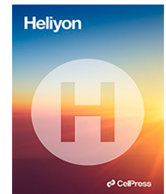
Heliyon 9 (2023) e16035

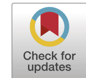
Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article



One-pot making of sequence-restricted DNA dumbbells

Ya-Hui Chang [a], Yu-Cheng Lin [b,c], Hong-Hsing Liu [a,d,*]

[a] Institute of Molecular and Genomic Medicine, National Health Research Institutes, Zhunan Town, Miaoli County, 35053, Taiwan
[b] Pediatrics, Far Eastern Memorial Hospital, Pan-Chiao District, New Taipei City, 220216, Taiwan
[c] Electronic Engineering, Oriental Institute of Technology, Pan-Chiao District, New Taipei City, 220216, Taiwan
[d] Pediatrics, En Chu Kong Hospital, Sanxia District, New Taipei City, 237414, Taiwan

ARTICLE INFO

Keywords:
Dumbbell DNA
Genome enrichment
Next-generation sequencing

ABSTRACT

A method to build sequence-restricted DNA dumbbells was developed. 5′-exonuclease converts end sequences of DNA targets into sticky ends. Self-looping oligonucleotides with complementary 3′-overhangs are ligated to form dumbbells by DNA polymerase and ligase in a sequence-restricted manner. These reactions proceed in one pot at one temperature. We demonstrated one use of this method to 'tunnel' sequencing libraries into dumbbells for the Pacific Biosciences (PacBio) platform. Readouts of an Illumina P5/P7-ended 16S library from a standard microbial community confirmed successful tunneling. Twelve fecal samples additionally showed significant correlations between standard and tunneled 16S sequence variants on the PacBio platform. We further extended the method at a genomic scale to build a ~0.45 Mbp giant dumbbell on chromosome 6. Sequences inside the dumbbells were successfully protected from a cocktail of exonucleases. Roughly 11-fold enrichment was achieved for the dumbbell-guarded region compared to the vicinity.

1. Introduction

DNA dumbbells are end-looped linear DNA. They are highly resistant to exonucleases and can serve as iterative templates for rolling-circle amplification [1] because of topological equivalence to DNA circles. Many modern sequencers use DNA dumbbells or circles as sequencing templates. Pacific Biosciences (PacBio) platforms, for example, derive long and highly accurate sequences *via* alignment of iterative 'subreads' into consensus 'HiFi' reads [2]. MGI Tech instead uses circular DNA to amplify sequencing signals *via* 'DNA nanoballs' [3].

DNA dumbbells can be made by PCR [4,5] or direct ligations of end loops [6]. PCR-dependent workflows are relative complex and suffer from the processive limits of DNA polymerase. The native status of input DNA is destroyed after PCR amplification. Direct ligations instead lack the control of end sequence specificity in dumbbell making. Here we describe a method to achieve one-pot looping of linear DNA with fixed or known end sequences into dumbbells. It resembles Gibson Assembly (GA) [7] by firstly exposing 3′-overhangs of input DNA with 5′ exonuclease but differs from GA in the assembled components. We use pre-designed self-looping oligonucleotides with complementary 3′-overhangs as the ligation targets to enhance end sequence specificity. The reaction completes with assistance from DNA polymerase and ligase in the same pot. Because processivity of 5′ exonuclease is limited by GA design [7], native status for the most part of input DNA is preserved.

* Corresponding author. Institute of Molecular and Genomic Medicine, National Health Research Institutes, Zhunan Town, Miaoli County, 35053, Taiwan.

E-mail address: hhliu@nhri.edu.tw (H.-H. Liu).

https://doi.org/10.1016/j.heliyon.2023.e16035
Received 21 January 2023; Received in revised form 27 April 2023; Accepted 3 May 2023
Available online 9 May 2023
2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

We demonstrated the first example use of this method in tunneling sequencing libraries across incompatible platforms. Short-read libraries for Illumina are linear DNA with fixed P5/P7 sequences on either end, which allow PCR amplification and hybridization to Illumina flow cells [8]. Long-read libraries for PacBio are instead dumbbell-shaped with fixed loop sequences [6]. These libraries require different workflows to prepare. By incorporating PacBio-compatible self-looping oligonucleotides onto P5/P7-ended libraries, we offered a 'tunneling' option in sequencing library preparation.

The same method was secondly tested at a genome-wide scale. A region of Human Leukocyte Antigen II (HLA II) on chromosome 6 between gene *BTNL2* and *PSMB9* was made into a giant dumbbell of ~0.45 Mbp in size. HLA II is a group of genes that encode proteins involved in antigen presentation and immune response. They are highly polymorphic and variable among individuals. The ends of HLA II dumbbells were created by CRISPR/Cas9 [9], which is a genome editing tool that uses a guide RNA and a nuclease to introduce targeted double-strand breaks in DNA. The cut ends were further processed by our sequence-restricted looping. Resultant dumbbells were resistant to a cocktail of exonucleases. Around 11-fold enrichment was achieved for the dumbbell-guarded region compared to

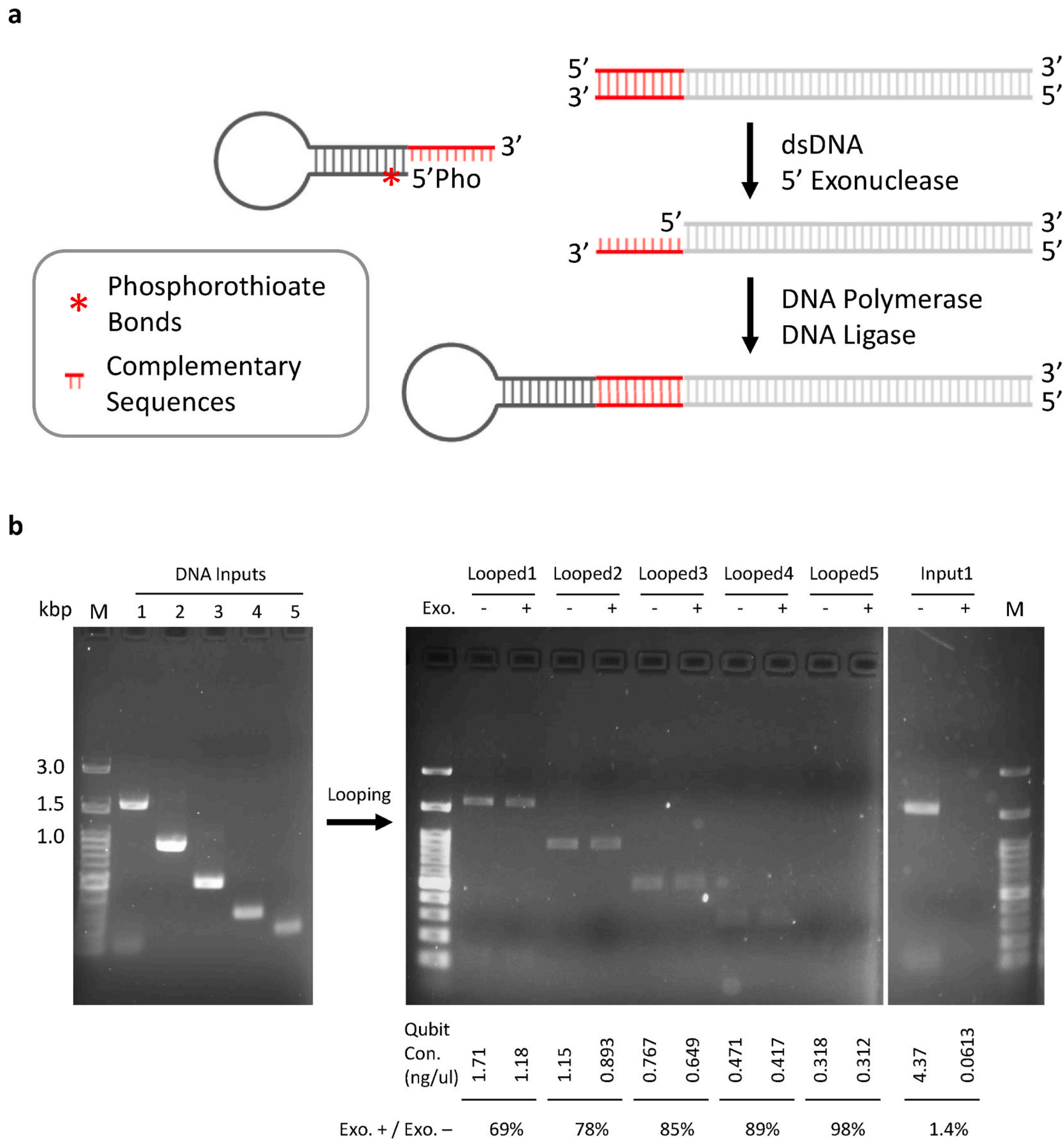


Fig. 1. Demonstration of protocol. (**a**) P5/P7 ends of input libraries are processed using a dsDNA 5′-exonuclease, as in Gibson Assembly. Only one end is illustrated. 3′-ssDNA is exposed by dsDNA 5′-exonuclease, which allows 5′-phosphorothioated and 5′-phosphorylated self-looping oligonucleotides to anneal with complementary 3′-overhangs. DNA polymerase and ligase complete the looping reaction. (**b**) Five DH5α 16S libraries of lengths varying by factors of two were subjected to conversion. Products were tested with a cocktail of exonucleases, which has no effect on successfully looped dumbbells. Conversion efficiencies varied from 69% for the longest input sequence to 98% for the shortest input sequence. Little unconverted input survived digestion by the exonuclease cocktail. The raw supplement image files are raw_1b_toLeft.tif and raw_1b_toRight.tif, respectively.

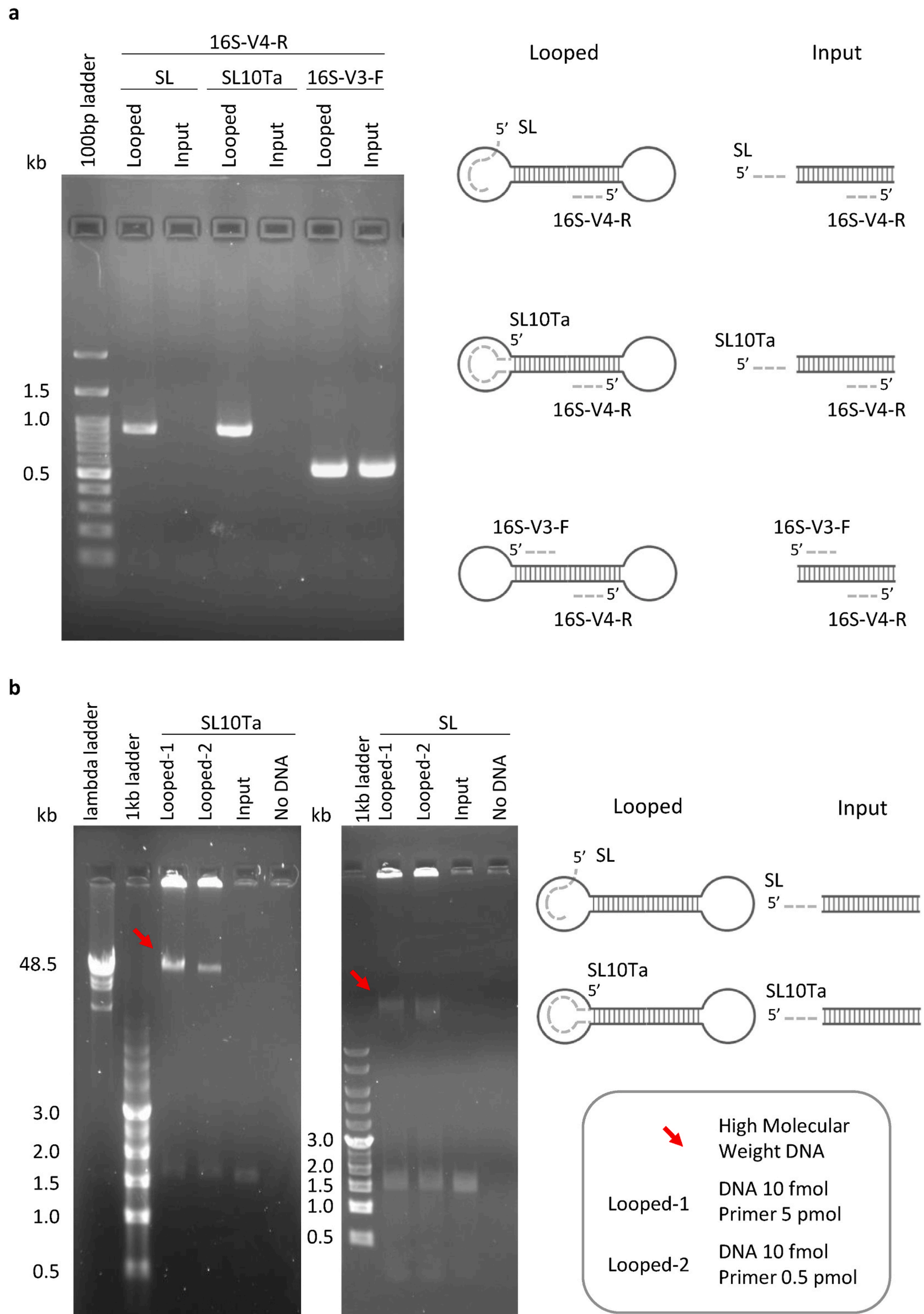


Fig. 2. Verification of protocol. (**a**) Two different primers capable of hybridizing with a loop paired with a reverse primer matching the input sequence. Only looped products, and not unconverted inputs, yielded positive PCR bands. The raw supplement image file is raw_2a_fromLeft.tif. (**b**) Rolling circle amplifications by Phi29 polymerase proceeded only on dumbbell templates primed with two different loop-hybridizing primers. No high molecular-weight products (arrow) were produced using unconverted inputs as templates. The raw supplement image files are raw_2-b_fromLeft_toLeft_SL10Ta.tif and raw_2b_toRight_SL.tif, respectively.

the vicinity.

2. Methods

2.1. Oligonucleotides

All PCR primers and self-looping oligos (SLOs) were synthesized by Protech Ltd. (Taipei, Taiwan). Nucleotide variants or bond modifications were incorporated commercially as indicated.

2.2. Preparation of DH5α 16S rDNA

16S rDNA of variable lengths for Fig. 1 were PCR amplified from the DH5α *E. coli* strain (ECOS101 DH5α competent cell, YB Biotech, Taipei, Taiwan) using the following thermocycle: 95 °C 3 min, 30 cycles of 98 °C 20 s + 60 °C 15 s + 72 °C 2 min, and 72 °C 7 min. Each 50 μL reaction contained either a 10 μL pellet from a 5-h culture or 10 ng template DNA, 0.2 μM each primer, and KAPA HiFi HotStart ReadyMix (Roche, Cape Town, South Africa). A full-length founder rDNA of 1.6 kbp was prepared with the primer pair 5′-AAGCAGTGGTATCAACGCAGAGT AGRGTTYGATYMTGGCTCAG-3′ and 5′-CAGACGTGTGCTCTTCCGATCT RGY-TACCTTGTTACGACTT-3'. P5 and P7-flanking derivatives of lengths varying by factors of two were subsequently amplified from the founder by a common primer 5′-GGCGACCACCGAGATCTACAC AGRGTTYGATYMTGGCTCAG-3′ coupled with 5′-GCAGAA-GACGGCATACGAGAT RGYTACCTTGTTACGACTT-3′, 5′-GCAGAAGACGGCATACGAGAT CCAAGTCGACATCGTTTACGG-3′, 5′-GCA-GAAGACGGCATACGAGAT CCTTCCTCCCCGCTGAAAGTA-3′, 5′-GCAGAAGACGGCATACGAGAT GCACATCCGATGGCAAGAGG-3′, and 5′-GCAGAAGACGGCATACGAGAT ACATTACTCACCCGTCCGCCACT-3′ for templates of full-length, ~800bp, ~400bp, ~200bp, and ~100bp, respectively. 16S rDNA for Fig. 2 was the same as the above founder rDNA.

2.3. Dumbbell formation of DH5α 16S rDNA

Pooled 16S rDNA inputs reacted with two SLOs in NEBuilder® HiFi DNA Assembly Master Mix (New England BioLabs®, Ipswich, Massachusetts, USA) at 45 °C for 2–3 h. Products in Fig. 1 were formed from the SLO pair Pho5′-G*G*C*C*A*GCAGC AGAGGAGGAC GCTGCTGGCC ATATG AATGATACGGCGACCACCGAGATCTACAC-3′ and Pho5′-G*A*A*G*T*GCGCTG TAAGTATTAC CAGCG-CACTTC ATAT CAAGCAGAAGACGGCATACGAGAT-3′, while those in Fig. 2 were formed from the SLO pair Pho5′-G*G*C*C*A*G-CAGC AGAGGAGGAC GCTGCTGGCC AAGCAGTGGTATCAACGCAGAGT-3′ and Pho5′-G*A*A*G*T*GCGCTG TAAGTATTAC CAGCGCACTTCT ATAATGG CAGACGTGTGCTCTTCCGATC-3'. Asterisks (*) denote phosphorothioate bonds.

2.4. Efficiency calculation for DH5α 16S dumbbells

To calculate efficiency, one half of each looping reaction was diluted three-fold and subjected to a cocktail of exonucleases at a final concentration of 0.2 U/μL T7 exonuclease, 0.2 U/μL truncated exonuclease VIII, 0.2 U/μL thermolabile exonuclease I, and 1 U/μL exonuclease III in rCutSmart buffer for 30 min at 25 °C, followed by 30 min at 37 °C. All reagents and buffer were from New England BioLabs® (Ipswich, Massachusetts, USA). DNA with or without exonuclease treatment was quantified using a DNA Qubit™ dsDNA HS Assay Kit (Invitrogen™, Eugene, Oregon, USA). Efficiency of dumbbell formation was derived accordingly.

2.5. Verifications for DH5α 16S dumbbells

Before verifications, dumbbell reactions in NEBuilder® were treated with a cocktail of exonucleases as described above, and cleaned up using MinElute PCR Purification Kit (QIAGEN, Hilden, Germany). For PCR verification in Fig. 2a, a common primer 16S–V4-R (5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GACTACHVGGGTATCTAATCC-3′) was paired with loop-hybridizing primer SL (5′-GTCCATCCTC GTCCTCCTC-3′) or SL10Ta (5′-AGCAGC GTCCTCCTCT GCTG-3′) to confirm dumbbell formation, and 16S–V3–F (5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT CCTACGGGNGGCWGCAG-3′) to control for inputs. PCR reactions were performed in Q5 HiFi polymerase ready mix (New England BioLabs®, Ipswich, Massachusetts, USA) using the following thermocycle: 98 °C 30 s; 5 cycles of 98 °C 10 s, 45 °C (SL) or 62 °C (SL10Ta or 16S–V3–F) 3 min; 25 cycles of 98 °C 10 s, 54 °C (SL) or 62 °C (SL10Ta or 16S–V3–F) 30 s, 72 °C 40 s; and 72 °C 2 min. For Fig. 2b Phi29-dependent rolling circle amplification on dumbbells, loops were primed with primer SL or SL10Ta. Specifically, a 10 μL mixture of 10 fmol dumbbell template, 0.5 or 5 pmol primer, and 1 mM dNTP was heated at 95 °C for 3 min before ramping down slowly to 25 °C. Bovine serum albumin at final 0.2 mg/mL and Phi29 polymerase (New England BioLabs®, Ipswich, Massachusetts, USA) were added and the reaction ran at 30 °C for 16 h before inactivation at 65 °C for 10 min.

2.6. Tunneled PacBio sequencing of microbial 16S dumbbells

15 samples (Supplementary Tab. 2) were sequenced on the PacBio Sequel platform by GENOMICS (New Taipei City, Taiwan). Two samples were from DH5α *E. coli* strain (ECOS101 DH5α competent cell, YB Biotech, Taipei, Taiwan); one sample was ZymoBIOMICS Microbial Community Standard (D6300, ZYMO RESEARCH, Irvine, CA, USA); the others were controls that were previously sequenced on the PacBio Sequel IIe platform. For the latter 12 controls, fecal DNA of sample #67 was extracted from one SCID mouse using

ZymoBIOMICS DNA Miniprep Kit (ZYMO RESEARCH, Irvine, CA, USA), and the other human fecal DNA was from a published study [10] using QIAamp PowerFecal Pro DNA Kit (QIAGEN, Hilden, Germany).

Sequence details of customized primers to amplify full-length 16S with P5/P7 ends are in Supplementary Tab. 1. PCR thermocycles were: 95 °C 3 min, 25–28 cycles of 95 °C 30 s + 57 °C 30 s (ramp rate ≤ −3 °C/sec) + 72 °C 1 min, and 72 °C 3 min. Each 25 μL reaction contained 2 ng template DNA, 0.375 μM each primer, and KAPA HiFi HotStart ReadyMix (Roche, Cape Town, South Africa). PCR products were cleaned up with MinElute PCR Purification Kit (QIAGEN, Hilden, Germany) and quantified by DNA Qubit™ dsDNA HS Assay Kit (Invitrogen™, Eugene, Oregon, USA). Purified amplicons receiving the same pair of SLOs were pooled equally before assembly. Most samples were subjected to asymmetric looping (Supplementary Tab. 2), which allowed docking of the PacBio sequencing primer on the P5 loop only. One sample received symmetric looping, which took the PacBio sequencing primer on both loops. SLOs for asymmetric looping are Pho5′-C*G*T*A*T*CATCTCTCTC TTT TCCTCCTCCTCC GTTGTTGTTGTT GAGAGAGATGATACG GCGACCACCGAGATCTACAC-3′ and Pho5′-G*A*A*G*T*GCGCTG TAAGTATTAC CAGCGCACTTC ATAT CAAGCAGAAGACGGCATACGAGAT-3’. The latter was replaced with Pho5′- C*T*C*T*C*TCAACAACAAC TCCTCCTCCTCCGTT GTTGTTGTTGAGAGAG ATCAA GCAGAAGACGGCATACGAGAT -3′ for symmetric looping. The assembly condition has been described above.

2.7. Bioinformatic analysis of tunneled PacBio sequencing

Detailed commands and Python scripts are described in the Supplementary file. In brief, adapters were recalled using PacBio's utility *recalladapters*. The output subreads.bam was processed by the *ccs* utility to get the fastq file. Both *recalladapters* and *ccs* ran with default parameters. *After* demultiplexing, records belonging to the standard microbial community or each fecal sample were pooled into corresponding fastq files before downstream analyses. Demultiplexed fastq files are deposited to European Nucleotide Archive under accession number PRJEB58751. BAM files are available on requests.

2.8. Dumbbell formation of human genomic DNA

3 ml whole blood was donated by a male adult volunteer. Genomic DNA (gDNA) was extracted with Puregene® Blood Core Kit (QIAGEN, Maryland, USA). Targets were prepared from CRISPR/Cas9-mediated cuts between exon 1 of *BTNL2* and exon 5 of *PSMB9* with crRNA 5′-GGAAAATCCACCATCCTCCC-3′ and 5′-CCGGAGCCACCAATGGC-3′, respectively. Each crRNA annealed to tracrRNA in Duplex Buffer (Integrated DNA Technologies, Newark, New Jersey, USA) and complexed to Cas9 nuclease to form ribonucleoprotein in 1× NEBuffer r3.1 buffer (New England BioLabs®, Ipswich, Massachusetts, USA). About 3 μg gDNA was cut at 37 °C for 2 h in an 18 μl reaction containing 56 nM ribonucleoprotein, which was stopped at 65 °C for 5 min and digested by 0.3 μl of Thermolabile Proteinase K (New England BioLabs®, Ipswich, Massachusetts, USA) at 37 °C for 15 min and 55 °C for 10 min. The CRISPR/Cas9-processed gDNA was then diluted to less than 20% of total volume in the following loop-assembly mixture.

In NEBuilder® HiFi DNA Assembly Master Mix (New England BioLabs®, Ipswich, Massachusetts, USA) biotin-loops were ligated one-sidedly at 45 °C for 3 h with either Pho5′-G*G*C*C*A*GCAGC AGAGG/iBiodT/GGAC GCTGCTGGCC CCCTGGAACAAAGACAAGGA-3′ or Pho5′-G*G*C*C*A*GCAGC AGAGG/iBiodT/GGAC GCTGCTGGCC AAAAGGCTGTCGAGTCAGCA-3’. DNA was then purified with 10 μg of Dynabeads™ MyOne™ Streptavidin C1 (Invitrogen™, Norway) in 1× B&W buffer (5 mM Tris-HCl pH8.0, 1 M NaCl, 0.5 mM EDTA) for 1 h with gentle rotation. After triple washes with 1× B&W buffer, DNA was resuspended in Buffer EB (QIAGEN, Hilden, Germany) before PCR verification.

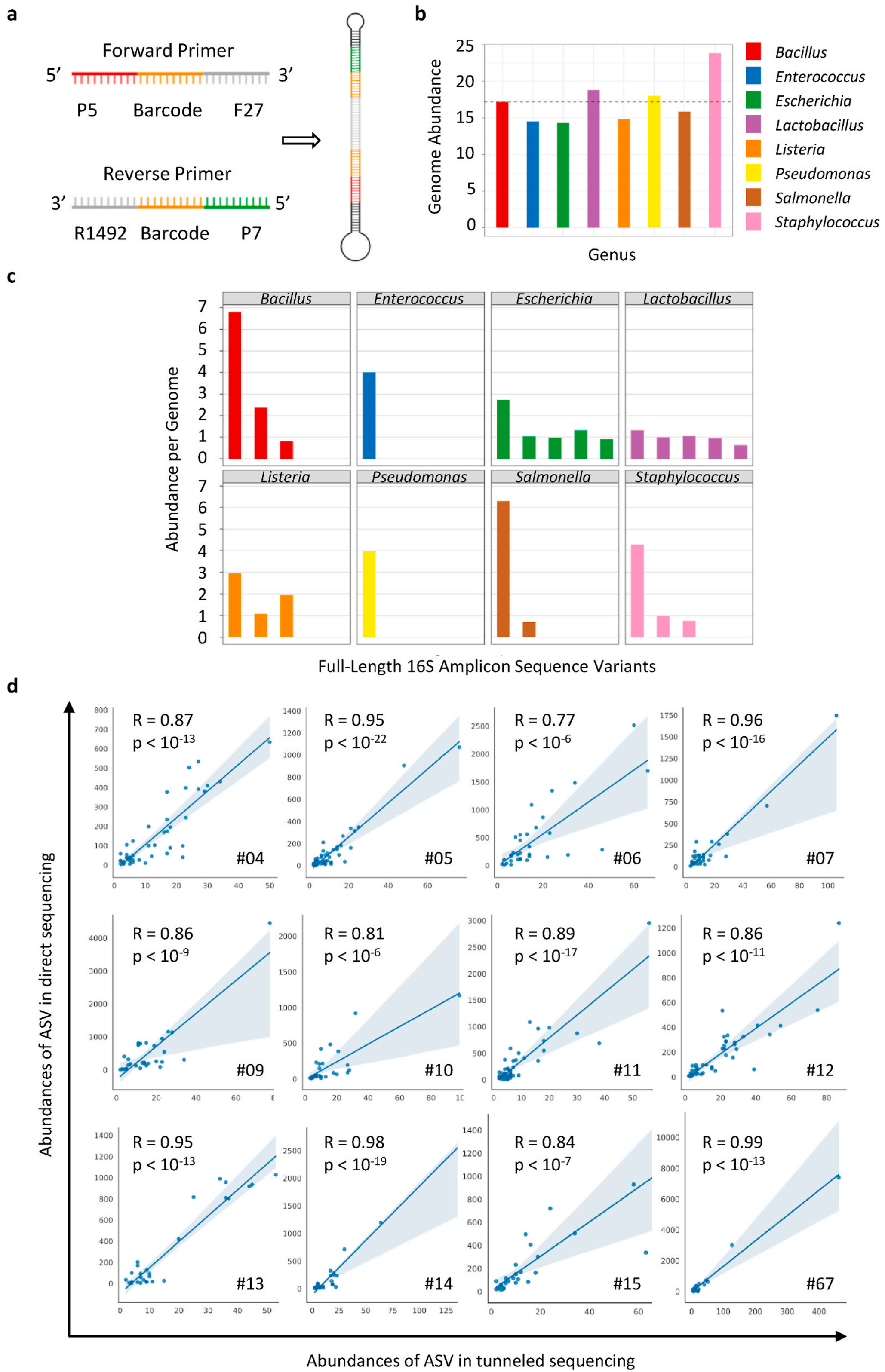
Genomic dumbbells for exonuclease challenges were similarly prepared with variants of the above two SLOs, whose biotin-dT was replaced with dA, to ligate both loops simultaneously. After completion the assembly mixture was diluted three-fold and subjected to digestion by a cocktail of exonucleases at a final concentration of 0.2 U/μL T7 exonuclease, 0.2 U/μL truncated exonuclease VIII, 0.2 U/μL Thermolabile exonuclease I, and 1 U/μL exonuclease III in NEBuffer 4 (New England BioLabs®, Ipswich, Massachusetts, USA) for 3 h at 37 °C and 30 min at 70 °C before PCR check.

2.9. PCR verification for genomic dumbbells

A series of 8 PCR reactions as illustrated in Fig. 4a were used for verification of genomic dumbbells. The sequence pairs are 5′-CAGTAAAGGGAGAAGGGAATCCT-3’ + 5′-TGGTGGATTTTCCAGGCTACA-3’ (PCR 1), 5′-CAGTAAAGGGAGAAGGGAATCCT-3’ + 5′-ATGGAAGCCATGGAGTGTGC-3’ (PCR 2), 5′-TGGCTTCCATTGGTTATGCAG-3’ + 5′-TAGCAAGCCGAAATGAGCAAG-3’ (PCR 3), 5′-CAGGGCTCTCAGGGAGACAGT-3’ + 5′-CTCAAGGGCTGCCAGGACTAT-3’ (PCR 4), 5′-CCAAGGACACTACCTCTGAAAGC-3’ + 5′-ATCAAGCCCTTCCTTCCCACT-3’ (PCR 5), 5′-TGAGGATCCCTTTCCCAGGTA-3’ + 5′-ACCCTTCCACTTGGCTGGTTA-3’ (PCR 6), 5′-CCGGCAGCACCTTTATCTATG-3’ + 5′-ACCCTTCCACTTGGCTGGTTA-3’ (PCR 7), and 5′-GCCAACACCTGTGGGTTAGAG-3’ + 5′-CATCCTGGAGTCAGGACTTTCC-3’ (PCR 8). Reactions proceeded in a final 5 μL of KAPA HiFi HotStart ReadyMix (Roche, Cape Town, South Africa) with 0.3 μM each primer at 95 °C 3 min, 30 (Fig. 4d upper two series) or 33 (Fig. 4d bottom series) cycles of 98 °C 20 s + 64 °C (PCR 4, 8) or 66 °C (PCR 1, 7) or 67 °C (PCR 2, 3, 5, 6) 15 s + 72 °C 10 s, and final 72 °C 1 min.

2.10. Real-time PCR for genomic dumbbells

We used QuantStudio™ 5 Real-Time PCR System with default settings of 'Fast Mode and Analyze Automatically'. Reactions went in duplicates, each running in a final 15 μL of KAPA SYBR® FAST qPCR Master Mix with 0.2 μM each primer. Sequences of primers are the same as above for PCR verification.



(caption on next page)

Fig. 3. Tunneling to PacBio sequencing. (**a**) A pair of primers able to amplify full-length 16S with P5/P7 end sequences is illustrated. (**b**) Genome abundances of eight bacterial genera were relatively equal, corresponding to the known composition of the microbial community standard. (**c**) Stoichiometric analyses of 16S ASV per bacterial genome revealed integral values and sums in line with expectations. (**d**) ASV abundances are significantly correlated between direct and tunneled sequencing.

3. Results

3.1. One-pot synthesis of DNA dumbbells

This method follows Gibson Assembly [7] in utilizing 5′-exonuclease to expose sticky ssDNA from input DNA (Fig. 1a). Introduction of self-looping oligonucleotides with complementary 3′-ends completes dumbbell formation in the same pot. Phosphorothioate bonds are artificially incorporated at the 5′-ends of looping oligonucleotides to resist the action of 5′ exonuclease.

As a pilot demonstration, five DNA libraries of DH5α 16S ended with Illumina P5/P7 sequences were prepared, with lengths

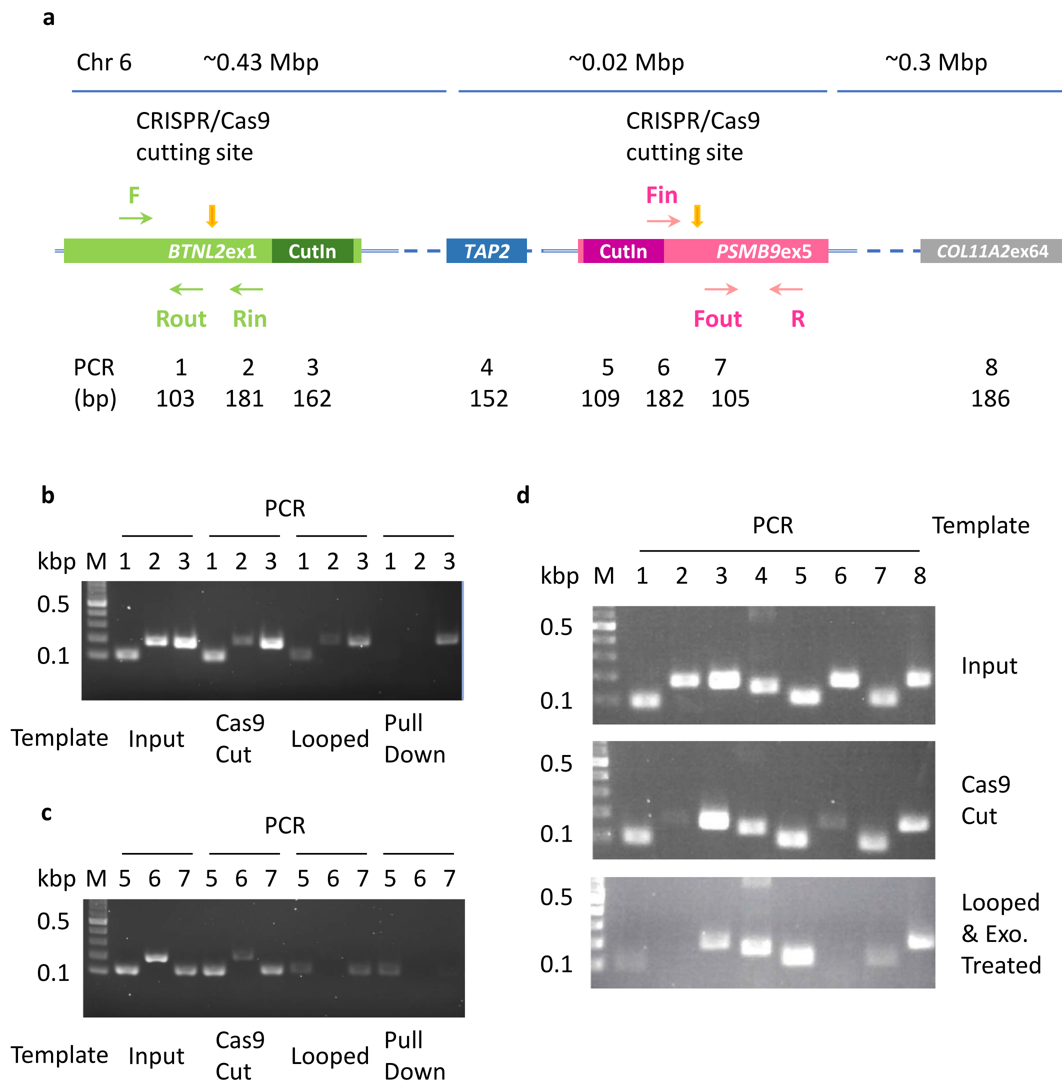


Fig. 4. Genomic dumbbells. (**a**) Eight PCR reactions were designed to verify the formation of genome-scale dumbbells between exon 1 of *BTNL2* and exon 5 of *PSMB9*. PCR 1, 7, and 8 are outside the dumbbells; PCR 3, 4, and 5 are inside the dumbbells; PCR 2 and 6 amplify across CRISPR/Cas9-mediated cuts which are made for sequence-restricted looping reactions. (**b**) PCR 1, 2, and 3 demonstrated successful ligations of biotin-tagged loops for the *BTNL2* cut. Only PCR 3 persisted after pull-down by Streptavidin beads. The raw supplement image file is raw_4b.tif. (**c**) A similar three-PCR profile verified the formation of the *PSBM9* loop. The raw supplement image file is raw_4c.tif. (**d**) Insider PCR 3, 4, and 5 successfully survived exonuclease treatments after the dumbbell formation. The raw supplement image files are raw_4d_fromLeft_Input.tif, raw_4d_fromRight_Cas9Cut.tif, and raw_4d_Looped.tif, respectively.

varying over four powers of two (Fig. 1b, left panel). Library conversions were performed using commercially-available ready-mix NEBuilder® kits, an augmented version of Gibson Assembly from New England BioLabs®. Products were subjected to digestion by a cocktail of exonucleases (Fig. 1b, right panel). Only successfully converted dumbbells were not degraded by the exonuclease cocktail, but DNA inputs with open ends were highly susceptible. Conversion efficiency decreased moderately with input length, ranging from 69% for the longest input to 98% for the shortest input. Only 1.4% of unconverted input DNA remained undegraded in the longest input (Fig. 1b, right panel).

3.2. Verification of dumbbell formation

Looping reactions were verified using PCR (Fig. 2a). Dumbbells with known loop sequences were prepared from the DH5α 16S sequence. Two primers, SL and SL10Ta, differ in 5′-end sequences but hybridize with the same loop of each dumbbell. Both were each paired with a primer matching the input sequence (16S–V4-R). Only dumbbells, but not unmodified inputs, yielded positive bands. SL and SL10Ta were also used for Phi29-based rolling circle amplification [1] (Fig. 2b). Only dumbbells, but not unmodified inputs, allowed Phi29-catalyzed polymerization of high molecular-weight products (Fig. 2b, arrow).

3.3. Tunneling to PacBio sequencing

DNA dumbbells are the template format of PacBio sequencers. Analyses of a standard microbial community on the PacBio platform are published in Callahan et al. [11]. For comparison, we prepared a full-length 16S library from the same standard community with customized PCR primers (Fig. 3a and Supplementary Tab. 1), along with other full-length 16S control libraries (Supplementary Tab. 2). These PCR oligos differed from PacBio standards on the 5' ends, where the forward primer had Illumina P5 sequences and the reverse primer had Illumina P7 sequences, respectively. After PCR amplification, these linear DNA libraries were converted to dumbbell derivatives.

Dumbbell libraries were directly sequenced on the PacBio Sequel platform with default parameters. Demultiplexed reads of the standard community were analyzed step-by-step following the pipeline from Callahan et al. [11]. 23 amplicon sequence variants (ASV) of full-length 16S were identified by DADA2 [12], 12 of which exactly matched reference sequences (Table 1). All eight bacterial genera, including *Bacillus, Enterococcus, Escherichia, Lactobacillus, Listeria, Pseudomonas, Salmonella, and Staphylococcus* were detected. Their genomic abundances are expected to be equal and our results recapitulated the prediction (Fig. 3b). Precise stoichiometry of ASV per genome (Fig. 3c) corresponding to Callahan's results [11] was successfully reproduced.

Twelve fecal full-length 16S libraries previously sequenced following PacBio's standard protocol (Supplementary Tab. 2), henceforth referred to as direct sequencing (DS), were re-sequenced as controls. ASV of each sample by DS and tunneled sequencing (TS) were both identified by DADA2 [12]. Although DS had higher HiFi read depths than TS (14.4 fold on average), ASV identified from DS were only 4.5 fold higher than TS (Table 2). Among TS ASV, 92.2 ± 2.1% (s.e.) overlapped with those from DS. Abundances of overlapped ASV showed significant correlations (Fig. 3d). Pearson coefficients ranged from 0.77 to 0.99 (0.89 ± 0.02, s.e.), with a significance level $<10^{-6}$ at least.

3.4. Genomic dumbbell

DNA dumbbells are highly resistant to exonuclease digestion. The principle could be applied to enrich genomic regions by looping targets into dumbbells for protection. Unguarded regions could be made degraded by exonucleases. For illustration, we aimed at a span of ~0.45 Mbp on chromosome 6 between exon 1 of *BTNL2* and exon 5 of *PSMB9* (Fig. 4a). *HLA-DRs, HLA-DQs*, *TAP2*, *PSMB8*, and *TAP1* are all included in the target [13]. Boundary cuts were made by sequence-guided CRIPSR/Cas9 nucleases [9].

Eight PCR reactions across the dumbbell span were used to verify the enrichment (Fig. 4a). PCR 1 (F/Rout) and 7 (Fout/R) targeted sequences right outside the dumbbell loops. PCR 2 (F/Rin) and 6 (Fin/R) covered areas across the indicated cuts on *BTNL2* and *PSMB9*, respectively. PCR 3 and 5 were in the immediate vicinities inside both loops. PCR 4 regarded the insider gene *TAP2*. PCR 8 instead tested the exon 64 of the outsider gene *COL11A2*, which is around 0.3 Mbp further from the *PSMB9* loop.

Looping reactions at a genome-wide scale were firstly tested with biotin-tagged loops. Streptavidin-mediated precipitates are expected to yield positive PCR results only for those areas inside the loop but not outside. Fig. 4b showed verifying results of PCR 1, 2, and 3 for the *BTNL2* loop. Inputs had all three reactions present, but PCR 2 decreased a lot after CRISPR/Cas9 cuts and dumbbell

Table 1
Matches of ASV to ZymoBIOMICS reference.

Genus	No	Yes
Bacillus	0	3
Enterococcus	0	1
Escherichia	5	0
Lactobacillus	5	0
Listeria	0	3
Pseudomonas	0	1
Salmonella	0	2
Staphylococcus	1	2

Table 2
Summary of HiFi reads and ASV from direct and tunneled sequencing.

Sample	HiFi Reads by Direct Sequencing	ASV by Direct Sequencing	HiFi Reads by Tunneled Sequencing	ASV by Tunneled Sequencing	Matched ASV between Both Sequencings
#04	11,009	184	864	44	43
#05	10,217	197	881	55	46
#06	29,282	295	889	38	33
#07	11,429	167	858	34	32
#09	10,193	318	866	32	30
#10	29,334	133	856	38	29
#11	10,146	431	822	53	51
#12	30,917	159	1183	40	40
#13	12,164	95	703	31	28
#14	11,543	167	747	31	28
#15	11,730	144	633	29	28
#67	11,694	55	1180	17	17

looping. After pull-down by Streptavidin beads only PCR 3 reacted positively. A similar three-PCR profile was demonstrated for the *PSMB9* loop, too (Fig. 4c). Only PCR 5 persisted after pull-down by Streptavidin beads.

We then tested if the giant dumbbell between *BTNL2* and *PSMB9* could survive exonuclease attacks after the looping reaction (Fig. 4d). In the beginning, all eight PCRs yielded expected results from the human genome input (Fig. 4d, top series). After CRISPR/Cas9 cuts, PCR 2 and 6 decreased significantly but not the others (Fig. 4d, middle series). The processed genome was further treated by a cocktail of exonucleases. Resembling the pull-down results, PCR 3, 4, and 5 (insiders) persisted well but not PCR 2, 6 (cut-site crossers), or PCR 1, 7 (outsiders) (Fig. 4d, lower series). A more distant PCR 8 was discernible, though. Quantitatively real-time PCR showed no significant differences of cycles of threshold (Ct) between insiders *vs.* outsiders (24.75 ± 0.20 *vs.* 24.62 ± 0.29 s.e., $p = 0.7 > 0.05$, *t*-Test) in the beginning. After exonuclease treatment, there was a significant difference of Ct between insiders and outsiders (28.93 ± 0.14 *vs.* 32.40 ± 0.27 s.e., $p < 10^{-3}$, *t*-Test), which corresponded to a roughly 11-fold enrichment.

4. Discussion

We present a method in forming DNA dumbbells without the needs of PCR [4,5]. The approach combines the principle of Gibson Assembly [7] and the unique topology of self-looping oligonucleotides to complete the reaction in one pot. Compared with PCR-dependent workflows, the method takes no multiple steps and preserves most native nucleotides of the input material because of limited processivity of the exonuclease in Gibson Assembly. With NEBuilder® as a specific example, the kit removes no more than 100 nucleotides on either end of input DNA in a 1-h reaction at 50 °C [14]. Therefore no more than 200 end nucleotides in total would be destroyed for each DNA fragment at that condition. In contrast to simple ligation [6], the conjugation of dumbbell loops is sequence-dependent in our design. Asymmetric dumbbells can be readily achieved, making oriented dumbbell tandems possible with proper extensions of our current working protocol.

We demonstrated a tunneling example by converting Illumina P5/P7-ended libraries to PacBio-compatible dumbbells (Fig. 3). Although read depths were suboptimal (Table 2), analyses of the standard microbial community (Fig. 3a–c) exhibited very good correspondence to previously published results [11]. We saw reasonably distributed abundances of bacterial genomes from the mock standard community. In total, 23 full-length 16S ASV were identified. The number was six fewer than identified by Callahan et al. [11]. Due to the significant difference in sequencing depths, this disparity is expected. Importantly, stoichiometry of ASV per genome matched Callahan's results exactly. The missing three ASV for *Bacillus*, two ASV for *Enterococcus*, and one ASV for *Staphylococcus* did not change overall integral sums of ASV for these genera, which remained 10, 4, and 6, respectively. Even for fecal samples with complex bacterial communities, reads of tunneled libraries correlated well (Fig. 3d). The suboptimal read depths apparently affected little the accuracy of tunneled sequencing.

In another example we used the method to create a roughly 0.45Mbp dumbbell on chromosome 6 (Fig. 4). The dumbbell exhibited expected properties in resisting the actions of exonucleases. This could be a valuable addition to the current approaches in genomic region enrichment. Prior arts include probe-based hybridization and CRISPR/Cas9-based treatments [9]. Published protocols for the latter category often require complex steps and/or expensive instruments to retrieve the desired material. Off-target cuts would be another concern. With our method, genomic regions could be readily enriched in one pot and selected in a sequence-restricted manner without the needs of complex steps or capital equipment. Suboptimal clearance of untargeted regions such as PCR 8 on Fig. 4d could be further optimized by customization of exonuclease buffer and compositions. For smaller dumbbells, secondary enhancement by rolling circle amplification [1] could be another choice to enrich targeted sequences. For bigger dumbbells, an additional step of DNA repair might help fix unexpected nicks and enhance exonuclease resistance.

Author contribution statement

Ya-Hui Chang: Performed the experiments.
Yu-Cheng Lin: Contributed reagents, materials, analysis tools or data.
Hong-Hsing Liu: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Funding statement

Hong-Hsing Liu was supported by National Health Research Institutes {MG-109-SP-05}

Data availability statement

Data associated with this study has been deposited at Demultiplexed fastq files are deposited to European Nucleotide Archive under accession number PRJEB58751.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Chih-Wei Joshua Liu (MIT Physics of Living Systems) for English editing. This work was supported by the National Health Research Institutes [MG-109-SP-05 to H.H.L.]

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.heliyon.2023.e16035.

References

[1] F.B. Dean, et al., Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification, Genome Res. 11 (6) (2001) 1095–1099.
[2] T. Hon, et al., Highly accurate long-read HiFi sequencing data for five complex genomes, Sci. Data 7 (1) (2020) 399.
[3] R. Drmanac, et al., Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays, Science 327 (5961) (2010) 78–81.
[4] M. Taki, et al., A direct and efficient synthesis method for dumbbell-shaped linear DNA using PCR in vitro, Nucleic Acids Res. Suppl. (3) (2003) 191–192.
[5] H. Yu, et al., Efficient production of superior dumbbell-shaped DNA minimal vectors for small hairpin RNA expression, Nucleic Acids Res. 43 (18) (2015) e120.
[6] K.J. Travers, et al., A flexible and efficient template format for circular consensus sequencing and SNP detection, Nucleic Acids Res. 38 (15) (2010) e159.
[7] D.G. Gibson, et al., Enzymatic assembly of DNA molecules up to several hundred kilobases, Nat. Methods 6 (5) (2009) 343–345.
[8] M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing, Cold Spring Harb. Protoc. 2010 (6) (2010) prot5448.
[9] P.D. Hsu, E.S. Lander, F. Zhang, Development and applications of CRISPR-Cas9 for genome engineering, Cell 157 (6) (2014) 1262–1278.
[10] Y.C. Lin, et al., Pathogenic effects of Desulfovibrio in the gut on fatty liver in diet-induced obese mice and children with obesity, J. Gastroenterol. 57 (11) (2022) 913–925.
[11] B.J. Callahan, et al., High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution, Nucleic Acids Res. 47 (18) (2019) e103.
[12] B.J. Callahan, et al., DADA2: high-resolution sample inference from Illumina amplicon data, Nat. Methods 13 (7) (2016) 581–583.
[13] T. Shiina, et al., The HLA genomic loci map: expression, interaction, diversity and disease, J. Hum. Genet. 54 (1) (2009) 15–39.
[14] Q.&A. Webinar, NEBuilder HiFi, The next generation of DNA assembly. https://www.neb.com/~/media/nebus/files/publication%20pdfs/nebuilderwebinarqa.pdf, 2020.