



# The insect-killing bacterium *Photorhabdus luminescens* has the lowest mutation rate among bacteria

Jiao Pan<sup>1</sup> · Emily Williams<sup>2</sup> · Way Sung<sup>3</sup> · Michael Lynch<sup>2</sup> · Hongan Long<sup>1</sup>

Received: 26 March 2020 / Accepted: 3 July 2020 / Published online: 31 August 2020  
© The Author(s) 2020

## Abstract

Mutation is a primary source of genetic variation that is used to power evolution. Many studies, however, have shown that most mutations are deleterious and, as a result, extremely low mutation rates might be beneficial for survival. Using a mutation accumulation experiment, an unbiased method for mutation study, we found an extremely low base-substitution mutation rate of  $5.94 \times 10^{-11}$  per nucleotide site per cell division (95% Poisson confidence intervals:  $4.65 \times 10^{-11}$ ,  $7.48 \times 10^{-11}$ ) and indel mutation rate of  $8.25 \times 10^{-12}$  per site per cell division (95% confidence intervals:  $3.96 \times 10^{-12}$ ,  $1.52 \times 10^{-11}$ ) in the bacterium *Photorhabdus luminescens* ATCC29999. The mutations are strongly A/T-biased with a mutation bias of 10.28 in the A/T direction. It has been hypothesized that the ability for selection to lower mutation rates is inversely proportional to the effective population size (drift-barrier hypothesis) and we found that the effective population size of this bacterium is significantly greater than most other bacteria. This finding further decreases the lower-bounds of bacterial mutation rates and provides evidence that extreme levels of replication fidelity can evolve within organisms that maintain large effective population sizes.

**Keywords** Neutral evolution · Mutation accumulation · Mutation spectrum · Drift-barrier hypothesis · Lower-limit of mutation rate

## Introduction

Mutations are the ultimate source for genetic variation and also contribute to diseases, cell senescence, and cancer (Alfred and Knudson 1971; Davies et al. 2002). Although mutations provide the primary source for evolutionary processes, the mostly deleterious nature of mutations requires that the ability for selection to refine the mechanisms driving

replication fidelity is limited by random genetic drift (Eyre-walker and Keightley 2007; Lynch et al. 2016; Sung et al. 2012a). Until now, the drift-barrier hypothesis provided the only universal explanation for mutation rate determination (Sung et al. 2016, 2012a, b).

Mutation accumulation (MA) technique combined with deep whole-genome sequencing is one of the most accurate and unbiased methods for estimating the rate and spectrum of spontaneous mutations (Halligan and Keightley 2009; Lynch et al. 2008). In a typical MA experiment involving bacteria, numerous inbred or clonal lines are cultured on rich growth media and single-colonies are transferred repeatedly. This bottlenecking process reduces the effective population size ( $N_e$ , the population genetic parameter reflecting the size of an ideal population that meets all Hardy–Weinberg assumptions) and lessens the efficiency of selection, allowing all but the most deleterious mutations to drift to fixation (Bateman 1959; Mukai 1964). Using MA experiments, previous researchers have found that extremely low mutation rates exist both in unicellular eukaryotes (*Tetrahymena thermophila*— $7.61 \times 10^{-12}$  per nucleotide site per cell

---

Edited by Jiamei Li.

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s42995-020-00060-0>) contains supplementary material, which is available to authorized users.

---

✉ Hongan Long  
longhongan@ouc.edu.cn

<sup>1</sup> Institute of Evolution and Marine Biodiversity, KLMME, Ocean University of China, Qingdao 266003, China

<sup>2</sup> Center for Mechanisms of Evolution, The Biodesign Institute, Arizona State University, Tempe, AZ 85281, USA

<sup>3</sup> Department of Bioinformatics and Genomics, University of North Carolina, Charlotte, NC 28223, USA

division (Long et al. 2016)) and bacteria (*Pseudomonas aeruginosa*— $7.92 \times 10^{-11}$  per nucleotide per generation (Dettman et al. 2016)), and both have extremely large  $N_e$  (Lynch et al. 2016).

*Photobacterium luminescens* (Gammaproteobacteria, Enterobacteriaceae) is a facultative symbiont of nematodes that can infect insects (Boemare et al. 1993). It has a complex life-cycle involving stages of free-living, symbiosis and parasitism. *Photobacterium luminescens* lives within the intestines of nematodes acting as a symbiont that provides luminescence and protection from pathogens (Derzelle et al. 2002; Richardson et al. 1988). It also assists the nematode in parasitizing insect larvae by releasing toxins into the blood of the insect after invasion by the nematode (Bowen et al. 1998; Daborn et al. 2001).

The genome assembly and annotation of *P. luminescens* has been published (Duchaud et al. 2003). Most previous studies have focused on mechanisms of toxicity and its use for pest control (Chalabaev et al. 2007; Derzelle et al. 2002, 2004). Other studies have explored its luminescence mechanism (Gregor et al. 2018; Winson et al. 1998). While not a human pathogen, *P. luminescens* glows at high density within human wounds, and the name “Angel’s glow” was given to the glowing of wounds of soldiers during the American Civil War (Mulley et al. 2015).

The unique biological properties and lifestyle of *P. luminescens* (symbiosis, parasitism, luminescence) raise interesting questions regarding how this organism mutates and evolves. Here, we report the findings from a 16-month *P. luminescens* mutation accumulation (MA) experiment. We estimated the genomic mutation rate and spectrum of *P. luminescens* ATCC29999 and found it has the lowest mutation rate among all bacteria studied to date.

## Results

A total of 73 *P. luminescens* MA lines were transferred every other day over the course of 16-months (232 single-cell bottlenecks over 464 days), which accounted for ~4517 cell divisions for each MA line. After sequencing and preliminary analysis, we removed 21 lines that were contaminants, genome library construction failures or with low sequencing coverage (minimum coverage requirement of 20×). The median depth of coverage for the remaining 52 lines was about 163×, and >99% of the genomic sites were covered with reads in all sequenced lines (Supplementary File S1).

## Mutation rate

Using a genome that we assembled de novo, we applied established methods of mutation analysis on 52 MA lines of *P. luminescens* ATCC29999 (Lynch et al. 2008) and identified 72 base-pair substitution mutations, yielding a base-substitution mutation rate of  $5.94 \times 10^{-11}$  per nucleotide site per cell division (95% Poisson confidence intervals =  $4.65 \times 10^{-11}$ ,  $7.48 \times 10^{-11}$ ), which is the lowest bacterial mutation rate to date (Table 1, Supplementary Files S1–S3). To maintain gene function, purifying selection operates to remove mutations that arise in coding regions. In order to ensure that selection was limited in our MA experiment, we examined the distribution of mutations within non-coding and coding regions. We found that the mutation rate of *P. luminescens* ATCC29999 in non-coding regions ( $5.87 \times 10^{-11}$  per nucleotide site per cell division; 95% confidence intervals:  $4.49 \times 10^{-11}$ ,  $7.55 \times 10^{-11}$ ) and coding regions ( $6.24 \times 10^{-11}$ ; 95% confidence intervals:  $3.12 \times 10^{-11}$ ,  $1.12 \times 10^{-10}$ ) are not significantly different (Fisher’s exact test,  $P > 0.05$ ), indicating that selection did not play a large role in this experiment.

In order to test that the extremely low mutation rate is not an artifact of the mutation analysis based on GATK, we

**Table 1** Genome size, base substitution rate ( $\mu$ ), mutation bias in the A/T direction ( $m$ ,  $m = \mu_{G:C \rightarrow A:T} + \mu_{G:C \rightarrow T:A} / \mu_{A:T \rightarrow G:C} + \mu_{A:T \rightarrow C:G}$ ) and the effective population size ( $N_e$ ) of different species of bacteria

Species	Genome size (Mb)	$\mu$	$\mu_{A:T > C:G}$	$m$	$N_e$
<i>Agrobacterium tumefaciens</i> C58	5.67	2.92	0.60	1.38	3.42
<i>Bacillus subtilis</i> NCBI 3610	4.29	3.28	0.44	1.27	0.61
<i>Burkholderia cenocepacia</i> HI2424	7.70	1.33	0.49	0.85	2.47
<i>Escherichia coli</i> K-12 MG1655	4.64	2.23	0.86	1.10	1.80
<i>Photobacterium luminescens</i> ATCC29999	5.17	0.59	0.00	10.28	7.00
<i>Pseudomonas aeruginosa</i> PA14	6.58	0.79	0.21	1.52	2.10
<i>Pseudomonas fluorescens</i> SBW25	6.72	0.93	0.29	1.17	–
<i>Staphylococcus aureus</i> ATCC25923	2.81	4.38	0.48	4.57	–
<i>Vibrio cholerae</i> 2740-80	3.95	1.07	0.33	2.30	4.78

Except for *P. luminescens* ATCC29999, data are from Dettman et al. (2016), Long et al. (2018c), and Lynch et al. (2016).  $\mu$  and  $\mu_{A:T > C:G}$  are in units of  $\times 10^{-10}$  per site per cell division,  $N_e$  is in  $10^8$

also applied the consensus approach, which has a low false-negative rate (Farlow et al. 2015; Keightley et al. 2015). The number of mutations detected by the GATK method is about 93% of that of the consensus method. The base-pair substitution mutation rate from the consensus method is  $6.34 \times 10^{-11}$  per nucleotide site per cell division vs.  $5.94 \times 10^{-11}$  (95% confidence interval:  $4.65 \times 10^{-11}$ ,  $7.48 \times 10^{-11}$ ) using the GATK method. Thus, the extremely low mutation rate of *P. luminescens* does not result from false negatives. The false positive rate of the mutation analysis, also used in previous studies, was < 1% with either Sanger sequencing or Miseq re-sequencing of mutation calls. Consequently we did not

implement this step for the bacterial MA in the present study (Farlow et al. 2015; Long et al. 2015; Nguyen et al. 2020).

There were in total four insertion and six deletion mutations across all lines, yielding an indel mutation rate of  $8.25 \times 10^{-12}$  per site per cell division (13.89% of the base-pair substitution rate; 95% confidence intervals:  $3.96 \times 10^{-12}$ ,  $1.52 \times 10^{-11}$ ), the lowest indel mutation rate in all studied bacteria to date. This finding is consistent with a ~ 10-fold reduction between base-substitution mutation rates and indel mutation rates in bacteria (Sung et al. 2016).

## Mutation spectrum

A/T→G/C and G/C→A/T mutation rates were  $1.00 \times 10^{-11}$  ( $0.40$ – $2.07 \times 10^{-11}$ ) and  $1.03 \times 10^{-10}$  ( $0.77$ – $1.35 \times 10^{-10}$ ) respectively, yielding a mutation bias to A/T of 10.28, one of the highest in bacteria (Tables 1, 2; Figs. 1, 2). This strong A/T bias is driven by the dominating G:C→A:T transition rate  $5.84 \times 10^{-11}$  and G:C→T:A transversion rate  $4.47 \times 10^{-11}$ , in contrast to the mutation rates in the G/C direction: A:T→G:C transition rate  $1.00 \times 10^{-11}$  and zero A:T→C:G transversions.

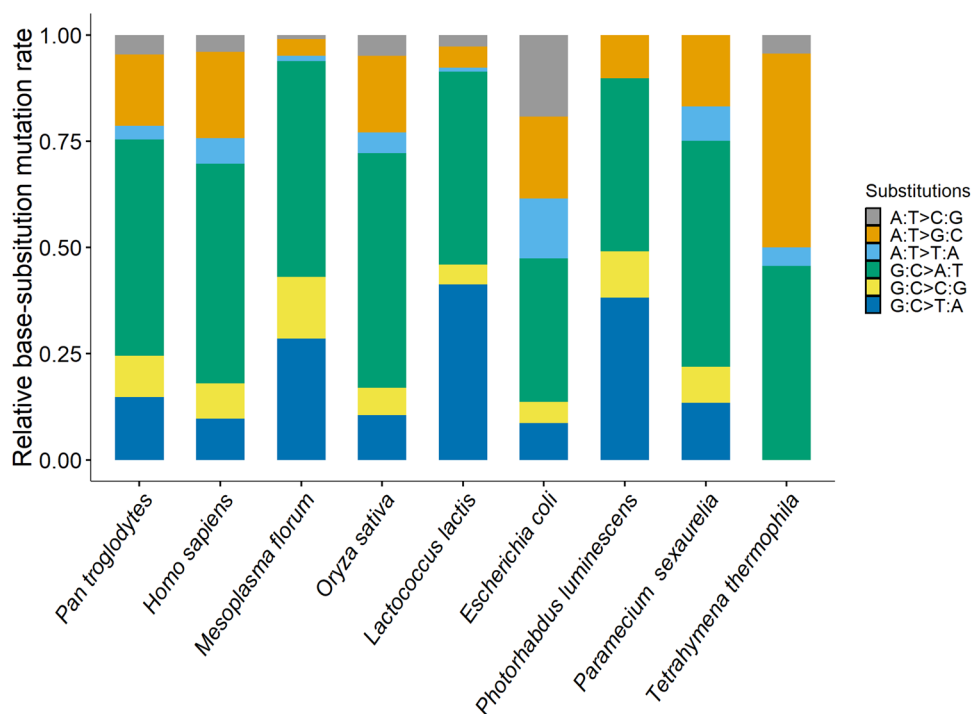
We also estimated the expected equilibrium A + T content of the genome to be 91.14%, the highest in bacteria (Long et al. 2018c), while the A/T content of the reference genome is 57.59%. Hypotheses regarding the evolution of genome-nucleotide composition suggest that lower than expected genomic A/T content from mutation alone could be caused by natural selection and gene conversion against mutations in the A/T direction (Long et al. 2018c). Due to the

**Table 2** Mutation spectrum of *P. luminescens* ATCC29999

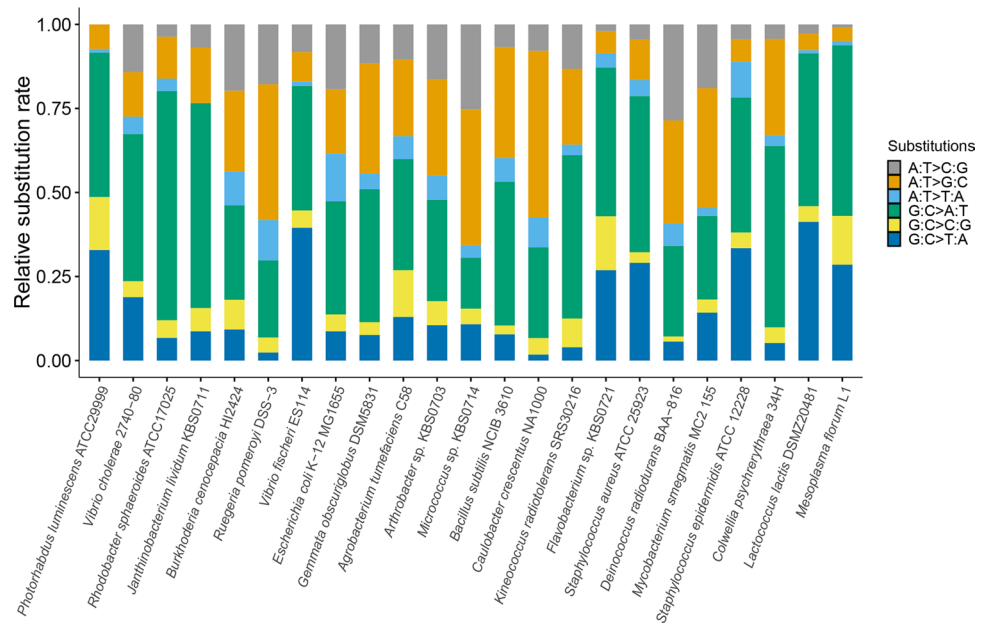
Substitutions	Count	$\mu$	CI
Transitions			
G:C→A:T	30	5.84	3.937, 8.330
A:T→G:C	7	1.00	0.403, 2.065
Transversions			
A:T→T:A	1	0.14	0.004, 0.798
G:C→T:A	23	4.47	2.836, 6.712
A:T→C:G	0	0	0.000, 0.528
G:C→C:G	11	2.14	1.068, 3.828
Insertions	4	0.33	0.090, 0.845
Deletions	6	0.49	0.182, 1.077

Mutation rates ( $\mu$ ) are in units of  $\times 10^{-11}$  per site per cell division  
CI 95% Poisson confidence intervals

**Fig. 1** The relative base substitution mutation rates of different organisms, which cover almost the full range of mutation rates of repair-functional organisms (Long et al. 2016, 2018a; Lynch et al. 2016). From left to right, the mutation rates are ordered from high to low and belong to *Pan troglodytes verus*, *Homo sapiens* CEU, YRI, Iceland, *Mesoplasma florum* L1, *Oryza sativa*, *Lactococcus lactis* DSMZ20481, *Escherichia coli* K-12 MG1655, *Photorhabdus luminescens* ATCC29999, *Paramecium sexaurelia* and *Tetrahymena thermophila*, respectively



**Fig. 2** The relative base substitution mutation rates of *P. luminescens* and other bacteria (Long et al. 2018a). From left to right, the mutation rates are ordered from low to high



extremely low mutation rate, only three types of transversions were observed (Table 2): G:C→T:A (23), G:C→C:G (11), and A:T→T:A (1). The transition to transversion ratio (ts/tv) of *P. luminescens* is 1.06, similar to that of many previously-reported wild-type bacteria.

### The effective population size

To estimate the effective population size of *P. luminescens*, we aligned the genome sequences of 12 *P. luminescens* strains available in NCBI, including the draft genome of *P. luminescens* ATCC29999 assembled here (see “Materials and methods”). We found that the nucleotide diversity at four-fold degenerate sites ( $\theta$ ; four-fold degenerate sites: 280,141, SNPs: 49,343) was  $8.31 \times 10^{-2}$ . Using mutation rate estimates from this study, we calculated the effective population size ( $N_e$ ) of *P. luminescens* to be approximately  $7.00 \times 10^8$ , one of the largest reported among all bacteria (Lynch et al. 2016) (Table 1).

### Discussion

Using MA experiments combined with deep whole-genome sequencing, we calculated the mutation rate of *Photobacterium luminescens* ATCC29999, which is  $5.94 \times 10^{-11}$  per site per cell division. This is the lowest known measurement of mutation rates in bacteria. According to the drift-barrier hypothesis (Sung et al. 2012a), the ability for selection to refine the genetic mechanisms responsible for DNA fidelity is positively correlated with effective population size. Thus, organisms with large effective population sizes have extremely efficient

DNA repair and synthesis, and consequently extremely low mutation rates. *Photobacterium luminescens* has the largest estimate of  $N_e$  seen in bacteria and this may arise from high genome plasticity and widespread horizontal gene transfer observed in this bacterium (Gaudriault et al. 2008; Tounsi et al. 2006). It is noteworthy that estimates of effective population size are very dependent on diversity estimates from sequenced strains, and including additional representative strains of *P. luminescens* when estimating nucleotide diversity may improve estimates of  $N_e$ . Nevertheless, our findings are consistent with the theoretical framework established by the drift-barrier hypothesis regarding the evolution of mutation rate across the tree of life.

*Photobacterium luminescens* mutations are more strongly A/T-biased than 87% of studied bacteria (Supplementary Table S1 in Long et al. (2018c); Table 1). This A/T bias is caused by the extremely low mutation rate in the G/C direction, especially 0 A:T→C:G transversion rate (Table 2). Besides regular oxidative repair systems such as *mutT/mutH* homologs (see the genome annotation in Supplementary File S4), *P. luminescens* can also express luciferase, which is an oxygen scavenger and thus reduces free 8-oxo-guanines in the cytoplasm (Ffrench-Constant et al. 2003). Unlike 8-oxo-guanines in the DNA strands, free 8-oxo-guanines in the nucleotide pool of the cytoplasm could be incorporated into newly synthesized DNA strand and pair with adenines during DNA replication. In the next replication round, after the incorporated 8-oxo-guanine is repaired to guanines and becomes a template, the original A:T pair becomes a C:G pair at the site, i.e., A:T→C:G transversion (Lynch 2007). We thus speculate that luciferase could be an important factor in decreasing the A:T→C:G transversion rate. If this is the case, other bacteria

with luciferase should also show low A:T→C:G transversion rates. We thus blasted the NCBI gene database using the *P. luminescens* luciferase sequence as a query and also required the hit species to have resolved genomic mutation spectra. We found one bacterium, i.e., *Vibrio cholerae* (luciferase blast results: 80% amino acid sequence identity with 95% coverage; BPS rate:  $1.07 \times 10^{-10}$ ), which does indeed have a low A:T→C:G transversion rate of  $0.33 \times 10^{-10}$  (re-analyzed from Dillon et al. (2017) and Sung et al. (2016); Table 1). Despite this, mutation data of other luminescent species are needed in order to demonstrate a definitive association between luciferase function and low A:T→C:G transversion rates in bacteria, since many pathways could be involved in decreasing this type of mutations and the relative potency between these pathways and luciferase remains to be tested.

With limited data points in a subset of model microbes, Drake (1991) proposed a constant 0.0033 mutation per genome per generation for DNA-based microbes. However, recent MA studies have revealed that bacteria exhibit a large variation in mutation rates with significantly higher and lower mutation rates per genome per generation than previously expected (Sung et al. 2012a) (Table 1). The present MA experiment on *P. luminescens* reveals a new lower limit for bacterial mutation rate suggesting that DNA fidelity can reach extreme refinement in prokaryotes. It is unlikely that *P. luminescens* has achieved the lowest limit attainable for bacteria, and it remains to be seen how low mutation rates can go (Sniegowski and Raynes 2013). Mutation rates from MA using more than one strain per species would be more representative since rates are known to vary within a species (Hamilton et al. 2017; Morgan et al. 2014). Unfortunately, we were only able to include one strain of *P. luminescens* due to the limitation of research resources and the large number of species that needed to be investigated in this project. The findings of the present study should also inspire investigations of the highest mutation rate limit. Previous studies show that if there is one mutation in one cell division in asexual populations, then mutational meltdown would occur and kill the organism (Eigen 1971; Lynch et al. 1993). Most DNA mismatch repair-deficient organisms have ~0.1 to 0.3 mutation per cell division per genome (Long et al. 2018b), which leaves plenty of scope for exploring high mutation rates and the mutagenesis mechanisms involved. These studies would help to understand the long-term evolution of mutation rates, factors affecting genomic stability, and mutagenesis patterns common in natural organisms and diseases.

## Materials and methods

### Cell lines and transfers

In total, 73 MA lines from a single progenitor cell of *Photobacterium luminescens* ATCC29999 (obtained from ATCC) were initiated and passed through single-cell bottlenecks

every other day on nutrient agar plates at 30 °C. Every month, single colonies from 10 randomly selected MA lines were cut from the nutrient agar plates and serially diluted to count colony forming units (CFU). The mean number of cell divisions that took place to form a colony from a single cell was estimated by  $\log_2(\text{CFU})$ . The total number of cell divisions of each MA line is the product of the mean (19.5) of all cell division estimates and the total number of transfers for each line. On average, each MA line experienced 232 transfers. The total time taken to complete the process of transferring each MA line was 464 days. The total number of cell divisions per MA line was 4517. We removed one MA line (photo41) due to cross contamination with photo42 and photo47, which was detected after genome sequencing. One line (photo49) only having part of the genome sequenced was also excluded. Other lines were removed from the final mutation analysis because of genome library construction failure or low coverage (< 20×). Eventually, 52 cell lines were used in the final mutation analysis.

### DNA extraction, library construction, and genome sequencing

We extracted DNA from the final MA lines using the Wizard® Genomic DNA Purification Kit (Promega, USA) and constructed libraries using the Nextera® DNA Library Preparation Kit (Illumina, USA). DNA libraries with insert size ~300 bp were sequenced by an Illumina HiSeq2500 sequencer at Hubbard Center for Genome Studies, University of New Hampshire, with 2×150 bp rapid run.

### de novo assembly and annotation of *P. luminescens* ATCC29999 genome

Initially, we used the reference genome of *Photobacterium luminescens* ATCC29999 (GenBank accession number: GCA\_900102985.1; with the minimum size of 1000 bp) for preliminary mutation analysis and obtained a base-pair substitution rate of  $4.39 \times 10^{-11}$  per nucleotide site per cell division. This is comparable to the mutation analysis using the de novo assembled genome (see below). However, we found the breadth of coverage of the MA lines was lower than 70%, i.e., proportion of the genome covered by reads. This was possibly caused by over-filtering when the reference genome was assembled, indicated by the low number of scaffolds in the assembly. We thus used Unicycler -v0.4.8 (Wick et al. 2017) to assemble a new draft genome of *P. luminescens* ATCC29999 to analyse the mutations using the paired-end reads of one MA line (photo7), which did not have any mutations after mapping to the above NCBI reference genome. It is noteworthy that NCBI does not allow



contigs in assemblies shorter than 200 bp to be uploaded whereas we needed to retain these in order to detect all mutations, so we could not upload this assembly to NCBI. We therefore present the assembly and annotations in the supplementary files. The draft genome (Supplementary File S2) has 280 scaffolds with a total size of 5,171,217 bp, GC content: 42.39%, N50: 92,679 bp and no gaps. Of the 280 scaffolds, 42 are shorter than 200 bp. The total length of these short contigs is about 0.12% of the genome size, mean sequencing depth of coverage 151, and 89% of them are covered with reads. Extremely short contigs might cause false negatives, due to mapping failure of reads. A comparison of the mutation rates of the two genomes showed that these extremely short contigs did not bias the mutation rate. For annotation, we used RAST (<https://rast.theseed.org>) (Aziz et al. 2008). The annotation file (Supplementary File S4) has 5,240 predicted protein-coding genes, with the average length of 854 bp. We used BUSCO-3.1.0 (Seppey et al. 2019) to assess the draft genome using the database of bacteria\_odb9 and found that the coverage rate of complete universal single-copy orthologs in the genome was 99.3%, indicating high completeness of the assembly. All analyses reported in this study were based on the draft genome we assembled for *P. luminescens* ATCC29999.

## Mutation analyses

We first used Trimmomatic 0.32 (Bolger et al. 2014) to trim the library adapters for the raw reads and BWA mem-ver. 0.7.12 (Li and Durbin 2009) to map the reads to the draft genome of *P. luminescens* ATCC29999. We used samtools-1.3.1 (Li et al. 2009) to transform sam files to the bam format. Duplicate reads were removed using picard-tools-2.5.0, and SNP/indel variants were called using HaplotypeCaller in GATK-4.0 using standard hard filters across all 52 MA lines (McKenna et al. 2010; Van der Auwera et al. 2013). We also visually verified every single mutation using IGV-2.4.10 (Robinson et al. 2011).

The mutation rate was calculated using the following formula:

$$\mu = \frac{M}{\sum_1^n N \times T},$$

where  $\mu$  is the mutation rate,  $M$  is the total number of mutations in all MA lines,  $n$  is the total number of lines,  $N$  is the analyzed sites in one line, and  $T$  is the number of cell divisions that took place during the mutation accumulation experiments in a single line. The 95% Poisson confidence intervals of the mutation rate were calculated by the cumulative Poisson probability approximated by the binomial distribution. We also used the consensus approach to perform

mutation analyses to test our false negative rate (Sung et al. 2012a).

We calculated the mutation bias in the A/T direction  $m$  by  $m = \mu_{G:C \rightarrow A:T + G:C \rightarrow T:A} / \mu_{A:T \rightarrow G:C + A:T \rightarrow C:G}$ , and the transition to transversion ratios (ts/tv) of *P. luminescens* with the following formula:

$$\frac{ts}{tv} = \frac{\sum_1^n \text{transitions}}{\sum_1^n \text{transversions}}.$$

We calculated the expected equilibrium A+T content of the genome  $p$  using the following formula:

$$p = \frac{v}{u + v},$$

where  $u$  is the mutation rate in the G/C direction, i.e., A/T  $\rightarrow$  G/C, and  $v$  is the G/C  $\rightarrow$  A/T mutation rate (Lynch 2007).

## Effective population size

The genome of *P. luminescens* ATCC29999 assembled in this study and those of another 12 *P. luminescens* strains with > 95% sequence identity in the 16S rRNA gene sequence were used to calculate the nucleotide diversity at four-fold degenerate sites (Table 3; *P. luminescens* DSM 3368 was not used as it was actually a synonym of ATCC29999, as shown in the ATCC website). These 13 genomes were aligned using Mugsy (Angiuoli and Salzberg 2010) and the average pairwise genetic distance at four-fold degenerate sites ( $\theta$ ) was calculated using:

$$\theta = \frac{n}{n-1} \times (1 - \Sigma p^2),$$

where  $p$  is the average allele frequency of nucleotides per four-fold degenerate site between DNA sequences in

**Table 3** Names and genome accession numbers of different strains of *Photobacterium luminescens* and 16S rRNA gene sequence identity with ATCC29999

Strain name	GenBank assembly accession	Identity (%)
HIM3	GCA_002204205.1	99.61
MEX47-22	GCA_004348775.1	99.53
Caborca	GCA_006239335.1	95.89
NBAII H75HRPL105	GCA_000826725.2	95.58
H1	GCA_002968995.1	95.58
LN2	GCA_000767775.1	95.50
H5	GCA_002969055.1	95.50
BA1	GCA_000612035.1	95.50
NBAII HiPL101	GCA_000798635.2	95.43
NBAII Hb105	GCA_000931955.2	95.43
H4	GCA_002969005.1	95.43
H3	GCA_002968975.1	95.43

all possible pairs of all strains,  $n$  is the number of strains,  $n/(n-1)$  is the calibration factor. The effective population size ( $N_e$ ) of *P. luminescens* was then calculated using the following formula:

$$N_e = \frac{\theta}{2\mu}.$$

## Data availability and statistics

Illumina reads of all MA lines in the final analysis were deposited in NCBI SRA (BioProject Number: PRJNA376572). All statistical tests and plotting were performed in R-3.5.1 (R Development Core Team 2012). All genome sequence analyses were done using the IEMB-1 computation cluster of Institute of Evolution and Marine Biodiversity, Ocean University of China.

**Acknowledgements** This work is supported by the Young Tais-han Scholars Program of Shandong Province (tsqn201812024), the Fundamental Research Funds for the Central Universities of China (201822020) to H.L., the Multidisciplinary University Research Initiative Award from the US Army Research Office (W911NF-09-1-0444 and W911NF-09-1-0411) and National Institutes of Health award (R35-GM122566) to M.L. We thank Wei Yang (IEMB-1 cluster administrator) and Kun Wu for their technical help.

**Author contributions** WS, HL designed this study; EW, WS, HL performed experiments; JP, HL analyzed data; JP, WS, ML, HL wrote the manuscript. All authors read and approved the final submitted version.

## Compliance with ethical standards

**Conflict of interest** The authors declare no conflict of interest.

**Animal and human rights statement** No animal or human materials were used in this study.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alfred G, Knudson J (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA* 68:820–823
- Angiuoli SV, Salzberg SL (2010) Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* 27:334–342
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD et al (2008) The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75
- Bateman A (1959) The viability of near-normal irradiated chromosomes. *Int J Radiat Biol Relat Stud Phys Chem Med* 1:170–180
- Boemare N, Akhurst R, Mourant R (1993) DNA relatedness between *Xenorhabdus* spp. (*Enterobacteriaceae*), symbiotic bacteria of entomopathogenic nematodes, and a proposal to transfer *Xenorhabdus luminescens* to a new genus, *Photorhabdus* gen. nov. *Int J Syst Evol Microbiol* 43:249–255
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Bowen D, Rocheleau TA, Blackburn M, Andreev O, Golubeva E, Bhartiya R (1998) Insecticidal toxins from the bacterium *Photorhabdus luminescens*. *Science* 280:2129–2132
- Chalabaev S, Turlin E, Charles JF, Namane A, Pagès S, Givaudan A, Brito-Fravallo E, Danchin A, Biville F (2007) The HcaR regulatory protein of *Photorhabdus luminescens* affects the production of proteins involved in oxidative stress and toxemia. *Proteomics* 7:4499–4510
- Daborn PJ, Waterfield N, Blight MA, Ffrench-Constant RH (2001) Measuring virulence factor expression by the pathogenic bacterium *Photorhabdus luminescens* in culture and during insect infection. *J Bacteriol* 183:5834–5839
- Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A et al (2002) Mutations of the BRAF gene in human cancer. *Nature* 417:949–954
- Derzelle S, Duchaud E, Kunst F, Danchin A, Bertin P (2002) Identification, characterization, and regulation of a cluster of genes involved in carbapenem biosynthesis in *Photorhabdus luminescens*. *Appl Environ Microbiol* 68:3780–3789
- Derzelle S, Turlin E, Duchaud E, Pages S, Kunst F, Givaudan A, Danchin A (2004) The PhoP-PhoQ two-component regulatory system of *Photorhabdus luminescens* is essential for virulence in insects. *J Bacteriol* 186:1270–1279
- Dettman JR, Sztepanacz JL, Kassen R (2016) The properties of spontaneous mutations in the opportunistic pathogen *Pseudomonas aeruginosa*. *BMC Genomics* 17:27
- Dillon MM, Sung W, Sebra R, Lynch M, Cooper VS (2017) Genome-wide biases in the rate and molecular spectrum of spontaneous mutations in *Vibrio cholerae* and *Vibrio fischeri*. *Mol Biol Evol* 34:93–109
- Drake JW (1991) A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci USA* 88:7160–7164
- Duchaud E, Rusniok C, Frangeul L, Buchrieser C, Givaudan A, Taourit S, Bocs S, Boursaux-Eude C, Chandler M, Charles J-F, Dassa E, Derose R, Derzelle S, Freyssinet G, Gaudriault S, Médigue C, Lanois A, Powell K, Siguier P, Vincent R et al (2003) The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat Biotechnol* 21:1307–1313
- Eigen M (1971) Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58:465–523
- Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. *Nat Rev Genet* 8:610–618
- Farlow A, Long H, Arnoux S, Sung W, Doak TG, Nordborg M, Lynch M (2015) The spontaneous mutation rate in the fission yeast *Schizosaccharomyces pombe*. *Genetics* 201:737–744
- Ffrench-Constant R, Waterfield N, Daborn P, Joyce S, Bennett H, Au C, Dowling A, Boundy S, Reynolds S, Clarke D (2003)

- Photorhabdus*: towards a functional genomic analysis of a symbiont and pathogen. *FEMS Microbiol Rev* 26:433–456
- Gaudriault S, Pages S, Lanois A, Laroui C, Teyssier C, Jumas-Bilak E, Givaudan A (2008) Plastic architecture of bacterial genome revealed by comparative genomics of *Photorhabdus* variants. *Genome Biol* 9:R117
- Gregor C, Gwosch KC, Sahl SJ, Hell SW (2018) Strongly enhanced bacterial bioluminescence with the *ilux* operon for single-cell imaging. *Proc Natl Acad Sci USA* 115:962–967
- Halligan DL, Keightley PD (2009) Spontaneous mutation accumulation studies in evolutionary genetics. *Annu Rev Ecol Evol Syst* 40:151–172
- Hamilton WL, Claessens A, Otto TD, Kekre M, Fairhurst RM, Rayner JC, Kwiatkowski D (2017) Extreme mutation bias and high AT content in *Plasmodium falciparum*. *Nucleic Acids Res* 45:1889–1901
- Keightley PD, Pinharanda A, Ness RW, Simpson F, Dasmahapatra KK, Mallet J, Davey JW, Jiggins CD (2015) Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Mol Biol Evol* 32:239–243
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079
- Long H, Sung W, Miller SF, Ackerman MS, Doak TG, Lynch M (2015) Mutation rate, spectrum, topology, and context-dependency in the DNA mismatch repair-deficient *Pseudomonas fluorescens* ATCC948. *Genome Biol Evol* 7:262–271
- Long H, Winter DJ, Chang YC, Sung W, Wu SH, Balboa M, Azevedo RBR, Cartwright RA, Lynch M, Zufall RA (2016) Low base-substitution mutation rate in the germline genome of the ciliate *Tetrahymena thermophila*. *Genome Biol Evol* 8:3629–3639
- Long H, Doak TG, Lynch M (2018a) Limited mutation-rate variation within the *Paramecium aurelia* species complex. *G3-Genes Genom Genet* 8:2523–2526
- Long H, Miller SF, Emily W, Michael L (2018b) Specificity of the DNA mismatch repair system (MMR) and mutagenesis bias in bacteria. *Mol Biol Evol* 35:2414–2421
- Long H, Sung W, Kucukyildirim S, Williams E, Miller SF, Guo W, Patterson C, Gregory C, Strauss C, Stone C, Berne C, Kysela D, Shoemaker WR, Muscarella ME, Luo H, Lennon JT, Brun YV, Lynch M (2018c) Evolutionary determinants of genome-wide nucleotide composition. *Nat Ecol Evol* 2:237–240
- Lynch M (2007) The origins of genome architecture. Sinauer Associates, Sunderland
- Lynch M, Bürger R, Butcher D, Gabriel W (1993) The mutational meltdown in asexual populations. *J Hered* 84:339–344
- Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, Thomas WK (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci USA* 105:9272–9277
- Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, Foster PL (2016) Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet* 17:704–714
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The genome analysis toolkit: a map reduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303
- Morgan AD, Ness RW, Keightley PD, Colegrave N (2014) Spontaneous mutation accumulation in multiple strains of the green alga, *Chlamydomonas reinhardtii*. *Evolution* 68:2589–2602
- Mukai T (1964) The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* 50:1–19
- Mulley G, Beeton ML, Wilkinson P, Vlisidou I, Ockendon-Powell N, Hapeshi A, Tobias NJ, Nollmann FI, Bode HB, Van Den Elsen J, French-Constant RH, Waterfield NR (2015) From insect to man: *Photorhabdus* sheds light on the emergence of human pathogenicity. *PLoS ONE* 10:e0144937
- Nguyen DT, Wu B, Long H, Zhang N, Patterson C, Simpson S, Morris K, Thomas WK, Lynch M, Hao W (2020) Variable spontaneous mutation and loss of heterozygosity among heterozygous genomes in yeast. *Mol Biol Evol*. <https://doi.org/10.1093/molbev/msaa150>
- R Development Core Team (2012) R: a language and environment for statistical computing, Vienna
- Richardson WH, Schmidt TM, Neelson K (1988) Identification of an anthraquinone pigment and a hydroxystilbene antibiotic from *Xenorhabdus luminescens*. *Appl Environ Microbiol* 54:1602–1605
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. *Nat Biotechnol* 29:24–26
- Seppy M, Manni M, Zdobnov EM (2019) BUSCO: assessing genome assembly and annotation completeness. *Method Mol Biol* 1962:227–245
- Sniegowski P, Raynes Y (2013) Mutation rates: how low can you go? *Curr Biol* 23:R147–R149
- Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M (2012a) Drift-barrier hypothesis and mutation-rate evolution. *Proc Natl Acad Sci USA* 109:18488–18492
- Sung W, Tucker AE, Doak TG, Choi E, Thomas WK, Lynch M (2012b) Extraordinary genome stability in the ciliate *Paramecium tetraurelia*. *Proc Natl Acad Sci USA* 109:19339–19344
- Sung W, Ackerman MS, Dillon MM, Platt TG, Fuqua C, Cooper VS, Lynch M (2016) Evolution of the insertion-deletion mutation rate across the tree of life. *G3-Genes Genom Genet* 6:2583–2591
- Tounsi S, Mark B, Samir J, Andréa P (2006) From insects to human hosts: identification of major genomic differences between entomopathogenic strains of *Photorhabdus* and the emerging human pathogen *Photorhabdus asymbiotica*. *Int J Med Microbiol* 296:521–530
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA (2013) From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43:11.10.1–11.10.33
- Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595
- Winson MK, Swift S, Hill PJ, Sims CM, Griesmayr G, Bycroft BW, Williams P, Stewart GS (1998) Engineering the luxCDABE genes from *Photorhabdus luminescens* to provide a bioluminescent reporter for constitutive and promoter probe plasmids and mini-Tn 5 constructs. *FEMS Microbiol Lett* 163:193–202