



A modified protocol with less clean-up steps increased efficiency and product yield of sequencing library preparation

Junman Huang^{1,2} · Chenhong Li^{1,2}

Received: 13 November 2021 / Accepted: 19 March 2022
© King Abdulaziz City for Science and Technology 2022

Abstract

Library preparation is an essential step for the next-generation sequencing, such as whole-genome sequencing, reduced-representation genome sequencing, exome sequencing and transcriptome sequencing. The library preparation often involves many steps, including DNA fragmentation, end repair, ligation and amplification. Each step involves different enzymes and buffer systems, so many washing steps are implemented in between to clean-up the enzymes and solutes from the previous step. Those extra washing steps not only are tedious and costly, but more importantly may introduce cross-contamination and reduce the final library yield. Here, we modified the common protocol of Illumina library prep to reduce the washing steps by deactivating the enzymes with high temperature. The modified protocol has two less washing steps than the original one, which can save more than 40 min of hands-on time and reduce potential risk of cross-contamination. We compared our protocol with the original one by constructing libraries using 200 ng DNA of *Tetraodon nigroviridis*. The results showed that libraries prepared with the modified protocol had higher yields than that using the original protocol (53.4 ± 16.8 ng/ml vs. 8 ± 0.7 ng/ml), whereas the coverage and PCR duplication rate were similar. Furthermore, we eliminated the very first washing step after DNA shearing to preserve short DNA fragments, which increased proportion of fragments less than 100 bp DNA from 0.82 to 2.99%. In conclusion, using the modified protocols not only can save time and money, but also can generate higher yield and keep more short DNA fragments.

Keywords Library prep · Clean-up · Cost-effectiveness · Cross-contamination · Short fragments

Introduction

Library preparation protocols can be divided into two types based on different DNA shearing mechanism. Mechanical shearing by ultrasound is the most common method, after which the sheared DNA libraries are made following steps of end repair, adaptor ligation and fill-in (Hess et al. 2020). The other strategies involve fragmenting DNA using transposase or endonuclease. Tagmentation using transposase may reduce the number of steps comparing to the

mechanical shearing (Syed et al. 2009). However, the enzymatic approaches are less commonly used than the mechanical shearing probably due to high cost of the enzymatic kits and possible sequencing bias introduced in low-GC regions using the Nextera XT kit (Sato et al. 2019).

In general, each library preparation step is followed by clean-up steps. Clean-up can be used to wash away the enzymes to terminate reaction and prevent them from affecting the next reaction step. It also can be used to remove the salts, so a new buffer can be applied to the next reaction system to achieve optimal results. Clean-up step not only can be used to filter short fragments out, such as PCR primers and dimers, and adapters, but also it can be used to concentrate the DNA samples. Additionally, “dual size selection” with the Solid-Phase Reversible Immobilization (SPRI) method can be used to clean-up both small and long DNA molecules (Deangelis et al. 1995).

Spin-column and SPRI beads are the two common options for clean-up steps, but both of them have some drawbacks. Part of the target DNA sample is inevitably lost

✉ Chenhong Li
chli@shou.edu.cn

¹ Shanghai Universities Key Laboratory of Marine Animal Taxonomy and Evolution, Shanghai Ocean University, Shanghai 201306, China
² Engineering Research Center of Environmental DNA and Ecological Water Health Assessment, Shanghai Ocean University, 999 Huchenghuan Rd, Pudong New District, Shanghai 201306, China

in each clean-up step. As the frequency of washing steps increases, the sample loss would get worse. The DNA loss often is beyond what we expect in short fragments (Aigrain et al. 2016). Furthermore, the more we open the lid of reaction tubes or pipetting the samples for the clean-up steps, the risk of cross-contamination may increase. Finally, if we can reduce the clean-up steps, consumables, such as beads, buffer, ethanol, pipette tips and PCR tubes can be saved.

Clean-up step plays an important role in library preparation, but what would happen if we skip some of the clean-up steps? Some assumptions can be made: (1) without clean-up steps, residual enzymes maybe overreact, and remaining buffer system would produce undesired effects by unbalancing pH and salt concentration to the new solution; (2) the use of consumables for clean-up steps can be reduced, and the risk of potential cross-contamination can also be lowered; (3) the yield of library may be significantly increased and short fragments also can be kept if no clean-up is used before the ligation step. Instead of clean-up, irreversible inactivation of enzyme could be achieved by heating without loss of nucleic acid (Wiame et al. 2000), but it is unknown for how would the buffer systems interfere each reaction steps if not replaced. Nonetheless, no one has tested heat denaturation of enzymes and eliminating all the clean-up steps in library preparation.

Here, we tested a modified protocol with reduced clean-up steps by inactivating enzyme with heat denaturation and keeping the solutes of the buffer system from previous reaction for the next step. Two less clean-up steps were used in the modified protocol than in the original one, that is, no clean-up for the end repair and ligation steps. Hereafter, we name the modified protocol as two-less-clean-up-step protocol (2LCS). For fresh DNA samples, high molecular weight DNA is accounted for a great majority. For degraded DNA samples, such as ancient DNA or archived DNA samples, the short DNA fragments are dominant and prone to be washed out during the clean-up steps. Therefore, we further eliminated the very first clean-up step before end repairing to keep the short fragments, and we name it as three-less-clean-up-step protocol (3LCS) (Tables 1, 4). We present the results of comparison between the three library preparation methods: original protocol (OP), 2LCS and 3LCS. All libraries were compared in terms of number of operation steps, hands-on time, consumables use, library yield, distribution of the library fragments and complexity of the method.

Materials and methods

DNA extraction

We used *Tetraodon nigroviridis* as the genetic material. The tissue sample was directly stored at -80°C . A small

Table 1 Workflow of three different protocols for library preparation

Treatment	Concentration method	Blunt-end	Clean-up	Ligation	Clean-up	Fill-in	Clean-up	PCR	Clean-up	Number of clean-up steps	Hands-on time (min)
1	MagNA bead clean-up	25°C (15 min) $\rightarrow$ 12°C (5 min)	MagNA bead clean-up	30°C (22 min)	MagNA bead clean-up	37°C (20 min)	MagNA bead clean-up	—	MagNA bead clean-up	5	115
2	MagNA bead clean-up	25°C (15 min) $\rightarrow$ 12°C (5 min)	Deactivate enzymes 80°C (15 min)	30°C (22 min)	Deactivate enzymes 70°C (5 min)	37°C (20 min)	Deactivate enzymes 95°C (2 min) $\rightarrow$ MagNA bead clean-up	—	MagNA bead clean-up	3	75
3	Concentrator (eppendorf)	25°C (15 min) $\rightarrow$ 12°C (5 min)	Deactivate enzymes 80°C (15 min)	30°C (22 min)	Deactivate enzymes 70°C (5 min)	37°C (20 min)	Deactivate enzymes 95°C (2 min) $\rightarrow$ MagNA bead clean-up	—	MagNA bead clean-up	2	50

piece of tissue was sampled, followed by DNA extraction according to instruction of Ezup Column Animal Genomic DNA Purification Kit (Sangon, Cat: B518251). DNA were sheared to between 50 and 500 bp (Peak Power 75.0, Duty Factor 20.0, Cycles/Burst 200, Duration 360 s) using a Covaris Focused-ultrasonicator M220. The short DNA fragments were generated to simulate the degraded DNA. Two hundred nanograms of shared DNA were added to each of the nine tubes for three treatment groups in triplicates.

Treatment 1 was carried following the OP of library preparation with MagNA bead clean-up (Meyer and Kircher 2010; Rohland and Reich 2012). Treatment 2 and 3 followed the 2LCS and 3LCS protocol, respectively, in which the clean-up step was replaced by deactivating enzymes using high temperature. A positive control was carried using a 300 bp DNA fragment in treatment1, and blank control was added in all treatments. In the step of ligation, a pair of inline indices were inserted between the insert DNA and adapter to trace the data from pooled sequencing samples (Wang et al. 2021). Sixteen cycles of indexing PCR were performed after fill-in step of the library preparation. The detailed experiment protocols of each treatment are listed as following.

Protocols for OP

Sample concentration

1. Add 40 µl MagNA beads (Fishersci, Cat: 09-981-123) in an empty PCR tube. Discard the supernatant while keep the beads using a magnetic plate.
2. Add 0.9× volume MagNA buffer and 1× volume sheared DNA or reaction product, vortex the tubes for several times.
3. Incubate for 5 min at room temperature (RT). Briefly centrifuge the tubes. Collect the beads using a magnet plate at least 5 min until beads completely separate from solution. Discard the supernatant without touching the beads.
4. Wash the bead pellet by 186 µl of fresh 70% ethanol for 1 min, repeat the process one more time (keep the tube on the magnetic plate). Air-dry the beads for 3 min at RT.

Blunt-end repair

1. Prepare master mix as below on an ice box for the required number of reactions. Add 20 µl of master mix to each sample.

Reagent	Volume (µl) per sample	Final concentration in 20 µl reaction
Buffer Tango (10 ×) (Thermo, Cat: BY5)	2	1 ×
dNTPs (10 mM each) (Invitrogen, Cat: 18427088)	0.2	100 µM each
ATP (100 mM) (Thermo, Cat: R0441)	0.2	1 mM
T4 polynucleotide kinase (10 U/µl) (Thermo, Cat: EK0031)	1	0.5 U/µl
T4 DNA polymerase (5 U/µl) (Thermo, Cat: EP0061)	0.4	0.1 U/µl
H ₂ O (TEDIA, Cat: WS2211-001)	16.2	
Total	20	

2. Mix and incubate the samples in a thermal cycler as following:

No	Steps	Temperature (°C)	Time (min)
1	DNA blunting	25	15
2	5' Phosphorylation of DNA	12	5

3. MagNA Clean up same as in the concentration step but skip the 1st step (no need to add new beads).

Adapter ligation

4. Prepare master mix as below on an ice box for the required number of reactions. Add 39 µl of master mix to each sample.

Reagent	Volume (µl) per sample	Final concentration in 40 µl reaction
T4 DNA ligase buffer (10 ×) (Invitrogen, Cat: 15224041)	4	1 ×
PEG-4000 (50%) (Invitrogen, Cat: 15224041)	4	5%
T4 DNA ligase (5 U/µl) (Invitrogen, Cat: 15224041)	1	0.125 U/µl
H ₂ O (TEDIA, Cat: WS2211-001)	30	
Total	39	

5. Add inline adapter mix IS1 and IS2, respectively, 0.5 µl of each to the sample tube in different combinations.

6. Mix and incubate the samples in a thermal cycler as following:

No	Steps	Temperature (°C)	Time (min)
1	Adapter ligation	30	22

7. MagNA Clean up same as in the concentration step but skip the 1st step (no need to add new beads).

Reagent	Volume (μl) per sample	Final concentration in 25 μl reaction
KAPA HiFi HotStart Ready Mix (2 ×) (KAPA BIOSYSTEMS, Cat: KM2602)	12.5	1 ×
P5 primer (10 μM)	0.5	100 μM each
P7 primer (with index) (10 μM)	0.5	1 mM
Total	13.5	

Fill-in

8. Prepare master mix as below on an ice box for the required number of reactions. Add 40 μl of master mix to each sample.

Reagent	Volume (μl) per sample	Final concentration in 40 μl reaction
Bsm buffer (10x) (Thermo, Cat: EP0691)	4	1 ×
dNTPs (10 mM each) (Invitrogen, Cat: 18427088)	1	250 μM each
Bsm polymerase, large fragment (8 U/μl) (Thermo, Cat: EP0691)	1.5	0.3 U/μl
H ₂ O (TEDIA, Cat: WS2211-001)	33.5	
Total	40	

9. Mix and incubate the samples in a thermal cycler as following:

No	Steps	Temperature (°C)	Time (min)
1	Fill-in	37	20

10. MagNA clean-up same as in the concentration step but skip the 1st step (no need to add new beads). After cleaning, re-suspend the bead pellet in 40 μl TE buffer.

Amplification

11. Add 11.5 μl product of “Fill-in” in empty sample tubes. Prepare master mix as below on an ice box for the required number of reactions. Add 13.5 μl of master mix to 11.5 μl of sample.

12. Mix and incubate the samples in a thermal cycler as following:

No	Steps	Temperature (°C)	Time (min)	Cycles
1	Pre-denaturation	98	45	1
2	Denaturation	98	15	12–16 cycles
3	Annealing	60	30	
4	Extension	72	45	
5	Extension	72	1	1
6	Storage	4	10	1

13. MagNA clean-up same as in the concentration step. Resuspend the bead pellet in 35 μl TE buffer. Incubate for 1 min at RT. Place the tube on magnetic plate for 1 min, and transfer the supernatant to a new tube.
14. Check the amplified products by electrophoresis on a 1.5% agarose gel.

Protocols for 2LCS

Sample concentration

Same as in the OP.

Blunt-end repair

1. Prepare master mix same as in the OP.
2. Mix and incubate same as in the OP.
3. Terminate reaction.

No	Steps	Temperature (°C)	Time (min)
1	Deactivating enzyme	80	15

Adapter ligation

1. Prepare master mix as below on an ice box for the required number of reactions. Add 9 µl of master mix to 20 µl of each sample.

Reagent	Volume (µl) per sample	Final concentration in 30 µl reaction
T4 DNA ligase buffer (10 ×) (Invitrogen, Cat: 15224041)	3	1 ×
PEG-4000 (50%) (Invitrogen, Cat: 15224041)	3	5%
T4 DNA ligase (5 U/µl) (Invitrogen, Cat: 15224041)	0.75	0.125 U/µl
H ₂ O (TEDIA, Cat: WS2211-001)	2.25	
Total	9	

2. Add inline adapter same as in the OP.
3. Mix and incubate same as in the OP.
4. Terminate reaction.

No	Steps	Temperature (°C)	Time (min)
1	Deactivating enzyme	70	5

Fill-in

1. Prepare master mix as below on an ice box for the required number of reactions. Add 10 µl of master mix to 30 µl of each sample.

Reagent	Volume (µl) per sample	Final concentration in 40 µl reaction
Bsm buffer (10 ×) (Thermo, Cat: EP0691)	4	1 ×
dNTPs (10 mM each) (Invitrogen, Cat: 18427088)	1	250 µM each (Which does not account for the part from the Blunt-End Repair)
Bsm polymerase, large fragment (8 U/µl) (Thermo, Cat: EP0691)	1.5	0.3 U/µl
H ₂ O (TEDIA, Cat: WS2211-001)	3.5	
Total	10	

2. Mix and incubate same as in the OP.
3. Terminate reaction.

No	Steps	Temperature (°C)	Time (min)
1	Deactivating enzyme	95	2

4. MagNA clean-up same as in the concentration step. Resuspend the bead pellet in 40 µl TE buffer.

Amplification

Same as in the OP.

Protocols for 3LCS**Sample concentration**

Add 200 ng sheared DNA to a volume of 16.2 µl. Concentrate the DNA or add water to adjust the volume to 16.2 µl if the DNA concentration is too low or too high.

Blunt-end repair

1. Prepare master mix as below on an ice box for the required number of reactions. Add 3.8 µl of master mix to 16.2 µl of concentrated sample.

Reagent	Volume (µl) per sample	Final concentration in 20 µl reaction
Buffer Tango (10 ×) (Thermo, Cat: BY5)	2	1 ×
dNTPs (10 mM each) (Invitrogen, Cat: 18427088)	0.2	100 µM each
ATP (100 mM) (Thermo, Cat: R0441)	0.2	1 mM
T4 polynucleotide kinase (10 U/µl) (Thermo, Cat: EK0031)	1	0.5 U/µl
T4 DNA polymerase (5 U/µl) (Thermo, Cat: EP0061)	0.4	0.1 U/µl
Total	3.8	

2. Mix and incubate same as in the OP.
3. Terminate reaction same as in the 2LCS.

Following steps, including adapter ligation, fill-in and amplification are same as in the 2LCS.

Product quantification and sequencing

We employed quantitative real-time PCR (qPCR) to quantify the yield of libraries, because the final product of library

preparation contained not only the targeted library fragments, but also non-targeted sequences, such as adapters and abnormal fragments. Positive control of 300 bp was quantified and diluted six times as the standards, respectively, which are $1.00\text{E}+01$, $1.00\text{E}+00$, $1.00\text{E}-01$, $1.00\text{E}-02$, $1.00\text{E}-03$, and $1.00\text{E}-04$ ng/ml. The standard curve was constructed from the diluted samples. Pure water was set as blank control group. Each sample library was diluted 100 times and performed with three replicates.

The final PCR products were quantified using a NanoDrop3300 (Thermo). All samples were pooled equimolarly. Gel-based fragment size selection was conducted using QIAquick Gel Extraction Kit (QIAGEN, #28706). DNA fragments ranged from 150 bp to 1 kb were extracted to keep short fragments as much as possible and assure good sequencing results. Then, the majority of DNA inserts would range from 14 to 864 bp without the adapter sequences. All the libraries were sequenced on an Illumina NovaSeq 6000 platform for paired-end 150 bp reads (Novogene, Beijing, China).

Sequence analyses

Raw sequence data were under the quality control using FastQC and trim_galore, to check and remove adapter sequence and low-quality bases. Qualified reads were assigned to each sample according to their inline index using a custom script (demultiplex_inline.pl) (Wang et al. 2021). We randomly selected 11,000,000 reads from the cleaned data generated by each protocol for comparison. PCR duplicates were removed using a custom script (rmrep.pl). Deduplicated reads were mapped to the reference genome of *T. nigroviridis* (https://www.ncbi.nlm.nih.gov/search/all/?term=GCA_000180735) using BWA v0.7.16a (Li and Durbin 2009). The coverage of sequencing reads was calculated using SAMtools (Li et al. 2009). CollectInsertSizeMetrics tool in GATK was employed to count the insert size. Distributions of insert size was built by a custom script (statistics_insert_size.pl).

Results

Yield of library prepared using different protocols

The amplified DNA samples were checked using gel electrophoresis. The size of visible DNA fragments was around 500–750 bp (Fig. 1). The yield of original protocol ranged from 7.2 to 8.6 ng/ml, 2LCS ranged from 40.8 to 72.5 ng/ml, and 3LCS, 12.1–43.7 ng/ml (Table 2). In average, the yield of 2LCS protocol (53.4 ± 16.8 ng/ml) was 6.7 times higher than the yield of OP method (8 ± 0.7 ng/ml), whereas the yield of 3LCS (27 ± 15.9 ng/ml) was 3.3 times higher

than that of OP. The library yield of both 2LCS and 3LCS methods were significantly higher than the product using the OP protocol ($p < 0.05$).

Comparison of sequencing results

The proportion of duplicated reads and read coverage in OP and 2LCS were similar, both slightly higher than that of 3LCS (Table 2). A total of 97–98% reads were mapped to the reference in all treatments. The size distribution of fragments produced using OP was between 201 and 400 bp similar to that of 2LCS. However, small fragments, less than 200 bp, accounted for almost 50% in product of 3LCS, much higher than that of OP (12%, Table S1; Fig. 1).

Reduction of cross-contamination, consumption and hands-on time

In our protocol, we replaced most of the bead clean-up steps by heat denaturation, to eliminate the washing steps, so to reduce the chance of cross-contamination as well. We also added inline index within adapter ligation step, which can be used to trace contamination (Wang et al. 2021). In total, the clean-up steps including open lid and pipetting had been lowered six times less in 2LCS than in OP, and ten times less in 3LCS than in OP (Table S2). OP can be used to produce sequencing library at a cost of 70.96 RMB per sample, whereas the total cost of 2LCS and 3LCS are slightly lower, 66.26 RMB and 63.63 RMB, respectively (Table 3). Less clean-up steps in 2LCS and 3LCS saved 40 min and 65 min hands-on time, respectively, comparing to OP (Table S2).

Discussion

In the step of adapter ligation, only half of the input DNA can be attached with validated adapters, that is, P5 in one end and P7 in the other end for the subsequent amplification step (Rohland and Reich 2012). Thus, providing more input DNA is important in library preparation to produce more usable product. Alternatively, reducing the number of clean-up steps should increase the yields of libraries. The 2LCS and 3LCS can be used to reduce the clean-up steps, and significantly increased the products (Table 2). Furthermore, the washing step is skipped in 3LCS, from the very beginning of library preparation step—sample concentration, which means no loss of input DNA in theory (Tables 1, 4). However, the yield of 3LCS was only half comparing to that of 2LCS. This may be due to more DNA loss in 3LCS using concentrator machine. Nonetheless, using our modified protocols can generate higher library yields than using the original protocol due to less washing steps.

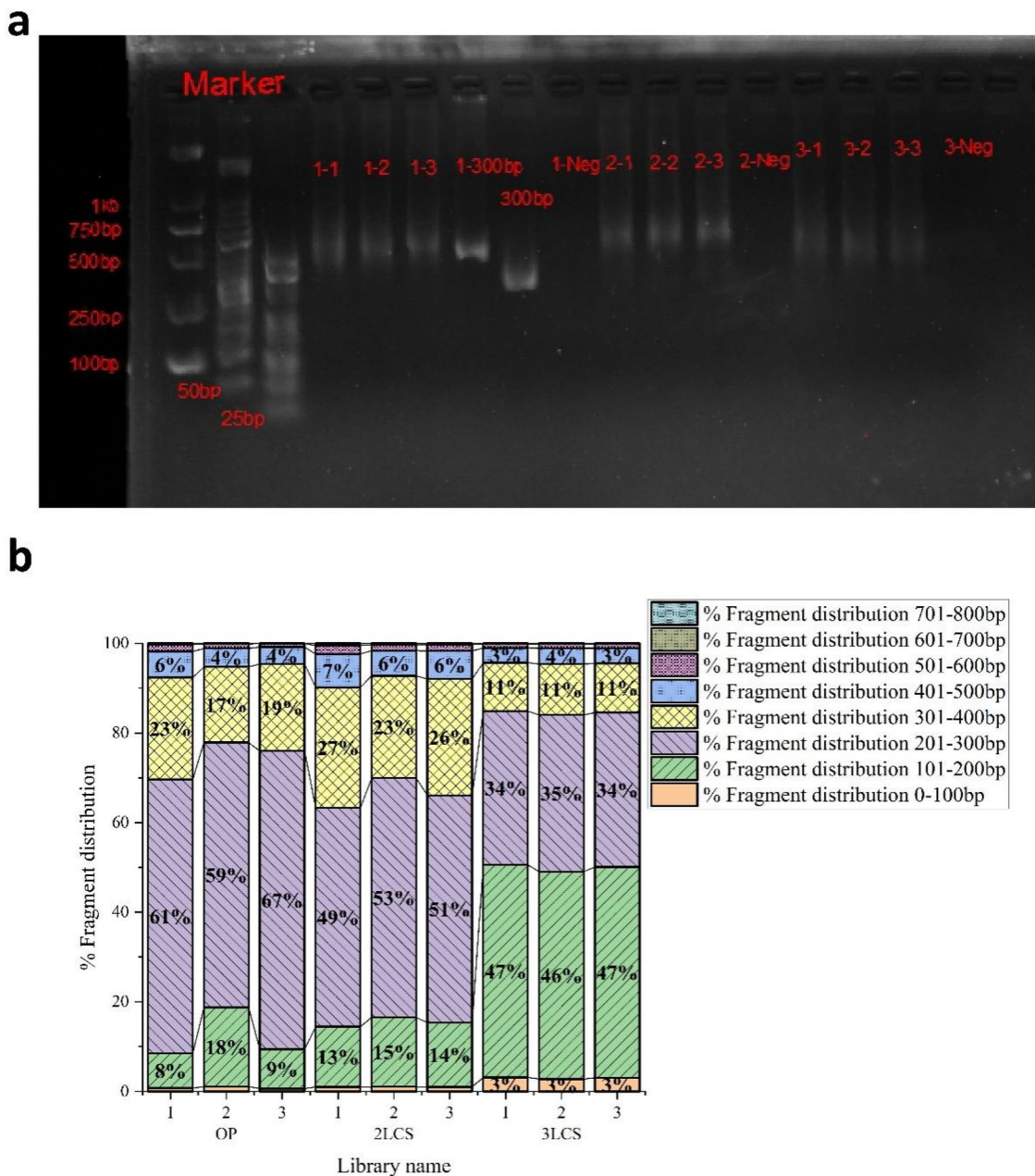


Fig. 1 Fragment distribution. **a** Image of PCR product of DNA libraries made with different protocols run on a 1.5% agarose gel, at 110 V for 60 min in 1×TBE buffer. 1-1 to 1-3 are three replicates of libraries made with the original protocol. 2-1 to 2-3 are three replicates of libraries made with the 2LCS protocol. 3-1 to 3-3 are three replicates of libraries made with the 3LCS protocol. 1–300 bp is the library of

positive control made with the original protocol, whereas 300 bp is the positive control DNA. 1-Neg, 2-Neg and 3-Neg are three replicates libraries of blank control made with three different protocol. **b** Size distribution of the library fragments made with different protocols. The percentages less than 3% are not shown

Our improved protocols also helped in reducing the risk of cross-contamination and operation time. Cross-contamination of experimental origin can potentially arise at multiple benchwork steps, especially when the samples are handled in parallel, such as DNA extraction, library preparation and amplification (Simion et al. 2018). Pipetting and transferring the amplicon carry a risk of cross-contamination

(Seitz et al. 2015). Opening the lid might pollute gloves and then contaminate the sample as well. Ancient DNA or eDNA laboratories often are set up to be less susceptible to contamination from exogenous DNA, and often contain at least three independent rooms (Miya et al. 2020). However, high lab standard also means expensive equipment and higher lab cost. We made modifications on the original protocol

Table 2 Sequencing results from libraries prepared with different protocols: original protocol (OP), 2-less-clean-up-step protocol (2LCS) and 3-less-clean-up-step protocol (3LCS)

Library name	Input DNA (ng)	Con. of lib (ng/ml)	Number of cleaned reads	Duplicated reads (removed) %	Coverage % ^a	Mapped reads % ^b
OP-1	200	7.2 ± 1.0	11,000,000	8.37	83.49	97.49
OP-2	200	8.6 ± 0.4	11,000,000	8.39	83.54	97.59
OP-3	200	8.1 ± 1.2	11,000,000	8.27	83.85	97.26
2LCS-1	200	46.9 ± 4.2	11,000,000	9.55	83.70	97.43
2LCS-2	200	40.8 ± 9.8	11,000,000	8.12	83.52	97.51
2LCS-3	200	72.5 ± 15.3	11,000,000	8.97	83.53	97.48
3LCS-1	200	43.7 ± 4.8	11,000,000	7.82	81.20	98.11
3LCS-2	200	25.1 ± 1.6	11,000,000	7.57	81.12	97.97
3LCS-3	200	12.1 ± 0.5	11,000,000	7.45	81.07	97.79

^aThe percentage of covered bases with depth > 1 in all positions^bThe proportion of reads that mapped to the reference regardless of primary alignment or supplementary alignment**Table 3** Comparison of cost estimation and cost efficiency between original protocol and modified protocols: 2-less-clean-up-step protocol (2LCS) and 3-less-clean-up-step protocol (3LCS)

Steps	Original protocol (¥ ^a)	Modified protocol (¥)	
		2LCS	3LCS
1. Sample concentration	Consumables ^b and MagNA clean-up kit ^c 3.13	Consumables and MagNA clean-up kit 3.13	– 0
2. Blunt-end repair and product clean-up	Consumables, blunt-end repair kit and MagNA clean-up kit 26.21	Consumables and blunt-end repair kit 23.86	23.86
3. Adapter ligation and product clean-up	Consumables, adapter ligation kit and MagNA clean-up kit 19.59	Consumables and adapter ligation kit 17.24	17.24
4. Fill-in and product clean-up	Consumables, fill-in kit and MagNA clean-up kit 12.2	Consumables and fill-in kit 12.2	12.7
5. Amplification and product clean-up	Consumables, amplification kit and MagNA clean-up kit 9.55	Consumables and amplification kit 9.55	9.55
Total cost	70.96	66.26	63.63
Yield (ng)	3.2	21.36	10.8
Cost per yield	22.18	3.1	5.9

^aCost refers to Chinese RMB per sample. The cost of PCR water is negligible^bConsumables: PCR tube and pipette tip^cMagNA clean-up kit: MagNA bead, MagNA buffer and ethanol**Table 4** Procedures comparison between original protocol and modified protocols

Steps	Original protocol	Modified protocol	
		2LCS	3LCS
1. Sample concentration	MagNA bead	MagNA bead	Concentrator
2. Clean-up step in “Blunt-end repair”	MagNA bead	Enzymes deactivation in 80 °C 15 min	Enzymes deactivation in 80 °C 15 min
3. Clean-up step in “Adapter ligation”	MagNA bead	Enzymes deactivation in 70 °C 5 min	Enzymes deactivation in 70 °C 5 min
4. Clean-up step in “Fill-in”	MagNA bead	Enzymes deactivation in 95 °C 2 min	Enzymes deactivation in 95 °C 2 min
5. Clean-up step in “Amplification”	MagNA bead	MagNA bead	MagNA bead

to lower the risk of cross-contamination by reducing pipetting and lid opening. Therefore, doing less washing steps is the most effective way to reduce the probability of contamination. The latest Nextera DNA Flex Library Prep Kit (Illumina, cat: 20018704) also involves less steps, that is, the DNA extraction, fragmentation, library preparation, and library normalization are integrated into one step using the kit (Syed et al. 2009). However, it is too costly for dealing with a large number of samples, and its DNA extraction steps are carried out with bead-based reagents by lysis protocols, only suitable for blood, saliva and dried blood spot. Our modified protocols (2LCS and 3LCS) not only can save 40 min and 65 min compared to OP, respectively, but also can lower six times and ten times risk of cross-contamination (Table S2). The washing step is the main difference in hands-on time and risk of contamination. Therefore, less hands-on time and contamination can be achieved in our modified protocols.

Besides the reduced hands-on time and lowered risk of cross-contamination, the modified protocols also are cost-efficient. We calculated expense of each step of library prep combined with clean-up step in three protocols (Table 3). The cost of cleaning step is low compared to the cost of enzymes and consumables, so the difference of total cost between three protocols is marginal, less than 3 RMB per sample. However, there was significant difference in the product yield of different protocols, which make a significant difference in unit cost. 2LCS and 3LCS only cost 3.1 and 5.9 RMB to produce 1 ng sequencing library, while the cost of OP is 22.18 RMB.

The original protocol with more clean-up steps was designed for optimal reaction condition or was thought so, but more operation would increase the hands-on time and risk of cross-contamination. Our modified protocols with less clean-up steps keep buffer to next reaction steps, but the effect of residual buffer is negligible (Table S3). The high yield of modified protocols suggested that less washing steps did not result in sub-optimal condition for reactions.

The components of residual buffer can be divided into three group: salt solution, enzyme stabilizer and surfactant. Salt solution includes Tris-acetate, Tris-HCl, magnesium acetate, potassium acetate, KCl, MgCl₂, (NH₄)₂SO₄, MgSO₄, EDTA and etc., which can increase enzyme activity in low concentration, while inhibit it in high concentration. Enzyme stabilizers include BSA and DTT. PEG-4000 is an inert polymer, which can stimulate ligation efficiency by the mechanism of macromolecular crowding (Zimmerman and Pheiffer 1983). Glycerol (50% v/v) can be used to prevent protein aliquot from freezing solid. Surfactant, such as Tween 20 and Triton X-100 are non-ionic detergents, which can be used to solubilize proteins.

In the step of adapter ligation, the first step that has carry-over solutes from the previous step, T4 DNA ligase enzyme

catalyzes the ligation of adapter to DNA fragment with a phosphodiester bond. In this process, ATP supplied energy is needed for the reaction from its chemical bond. The concentration of KCl is four times higher than that in the original recipe due to the carryover effect. The concentration of EDTA is two times higher, while other solutes are increased a little (Table S3). Although high-level metal cation, like K⁺ can inhibit the enzyme, but the results showed that the reaction was not affected. The change of salt concentration might be mild as a whole, so no negative effect was seen on the ligation step. Deoxyadenosine triphosphate (dATP) is a competitive inhibitor with ATP (Rossi et al. 1997), but the carryover of dNTP from last step is less than 2×10^{-6} mmol, one order of magnitude lower than ATP used in the ligation step, 3.5×10^{-5} mmol.

In the fill-in step, the second step having carryover solutes from the previous one, the concentration of DTT is increased by two orders of magnitude, up to 200 times higher. Other solutes are increased less than three times higher than the original protocol (Table S3). DTT becomes a less potent reducing agent with high temperature, and powerless to stabilize enzyme. PEG4000 degrades into acid dissolved in solution when heated (Glastrup 1996), so they have little impact interfering the fill-in.

Our preliminary experiment suggested that duration and temperature should be long and high to completely deactivate the enzyme. T4 DNA Polymerase and T4 Polynucleotide Kinase can be used in the same reaction buffer of T4 DNA Ligase (<https://international.neb.com/faqs/0001/01/01/can-t4-dna-polymerase-be-used-in-other-nebuffers>, <https://international.neb.com/faqs/2011/08/13/can-i-use-t4-polynucleotide-kinase-and-t4-dna-ligase-in-the-same-reaction-buffer>). If the T4 Polynucleotide Kinase had not been deactivated completely, it would provide 5' phosphate groups to the adapters and cause the self-ligation of adapters. Ligated adapters would complete the normal library construction steps, which may produce adapter dimers and dominate the sequencing data. Fortunately, all the enzyme including T4 DNA Polymerase, T4 Polynucleotide, T4 DNA ligase and Bsm polymerase can be heat inactivated without affecting other components of the reaction. Finally, doing a clean-up step before the PCR amplification step is necessary, because residual buffer can interfere downstream reaction, particularly for the highly sensitive assay, such as qPCR.

The general principle of library preparation was originally developed for 454 sequencing (Margulies et al. 2005). Subsequently, library preparation protocol was improved for multiple samples on a single run of experiment using “indexing” system to substitute the expensive commercial kits (Meyer and Kircher 2010). Pooling samples prior to target enrichment and using inexpensive paramagnetic beads can generate libraries for thousands of samples and save cost (Rohland and Reich 2012). In addition, novel library

preparation kit and protocol—Nextera are efficient and flexible, but at a high cost (Syed et al. 2009). Our modified protocols can be used to generate high yield and keep short DNA fragments, while save operation time and lab cost.

Conclusion

We replaced the clean-up steps with enzyme deactivation using high temperature in our modified library preparation protocols (2LCS and 3LCS), which can yield higher library product than the original protocol, as well as shorten hands-on time, lower potential risk of cross-contamination and reduce cost of enzymes and consumables. Besides, the 3LCS protocol eliminated the very first washing step, can be used to keep more short DNA fragments, which is important for working with degraded DNA samples, such as ancient DNA or archived samples. We recommend the two modified protocols over the traditional protocol for Illumina library preparation, which could produce better results and be more environment friendly for using less consumables.

Accession numbers

Sequencing data have been submitted to the NCBI under Bioproject number PRJNA776692.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13205-022-03168-5>.

Acknowledgements We thank Ms. Lanxin Ling, Mr. Li Wang for their help with preparing samples for sequencing, Ms. Jingtong Zhou for her help with administering and preparing reagents, and Mr. Wenjun Chen for testing in preliminary experiment.

Author contributions JH and CL conceived the research plan. JH did the experiments, collected and analyzed the data. JH and CL wrote and approved the final version of the manuscript.

Funding This work was supported by “Science and Technology Commission of Shanghai Municipality (19050501900)” and “Shanghai Academy of Environmental Sciences”.

Code availability The custom codes can be found in online version of this article.

Declarations

Conflict of interest The authors declared no financial conflict of interest.

Ethics approval We followed the guidelines approved by the Ethical Committee of Shanghai Ocean University, China.

Consent to participate All authors consent to participate.

Consent for publication All authors consent for publication.

References

- Aigrain L, Gu Y, Quail MA (2016) Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays—a systematic comparison of DNA library preparation kits for Illumina sequencing. *BMC Genom*. <https://doi.org/10.1186/s12864-016-2757-4>
- Deangelis MM, Wang DG, Hawkins TL (1995) Solid-phase reversible immobilization for the isolation of Pcr products. *Nucleic Acids Res* 23:4742–4743. <https://doi.org/10.1093/nar/23.22.4742>
- Glastrup J (1996) Degradation of polyethylene glycol. A study of the reaction mechanism in a model molecule: Tetraethylene glycol. *Polym Degrad Stabil* 52:217–222. [https://doi.org/10.1016/0141-3910\(95\)00225-1](https://doi.org/10.1016/0141-3910(95)00225-1)
- Hess JF, Kohl TA, Kotrova M, Ronsch K, Paprotka T, Mohr V, Hutzenlaub T, Bruggemann M et al (2020) Library preparation for next generation sequencing: a review of automation strategies. *Biotechnol Adv*. <https://doi.org/10.1016/j.biotechadv.2020.107537>
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G et al (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380. <https://doi.org/10.1038/nature03959>
- Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc*. <https://doi.org/10.1101/pdb.prot5448>
- Miya M, Gotoh RO, Sado T (2020) MiFish metabarcoding: a high-throughput approach for simultaneous detection of multiple fish species from environmental DNA and other samples. *Fish Sci* 86:939–970. <https://doi.org/10.1007/s12562-020-01461-x>
- Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res* 22:939–946. <https://doi.org/10.1101/gr.128124.111>
- Rossi R, Montecucco A, Ciarrocchi G, Biamonti G (1997) Functional characterization of the T4 DNA ligase: a new insight into the mechanism of action. *Nucleic Acids Res* 25:2106–2113. <https://doi.org/10.1093/nar/25.11.2106>
- Sato MP, Ogura Y, Nakamura K, Nishida R, Gotoh Y, Hayashi M, Hisatsune J, Sugai M et al (2019) Comparison of the sequencing bias of currently available library preparation kits for Illumina sequencing of bacterial genomes and metagenomes. *DNA Res* 26:391–398. <https://doi.org/10.1093/dnares/dsz017>
- Seitz V, Schaper S, Droge A, Lenze D, Hummel M, Hennig S (2015) A new method to prevent carry-over contaminations in two-step PCR NGS library preparations. *Nucleic Acids Res*. <https://doi.org/10.1093/nar/gkv694>
- Simion P, Belkhir K, Francois C, Veyssier J, Rink JC, Manuel M, Philippe H, Telford MJ (2018) A software tool “CroCo” detects pervasive cross-species contamination in next generation sequencing data. *Bmc Biol*. <https://doi.org/10.1186/s12915-018-0486-7>
- Syed F, Grunenwald H, Caruccio N (2009) Optimized library preparation method for next-generation sequencing. *Nat Methods* 6:I–II. <https://doi.org/10.1038/nmeth.f.269>
- Wang Y, Yuan H, Huang J, Li C (2021) Inline index helped in cleaning up data contamination generated during library preparation

- and the subsequent steps. Mol Biol Rep. <https://doi.org/10.1007/s11033-021-06884-y>
- Wiame I, Remy S, Swennen R, Sagi L (2000) Irreversible heat inactivation of DNase I without RNA degradation. Biotechniques 29:252. <https://doi.org/10.2144/00292bm11>
- Zimmerman SB, Pfeiffer BH (1983) Macromolecular crowding allows blunt-end ligation by DNA ligases from rat liver or *Escherichia coli*. Proc Natl Acad Sci USA 80:5852–5856. <https://doi.org/10.1073/pnas.80.19.5852>