# A single-molecule approach to DNA replication in *Escherichia coli* cells demonstrated that DNA polymerase III is a major determinant of fork speed

Tuan Minh Pham [1], Kang Wei Tan [1], Yuichi Sakumura [1, 2], Katsuzumi Okumura [3], Hisaji Maki [1] and Masahiro Tatsumi Akiyama [1, *]

[1] Division of Systems Biology, Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan
[2] Department of Information Science and Technology, Aichi Prefectural University, Nagakute, Aichi 480-1198, Japan
[3] Department of Life Science, Graduate School of Bioresources, Mie University, Tsu, Mie 514-8507, Japan

* Corresponding author. Tel.: +81-743-72-5491; fax: +81-743-72-5499.
E-mail address: akiyamam@bs.naist.jp (M. T. Akiyama)

**Key words**: DNA combing; DNA replication fork; driving force; replisome dynamics; single-molecule studies

**Running title**: Replication fork speed in *E. coli* cells

## SI Experimental Procedures

**Cell shape and nucleoids.** Cells were grown to an $OD_{600}$ of 0.3 at 37°C in LB or 56/2 containing 2 μg $ml^{-1}$ thymidine, mixed with 4% paraformaldehyde, and incubated on ice for 1 h. After the fixed cells were rinsed with PBS, nucleoids and membrane were stained with DAPI and FM4-64, respectively. Cell shape and nucleoids were visualized with a 100x objective on a DMRE-HC fluorescence microscope (Leica Microsystems, Germany) with the appropriate filters. Cell images were captured and analyzed by MetaMorph software (Universal Imaging, USA) and processed by Photoshop software (Adobe, USA).

**Western blot analysis of the SOS response.** The cells equivalent to 1 ml of a suspension at an $OD_{600}$ = 0.25 were harvested by centrifugation at 4°C, suspended in 50 μl of 1x SDS-gel loading buffer (Sambrook & Russel, 2001), and heated at 99°C for 5 min. An appropriate amount of the total cellular proteins from the cells was loaded in each lane and separated by SDS-PAGE (Sambrook & Russel, 2001). The resolved proteins were transferred to nitrocellulose membranes (Schleicher & Schuell, Germany) and probed with rabbit anti-RecA antibodies as described elsewhere (Harlow & Lane, 1988). Immunoblots were developed with ECL reagents (GE Healthcare) and visualized by the LAS-4000 Mini luminescence image analyzer (GE Healthcare). For quantification of the SOS induced levels, the linear range for the RecA protein signals from SOS-constitutive SMR7623 (*lexA51*(Def)) cells was established by serial dilution (Pennington & Rosenberg, 2007). Within the linear portion of the standard curve, the amount of cellular RecA protein relative to that in the fully SOS-induced SMR7623 cells was determined.

**Statistical analysis.** For Figure S5A, the bootstrap dataset of the fork speed was obtained by applying a linear regression (*length* = *velocity* × *time*) to the observed dataset of Figure 2D. The confidence interval shown in Figure S5B was computed from the bootstrap samples.

**SI References**

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., *et. al*. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**: 2006.0008.

Burnham, K.P., and Anderson, D.R. (1998) *Model selection and inference: A practical information-theoretical approach*. Springer-Verlag, New York.

Datsenko, K.A., and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* **97**: 6640-6645.

Guyer, M.S., Reed, R.R., Steitz, J.A., and Low, K.B. (1981) Identification of a sex-factor-affinity site in *E. coli* as γδ. *Cold Spring Harb Symp Quant Biol* **45 Pt 1**: 135-140.

Harlow, E., and Lane, D. (1988) *Using Antibodies, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.

Pennington, J.M., and Rosenberg, S.M. (2007) Spontaneous DNA breakage in single living *Escherichia coli* cells. *Nat Genet* **39**: 797-802.

Roepke, R.R., Libby, R.L., and Small, M.H. (1944) Mutation or variation of *Escherichia coli* with respect to growth requirements. *J Bacteriol* **48**: 401-412.

Saenger, W. (1984) *Principles of Nucleic Acids Structure*. Springer-Verlag, New York.

Sambrook, J., and Russel, D.W. (2001) *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.

**Table S1. Bacterial strains used in this study**

| Name | Bacterial genotype | Reference or source |
|---|---|---|
| 15T$^-$ | *thyA42 deoB20* | Roepke *et al*., 1944 [a]; not a K12 strain |
| BW25113 | *lacI*$^q$ *rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ *hsdR514* Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$ | Datsenko & Wanner, 2000 |
| eCOMB | MG1655 except Δ*thyA* Δ(*yjjG-deoB*) | this work [b] |
| JWK2795 | BW25113 except Δ*thyA*::*kan* | Baba *et al*., 2006; Keio collection [c] |
| JWK4336 | BW25113 except Δ*yjjG*::*kan* | Baba *et al*., 2006; Keio collection [c] |
| JWK4346 | BW25113 except Δ*deoB*::*kan* | Baba *et al*., 2006; Keio collection [c] |
| MK935 | MG1655 except *dnaE173 zae-502*::Tn*10* | laboratory stock |
| MK7158 | MG1655 except Δ*thyA* | this work; P1(JWK2795) x MG1655 [d] |
| MK7167 | MG1655 except Δ*thyA* Δ*deoB* | this work; P1(JWK4346) x MK7158 [d] |
| MK7426 | MG1655 except Δ*thyA* Δ*deoB* Δ*yjjG*::*kan* | this work [b] |
| MK7927 | eCOMB except *zae-502*::Tn*10* | this work; P1(MK935) x eCOMB |
| MK7928 | eCOMB except *dnaE173 zae-502*::Tn*10* | this work; P1(MK935) x eCOMB |
| MG1655 | sequenced wild-type *E. coli* K12 | Guyer *et al*., 1981 [a] |
| SMR7467 | MG1655 except *lexA3*(Ind$^-$) *malB*::Tn*9* Δ*attλ*::P$_{sulA}$Ω*gfp-mut2* | Pennington & Rosenberg, 2007 |
| SMR7623 | MG1655 except *lexA51*(Def) *malB*::Tn*9* Δ*attλ*::P$_{sulA}$Ω*gfp-mut2 sulA211* | Pennington & Rosenberg, 2007 |

All strains except 15T$^-$ are derivatives of *E. coli* K12. For P1 transduction, "P1(A) x B" represents that strain B was infected by P1 phage grown on strain A. [a] MG1655 and 15T$^-$ (CGSC#4906) were purchased from The Coli Genetic Stock Center (CGSC) at Yale University, USA. [b] See Experimental Procedures for details. [c] Keio collection strains were obtained from the National BioResource Project: *E. coli* (National Institute of Genetics, Japan). [d] After P1(*vir*) transduction, the *kan* gene was deleted by flippase-mediated recombination.

**Table S2. Plasmids and oligonucleotides used in this study**

| Name | Plasmid properties | Reference |
|---|---|---|
| pCP20 | flippase helper plasmid | Datsenko & Wanner, 2000 |
| pKD46 | Red recombinase expressing plasmid | Datsenko & Wanner, 2000 |
| **Name** | **DNA sequences of oligonucleotides** | **Purpose** |
| yjjG-F | CCGCCATTGC CCTGTACGAA G | amplification of Δ*yjjG*::*kan* of JWK4336 |
| yjjG-R | CTTCTTGAGT AAGCGGCATC GC | amplification of Δ*yjjG*::*kan* of JWK4336 |
| mioC-rF1 | TTGAGTAAATTAACCCACGATCC | amplification of *mioC* for quantitative PCR |
| mioC-rR1 | AACATTCTTGATCACGACATTCC | amplification of *mioC* for quantitative PCR |
| tus-rF1 | TGAAATCACCACGCAGTGTC | amplification of *tus* for quantitative PCR |
| tus-rR1 | TCCTGATACTCTCGCTCCAGT | amplification of *tus* for quantitative PCR |

Nucleotides sequences are complementary to their genomic target sequences. DNA sequences of oligonucleotides are shown form the 5' to 3' direction. Abbreviation: *kan*, kanamycin resistant gene.

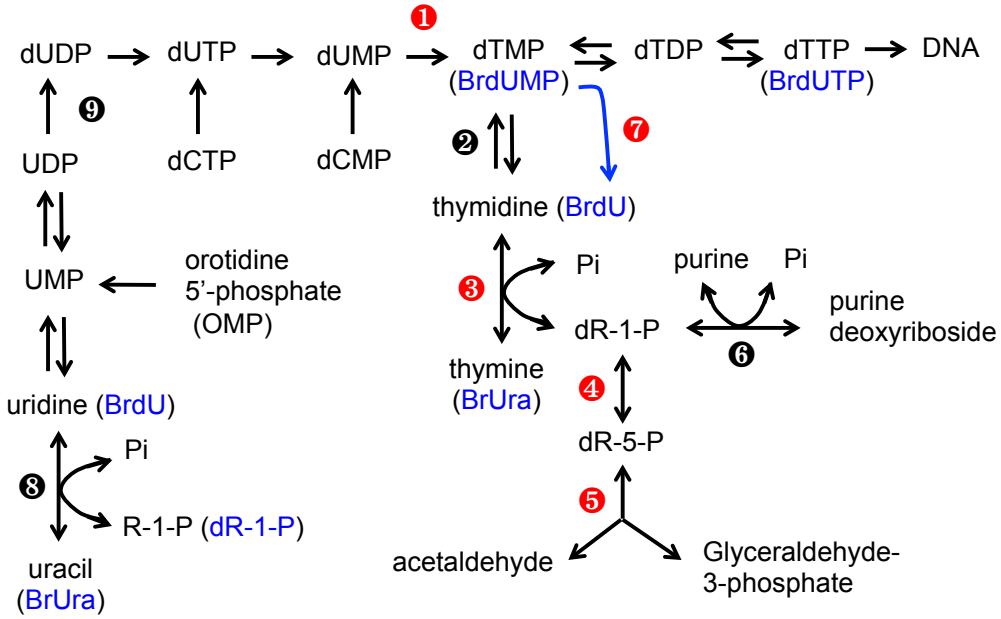
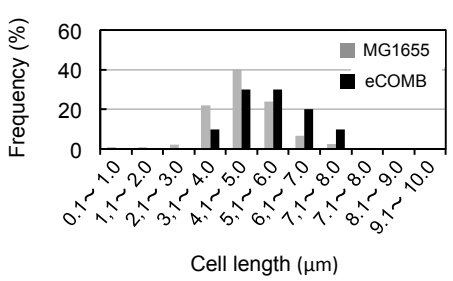


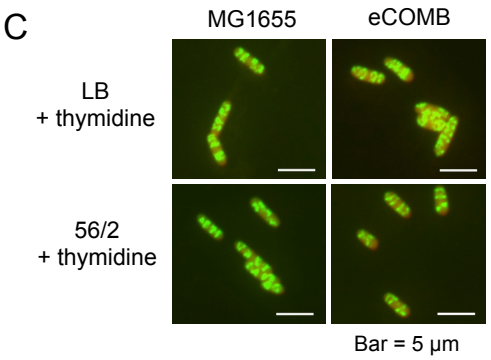

**Fig. S1. Salvage and *de novo* pathways involved in thymidylate (dTMP) biosynthesis of *E. coli***

Numbers represent enzymes catalyzing individual steps; the corresponding genes are in parentheses as follows: 1, thymidylate synthase (*thyA*); 2, thymidine kinase (*tdk*); 3, thymidine phosphorylase (*deoA*); 4, phosphodeoxyribomutase (*deoB*); 5, phosphodeoxyriboaldolase (*deoC*); 6, purine nucleoside phosphorylase (*deoD*); 7, nucleoside monophosphate phosphohydrolase (*yjjG*); 8, uridine phosphorylase (*udp*); 9, ribonucleotide diphosphate reductase (*nrdAB*). The genes encoding the enzymes for the pathways with the red numbers were deleted in eCOMB cells. The route specific to BrdU is shown in blue. Abbreviations are as follows: BrdU, bromodeoxyuridine; BrUra, bromouracil; BrdUMP, bromodeoxyuridine monophosphate; BrdUTP, bromodeoxyuridine triphosphate; dR-1-P, deoxyribose-1-phosphate; dR-3-P, deoxyribose-3-phosphate; R-1-P, ribose-1-phosphate.

A

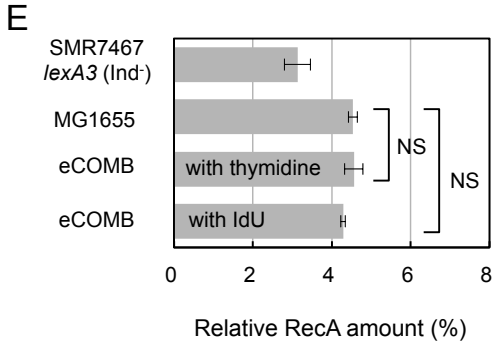
| | Cell size in μm | |
|---|---|---|
| | MG1655 | eCOMB |
| width | 1.5 ± 0.2 (219) | 1.6 ± 0.2 (248) |
| length | 4.7 ± 1.1 (382) | 5.3 ± 1.2 (200) |




D

| | Generation time (min) | | |
|---|---|---|---|
| | MG1655 | eCOMB | MK7928 (*dnaE173*) |
| $OD_{600}$ | 50.5 ± 1.6 | 46.2 ± 6.2 | 58.1 ± 3.0 |
| CFU | 35.0 ± 3.7 | 31.3 ± 3.7 | 28.0 ± 3.4 |




F

| | Cell size in μm | |
|---|---|---|
| | MK7927 (*dnaE*$^+$) | MK7928 (*dnaE173*) |
| width | 1.4 ± 0.2 (356) | 1.3 ± 0.2 (316) |
| length | 3.7 ± 0.8 (384) | 3.8 ± 0.7 (302) |

**Fig. S2. Physiological properties of eCOMB cells**
Cells were exponentially grown in 56/2 medium containing 2 μg ml$^{-1}$ thymidine at 37°C unless otherwise noted. The mean values and corresponding their SD values were determined by three independent experiments except in (E). The cells were fixed with paraformaldehyde and analyzed with a microscope except in (D) and (E). (A) Mean sizes of eCOMB cells. The MG1655 and eCOMB cells were visualized using a phase-contrast microscope. The number of cells observed is in parenthesis. (B) The size distributions of cell length. The graph shows cell length distribution of MG1655 (gray bars) and eCOMB (black bars) cells with the same data as in (A). (C) Cell shapes and nucleoid structures. The cells grown in LB or 56/2 containing 2 μg ml$^{-1}$ thymidine were stained simultaneously with FM4-64 and DAPI. The cell shapes (red) and nucleoids (green) were visualized using a fluorescence microscope. White bars represent 5 μm. (D) Average generation time. The generation time was evaluated by measuring OD at 600 nm and determining the number of colony forming units (CFU); CFU values were verified on LB agar plates containing 2 μg ml$^{-1}$ thymidine. (E) SOS response. After the eCOMB cells were grown in 56/2 containing 2 μg ml$^{-1}$ thymidine, they were cultured for 6 min in the medium with either 50 μg ml$^{-1}$ thymidine or IdU. The RecA protein was detected in the total protein fraction of these cells by western blotting with anti-RecA antibodies, and ECL signals were quantified using a luminescence image analyzer. The bar graph shows the amount of RecA relative to that in the fully SOS-induced SMR7623 (*lexA51*) cells. Black bars show the SEM of three independent experiments. NS: not significant ($p > 0.05$) (two-tailed Student's *t*-test). (F) Mean sizes of eCOMB cells. The MK7927 (*dnaE*$^+$) and MK7928 (*dnaE173*) cells were analyzed as described in (A).

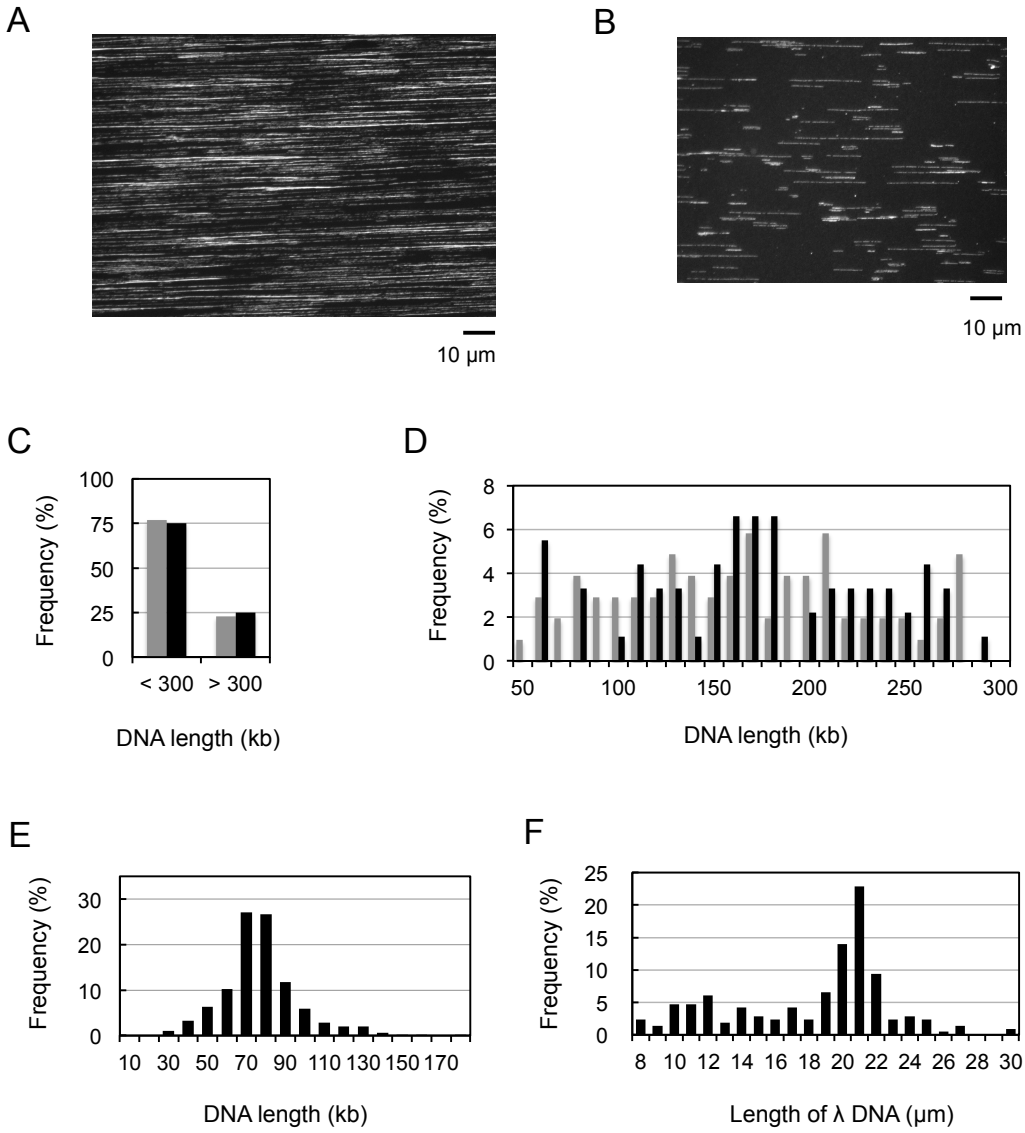



**Fig. S3. Distribution of DNA length combed on cover slips**
(A, B) A typical image of DNA on a glass surface. (A) Chromosomal DNA of eCOMB cells and (B) λ DNA. DNA was stretched on the glass surface, stained with YOYO-1, and visualized using a fluorescence microscope. The black bar represents 10 μm. (C–F) Size distributions of combed DNA molecules: (C–E) chromosomal DNA of eCOMB cells and (F) 48.5-kb λ DNA. The relative numbers of DNA molecules (%) are presented as a histogram of length. (C, D) The eCOMB cells were grown in 56/2 medium containing 50 μg $ml^{-1}$ IdU for 2 min (gray bars) and 3 min (black bars). The chromosomal DNA was extracted, stretched on glass, and stained with YOYO-1. DNA molecules larger than the field of view in a microscope were classified as > 300 kb in (C). DNA molecules smaller than 300 kb were measurable, and their size distribution is shown in (D). The number of molecules analyzed was 103 for 2 min and 91 for 3 min. (E) The graph shows the size distribution of the IdU tracks used to determine the fork speed in Figure 2D. (F) Under our experimental conditions, the length of λ DNA was 21 μm (mode value for the 214 molecules observed). To remove negative effects of breakdown products in the dataset, the mode length of the histogram was used. This value provided a direct relation between the measured length of combed DNA (in μm) and the number of nucleotides in kb (where 1 μm equals to 2.3 kb). Since one nucleotide pair corresponds to 3.4 Å (Saenger, 1984), the crystallographic size of the λ DNA is 16.5 μm, indicating that under our experimental conditions the molecules were stretched 1.3-fold.

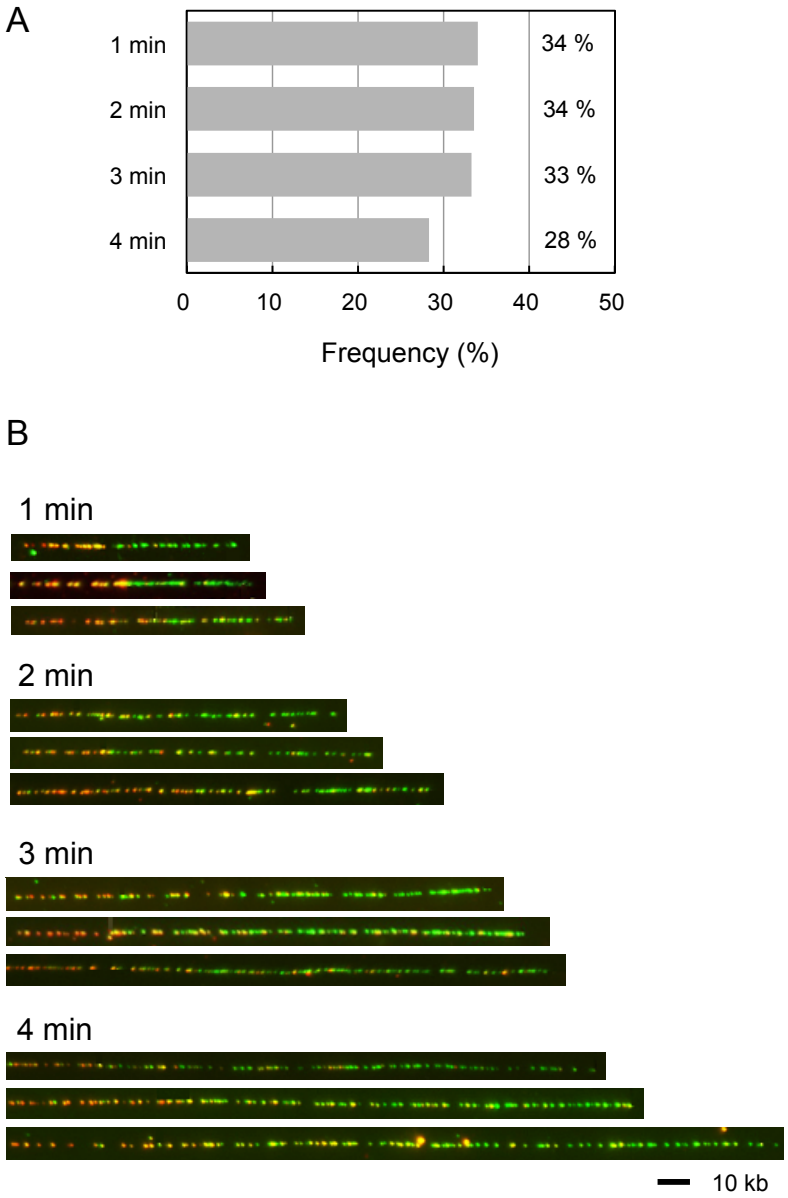



**Fig. S4. Dual-labelled molecules on cover slips**
(A) Frequencies of the dual-labelled molecules in double-labelling experiments. eCOMB cells were pulse-labelled first with CldU for 2 min and then with IdU for 1, 2, 3, and 4 min. The frequency of the dual-labelled DNA fibers at each incubation time with IdU was calculated by dividing the number of the DNA fiber containing both CldU and IdU signals by the total number of the DNA fibers detected with antibodies. (B) Images of the CldU- and IdU-labelled DNA molecules. Among the dual-labelled DNA molecules, the DNA fibers with an IdU stretch lined up end-to-end with a CldU stretch were selected to determine the fork speed. The three DNA molecules shown at each incubation time with IdU are representative of those used for determination of the fork speed. The black bar corresponds to 10 kb.

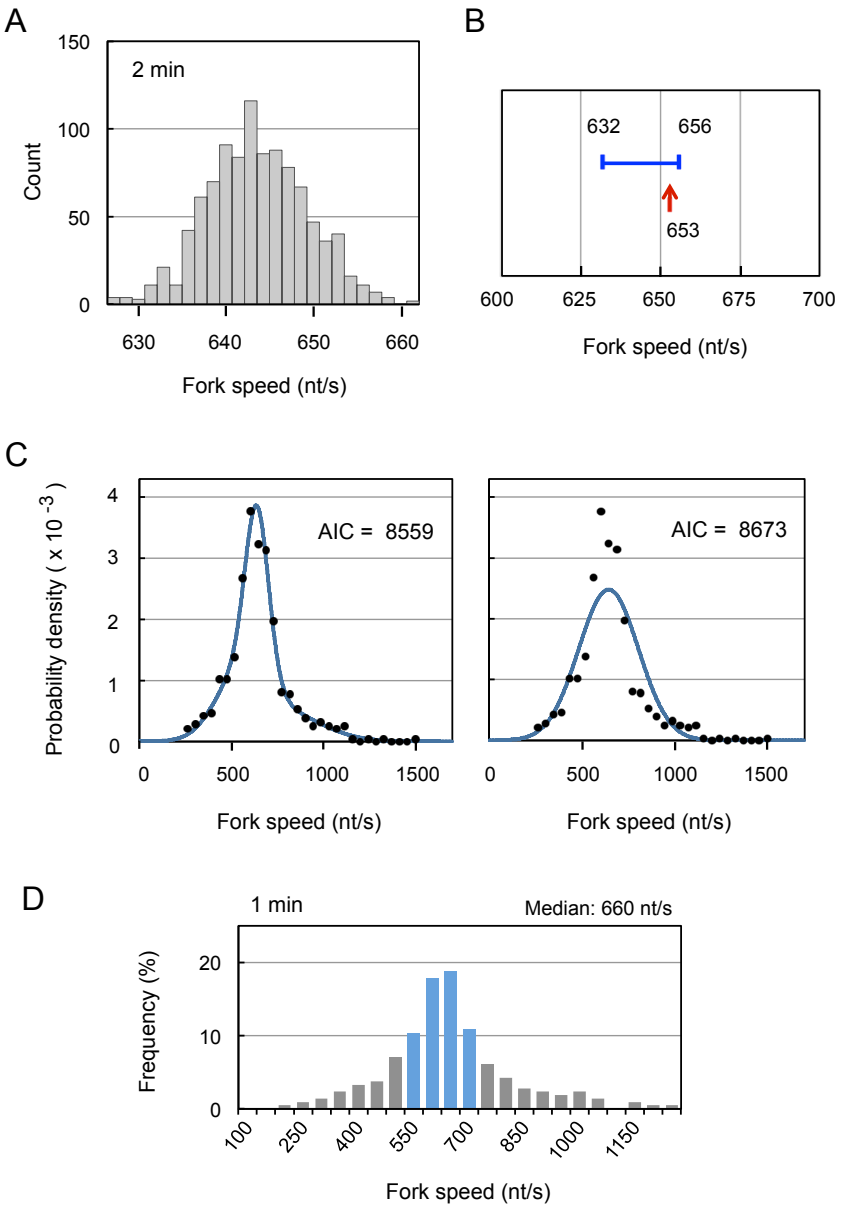
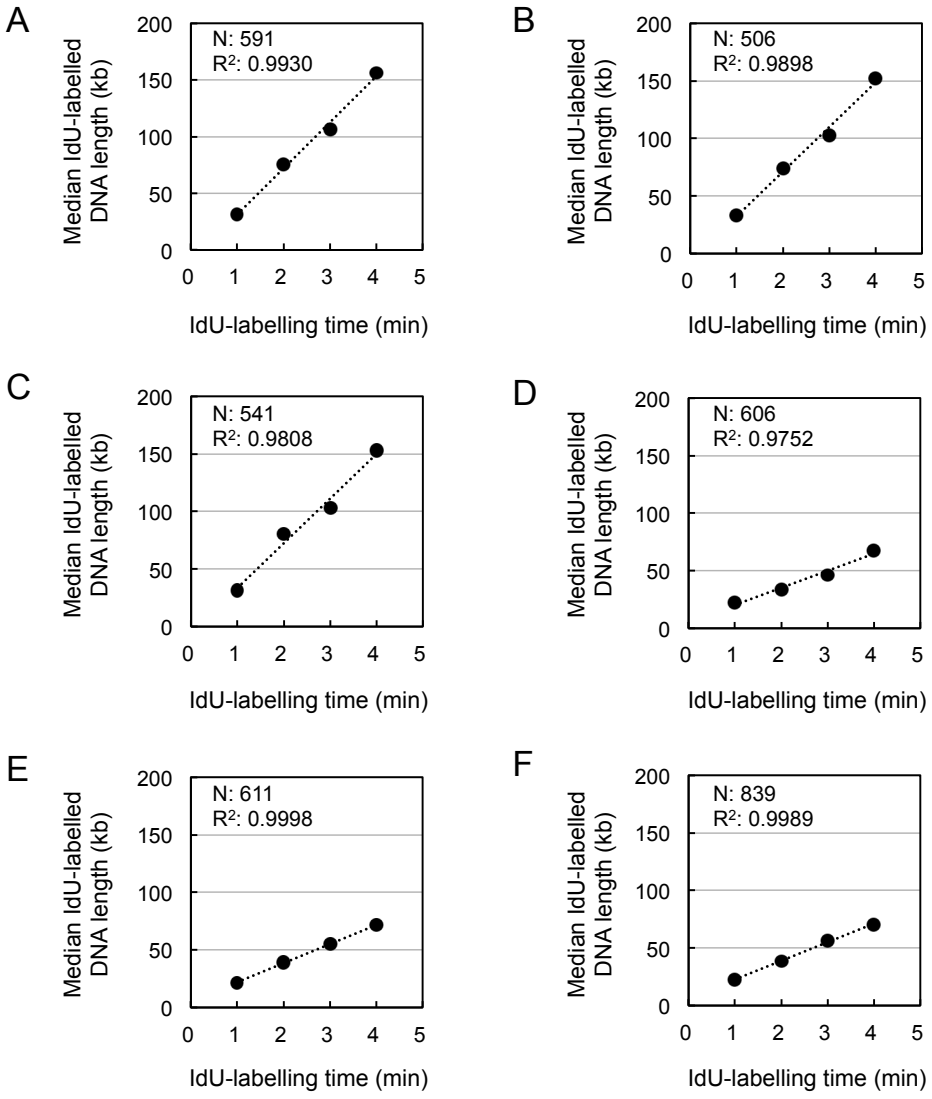



**Fig. S5. Statistical analysis of the fork speed distribution**
(A) Histogram of bootstrap samples of the average fork speed. The samples were computed from the data in Figure 2D by the bootstrap algorithm (n=1,000). (B) Comparison of the average fork speed. The confidence interval for the average fork speed shown in Figure 2D was determined from the distribution in (A). The blue line shows the 95% confidence interval (632–656 nt $s^{-1}$). The red arrow represents the accurate fork speed (653 nt $s^{-1}$) determined by the time-course experiments shown in Figure 4. (C) The Gaussian curve of the fork speed distribution. The fork speed distribution in Figure 2D was converted to probability density (filled circles) and fitted to the mixed-Gaussians curve (left) and the single-Gaussians curve (right). The Akaike Information Criterion (AIC) values for the mixed and single Gaussian curves are 8559 and 8673, respectively. The three-Gaussian model (left) has the smallest AIC, and the four-Gaussian model has the next smallest AIC. In the model selection using the AIC, the model having the smallest AIC is strongly supported if the difference between the AICs for the models is more than four (Burnham and Anderson, 1998). (D) Distribution of fork speed at the 1-min time point. Fork speeds were calculated from the DNA length at the 1-min time point in Fig. 4C, and the median value is shown above the panel. Nearly 60% (57.7%) of the replication forks had speeds between 550 and 750 nt $s^{-1}$ (blue bars). The total number of DNA fibers observed was 213.




G

| Experiments | *dnaE* + | | | *dnaE173* | | |
|---|---|---|---|---|---|---|
| | Slope (kb/min) | Fork speed (nt/s) | | Slope (kb/min) | Fork speed (nt/s) | |
| | | Mean | SEM | | Mean | SEM |
| (A) #1 | 40.59 | 656 | 10 | 14.81 | 264 | 9 |
| (B) #2 | 38.78 | | | 16.54 | | |
| (C) #3 | 38.75 | | | 16.17 | | |

**Fig. S6. Determination of the fork speed in *dnaE173* cells**
The MK7927 (A-C; eCOMB, *zae-502*::Tn*10*) and MK7928 (D-F; eCOMB, *dnaE173*, *zae-502*::Tn*10*) cells were pulse-labelled with CldU for 2 min and then with IdU for 1, 2, 3, and 4 min in triplicate: (A, D) Experiment #1, (B, E) Experiment #2, and (C, F) Experiment #3. (G) The mean fork speed and SEM. The data were analyzed and presented as described in the legend for Figure 4. The total number of DNA fibers observed was 1638 for MK7927 and 2056 for MK7928.