Table 3 Details of the 160 base substitutions called at the first screening in the ENU-treated sample *(Continued)*

2647267	T	C	6	5	1.1E-05	C	7	7	6.3E-09	Mutation	1.1E-05
1579929	G	A	6	6	1.0E-07	A	6	5	1.1E-05	Mutation	1.1E-05
2458998	C	T	6	6	1.0E-07	T	6	5	1.1E-05	Mutation	1.1E-05
3936247	C	T	14	13	3.0E-15	T	14	8	1.2E-05	Mutation	1.2E-05
4121383	G	A	14	8	1.2E-05	A	14	9	1.5E-07	Mutation	1.2E-05
1511517	G	A	5	5	1.7E-06	A	6	5	1.1E-05	Mutation	1.2E-05
2963125	G	A	5	5	1.7E-06	A	6	5	1.1E-05	Mutation	1.2E-05
2953567	G	A	9	8	2.4E-09	A	9	6	2.1E-05	Mutation	2.1E-05
4521210	G	A	9	6	2.1E-05	A	10	7	1.3E-06	Mutation	2.2E-05
1066165	C	T	4	4	3.0E-05	T	5	5	1.7E-06	Mutation	3.2E-05
4377924	G	A	5	5	1.7E-06	A	4	4	3.0E-05	Mutation	3.2E-05
655040	C	T	12	7	3.6E-05	T	12	11	6.7E-13	Mutation	3.6E-05
3801057	G	A	12	7	3.6E-05	A	12	8	4.5E-07	Mutation	3.6E-05
1064555	G	A	4	4	3.0E-05	A	6	5	1.1E-05	Mutation	4.1E-05
3156134	C	T	15	8	5.6E-05	T	12	8	4.5E-07	Mutation	5.7E-05
1090650	C	T	7	5	6.1E-05	T	8	8	4.0E-10	Mutation	6.1E-05
4836541	T	C	7	5	6.1E-05	C	8	8	4.0E-10	Mutation	6.1E-05
3417592	G	A	7	5	6.1E-05	A	8	7	3.9E-08	Mutation	6.1E-05
3188210	G	A	6	6	1.0E-07	A	7	5	6.1E-05	Mutation	6.1E-05
700494	C	T	7	5	6.1E-05	T	7	6	6.3E-07	Mutation	6.2E-05
3298937	C	T	7	6	6.3E-07	T	7	5	6.1E-05	Mutation	6.2E-05
496768	A	G	7	5	6.1E-05	G	5	5	1.7E-06	Mutation	6.3E-05
630974	G	A	7	5	6.1E-05	A	7	5	6.1E-05	Mutation	1.2E-04
4169252	G	A	5	4	1.8E-04	A	6	6	1.0E-07	Mutation	1.8E-04
2123568	G	A	5	4	1.8E-04	A	6	5	1.1E-05	Mutation	1.9E-04
3795698	G	A	6	5	1.1E-05	A	5	4	1.8E-04	Mutation	1.9E-04
1779923	G	A	7	5	6.1E-05	A	5	4	1.8E-04	Mutation	2.4E-04
3668382	G	A	19	9	3.1E-04	A	20	16	2.0E-16	Mutation	3.1E-04
2989782	G	A	8	5	3.4E-04	A	8	8	4.0E-10	Mutation	3.4E-04
1297655	C	T	8	5	3.4E-04	T	8	7	3.9E-08	Mutation	3.4E-04
3625847	A	G	8	5	3.4E-04	G	10	7	1.3E-06	Mutation	3.4E-04
4660505	G	A	17	8	9.7E-04	A	20	13	8.6E-11	Mutation	9.7E-04
4586383	C	T	6	4	1.0E-03	T	6	6	1.0E-07	Mutation	1.0E-03
2325510	C	T	5	4	1.8E-04	T	4	3	3.2E-03	Mutation	3.4E-03
4111137	G	A	5	4	1.8E-04	A	4	3	3.2E-03	Mutation	3.4E-03
2901163	C	T	8	5	3.4E-04	T	7	4	5.7E-03	Mutation	6.0E-03
3551802	G	A	7	4	5.7E-03	A	8	5	3.4E-04	Mutation	6.0E-03
4469079	C	T	10	8	1.5E-08	T	10	5	9.1E-03	Mutation	9.1E-03
4539546	G	A	5	3	0.02	A	6	4	1.0E-03	Mutation	0.02
4539738	G	A	5	5	1.7E-06	A	3	2	0.06	Mutation	0.06
2955452	G	A	6	3	0.10	A	7	7	6.3E-09	Mutation	0.10
4153066	G	A	8	6	3.7E-06	A	6	3	0.10	Mutation	0.10
4767697	C	T	6	5	1.1E-05	T	6	3	0.10	Mutation	0.10
4128014	C	T	6	3	0.10	T	6	4	1.0E-03	Mutation	0.10
2410269	G	A	7	3	0.23	A	8	5	3.4E-04	Mutation	0.23

Table 3 Details of the 160 base substitutions called at the first screening in the ENU-treated sample *(Continued)*

3010834	C	C	18	16	5.8E-18	T	15	15	2.1E-18	Mismatch	0
3615885	C	C	21	21	2.1E-25	T	21	17	1.4E-17	Mismatch	0
4623405	A	T	71	63	3.4E-67	A	71	66	3.6E-73	Mismatch	0
2499952	G	A	15	14	2.0E-16	G	15	15	2.1E-18	Mismatch	2.2E-16
4452587	G	A	10	10	1.6E-12	G	10	9	1.6E-10	Mismatch	1.6E-10
3911612	C	C	21	13	4.2E-10	T	20	15	1.6E-14	Mismatch	4.2E-10
2608981	C	C	7	6	6.3E-07	T	8	8	4.0E-10	Mismatch	6.3E-07
128045	G	C	14	12	2.7E-13	G	12	7	3.6E-05	Mismatch	3.6E-05
3890711	G	A	7	4	5.7E-03	G	7	4	5.7E-03	Mismatch	0.01
2691271	C	T	7	7	6.3E-09	C or -	8	2		Damage	
2750772	G	-	4	3	3.2E-03	A	5	5	1.7E-06	Damage	
3648312	A	A or -	6	2		G	5	4	1.8E-04	Damage	
4329658	C	T	5	4	1.8E-04	T,G,C,-	4	1		Damage	
1412330	G								original allele	No	
2298627	A								original allele	No	
2846790	G								original allele	No	
3386508	G								original allele	No	
3386511	T								original allele	No	
171648	C								No mutation	No	
357352	A								No mutation	No	
992221	G								No mutation	No	
4831490	C								No mutation	No	
1291152	G								edge of map	No	
3591684	C								edge of map	No	
4010951	G								edge of map	No	
4065759	G								edge of map	No	
4314300	G								edge of map	No	

*Probability that the real allele is not the most dominant allele
**Probability that the Judgement is not correct

and minus strands of a double-stranded DNA molecule are read alternately, thus almost equivalent numbers of forward and reverse reads were obtained. In cases of real mutations, the same base substitutions will be called in both the forward and reverse reads. In cases where different base substitutions were called between the forward and reverse reads, these must be templates bearing a mismatch. In cases where a specific base was clearly called for on one strand but a variety of bases was called for the opposite strand, this may indicate the existence of persistent DNA damage.

After carefully checking the raw data, the base substitution mutations called in Table 1 were counted again and shown in Tables 2, 3 and 4. After recalculation, the numbers of 'real' base substitution mutations were 0 and 132 in the control and ENU-treated samples, respectively (Table 4). The rest were likely due to mismatches, DNA damage, SNPs that the strain originally possessed, calls at the edges of the mapped read which did not have sufficient coverage, and so on.

We compared the mutation data by this method (SMRT method) with our previous result from colony isolation and whole-genome sequencing (Colony-NGS method). In the ENU-treated samples, the mutation frequencies estimated by the SMRT method (15.4/Mbp) and the Colony-NGS method (12.7/Mbp) were very similar and not significantly different by the binomial test (Fig. 3a). The mutation spectrum obtained by the SMRT method showed that 95% were G:C → A:T transitions and 5% were A:T → G:C transitions (Table 3 and

Table 4 No. of base substitutions after checking original fastq files

	Control	ENU
Real mutation	0	132
Mismatch	7	9
Possibly DNA damage	0	4
No mutation	12	15
Total	19	160

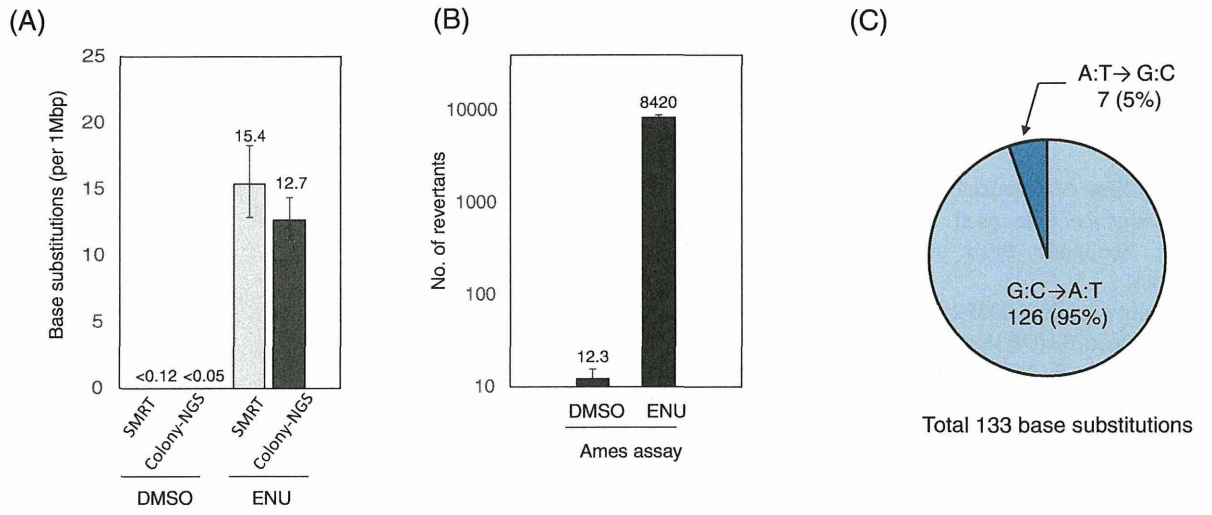

Fig. 3 The SMRT method successfully detects ENU-induced base substitution mutations, with a very low background level. **a** Frequency of base substitutions estimated by the SMRT method and the Colony-NGS method. **b** Results of the Ames assay. **c** Mutation spectrum of ENU-induced base substitutions estimated by the SMRT method

Fig. 3c). This mutation spectrum is well consistent with the ENU signature shown in a previous report [10] and our previous data obtained by the Colony-NGS method (unpublished observations). As for the control (DMSO treated) samples, no mutation was observed in both the SMRT and Colony-NGS methods, thus the mutation frequency was calculated as less than 0.12 per Mbp (1 mutation/8.09 Mbp) and less than 0.05 per Mbp (1 mutation/19.6 Mbp), respectively (Fig. 3a).

Discussions

In this paper, we successfully detected ultra-low frequency base substitution mutations by using a single-molecule real-time sequencer with the SMRTbell strategy. In principle, this strategy is applicable to any DNA samples such as from bacteria, cell lines, tissues of experimental animals, specimens from patients, and enables us to quantify the mutation frequency and the mutation signature of such DNA samples.

The significant merit for using SMRTbell strategy is that we can sequence each plus and minus strand of a double stranded DNA, thus we are able to distinguish 'real mutations' from 'mismatches' or 'DNA damages'. Intriguingly, we could detect not only fixed mutations but also mismatches in the *Salmonella* DNA. In this current procedure, a half of the total mismatches are expected to be detected. From our data, the occurrence of the mismatches in the *Salmonella* genome was roughly estimated as 8 - 10. However, to quantify mismatches absolutely, a new bioinformatics tool should be developed. We also detected 4 possible 'DNA damages' only in the ENU-treated sample (Table 4). In Table 3, the raw read judged as 'Damage' seems to have lower coverage number than 'mutation' or 'mismatch'. This would reflect the presence of the DNA damages in the SMRTbell templates. Note that, the current procedure is not designed for detection of the DNA damages, thus the detected number would be far less than that of real DNA damages.

The background mutation frequency of the SMRT method in this study was less than 0.12 per Mbp which was comparable to the background level of 'Duplex Sequencing' methodologies [2,3]. The background level would depend on the threshold of pass time and accuracy of the CCS. The threshold values used in this study were the most strict values in the current version of PacBio's instrument control and SMRT Analysis software. The real mutation frequency of the control sample was estimated by combining the Colony-NGS and Ames assay results. In the Ames assay using the same exposure procedure, the mutation frequency of the control sample was 1/685 of that of the ENU-treated sample (Fig. 3b), thus the mutation frequency of the control sample was estimated as 12.7/685 = 0.02 per Mbp. Therefore, more sequencing data (at least 50 Mbp) are required to detect mutations in the control sample.

As for insertion and deletion type mutations, this strategy cannot be used at present because of the very high background level of indels. The reason why more deletions were observed in the ENU-treated sample may be because remaining DNA damages influenced the sequence reaction. Ongoing improvements to the hardware and software of the SMRT sequencer and to the bioinformatics of mutation detection will likely overcome this problem in the near future.

Conclusion

Ultra-low frequency base-substitution mutations can be detected directly by using the SMRT DNA sequencer, and this technology provides a useful phenotype-independent mutation assay.

Availability of supporting data

The sequence data used in this study are available at 'DDBJ Sequence Read Archive' with the following accounts.

Submission: DRA003525
BioProject: PRJDB3888
BioSample: SAMD00029313 (data of DMSO-treated sample), SAMD00029314 (data of ENU-treated sample).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TM conceived and designed the experiments and wrote the manuscript. SM carried out statistical evaluation. MY designed and carried out the experiments. All authors read and approved the final manuscript.

Acknowledgments

We thank Ken Osaki, Kenichi Dedachi and Nishiho Makino of TOMY DIGITAL BIOLOGY CO., LTD for discussions about data analysis, and Makiko Takamune for her technical assistance. This research was supported by KAKENHI (2321006) and Health and Labour Sciences Research Grants for Research on Development of New Drugs (KHB1209 and H26-souyaku-ippan-008).

Author details

[1]Research Center for Environmental Quality Management, Kyoto University, Shiga, Japan. [2]National institute of Health Sciences, Tokyo, Japan. [3]Tomonari Matsuda, Research Center for Environmental Quality Management, Kyoto University, 1-2 Yumihama, Otsu, Shiga 520-0811, Japan.

Received: 19 May 2015 Accepted: 9 July 2015
Published online: 01 September 2015

References

1. Matsuda T, Takamune M, Matsuda Y, Yamada M. A pilot study for the mutation assay using a high-throughput DNA sequencer. Genes and Environ. 2013;35:53–6.
2. Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. Detection of ultra-rare mutations by next-generation sequencing. Proc Natl Acad Sci U S A. 2012;109:14508–13.
3. Kennedy SR, Schmitt MW, Fox EJ, Kohrn BF, Salk JJ, Ahn EH, et al. Detecting ultralow-frequency mutations by duplex sequencing. Nat Protoc. 2014;9:2586–606.
4. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323:133–8.
5. Matsuda T. Anticipated mutation assay using single-molecule real-time (SMRT TM) sequencing technology. Genes and Environ. 2010;32:21–4.
6. Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. 2010;38, e159.
7. Jiao X, Zheng X, Ma L, Kutty G, Gogineni E, Sun Q, et al. A Benchmark Study on Error Assessment and Quality Control of CCS Reads Derived from the PacBio RS. J Data Mining Genomics Proteomics. 2013;4.
8. Yamada M, Matsui K, Sofuni T, Nohmi T. New tester strains of Salmonella typhimurium lacking O6-methylguanine DNA methyltransferases and highly sensitive to mutagenic alkylating agents. Mutat Res. 1997;381:15–24.
9. Maron DM, Ames BN. Revised methods for the Salmonella mutagenicity test. Mutat Res. 1983;113:173–215.
10. Richardson KK, Richardson FC, Crosby RM, Swenberg JA, Skopek TR. DNA base changes and alkylation following in vivo exposure of Escherichia coli to N-methyl-N-nitrosourea or N-ethyl-N-nitrosourea. Proc Natl Acad Sci U S A. 1987;84:344–8.