



OPEN ACCESS

EDITED BY

Simon J. Moore,
Queen Mary University of London,
United Kingdom

REVIEWED BY

Thomas Schoenfeld,
Tamarack Bioscience, Inc., United States
Hung-En Lai,
Victoria University of Wellington,
New Zealand

*CORRESPONDENCE

Shuichi Asakawa,
✉ asakawa@g.ecc.u-tokyo.ac.jp

SPECIALTY SECTION

This article was submitted
to Synthetic Biology,
a section of the journal
Frontiers in Bioengineering
and Biotechnology

RECEIVED 03 December 2022

ACCEPTED 27 February 2023

PUBLISHED 09 March 2023

CITATION

Bhuiyan A and Asakawa S (2023),
Synthesis and cloning of long repeat
sequences using single-stranded
circular DNA.
Front. Bioeng. Biotechnol. 11:1115159.
doi: 10.3389/fbioe.2023.1115159

COPYRIGHT

© 2023 Bhuiyan and Asakawa. This is an
open-access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Synthesis and cloning of long repeat sequences using single-stranded circular DNA

Afsana Bhuiyan and Shuichi Asakawa*

Laboratory of Aquatic Molecular Biology and Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

Non-coding repeat expansion causes several neurodegenerative diseases, such as fragile X syndrome, amyotrophic lateral sclerosis/frontotemporal dementia, and spinocerebellar ataxia (SCA31). Such repetitive sequences must be investigated to understand disease mechanisms and prevent them, using novel approaches. However, synthesizing repeat sequences from synthetic oligonucleotides is challenging as they are unstable, lack unique sequences, and exhibit propensity to make secondary structures. Synthesizing long repeat sequence using polymerase chain reaction is often difficult due to lack of unique sequence. Here, we employed a rolling circle amplification technique to obtain seamless long repeat sequences using tiny synthetic single-stranded circular DNA as template. We obtained 2.5–3 kbp uninterrupted TGGAA repeats, which is observed in SCA31, and confirmed it using restriction digestion, Sanger and Nanopore sequencing. This cell-free, *in vitro* cloning method may be applicable for other repeat expansion diseases and be used to produce animal and cell culture models to study repeat expansion diseases *in vivo* and *in vitro*.

KEYWORDS

repeat sequence, cell-free synthetic biology, *in vitro* cloning, neurodegenerative diseases, spinocerebellar ataxia (SCA) 31, benign adult familial myoclonic epilepsy (BAFME)

Introduction

The discovery that simple tandem repeats or microsatellites can cause neurological diseases was revolutionary in the field of neurodegenerative disorders. Nearly 50 neurological diseases have been identified so far, of which 26 diseases are related to the repeat expansion in coding, non-coding, intron and 5' and 3' UTR regions (Rohilla and Gagnon, 2017; Paulson, 2018; Chintalaphani et al., 2021). Among the repeat expansion-related diseases, spinocerebellar Ataxia type 31 (SCA31) is caused by the repeat expansion of 2.5 to 3.8 kbp pentanucleotide TGGAA, where a pure (TGGAA)_n extended for at least 110 repeats in the intron region of the BEAN gene (Sato et al., 2009), whereas benign adult familial myoclonic epilepsy (BAFME1) is associated with the repeat expansion of 105 to 3,680 units of TTTCA (Cen et al., 2018; Ishiura et al., 2018). Such diseases can be classified based on the DNA sequences of repeat units (trinucleotide, tetranucleotide, pentanucleotide, or hexanucleotide). To develop transgenic models of these diseases, we need to obtain disease sequences from patients, except for some special cases (Mizielinska et al., 2014; Swinnen et al., 2018). However, this is often difficult due to ethical regulations and unavailability of patient-derived genomic DNA. In addition, the sizes of disease sequences are limited to that of the patients. Yet, longer DNAs are desirable for developing disease models because of the anticipation found in the repeat diseases (Wells, 1996). In the case of non-repeat mutation

TABLE 1 Sequences for oligonucleotides used in ssDNA template preparation, rolling circle amplification, PCR amplification check and Sanger sequencing.

Oligonucleotides	Sequence (5'–3')
80nt TGGAA ssDNA	TGGAATGGAATGGAATGGAATGGAATG GAATGGAATGGAATGGAA TGGAAATGGA ATGGAATGGAATGGAATGGAATGGAA
Sca31_RCA_F	TGGAATGGAATGGAA
Sca31_RCA_R	TTCCATTCCATTCCA
pIRES_F	GTAACAACCTCCGCCCAT
pIRES_R	GGTACCGTCGACTGCAGAA
M13_90mer ssDNA (Control for RCA)	GTAAAACGACGGCCAGTAAACAGTGAC CATGATAGTGGCCACCCTGCAACCGTG TTGTTTGTTCAGGTTTCATTGTTCATAGCTGT TTCTTG
M13_RCA_F	CTGCAACCGTGTGT
M13_RCA_R	GGTGGCCACTATCAT
80 nt TTTCA ssDNA	TTTCATTTCATTTCATTTCATTTCATTTC TTTCATTTCATTTCATTTC ATTTCATTTC ATTTCATTTCATTTCATTTCATTTC
BAFME1_RCA_F	TTTCATTTCATTTC
BAFME1_RCA_R	TGAAATGAAATGAAA

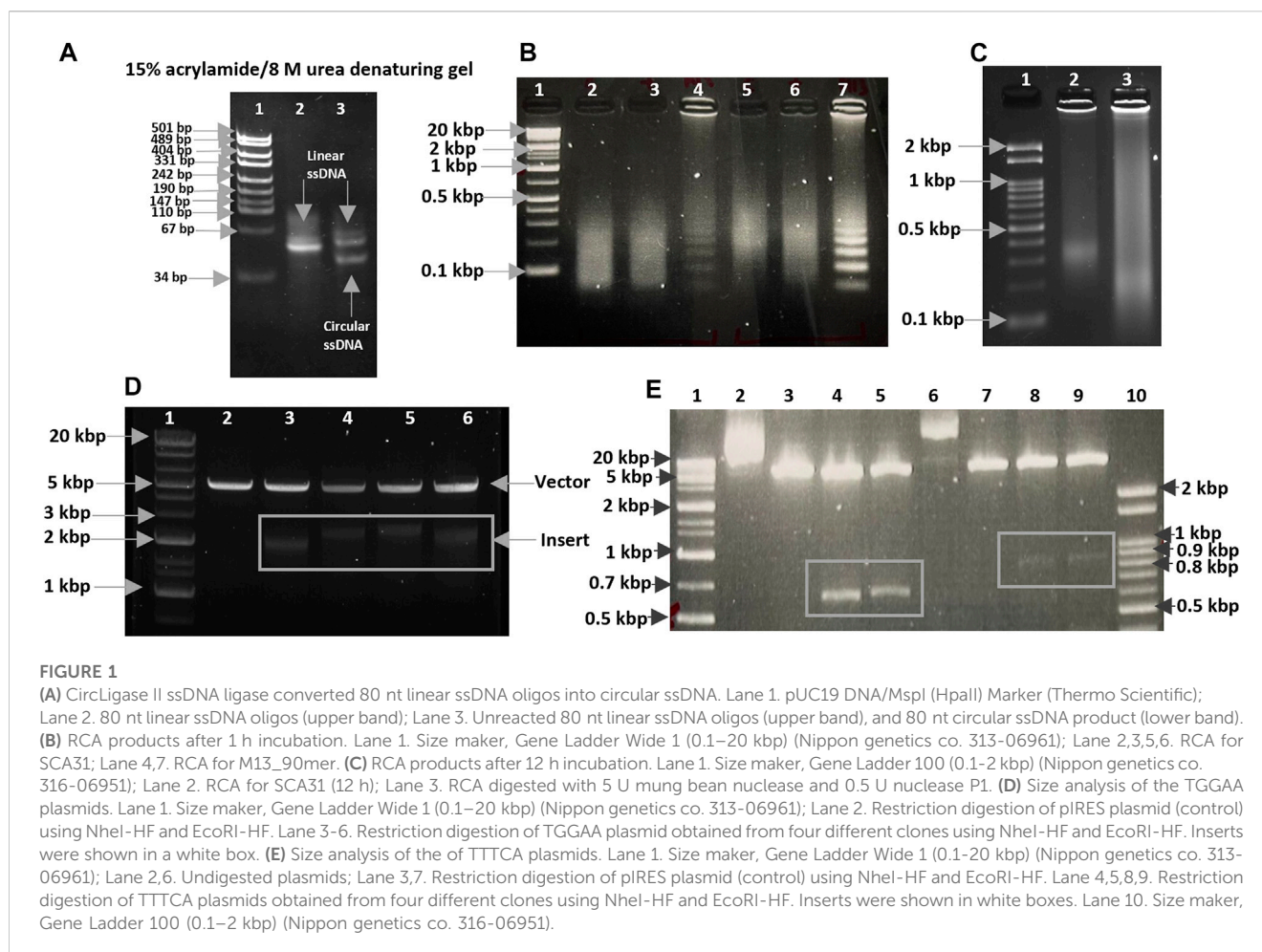
disorders, disease genes can be synthesized and studied without patient sources. Moreover, methods to obtain uninterrupted long repeats are lacking owing to the technical obstacles associated with amplification and cloning. Recombinant vectors containing these synthetic repeats can have numerous applications in biological, medical, and bioengineering research. These vectors can be used for studying repeat associated-non-AUG translation (RAN polypeptides) as well as formation of RNA foci and its interaction with RNA binding proteins (Lee et al., 2013; Malik et al., 2021).

Several methods have been described for trinucleotide repeat synthesis using synthetic oligonucleotides. For example, 20 bp trinucleotide repeat has been used as template as well as primer for conventional polymerase chain reaction (PCR) and cloned in a vector (Ordway and Detloff, 1996). Synthesis of long iterative polynucleotide (SLIP) and non-template PCR method for trinucleotide repeat synthesis was developed based on the theory that filling gaps leads to repeat expansion (Takahashi et al., 1999). Another ligation-based method required iterative ligation reactions to obtain expanded repeats (Kim et al., 2005). Alternatively, concatenated DNA was obtained by random insertion of restriction sites using ligation and PCR (Jiang et al., 1996). PCR has also been used to obtain repeat sequences from DNA sources for polyglutamine expansions disease like Huntington's disease and spinocerebellar ataxia type 10 (Laccone et al., 1999; Peters and Ross, 1999; Michalik et al., 2001; Matsuura and Ashizawa, 2002). Amplification of dimerized expanded repeats (ADER) method was developed to obtain 2,000 CTG repeats using phi29 DNA polymerase in a cell free system (Osborne and Thornton, 2008). In rolling circle amplification or RCA (also known as hyperbranched amplification), small single stranded circular DNA is used as a

template for amplification (Fire and Xu, 1995). This method has a low error rate, strong strand displacement activity, high processivity, and uses circular DNA for isothermal amplification (Hafner et al., 2001). Here we describe a cell-free synthetic method taking SCA31 and BAFME1 as examples for the synthesis and cloning of long repeat sequences using rolling circle amplification. As we were able to reproducibly obtain long repeat sequences in a cell-free manner, we regarded this method as a kind of "in vitro cloning".

Material and methods

Table 1 shows all oligonucleotide sequences used in the experiment. Repeat single-stranded DNA (ssDNA) with phosphorylated 5' end (80 nt, (TGGAA)₁₆, HPLC purified) was obtained from Shanghai Generay Biotech Co., Ltd. (Generay, Shanghai, China). For RCA control, M13 DNA (90 nt) without repeat sequence, was obtained from Hokkaido System Science Co., Ltd. (Hokkaido, Japan). The 20 pmol ssDNA was incubated overnight at 60°C with CircLigase II (Lucigen CL9021K) according to the manual condition to obtain single-stranded circular DNA. The ligase was inactivated by heating at 80°C for 10 min. To remove excess ssDNA, exonuclease T (NEB M0265) was used at 25°C for 30 min and inactivated at 65°C for 20 min. Alternatively Exonuclease I (NEB M0293) can be used (See the results & discussion). This clean circular DNA (2 pmol) was used as a template for RCA, which was performed using Bst DNA polymerase, large fragment (NEB M0275) with 5'phospholated forward and reverse primers (Sca31_RCA_F, Sca31_RCA_R) at 60°C for 12 h. The amplified products were treated with 5 U mung bean nuclease (Takara Bio, 2420) and 0.5 U nuclease P1 (Wako 145-08221) at 37°C for 10 min. After digestion, the DNA was run on a 1.5% agarose gel and 2-3 kbp DNA samples were excised from the gel and purified with the Fastgene gel/PCR extraction kit (Nippon genetics co. FG-91302). The pIRES2 DsRed-Express2 vector (Takara Bio, 632540) was digested by AfeI (NEB R0652) and Shrimp Alkaline Phosphatase (rSAP) (NEB M0371) to prepare blunt ended vector. pIRES and the purified 2-3 kbp DNA insert were ligated for 2 h using T4 DNA Ligase (NEB M0202) and electroporated in Stable Competent *E. coli* (NEB C3040) cells. After 30 min recovery of competent cells, they were plated on kanamycin-supplemented culture plates and incubated at 30°C overnight. The insert lengths of selected clones were checked by PCR amplification using Phusion High-Fidelity DNA Polymerase (NEB M0530) and vector primers (pIRES_F pIRES_R) and by restriction digestion with NheI-HF & EcoRI-HF whose sites are located close to the cloning site, and by Sanger sequencing. The same process was applied to another repeat TTTCA (BAFME1) (80 nt, (TTTCA)₁₆, HPLC purified) to check the versatility of the method. The whole length of one SCA31 plasmid was sequenced using Oxford Nanopore Technology (ONT) (Stevanovski et al., 2022). Library preparation with the ligation sequencing kit (SQK-LSK110) was followed by sequencing on Flongle Flow Cell (FLO-FLG001). Multiple alignment using fast fourier transform (MAFFT) on raw sequence data and consensus sequence polishing by the Medaka tool (Lee et al., 2021) were employed to construct a complete plasmid map.



Results & discussion

The initial oligonucleotide containing 80 nucleotides was selected for our experiment, as the circularized template exhibits maximum amplification efficiency for this size (Joffroy et al., 2018). The formation of circular DNA was confirmed using 15% acrylamide/8 M urea denaturing gel (Figure 1A; Supplementary Figure S1A, B). The size discrepancy between marker and ssDNA (Figure 1A; Supplementary Figure S1A) may be due to base composition bias of ssDNA causing differential mobility. The reaction mixture of circularization was treated with a single-strand specific nuclease, Exonuclease T. Unreacted linear ssDNA found in lane 3 of Supplementary Figure S1B was digested and not found in lane 5, which showed only circular DNA was left. Alternatively, Exonuclease I may be more effective for ssDNA containing C nucleotide(s). After removing linear ssDNA, RCA was employed for 1 h to SCA31 (repeat unit is TTGGA) and M13_90mer circular DNA (control). Ladder like pattern (Kuhn and Frank-Kamenetskii, 2005) was observed for control (M13_90mer) and smear was observed for concatenated repeat sequence showing the success of RCA (Figure 1B). In order to produce a longer repeat sequence, the RCA reaction was extended to 12 h; consequently, extremely large RCA

products were obtained, which were concatemers of the repeat DNA (Figure 1C). DNA ranging from 2 to 3 kbp was cut out from the randomly elongated RCA products and cloned into pIRES vector (Figure 1D). Sanger sequence result for one clone confirmed at least 90-unit repeats of TGGAA from the 5' end (Figure 2A) and 180-unit repeats of TTCCA from the 3' end (Figure 2B). The same procedure was applied for the BAFME1 repeat (repeat unit is TTTCA). Figure 1E white boxes shows that inserts of about 600 bp (Lane 4–5) and about 900 bp (Lane 8–9) were cloned in the pIRES vector. We confirmed the sequence of one clone and found that the insert contained at least 80-unit repeats of TTTCA (Figure 2C). These results show the reproducibility of this method; repetitive sequences of different repeat sequences with various unit lengths could be generated by RCA and be cloned into plasmid vectors.

To characterize the whole sequence composition and repeat unit length, Oxford Nanopore Technology (ONT) was attempted on a SCA31 plasmid containing a 2.5 kbp repeat. From 8,000 reads (N50: 6,833bp), 3,831 reads were extracted with 6,000–8,000 bp length (almost full length of the plasmid), which were assembled to form the plasmid containing the repeat and vector (Supplementary Figure S2A). Supplementary Figure S2B shows 50 random reads aligned onto the consensus

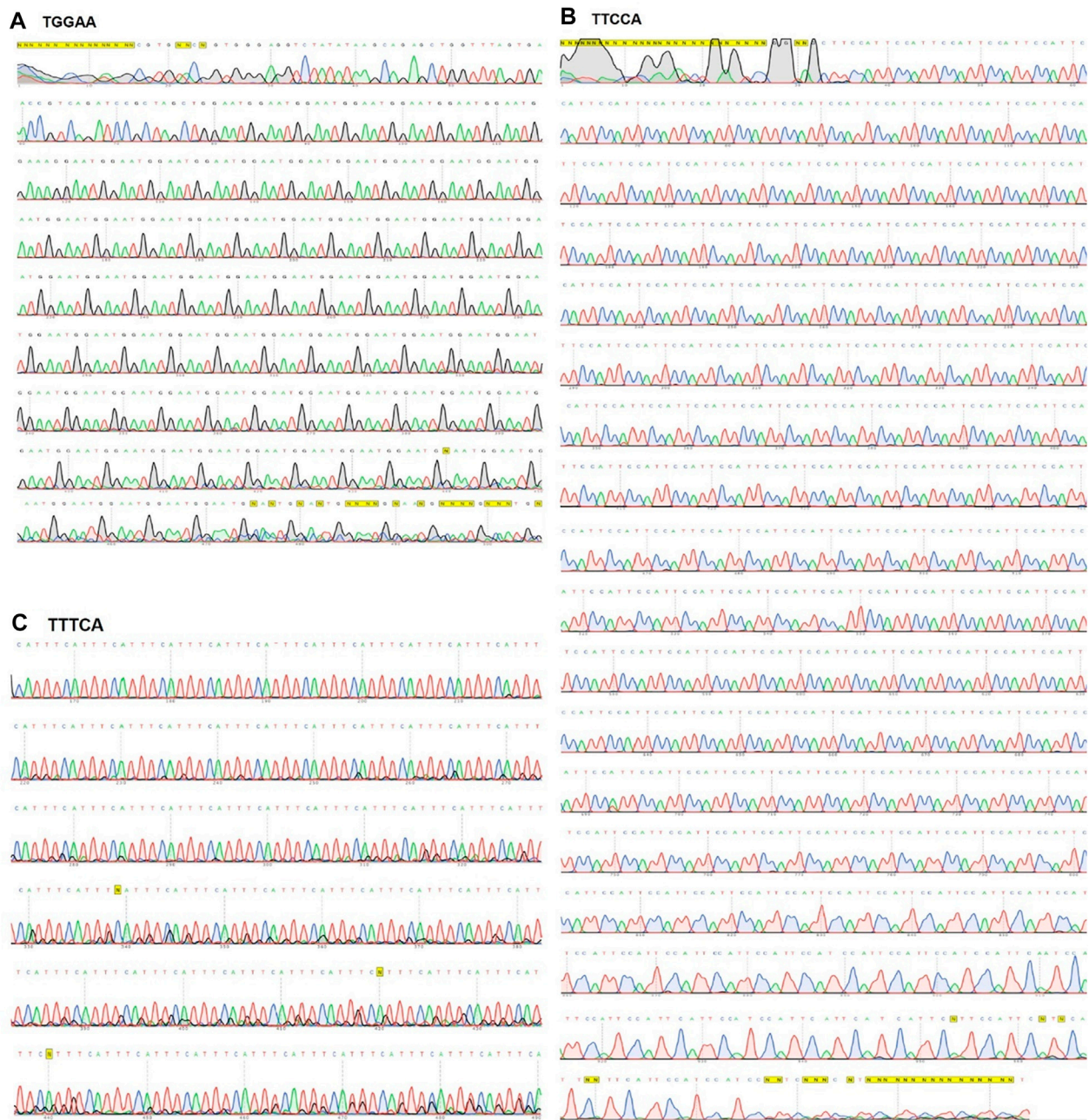


FIGURE 2
 Sanger sequencing result for SCA31 and BAFME1 repeats. **(A)** TGGAA—Up to 500 bp sequence from the 5′side of a plasmid containing 2.5 kb SCA31 insert was confirmed by Sanger sequencing. **(B)** TTCCA—Up to 1 kbp sequence from the 3′side of the SCA31 plasmid was confirmed by Sanger sequencing. **(C)** TTTCA—About 180 units of the BAFME1 repeat of was inserted into the vector. Up to 400 bp sequence from the 5′side was confirmed by Sanger sequencing.

sequence. After polishing with the Medaka tool, full length sequence of the plasmid was obtained (Supplementary Figure S2C), containing 445 TGGAA repeats that included four mutated sequences—TGTA A, TTAAA, TGGAGG, TGGAT. Although these point mutations were found in the sequence, the longest stretch of a perfect TGGAA repeat was 1960 bp (392 TGGAA), which covered the pathogenic size. In summary, our method successfully obtained uninterrupted long repeat sequences using

Bst DNA polymerase. However, our trial with phi29 polymerase was not successful despite trying with various concentrations and different reaction conditions (Supplementary Figure S3). RCA products did not migrate from the loading well into the gel and did not run after treating with various concentrations of mung bean nuclease. As a result, the DNA could not be extracted out from the gel. Our method can be used for generating long tandem repeats of tailored size in a cell-free manner and combined with

the ADER method (Osborne and Thornton, 2008) to stabilize the repeat units in *E. coli*. Moreover, this *in vitro* cloning method may be applicable for other types of repeat expansion, including those for neurodegenerative diseases, undiscovered disease-causative repeats or artificial repeats and may be used to produce animal and cell culture models to study repeat expansion diseases *in vivo* and *in vitro*. Lastly, we believe that this method can also be used for studying the function of repeats in the genome, such as alpha-satellite, centromeres and telomeres.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

Author contributions

SA and AB contributed to conception and design of the study. AB performed all experiments and analysis. AB wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was supported by the Japan Society for the Promotion of Science (grant no. JP20H00429).

References

- Cen, Z., Jiang, Z., Chen, Y., Zheng, X., Xie, F., Yang, X., et al. (2018). Intronic pentanucleotide TTCA repeat insertion in the SAMD12 gene causes familial cortical myoclonic tremor with epilepsy type 1. *Brain* 141 (8), 2280–2288. doi:10.1093/brain/awy160
- Chintalaphani, S. R., Pineda, S. S., Deveson, I. W., and Kumar, K. R. (2021). An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta neuropathol. Commun.* 9 (1), 98–20. doi:10.1186/s40478-021-01201-x
- Fire, A., and Xu, S. Q. (1995). Rolling replication of short DNA circles. *Proc. Natl. Acad. Sci.* 92 (10), 4641–4645. doi:10.1073/pnas.92.10.4641
- Hafner, G. J., Yang, I. C., Wolter, L. C., Stafford, M. R., and Giffard, P. M. (2001). Isothermal amplification and multimerization of DNA by Bst DNA polymerase. *Biotechniques* 30 (4), 852–867. doi:10.2144/01304rr03
- Ishiura, H., Mitsui, J., Yoshimura, J., Matsukawa, M. K., Fujiyama, A., Toyoshima, Y., et al. (2018). Expansions of intronic TTTCa and TTTTA repeats in benign adult familial myoclonic epilepsy. *Nat. Genet.* 50 (4), 581–590. doi:10.1038/s41588-018-0067-2
- Jiang, S. W., Trujillo, M. A., and Eberhardt, N. L. (1996). An efficient method for generation and subcloning of tandemly repeated DNA sequences with defined length, orientation and spacing. *Nucleic acids Res.* 24 (16), 3278–3279. doi:10.1093/nar/24.16.3278
- Joffroy, B., Uca, Y. O., Prešern, D., Doye, J. P., and Schmidt, T. L. (2018). Rolling circle amplification shows a sinusoidal template length-dependent amplification bias. *Nucleic acids Res.* 46 (2), 538–545. doi:10.1093/nar/gkx1238
- Kim, S. H., Cai, L., Pytlos, M. J., Edwards, S. F., and Sinden, R. R. (2005). Generation of long tracts of disease-associated DNA repeats. *BioTechniques* 38 (2), 247–253. doi:10.2144/05382rr01
- Kuhn, H., and Frank-Kamenetskii, M. D. (2005). Template-independent ligation of single-stranded DNA by T4 DNA ligase. *FEBS J.* 272 (23), 5991–6000. doi:10.1111/j.1742-4658.2005.04954.x

Acknowledgments

We acknowledge Mariko Kondo for editing the manuscript. We acknowledge Kazutoshi Yoshitake, Ryo Yonezawa and Hideaki Mizobata for various suggestions regarding the Nanopore experiment.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2023.1115159/full#supplementary-material>

- Laccone, F., Maiwald, R., and Bingemann, S. (1999). A fast polymerase chain reaction-mediated strategy for introducing repeat expansions into CAG-repeat containing genes. *Hum. Mutat.* 13 (6), 497–502. doi:10.1002/(sici)1098-1004(1999)13:6<497:aid-humu10>3.0.co;2-6
- Lee, J. Y., Kong, M., Oh, J., Lim, J., Chung, S. H., Kim, J. M., et al. (2021). Comparative evaluation of Nanopore polishing tools for microbial genome assembly and polishing strategies for downstream analysis. *Sci. Rep.* 11 (1), 20740. doi:10.1038/s41598-021-00178-w
- Lee, Y. B., Chen, H. J., Peres, J. N., Gomez-Deza, J., Attig, J., Štálekár, M., et al. (2013). Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell. Rep.* 5 (5), 1178–1186. doi:10.1016/j.celrep.2013.10.049
- Malik, I., Kelley, C. P., Wang, E. T., and Todd, P. K. (2021). Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat. Rev. Mol. Cell. Biol.* 22 (9), 589–607. doi:10.1038/s41580-021-00382-6
- Matsura, T., and Ashizawa, T. (2002). Polymerase chain reaction amplification of expanded ATTCT repeat in spinocerebellar ataxia type 10. *Ann. Neurol.* 51 (2), 271–272. doi:10.1002/ana.10049
- Michalik, A., Kazantsev, A., and Broeckhoven, C. V. (2001). Method to introduce stable, expanded, polyglutamine-encoding CAG/CAA trinucleotide repeats into CAG repeat-containing genes. *Biotechniques* 31 (2), 250–254. doi:10.2144/01312bm02
- Mizielinska, S., Grönke, S., Niccoli, T., Ridler, C. E., Clayton, E. L., Devoy, A., et al. (2014). C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins. *Science* 345 (6201), 1192–1194. doi:10.1126/science.1256800
- Ordway, J. M., and Detloff, P. J. (1996). *In vitro* synthesis and cloning of long CAG repeats. *Biotechniques* 21 (4), 609–612. doi:10.2144/96214bm08
- Osborne, R. J., and Thornton, C. A. (2008). Cell-free cloning of highly expanded CTG repeats by amplification of dimerized expanded repeats. *Nucleic acids Res.* 36 (4), e24. doi:10.1093/nar/gkn025

- Paulson, H. (2018). Repeat expansion diseases. *Handb. Clin. neurology* 147, 105–123. doi:10.1016/B978-0-444-63233-3.00009-9
- Peters, M. F., and Ross, C. A. (1999). Preparation of human cDNAs encoding expanded polyglutamine repeats. *Neurosci. Lett.* 275 (2), 129–132. doi:10.1016/S0304-3940(99)00758-2
- Rohilla, K. J., and Gagnon, K. T. (2017). RNA biology of disease-associated microsatellite repeat expansions. *Acta neuropathol. Commun.* 5 (1), 63–22. doi:10.1186/s40478-017-0468-y
- Sato, N., Amino, T., Kobayashi, K., Asakawa, S., Ishiguro, T., Tsunemi, T., et al. (2009). Spinocerebellar ataxia type 31 is associated with “inserted” penta-nucleotide repeats containing (TGAA)_n. *Am. J. Hum. Genet.* 85 (5), 544–557. doi:10.1016/j.ajhg.2009.09.019
- Stevanovski, I., Chintalaphani, S. R., Gamaarachchi, H., Ferguson, J. M., Pineda, S. S., Scriba, C. K., et al. (2022). Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. *Sci. Adv.* 8 (9), eabm5386. doi:10.1126/sciadv.abm5386
- Swinnen, B., Bento-Abreu, A., Gendron, T. F., Boeynaems, S., Bogaert, E., Nuyts, R., et al. (2018). A zebrafish model for C9orf72 ALS reveals RNA toxicity as a pathogenic mechanism. *Acta neuropathol.* 135 (3), 427–443. doi:10.1007/s00401-017-1796-5
- Takahashi, N., Sasagawa, N., Suzuki, K., and Ishiura, S. (1999). Synthesis of long trinucleotide repeats *in vitro*. *Neurosci. Lett.* 262 (1), 45–48. doi:10.1016/S0304-3940(99)00031-2
- Wells, R. D. (1996). Molecular basis of genetic instability of triplet repeats (*). *J. Biol. Chem.* 271 (6), 2875–2878. doi:10.1074/jbc.271.6.2875