Molecular Systems Biology 5; Article number 335; doi:10.1038/msb.2009.92
Citation: ***Molecular Systems Biology*** **5:335**

www.molecularsystemsbiology.com

molecular systems biology

CORRESPONDENCE

Update on the Keio collection of *Escherichia coli* single-gene deletion mutants

Molecular Systems Biology **5**: 335; published online 22 December 2009; doi:10.1038/msb.2009.92

The Keio collection (Baba *et al*, 2006) has been established as a set of single-gene deletion mutants of *Escherichia coli* K-12. These mutants have a precisely designed deletion from the second codon from the seventh to the last codon of each predicted ORF. Further information is available at http://sal.cs.purdue.edu:8097/GB7/index.jsp or http://ecoli.naist.jp/. The distribution is now being handled by the National Institute of Genetics of Japan (http://www.shigen.nig.ac.jp/ecoli/pec/

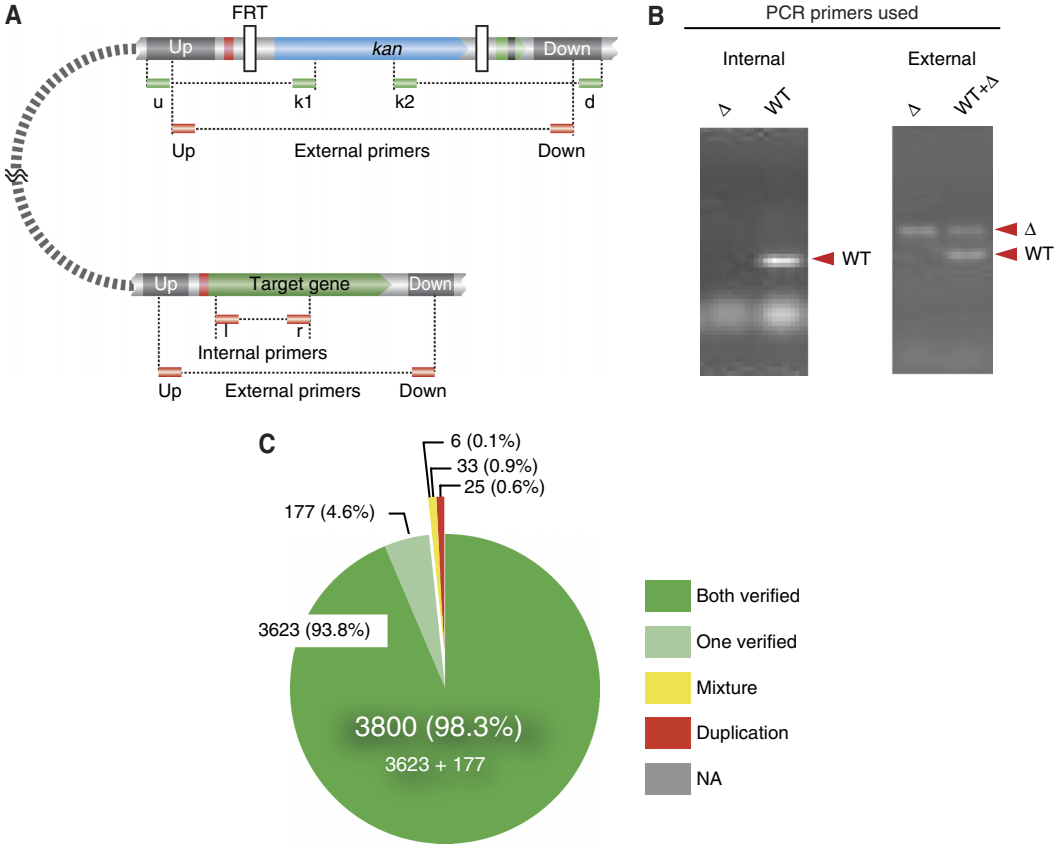

Figure 1 Identification of Keio collection mutants with partial duplications. (**A**) Primer design. The upper branch shows the expected structure of a single-gene mutant. The targeted ORF is replaced with the kanamycin resistance gene (*kan*). The lower branch shows the structure of the targeted gene. Small green boxes show the priming sites used to validate the genomic structures. Small red boxes represent priming sites for external primers (up, down) and internal primers (l, r) that were used to validate the deletion structure and presence or absence of partial duplications. (**B**) Confirmation by electrophoresis. Electrophoregrams showing the amplified fragments using internal primers (first PCR screen) and external primers (second PCR screen). (**C**) Results of evaluation. A total of 7728 mutants (two for each of the 3864 ORFs) were evaluated, as described in the text. Both isolates were validated for 3623 mutants (green). For 177, one was validated and the other was not further examined (pale green). The first set of PCR reactions failed for six ORFs (grey), for which multiple copies exist for the respective gene, e.g., an IS-encoded gene (NA, not available). For 33 mutants (yellow), the microplate sample wells contain mixtures (due to contamination), from which the correct mutant can be isolated by colony purification. Partial duplications were found for 25 mutants (red). Mutants with partial duplications are listed in Table I.

index.jsp). To date more than 4 million samples have been distributed worldwide. As we described earlier (Baba *et al*, 2006), gene amplification during construction is likely to have led to a small number of mutants with genetic duplications.

The design of the Keio deletions was based on annotations that are now outdated. Of 4288 ORFs targeted, mutants were obtained for 3985 (Baba *et al*, 2006). Re-annotation based on highly accurate sequencing of *E. coli* K-12 (Hayashi *et al*, 2006) led to changing many coding regions and the total number of ORFs to 4296, including pseudogenes (Riley *et al*, 2006) (Supplementary Table I). The recent *E. coli* K-12 MG1655 GenBank record (U0096, released in December 2008) has an additional 97 ORFs (exclusive of the ORFs in IS elements, Supplementary Table II) that were not targeted. Of these 4214 annotated ORFs, 4186 were targeted for deletion and 28 were not (Supplementary Table III), which resulted in the isolation of two independent mutants for 3864 targeted ORFs. No deletion was found for 299 ORFs, which are candidates for essential genes. Deletions were also isolated for 23 other ORFs; however, re-annotation led to re-classification of these ORFs as 'split ORFs', because their coding regions are interrupted by an IS element or some other mutation (Supplementary Table IV).

To identify mutants with partial duplications, we performed two sets of PCR reactions on both representatives of all 3864 mutants. In the first set, we tested for the presence of the targeted gene by using a pair of internal gene-specific primers (Figure 1A and B). With the parental strain *E. coli* K-12 BW25113, we were able to amplify 3803 ORFs, as indicated by the presence of PCR products of the expected sizes. For 61 ORFs, we used a pair of external primers that flanked the targeted gene either because the length of the initial PCR product was too short or because the internal primer pair failed to amplify fragments of the predicted sizes for the parental control strain. Results from testing 7728 strains (3864 ORFs) showed that the vast majority (96.1%, 7428/7728) are correct; results in Supplementary Table V show that one or both isolates are correct for 98.3% (3800/3864) of the Keio mutants (Figure 1C). As one isolate is correct for 177 ORFs for which the other isolate is ambiguous, no further tests were done with the other isolate of these mutants.

Mutants of the remaining 58 ORFs (33 with mixtures and 25 with duplications; Figure 1C) were tested in a second set of PCR reactions, which was carried out using external primers flanking the targeted gene (Figure 1A and B). A positive result in the first PCR test can occur not only from mutants with a partial duplication but also from ones that have been cross-contaminated from a nearby microplate well. Therefore, the second set of PCR tests was performed on three colonies after colony purification. In the second PCR test, colonies with the correct deletion or from a cross-contaminant mutant were expected to yield a single PCR product of length corresponding to the expected structure of the respective single-gene mutant or the structure of the targeted gene, respectively. In contrast, mutants with both the respective single-gene deletion and a genetic duplication were expected to yield both PCR products. In cases wherein the sizes of the predicted PCR products were indistinguishable for the deletion and wild-type structures, the PCR products were digested with *Xba*I before size separation by electrophoresis, which cuts within the *kan* (kanamycin resistance gene) replacement gene.

Table I Keio mutants with partial duplications for both isolates

Gene	ECK	JW	b	PEC[a]	FC[b]	Description
(A) New essential gene candidates						
alaS	ECK2692	JW2667	b2697	E	2/8	Alanyl-tRNA synthetase
coaA	ECK3966	JW3942	b3974	E	2/8	Pantothenate kinase
coaE	ECK0103	JW0100	b0103	E	2/8	Dephospho-CoA kinase
dnaG	ECK3056	JW3038	b3066	E	2/8	DNA primase
glmM	ECK3165	JW3143	b3176	E	7/8	Phosphoglucosamine mutase
glyS	ECK3547	JW3530	b3559	E	2/8	Glycine tRNA synthetase, beta subunit
groEL	ECK4137	JW4103	b4143	E	2/8	Cpn60 chaperonin GroEL
ileS	ECK0027	JW0024	b0026	E	2/8	Isoleucyl-tRNA synthetase
parC	ECK3010	JW2987	b3019	E	2/8	DNA topoisomerase IV, subunit A
prfB	ECK2886	JW5847	b2891	E	5/8	Peptide chain release factor RF-2
polA	ECK3855	JW3835	b3863	E	2/8	DNA polymerase I
rho	ECK3775	JW3756	b3783	E	6/8	Transcription termination factor
rpoD	ECK3057	JW3039	b3067	E	5/8	RNA polymerase, sigma 70 (sigma D) factor
yhbG	ECK3190	JW3168	b3201	E	2/8	Lipopolysaccharide export system ATP-binding protein
(B) Genes whose essentiality is uncertain						
btuB	ECK3958	JW3938	b3966	N	4/8	Vitamin B12/cobalamin outer membrane transporter
djlB	ECK0639	JW0641	b0646	N	8/8	Predicted chaperone
folP	ECK3166	JW3144	b3177	N	7/8	7,8-Dihydropteroate synthase
hemE	ECK3989	JW3961	b3997	N	3/8	Uroporphyrinogen decarboxylase
priB	ECK4197	JW4159	b4201	N	3/8	Primosomal protein N
rplK	ECK3974	JW3946	b3983	N	4/8	50S ribosomal subunit protein L11
rplY	ECK2179	JW2173	b2185	N	8/8	50S ribosomal subunit protein L25
rpsO	ECK3154	JW3134	b3165	N	2/8	30S ribosomal subunit protein S15
rpsU	ECK3055	JW3037	b3065	N	3/8	30S ribosomal subunit protein S21
tpr	ECK1224	JW1219	b1229	—	2/8	Predicted protamine-like protein
yiaD	ECK3539	JW5657	b3552	N	2/8	Predicted outer membrane lipoprotein

[a]Essentiality from PEC (Profiling of *E. coli* chromosome database, http://www.shigen.nig.ac.jp/ecoli/pec/index.jsp).
[b]Fraction of correct single-gene deletion candidates upon Keio collection construction (Baba *et al*, 2006).

For 33 of the 58 ORFs, one or more colonies yielded a single PCR product of size corresponding to the single-gene deletion, indicating that the wells for these mutants were cross-contaminated (Supplementary Table V). For the 25 other mutants, purified colonies consistently produced PCR fragments corresponding to structures for both the single-gene deletion and targeted, indicating that these mutants have partial duplications (Figure 1C and Table I). As mentioned above, our PCR tests also revealed 177 mutants, for which we showed that only one isolate is correct. Further testing of these ambiguous mutants by our second PCR test revealed that most of them do not carry a partial duplication.

The 25 ORFs for which both isolates have duplications are candidates for essential genes (Table I). Fourteen of these have been reported to be essential in the PEC (Profiling of *E. coli* Chromosome) database (http://www.shigen.nig.ac.jp/ecoli/pec/index.jsp; Table IA). Thus, it is likely that these 14 genes are essential. The other 11 with partial duplications have been designated as non-essential genes in the PEC database (Table IB). Further tests are required to validate their essentiality. We also carefully evaluated all single-gene deletion mutants in the Keio collection, which were classified as essential in the PEC database. None provided evidence of a partial duplication. Thus, some ORFs reported as essential in the PEC database are nonessential, at least not in the genetic background of our host *E. coli* K-12 BW25113 during aerobic growth at 37°C on LB agar. It should be noted that no evidence exists that the Red system that we used to generate the Keio collection is responsible for causing duplications. Besides, other authors have shown that genetic duplications can occur during DNA replication (Anderson and Roth, 1981). As a cautionary note, partial duplications can occur not only during the construction of single-gene deletion but also upon transfer of the deletion into a new host, e.g., by PCR or transduction as reported previously (Zhou *et al*, 2003).

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

This work was supported by a Grant-in-Aid for Scientific Research (A) and KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas 'System Genomics' from the Ministry of Education, Culture, Sports, Science and Technology of Japan to NAIST and by funds from the Yamagata Prefectural Government and Tsuruoka City to Keio University. BLW was supported by NIH GM62662.

Conflict of interest

The authors declare that they have no conflict of interest.

Natsuko Yamamoto[1], Kenji Nakahigashi[2], Tomoko Nakamichi[1], Mihoko Yoshino[1], Yuki Takai[2], Yae Touda[1], Akemi Furubayashi[1], Satoko Kinjyo[1], Hitomi Dose[1], Miki Hasegawa[2], Kirill A Datsenko[3], Toru Nakayashiki[1], Masaru Tomita[2], Barry L Wanner[3] and Hirotada Mori[1,2]

[1]Graduate School of Biological Sciences, Nara Institute of Science and Technology (NAIST), Nara, Japan, [2]Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, Japan and [3]Department of Biological Science, Purdue University, West Lafayette, IN, USA

References

Anderson P, Roth J (1981) Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (*rrn*) cistrons. *Proc Natl Acad Sci USA* **78:** 3113–3117

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2:** 2006 0008

Hayashi K, Morooka N, Yamamoto Y, Fujita K, Isono K, Choi S, Ohtsubo E, Baba T, Wanner BL, Mori H, Horiuchi T (2006) Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol Syst Biol* **2:** 2006 0007

Riley M, Abe T, Arnaud MB, Berlyn MK, Blattner FR, Chaudhuri RR, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, Mori H, Perna NT, Plunkett III G, Rudd KE, Serres MH, Thomas GH, Thomson NR, Wishart D, Wanner BL (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005. *Nucleic Acids Res* **34:** 1–9

Zhou L, Lei XH, Bochner BR, Wanner BL (2003) Phenotype microarray analysis of *Escherichia coli* K-12 mutants with deletions of all two-component systems. *J Bacteriol* **185:** 4956–4972