 2010. The population genetics of mutations: good, bad and indifferent. Philos Trans R Soc Lond B Biol Sci. 365: 1153–1167. doi:10.1098/rstb.2009.0317.

Lynch M, Sung W, Morris K, Coffey N, Landry CR, *et al.* 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl Acad Sci USA. 9272–9277. 105:

Mahenthiralingam E, Vandamme P. 2005. Taxonomy and pathogenesis of the *Burkholderia cepacia* complex. Chron Respir Dis. 2: 209–217.

McDermott GP, Do D, Litterst CM, Maar D, Hindson CM, *et al.* 2013. Multiplexed target detection using DNA-binding dye chemistry in droplet digital PCR. Anal Chem. 85:11619–11627. doi:10.1021/ac403061n.

Morrison T, Hurley J, Garcia J, Yoder K, Katz A, *et al.* 2006. Nanoliter high throughput quantitative PCR. Nucleic Acids Res. 34:e123.

Muller HJ. 1964. The relation of recombination to mutational advance. Mutat Res. 1:2–9.

Ottesen EA, Hong JW, Quake SR, Leadbetter JR. 2006. Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. Science. 314:1464–1467.

Perfeito L, Sousa A, Bataillon T, Gordo I. 2014. Rates of fitness decline and rebound suggest pervasive epistasis. Evolution. 68:150–162. doi:10.1111/evo.12234.

Pinheiro LB, Coleman VA, Hindson CM, Herrmann J, Hindson BJ, *et al.* 2012. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. Anal Chem. 84: 1003–1011. doi:10.1021/ac202578x.

Poledne R, Zicha J. 2018. Human genome evolution and development of cardiovascular risk factors through natural selection. Physiol Res. 67:155–163. doi:10.33549/physiolres.933885.

Robert L, Ollion J, Robert J, Song X, Matic I, *et al.* 2018. Mutation dynamics and fitness effects followed in single cells. Science. 359: 1283–1286.

Downloaded from https://academic.oup.com/genetics/article/219/2/iyab117/6325026 by UNC Charlotte user on 29 July 2022

Sharp NP, Agrawal AF. 2016. The decline in fitness with inbreeding: evidence for negative dominance-by-dominance epistasis in *Drosophila melanogaster*. J Evol Biol. 29:857–864.

Shaw FH, Geyer CJ, Shaw RG. 2002. A comprehensive model of mutations affecting fitness and inferences for *Arabidopsis thaliana*. Evolution. 56:453–463.

Sniegowski PD, Gerrish PJ, Johnson T, Shaver A. 2000. The evolution of mutation rates: separating causes from consequences. Bioessays. 22:1057–1066. doi:10.1002/1521-1878(200012)22:12<1057::AID-BIES3>3.0.CO;2-W.

Sniegowski PD, Gerrish PJ, Lenski RE. 1997. Evolution of high mutation rates in experimental populations of *E. coli*. Nature. 387: 703–705. doi:10.1038/42701.

Sprouffske K, Aguilar-Rodriguez J, Sniegowski P, Wagner A. 2018. High mutation rates limit evolutionary adaptation in *Escherichia coli*. PLoS Genet. 14:e1007324.doi:10.1371/journal.pgen.1007324.

Sung W, Ackerman MS, Dillon MM, Platt TG, Fuqua C, *et al.* 2016. Evolution of the Insertion-Deletion Mutation Rate Across the Tree of Life. G3 (Bethesda). 6:2583–2591. doi:10.1534/g3.116.030890.

Sung W, Ackerman MS, Gout JF, Miller SF, Williams E, *et al.* 2015. Asymmetric context-dependent mutation patterns revealed through mutation-accumulation experiments. Mol Biol Evol. 32: 1672–1683. doi:10.1093/molbev/msv055.

Sung W, Tucker AE, Doak TG, Choi E, Thomas WK, *et al.* 2012. Extraordinary genome stability in the ciliate *Paramecium tetraurelia*. Proc Natl Acad Sci USA. 109:19339–19344. doi:10.1073/pnas.1210663109.

Taylor SC, Laperriere G, Germain H. 2017. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. Sci Rep. 7: 2409.doi:10.1038/s41598-017-02217-x.

Tenaillon O, Toupance B, Nagard HL, Taddei F, Godelle B. 1999. Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. Genetics. 152:485–493.

Traverse CC, Mayo-Smith LM, Poltak SR, Cooper VS. 2013. Tangled bank of experimentally evolved Burkholderia biofilms reflects selection during chronic infections. Proc Natl Acad Sci USA. 110: E250–E259.

Trindade S, Perfeito L, Gordo I. 2010. Rate and effects of spontaneous mutations that affect fitness in mutator *Escherichia coli*. Philos Trans R Soc Lond B Biol Sci. 365:1177–1186.

Vanlaere E, Baldwin A, Gevers D, Henry D, De Brandt E, *et al.* 2009. Taxon K, a complex within the *Burkholderia cepacia* complex, comprises at least two novel species, Burkholderia contaminans sp. nov. and Burkholderia lata sp. nov. Int J Syst Evol Microbiol. 59: 102–111.

Vassilieva LL, Hook AM, Lynch M. 2000. The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. Evolution. 54: 1234–1246.

Vogelstein B, Kinzler KW. 1999. Digital PCR. Proc Natl Acad Sci USA. 96:9236–9241.

Whale AS, Cowen S, Foy CA, Huggett JF. 2013. Methods for applying accurate digital PCR analysis on low copy DNA samples. PLoS One. 8:e58177.doi:10.1371/journal.pone.0058177.

Whale AS, Huggett JF, Tzonev S. 2016. Fundamentals of multiplexing with digital PCR. Biomol Detect Quantif. 10:15–23. doi:10.1016/j.bdq.2016.05.002.

Wloch DM, Szafraniec K, Borts RH, Korona R. 2001. Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. Genetics. 159: 441–452.

Zeyl C, DeVisser JA. 2001. Estimates of the rate and distribution of fitness effects of spontaneous mutation in *Saccharomyces cerevisiae*. Genetics. 157:53–61.

Zhang XS, Hill WG. 2005. Genetic variability under mutation selection balance. Trends Ecol Evol. 20:468–470. doi:10.1016/j.tree.2005.06.010.

Communicating editor: P. J. Wittkopp

Downloaded from https://academic.oup.com/genetics/article/219/2/iyab117/6325026 by UNC Charlotte user on 29 July 2022