Development and Verification of an Economical Method of Custom Target Library Construction

Xinyao Miao, Bowen Li, Yuesheng Shen, Huiyun Yu, Guoqiang Zhu, Chen Liang, Xiao Fu, Chu Wang, Shengbin Li,* and Bao Zhang*

Cite This: ACS Omega 2020, 5, 13087−13095

ABSTRACT: Although technological advances have greatly reduced the cost of DNA sequencing, sample preparation time and reagent costs remain the limiting factors for many studies. Based on low-cost targeted amplification, we developed an economical method for custom target library construction based on DNA nanoball (DNB) technology and two-step polymerase chain reaction (PCR). Here, we refer to this method as the two-step PCR, which was compared to traditional multiplex PCR methods in three aspects, data quality, efficiency, and specificity to humans. The results confirmed that two-step PCR reduces to finishing 128 sequencing libraries in only 2 h 24 min 59 s of the total PCR time and at a data utilization rate of 0.44 at a cost of approximately $1.70 per sample for targeted sequencing via the two-step PCR. The replacement of traditional multiplex PCR methods with this strategy makes the sample preparation process before sequencing relatively more cost-effective and further reduces the cost of next-generation sequencing (NGS). This method may also be free from the interference of other species and the limitations of sample type and DNA content. These findings reveal possibilities for broad applications of this approach in forensic research.

1. INTRODUCTION

Since the introduction of the Sanger sequencing method for genome sequencing, demands for the transmission of fast, accurate, and inexpensive genomic information have continued to increase.1,2 For this reason, revolutionary next-generation sequencing (NGS) technology has emerged, which has revolutionized almost all areas of biology, agriculture, and medicine and has been widely used to analyze genetic variation.3 The simplicity of the NGS pipeline could still be improved to further decrease the overall cost, although the cost of sequencing per base has decreased 5-fold in the last 10 years.4,36 Prior to sequencing, DNA extraction and library preparation steps are required, which can be laborious and time-consuming processes.5 Compared to whole-genome sequencing and whole-exon sequencing, the method of selectively capturing genomic regions from DNA samples prior to sequencing (targeted NGS) based on polymerase chain reaction (PCR) or hybridization capture can generate smaller and easier-to-manage data and help analyze data sets for the target locus, thereby reducing the difficulty of data analysis and saving time, cost, and effort.6−9 However, the hybridization-capture-based targeted NGS approach has drawbacks such as limited design flexibility, high cost, and protocol complexity, which limit its application to sequencing analyses requiring low cost and high efficiency.6 NGS usually needs to be combined with simplified DNA extraction and library preparation methods to improve efficiency and reduce economic costs.10

The traditional NGS library preparation process consists of three primary steps: fragmentation, adapter ligation, and amplification. The species specificity of PCR mainly depends on the primer specificity. In addition, the sequence similarity of similar species is very high, especially within the same genus, and gene penetration may occur among different species.11,12 Selecting the appropriate gene fragments according to the order of the species to be distinguished should be highly conservative within species and highly variable between species.3 When the number of related species for identification increases and the phenomenon of gene penetration among related species occurs, finding suitable genes and designing specific primers that amplify only one species and accurately distinguish species are difficult. Thus, in this process, the multiplex PCR method has been widely used because the efficiency and species specificity of multiplex PCR are higher than those of traditional PCR.
methods. In forensic science, due to the value of each sample and the high requirements for testing time, scholars have developed direct PCR, which enables PCR to be performed directly from samples such as whole blood, blood cards, and saliva. Based on existing PCR and PCR-derived techniques (normal PCR, multiplex PCR, direct PCR, etc.), targeted sequencing has increasingly been applied. There are a number of commercial solutions for applications in forensic science that have been developed by manufacturers based on high multiplex PCR (high-heavy PCR) library construction, such as Illumina MiSeq FGx, which is the first system developed for the preparation and sequencing of targeted libraries for forensic genomes. MiSeq FGx still has potential shortcomings in some respects: (1) although MiSeq FGx supports the input of a 1.2 mm fluorescent treponemal antibody (FTA) blood card, it requires preprocessing; (2) more than 9 h of library preparation time is required; and (3) for the Chinese market, each sample requires a cost of USD 94.26.15,16

Herein, a simple but robust approach for direct library construction is described, which may increase the efficiency of the NGS pipeline. Thus, we can directly and cost-effectively complete the cumbersome library construction process while taking into consideration accuracy and uniformity. At the same time, the stability of the two-step PCR approach in different species and types of samples is equally important. The above properties made it possible to improve the scope of PCR in the preparation of NGS libraries via this approach.

2. RESULTS AND DISCUSSION

Results. As a direct library construction method of presequencing, the performance of the method was verified in three aspects (data quality, efficiency, species specificity) and compared with traditional library construction methods of presequencing (multiplex PCR library preparation). In addition, DNA extraction and sequencing processes were also included in the comparison.

2.1. Quality of the Sequencing Data. 2.1.1. Loci Detection Rates. We calculated the detection rates of correct loci for the sequencing data generated by our method and the traditional library construction approach. Two identical human samples (six samples) were used in both approaches. The results showed that the average loci detection rates of the two approaches were almost the same, approaching 100% (Table 1). There was no significant difference in the loci detection rate between the two approaches (P > 0.05).

Table 1. Quality of the Sequencing Data from the Two Different Approaches

| approach              | correct loci detection rate (%) | DOC (filtered) | CV (filtered) | uniformity frequency |
|-----------------------|--------------------------------|----------------|---------------|----------------------|
| traditional library construction | 99.84                          | 14 557.31      | 1.217         | 0.726                |
| our approach          | 99.68                          | 9414           | 0.684         | 0.851                |

2.1.2. Depth Uniformity. The uniformity of the sequencing data generated by our approach and by traditional multiplex PCR library construction was measured in different aspects (Table 1): (1) the depths of coverage (DOCs) of the two methods were 14 557.31 and 9414, respectively, and relatively higher DOCs guaranteed the quality of sequencing; (2) the coefficient of variation (CV) values were significantly different in these two approaches (P = 0.0099, P ≤ 0.01); (3) uniformity frequencies were also significantly different between the two groups (P = 0.0277, P ≤ 0.05); and (4) loci depth distribution of the sequencing data obtained by the two methods exhibited some differences. For the two different methods, the slopes of the fitting curve of the loci depth distribution were −0.0024 and −0.0035 (Figure 1).

Figure 1. Loci depth distribution of the two-step and traditional multiplex PCR library construction methods. The abscissa indicates the number of amplicons, and the ordinate indicates the log value of the depth. The depth distribution of the data generated by the two-step PCR approach is represented by a black line, and the gray line represents the data generated by the multiplex PCR approach. The corresponding dashed lines represent different linear fits of the two approaches. The slopes of the two-step PCR method and the method of constructing the library after ordinary PCR are 0.0024 and −0.0035, respectively.

2.2. Cost-Effectiveness. For efficiency reasons, the temporal cost, the economic cost, and data utilization were the primary considerations for the two-step PCR method. The use of two-step PCR not only shortened the time spent on library construction and DNA extraction but also reduced the monetary cost and data utilization.

2.2.1. Temporal Cost. Temporal Cost of PCR

Under the same targeted amplification step, according to the respective protocol, we compared the total PCR time required for the two library preparation methods with all objective influencing factors being excluded, such as laboratory, equipment, operator, and DNA samples. The PCR times for the two-step PCR and the traditional multiplex PCR methods (based on standard BGISEQ-500 library preparation) were 2 h 24 min 59 s and 4 h 7 min 22 s, respectively (Table 2). Two-step PCR is a direct way to construct a target library by targeted and undifferentiated amplification. Traditional library construction did not require undifferentiated amplification but three steps of
library preparation: end repair, adapter ligation, and pre-PCR. The time spent on reagent preparation, purification, and quantification was not included in the statistical scope because of individual differences. However, because traditional library construction requires more steps than the two-step PCR method, the total time required is correspondingly higher.

Temporal Cost of DNA Extraction
DNA extraction, which happens before library preparation, is also a time-consuming step. Therefore, the elimination of the DNA extraction step may be a potential method to improve the efficiency of NGS. We used six dry blood cards (diameter = 3.0 mm) as DNA sources without any preprocessing. NH9719 (Nuhighbio, Suzhou, China) was selected and tested for the PCR buffer system, which can effectively reduce the impact of various PCR inhibitors, such as proteins and urban dust. To reflect the depth of each locus, we showed sequencing alignment data in detail (Figure 2). The average detection rates of correct loci and DOCs were 94.25% and 18,034, respectively. Detailed genotyping results can be found in the Supporting Information 1). Additionally, there were no significant differences in depth distribution between different samples (P > 0.05). The results demonstrated that the two-step PCR method could maintain stability even without DNA extraction.

2.2.2. Monetary Cost. The economic cost of a target library construction method is also not negligible in the efficiency evaluation. We calculated the cost of reagents and primers for library preparation by two methods. The costs of a single reaction of the two-step and traditional library construction methods were $1.70 and $6.15, respectively (Table 3). Therefore, it may be more economical to construct a target library using two-step PCR than using traditional multiplex PCR.

2.2.3. Data Utilization Rate. The utilization rate of sequencing data reflects efficiency. The utilization rate of sequencing data is an indispensable indicator to reflect efficiency. We calculated the data utilization rates in two groups of data generated by the two-step and traditional multiplex PCR library construction approaches with the same samples. As displayed in Figure 3, the average data utilization rates were 0.36 and 0.44, respectively (Figure 3). There was a significant
difference in the data utilization rate of the two different approaches \( (P = 0.0433, P \leq 0.05) \).

Previous studies have shown that uniform sequencing data might reduce data storage and increase data utilization. The addition of 0.2 U of Platinum Pfx DNA polymerase (Pfx) to the second-step PCR may be an important reason for obtaining sufficient depth uniformity. Targeted library preparation was performed using a two-step PCR method for a group of identical samples (six human DNA blood samples). In the enzyme system of the second-step PCR, Pfx-containing and non-Pfx-containing operations were performed (Figure 4). Sequencing data produced from samples supplemented with Pfx had obviously lower CV values (0.67, \( P = 0.02, P \leq 0.05 \)) and higher uniformity frequencies (0.88, \( P = 0.02, P \leq 0.05 \)) than the data produced from samples without Pfx supplementation. Based on this result, the considerable data utilization posed by the two-step PCR method could be easily explained.

2.3. Species Specificity. In forensic research, DNA from other species can affect sequencing data, especially when a library is constructed using multiplex PCR. The two-step PCR method designed in this study aimed to address the problem of DNA contamination from other species.

All primers of the two-step PCR were designed based on the Homo sapiens hg38 genome, which was species specific for human DNA. Fecal DNA of nine primates (gibbons) was taken as test samples. With the same amount of DNA input, the products of the traditional multiplex PCR library were 48.47 ng, and the products of the two-step PCR library were lower than the detection ranges of Qubit 3.0 (Table 4). The PCR amplification efficiency and DOC of the two-step PCR data decreased by 86.04 and 91.74%, respectively, compared with those of the traditional multiplex PCR data. The specific genotyping results generated by traditional multiplex PCR and two-step PCR approaches are discussed in the Supporting Information2 and 3. There were obvious significant differences between these two approaches in PCR amplification efficiency and in DOC \( (P < 0.0001) \). The above results suggested that the use of a two-step PCR method for NGS library construction might increase species specificity.

**DISCUSSION**

Data-quality control is a key step in ensuring successful and meaningful research. However, there is currently no quality control standard for NGS-multiplex PCR data. Based on the Ion Torrent PGM platform, only the proportion of accurate profiles \( (i.e., \) the correct loci detection rate) is generally evaluated.

Therefore, by introducing the corresponding indicators (loci detection rate, DOC, CV, uniformity frequency), we could evaluate the quality of the multiplex PCR data from different perspectives (accuracy and uniformity). The detection rates of correct loci in the sequencing data produced by the two libraries were close to 100\%, which indicated that the accuracy of the two-step PCR method was similar to that of the traditional multiplex PCR library construction. In addition, CV, uniformity frequencies, and the slopes of the fitting curve of the loci depth distribution were all indicators of uniformity. Two-step PCR data had reduced CV values and uniformity frequencies, which indicated reduced dispersion \( (P \leq 0.01, P \leq 0.05) \). In addition, the data produced by two-step PCR had a smaller slope in the locus depth distribution than the data produced by traditional multiplex PCR. The above indicators suggested that the data produced by two-step PCR might be uniform.

The cost-effectiveness of the experimental methodology, data storage, and report generation in NGS needs to be considered. Therefore, the limiting factor for many experiments is the temporal and monetary cost of sample preparation. Sample preparation before sequencing generally consists of two parts: DNA extraction and library construction. For DNA extraction, the operation time of different DNA extraction methods is 1.15–3.15 h, and the average cost per sample is $2.75 to $20.31 (approximately €2.5 to €18.45). Moreover, research on direct amplification is usually limited to the capillary electrophoresis (CE) technology, which is rarely

---

**Table 4. Comparison of the Species Specificity of the Two Library Construction Methods Using Nonhuman DNA (Gibbon)**

| approach       | PCR amplification efficiency (%) | average DOC | average amount of DNA product after PCR |
|----------------|----------------------------------|-------------|----------------------------------------|
| multiplex PCR  | 79.07                            | 1790.47     | 48.47 ng                               |
| two-step PCR   | 11.04                            | 147.83      | below detection range                  |

**Figure 3.** Data utilization rates of the two different methods. Black and gray columns represent the data utilization of the traditional PCR method and the two-step PCR method, respectively. The data utilization rates of the two methods are 36 and 44%, respectively, which are significantly different \( (P \leq 0.05) \).

**Figure 4.** Effect of adding/not adding Pfx DNA polymerase on the uniformity of the sequencing data. The gray and black columns represent the data results of the addition and absence of Pfx DNA polymerase in the second-step PCR system, respectively. The average CV values of the two methods are 1.1 and 0.67 \( (P \leq 0.05) \), respectively. The uniformity frequencies of the two methods are 0.82 and 0.88 \( (P \leq 0.05) \), respectively.
involved in NGS. Although the FGx operating manual claims that it can process blood cards directly, there is still preprocessing, including reagent preparation and multiple centrifugation. The PCR method itself is not too limited by preprocessing, including reagent preparation and multiple centrifugation. Based on the above, we tried to use alkaline buffer in the two-step PCR, which may enable our method to achieve direct amplification using only blood cards. As we expected, the results showed that the DNA extraction process in sample preparation can be omitted under special circumstances (only blood cards as the only source of template DNA). Omittable DNA extraction steps can enable two-step PCR to be applied to complex forensic evidence studies and has considerable cost-effectiveness in terms of time and money.

For library construction, compared with the traditional multiplex PCR method based on the BGISEQ-500 platform, the two-step PCR method saved 41.39% of the temporal cost and 72.36% of the monetary cost. According to our market research, Illumina’s ForenSeq DNA Signature Kit has a high market price of $94.26 per reaction and requires more than 9 h of hands-on time. In addition to our two-step PCR approach, there are many PCR-based barcoding procedures that use large fusion primers for library construction. However, with the second PCR in our approach, which is undifferentiated amplification, barcode primers can be used repeatedly for barcoding diverse amplicons of interest; thus, costly investment into individual barcoded primer sets for each target gene is not required. As a result, the burden of library construction can be alleviated.

BGISEQ-500 is based on the DNA nanoball (DNB) sequencing technology. DNBs were generated from the single-stranded DNA (ssDNA) circle using rolling circle amplification (RCA) to enhance the fluorescent signals in the sequencing process. Unlike PCR index amplification, RCA amplification errors do not accumulate and there is no PCR bias because the original ssDNA circle is the only template during the entire amplification process. To some extent, BGISEQ-500 is more suitable for the sequencing of targeted PCR libraries.

The quality, capacity, and storage of NGS sequencing data also affect cost-effectiveness. Compared with traditional methods, two-step PCR improved data utilization by 18.18%. In addition, most of the technology patents for multiplex PCR require Pfx DNA polymerase. In our previous tests, Pfx (compared to NH9007) was not suitable for multiplex PCR; therefore, we chose to add Pfx in the second-step PCR instead of the first-step PCR. More uniform sequencing data suggested that the improved two-step PCR may be more cost-effective than the traditional multiplex PCR method.

In general, a novel library construction method for target sequencing based on multiplex PCR was reported, which was first changed by simplifying the steps for library preparation. After simplification, only two PCR steps were required, and the DNA-processing step could be skipped simultaneously. These steps make the method more cost-effective in terms of reagents and time consumption. Significantly reduced sequencing costs stem from the determination of sequencing directions through the introduction of universal sequences. Other changes were observed in the enzyme system of the second-step PCR. Due to high fidelity, the addition of Pfx improved the uniformity in depth by a certain degree, which may be the main reason for the increase in data utilization for cost-effectiveness. The introduction of phosphate groups and the maintenance of an alkaline environment in the reactions for DNB technology may give the two-step PCR method an advantage in accuracy compared to other sequencing platforms.

The species specificity was studied to testify that other biological sources did not interfere with the ability to obtain reliable results on samples recovered from crime scenes, which provides an important support for a robust approach. In a developmental validation study of the MiSeq FGx forensic genomics system, species specificity tests were carried out for rhesus monkeys and baboons, and the loci detection rates were 11 and 25%, respectively. However, rhesus monkeys and baboons belong to the macaque family, which is more distantly related to humans than the gibbon samples we used. In our study, the PCR amplification efficiency of two-step PCR using gibbon samples (more similar to Homo sapiens) was 11.04%, which indicated the species specificity of this method. Thus, the two-step PCR method might maintain the stability of sequencing data in the presence of primate DNA contamination.

3. CONCLUSIONS

Based on the increasing demand for cost reduction in NGS technologies, we developed and evaluated an approach for economical library construction. The results showed that the two-step PCR method designed in this study can solve the problems of common multiplex PCR methods, such as low time efficiency, high monetary expense, and data utilization. In addition, the use of two-step PCR could also eliminate pollution sources from nonhuman samples to the greatest extent. Admittedly, this method could be improved in several aspects: (1) we need to design a two-step PCR system based on other NGS platforms; (2) we will continue to work to reduce the time and economic cost required for targeted library construction; and (3) we will also design two-step PCR systems for different species.

4. EXPERIMENTAL SECTION

4.1. Target Loci Selection and Primer Design. The sequences of 188 and 128 target regions containing single nucleotide polymorphism (SNP) and short tandem repeat (STR) loci were downloaded from the National Center for Biotechnology Information dbSNP database and STRBase (https://strbase.nist.gov), respectively. In addition to conventional requirements, the two-step PCR primers met the following requirements: (1) the length was between 25 and 30 bp; (2) the $T_m$ value was between 58 and 64 °C; and (3) the GC content was <65%.

We tested 33 samples: 24 human samples and 9 gibbon stool samples. The kits used for DNA extraction were the Genomic DNA Extraction Kit and the Stool DNA extraction Kit (Tiangen, Beijing, China). Notably, although our two-step PCR approach was designed based on a BGISEQ-500 sequencer (BGI, Shenzhen, China), the principle of this approach could be applied to each sequencing platform. The protocol for the two-step PCR approach is displayed in Figure 5.

4.2. First-Step PCR (Targeted Amplification). In the first-step PCR of our approach, two different universal sequences (20 bp) (“ACATGGCTACGATCCGACTTT” and “GACCCTTGGCCTCCGACTTT”) were added to the 5′ ends of the specific forward and reverse primers.

The design of the universal sequences determined the direction of sequencing. Single-direction sequencing reduces economic costs and improves data utilization. The sequencing
direction of this method was determined by universal sequence 1. Forward primers were the starting point of sequencing.

In Figure 5, for the first-step PCR, target loci were amplified in 25 μL PCRs, which contained 2.5 μL of the primer mix, 2–4 U of the NH9007 DNA polymerase, 10 μL of the 1× PCR buffer (Nuhighbio, Suzhou, China), and input DNA (78 pg to 10 ng). To ensure the fidelity of PCR and the uniformity of the subsequent step of DNA cyclization. Therefore, the first base at the S’ end of adapter 1 was phosphorylated.

The second-step PCR system contained a purified product after the first-step PCR, 3 μL of the Adapter Mix, 10 μL of the 1× PCR buffer (Nuhighbio, Suzhou, China), 1 U of NH9007 the DNA polymerase (Nuhighbio, Suzhou, China), and 0.2 U of the Platinum Pfx DNA polymerase (ThermoFisher Scientific, MA). The reaction conditions of the second-step PCR designed here are shown in Table 6.

The second-step PCR was differentiated Ampli
ty of PCR and the uniformity of the

![Diagram](https://example.com/diagram.png)

Figure 5. Procedure for the two-step PCR approach. The two-step PCR approach includes the addition of DNA or blood cards, first-step PCR, product purification twice, second-step PCR, product purification, and quantitation. If the total amount of PCR products quantified was higher than 20 ng, the next step, named the fixed ratio hybrid library, was performed until the final sequencing and bioinformatics analyses steps; if not, the experiment was restarted from the first-step PCR step.

### Table 5. Details of the First-Step PCR Conditions

| temperature (°C) | time | number of cycles |
|------------------|------|------------------|
| 95               | 15 min | 9*               |
| 95               | 20 s   | 9*               |
| 60               | 2 min 30 s | 1                |
| 72               | 1 min  | 1                |
| 95               | 20 s   | 1                |
| 62.5             | 2 min 30 s | 1                |
| 72               | 1 min  | 1                |
| 95               | 20 s   | 8*               |
| 60               | 2 min 30 s | 8*              |
| 72               | 1 min  | 8                |
| 72               | 10 min | 8                |
| 4                | ∞      | 8                |

*It is appropriate to increase the number of cycles by 1–2 cycles at *.

To protect the S’-terminal phosphoric acid, the buffer of the second-step PCR was alkaline (pH > 8.0), which is NH9719. In addition, we replaced nuclease-free water with the Tris–EDTA (TE) buffer (ThermoFisher Scientific, MA), which maintained alkalinity.

### Table 6. Details of the Second-Step PCR Conditions

| temperature (°C) | time | number of cycles |
|------------------|------|------------------|
| 95               | 10 min |                 |
| 95               | 20 s   | 8                |
| 58               | 30 s   |                 |
| 72               | 1 min  |                 |
| 4                | ∞      |                 |

To protect the S’-terminal phosphoric acid, the buffer of the second-step PCR was alkaline (pH > 8.0), which is NH9719. In addition, we replaced nuclease-free water with the Tris–EDTA (TE) buffer (ThermoFisher Scientific, MA), which maintained alkalinity.

### 4.3. Purification

The PCR products generated by the first-step PCR were purified twice with Agencourt AMPure XP beads (Beckman Coulter, CA) at 1.5 times the product volume.

### 4.4. Second-Step PCR (Undifferentiated Amplification)

Our second-step PCR was undifferentiated amplification, and the ends of the amplified products were the adapter sequences of BGISEQ-500, wherein the sequences of adapter 1 and adapter 2 were “GAACGACATGGCTACGATCCGACTTTTTCCTAAGACCCTTGGCCCTCCGACTT” and “TGTGAGCCAAGGAGTTGGTATTGTC-TTCTAAGACCCTTGGCCCTCCGACTT”, respectively. ** is the barcode sequence of BGISEQ to distinguish between different samples. Barcode details of the BGISEQ sequencer can be viewed in the Supporting Information 4, with a total of 128 different barcode sequences (approximately 10 bp) used. Based on the DNB technology, a single strand of DNA containing hydroxyl and phosphoric acid is required for the subsequent step of DNA cyclization. Therefore, the first base at the S’ end of adapter 1 was phosphorylated.

### 4.5. Purification and Quantitation

PCR products generated from second-step PCR were purified once with 1.5× Agencourt AMPure XP beads. Then, the concentration of the purified products was determined by Qubit (Applied Biosystems, MA). When the total amount of the final product was less than 20 ng, it was implied that the experiment failed (Figure 5).

### 4.6. Sequencing and Bioinformatics Analyses

After quantification of the products, we mixed them proportionally. PCR products were cyclized to form ssDNA, and then, the single-stranded cyclic DNA was amplified by 2–3 orders of magnitude using RCA. Amplified products were called DNBs. Finally, the DNBs were fixed on an arrayed silicon chip by the DNB loading technology.
For the sequencing strategy, we chose paired-end 50 (PE50) and single-end 200 (SE200). Raw sequencing data were processed with GATK-BWA. The subsequent data analysis was conducted using statistical software R (version 3.2.1), and a series of proprietary bioinformatics procedures based on Perl (https://www.perl.org). Then, a threshold of 50 was applied for the depth of loci, and loci below the threshold were considered as nondetected loci in the subsequent statistical analyses.

4.7. Quality of the Sequencing Data. Two different presequencing methods were used in the same group of human DNA samples (Figure 6). We calculated the following parameters that were used to evaluate sequencing quality. (1) Locus detection rate: It is the ratio of the number of typing loci within each locus or sample.40 We calculated the DOC of a single sample or a single group of samples. However, for human dry blood cards, direct amplification was performed. From the data, we calculated the depth of a single locus. (3) Coefficient of variation (CV): CV, which is a unit-free and effective normalized measure of dispersion and monitoring the CV, is a normal approach in statistical process control. The CV was expressed as the ratio of the standard deviation (σ) to the mean (μ), i.e., \( y = \sigma / \mu \).41 In this calculation, we removed the maximum and minimum values. (4) Uniformity frequency: We artificially set the percentage of loci in the range of 90% above and below the median as the uniformity frequency. (5) Loci depth distribution: We used the number of amplicons as the abscissa and the log median as the uniformity frequency. (5) Loci depth distribution: We used the number of amplicons as the abscissa and the log median as the uniformity frequency.

4.8. Cost-Effectiveness of Two-Step PCR. 4.8.1. Temporal Cost.

- We compared the amount of time required for PCR for the two approaches before sequencing. One approach was our two-step PCR, and the other was traditional multiplex PCR combined with BGISEQ-500 library construction. For fairness in the comparison of time consumption, all objective influencing factors were excluded, such as laboratory, equipment, operator, and DNA samples. The PCR time is based on the display time of the same model of the instrument (ABI VERITI Gradient PCR Instrument).

- Since DNA extraction takes a long time and requires manual manipulation, we explored the omission of DNA extraction steps using blood cards as DNA samples. We selected six human blood cards, including four males and two females, with a blood card diameter of 3.0 mm.

4.8.2. Monetary Cost. We calculated the cost of the two different methods using the Chinese manufacturer’s reagents. The costs mentioned here did not include the cost of pretesting, consumables, and instrument depreciation.

4.8.3. Data Utilization. First, the two methods were used to test the same batch of human DNA samples (six samples each). The difference in data utilization between the two groups was calculated by the t-test. Based on the filtered data, which had a Phred-like consensus quality \( \geq 20 \), \( 35 \) data utilization was calculated as follows: total number of reads at all target loci/total number of reads of filtered data.

Additionally, the first-step PCR products (purified) of the same two groups of human blood DNA samples (three males and three females) were put in the two different enzyme systems of second-step PCR, which were 1 U of NH9007 DNA polymerase and 1 U of NH9007 + 0.3 U Platinum Pfx DNA polymerase. Then, the uniformity indicators of the two groups of sequencing data were calculated and analyzed.

4.9. Species Specificity. To avoid the effects of DNA incorporation by other species on the sequencing results, we tested nine gibbon stool samples (primates). All DNA starting amounts were the same, and all library products were used for sequencing. We compared the data quality and amplification efficiency of the two groups of data after sequencing.

**ASSOCIATED CONTENT**

**Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.0c01014.

Supporting Information of this article is presented in ESM 1—4 separately for clarity and space limitations (ZIP)
AUTHOR INFORMATION

Corresponding Authors
Shengbin Li — School of Forensic Sciences, Xi’an Jiaotong University, 710049 Xi’an, P. R. China; orcid.org/0000-0002-4093-679X; Email: shengbinli_xju@126.com
Bao Zhang — School of Forensic Sciences, Xi’an Jiaotong University, 710049 Xi’an, P. R. China; Email: zhangbao_814@mail.xjtu.edu.cn

Authors
Xinyao Miao — School of Forensic Sciences, Xi’an Jiaotong University, 710049 Xi’an, P. R. China
Bowen Li — School of Life Sciences, Sichuan University, 610207 Chengdu, P. R. China
Yuesheng Shen — School of Life Sciences, Northwest University, 710069 Xi’an, P. R. China
Huiyun Yu — School of Life Sciences, Northwest A & F University, 712100 Yangling, P. R. China
Guoqiang Zhu — Key Laboratory of Bio-Resources and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, 610065 Chengdu, P. R. China
Chen Liang — School of Mechanical Engineering, Xi’an Jiaotong University, 710049 Xi’an, P. R. China
Xiao Fu — The Beijing Genomics Institute (BGI)—Tianjin, 301700 Tianjin, P. R. China
Chu Wang — School of Life Sciences, Xiamen Medical College, 361023 Xiamen, P. R. China

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.0c01014

Author Contributions
X.M., B.L., Y.S., and H.Y. contributed equally to this work. The manuscript was written with contributions from all authors. All authors have given approval to the final version of the manuscript.

Funding
This study was supported by China Postdoctoral Science Foundation (Grant No. 2018M633526).

Notes
The authors declare no competing financial interest. All works were approved by the Institutional Review Board-Environment of Ministry of Education, College of Life Sciences, Sichuan University—Ethical Review and obtained informed consent of the participants.

ACKNOWLEDGMENTS

The achievement of this study was inseparable from the efforts of volunteers and laboratory staff.

ABBREVIATIONS USED

NGS next-generation sequencing
PCR polymerase chain reaction
CV coefficient of variation
CE capillary electrophoresis
STR short tandem repeat
SNP single nucleotide polymorphism
DOC depth of coverage
T_m melting temperature
RCA rolling circle amplification
DNB DNA nanoball
PE50 pair-end 50

REFERENCES

(1) Metzker, M. L. Sequencing technologies—the next generation. Nat. Rev. Genet. 2010, 11, 31−46.
(2) Butler, J.; Buel, E.; Crivellente, F.; McCord, B. R. Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. Electrophoresis 2004, 25, 1397−1412.
(3) Feng, K.; Costa, J.; Edwards, J. S. Next-generation sequencing library construction on a surface. BMC Genomics 2018, 19, 416.
(4) Head, S. R.; Komori, H. K.; LaMere, S. A.; Whisenant, T.; Van Nieuwkergh, F.; Salomon, D. R.; Ordoukhanian, P. Library construction for next-generation sequencing: Overviews and challenges. Biotechniques 2014, 56, 61.
(5) Quail, M. A.; Kozarewa, I.; Smith, F.; Scally, A.; Stephens, P. J.; Durbin, R.; et al. A large genome center’s improvements to the Illumina sequencing system. Nat. Methods 2008, 5, 1005−1010.
(6) Nguyen-Dumont, T.; Pope, B. J.; Hammet, F.; Southey, M. C.; Park, D. J. A high-plex PCR approach for massively parallel sequencing. Biotechniques 2013, 55, 69−74.
(7) Meldrum, C.; Doyle, M. A.; Tothill, R. W. Next-generation sequencing for cancer diagnostics: a practical perspective. Clin. Biochem. Rev. 2011, 32, 177−195.
(8) Moorhthie, S.; Mattocks, C. J.; Wright, C. F. Review of massively parallel DNA sequencing technologies. HUGO J. 2011, 5, 1−12.
(9) Mamanova, L.; Coffey, A. J.; Scott, C. E.; Kozarewa, I.; Turner, E. H.; Kumar, A.; Howard, E.; Shendure, J.; Turner, D. J. Target-enrichment strategies for next-generation sequencing. Nat. Methods 2010, 7, 111−118.
(10) Lee, E. Y.; Lee, H. Y.; Oh, S. Y.; Jung, S.-E.; Yang, I. S.; Lee, Y.-H.; et al. Massively parallel sequencing of the entire control region and targeted coding region SNPs of degraded mtDNA using a simplified library preparation method. Forensic Sci. Int.: Genet. 2016, 22, 37−43.
(11) Erlich, H.; Gelfand, D.; Sninsky, J. Recent advances in the polymerase chain reaction. Science 1991, 252, 1643−1651.
(12) Markoulatos, P.; Siafakas, N.; Moncay, M. Multiplex polymerase chain reaction: A practical approach. J. Clin. Lab. Anal. 2002, 16, 47−51.
(13) Fattorini, P.; Ciofuli, R.; Cossutta, F.; Giuliani, P.; Edomi, P.; Furlanut, M.; Previdere, M. Fidelity of polymerase chain reaction-direct sequencing analysis of damaged forensic samples. Electrophoresis 1999, 20, 3349−3357.
(14) Park, S. J.; Kim, J. Y.; Yang, Y. G.; Lee, S. H. Direct STR Amplification from Whole Blood and Blood- or Saliva-Spotted FTA without DNA Purification. J. Forensic Sci. 2008, 53, 335−341.
(15) Guo, F.; Yu, J.; Zhang, L.; Li, J. Massively parallel sequencing of forensic STRs and SNPs using the Illumina ForenSeq DNA Signature Prep Kit on the MiSeq FGX Forensic Genomics System. Forensic Sci. Int.: Genet. 2017, 31, 135−148.
(16) MiSeq FGX Applications and MethodsDesigned for Forensic Genomics, 2019. https://www.illumina.com/systems/sequencing-platforms/miseq-fgx/applications.html.
(17) Desai, A.; Marwah, V. S.; Yadav, A.; Jha, V.; Dhaygude, K.; Bangar, U.; et al. Identification of Optimum Sequencing Depth Especially for De Novo Genome Assembly of Small Genomes Using Next Generation Sequencing Data. PLoS One 2013, 8, No. e60204.
(18) Guo, Y.; Ye, F.; Sheng, Q.; Clark, T.; Samuels, D. C. Three-stage quality control strategies for DNA re-sequencing data. Briefings Bioinf. 2013, 15, 879−889.
(19) Gettings, K. B.; Kiesler, K. M.; Vallone, P. M. Performance of a next generation sequencing SNP assay on degraded DNA. Forensic Sci. Int.: Genet. 2015, 19, 1−9.
(20) Guo, F.; Zhou, Y.; Liu, F.; Yu, J.; Song, H.; Shen, H.; et al. Evaluation of the Early Access STR Kit v1 on the IonTorrent PGM platform. Forensic Sci. Int.: Genet. 2016, 23, 111−120.
(21) Bersting, C.; Fordye, S. L.; Olofsson, J.; Mogensen, H. S.; Morling, N. Evaluation of the Ion Torrent HID SNP 169-plex: A SNP
typing assay developed for human identification by second generation sequencing. Forensic Sci. Int.: Genet. 2014, 12, 144–154.

(22) Weber, E. U.; Shafir, S.; Blais, A.-R. Predicting Risk Sensitivity in Humans and Lower Animals: Risk as Variance or Coefficient of Variation. Psychol. Rev. 2004, 111, 430–445.

(23) Wordsworth, S.; Doble, B.; Payne, K.; Buchanan, J.; Marshall, D.; McCabe, C.; Begier, D. Using “Big Data” in the Cost-Effectiveness Analysis of Next-Generation Sequencing Technologies: Challenges and Potential Solutions. Value Health 2018, 21, 1048–1053.

(24) Next Generation DNA Sequencing: A Review of the Cost Effectiveness and Guidelines; PubMedHealth, 2014.

(25) Rohland, N.; Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. 2012, 22, 939–946.

(26) Psilidi, A.; Dovas, C. I.; Bramis, G.; Lazou, T.; Russel, C. L.; Arsenos, G.; Banos, G. Comparison of Eleven Methods for Genomic DNA Extraction Suitable for Large-Scale Whole-Genome Genotyping and Long-Term DNA Banking Using Blood Samples. PLoS One 2015, 10, e0115960.

(27) Gouveia, N.; Brito, P.; Serra, A.; Balsa, F.; Andrade, L.; Bento, M.; et al. Direct amplification of reference samples with GlobalFliper PCR Amplification Kit. Forensic Sci. Int.: Genet. Suppl. Ser. 2015, S, e135–e137.

(28) Uribe-Convers, S.; Settles, M. L.; Tank, D. C.; Berthold, H. A phylogenomic approach based on PCR target enrichment and high throughput sequencing: resolving the diversity within the south American species of bartsia l. (orobanchaceae). PLoS One 2016, 11, No. e0148203.

(29) Kozarewa, I.; Ning, Z.; Quail, M. A.; Sanders, M. J.; Berriman, M.; Turner, D. J. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat. Methods 2009, 6, 291–295.

(30) Li, Q.; Zhao, X.; Zhang, W.; Wang, L.; Jiang, Y. Reliable multiplex sequencing with rare index mis-assignment on dnb-based NGS platform. BMC Genomics 2019, 20, No. 215.

(31) Guo, J.; Rongrong, G. U. O.; Meiyan, L. I.; Chunyu, G. E. N. G.; Jiang, H. US15/537,3962018.

(32) Wang, Y.; Peng, Q. US15/544,7642018.

(33) Jäger, A. C.; Alvarez, M. L.; Davis, C. P.; Guzmán, E.; Han, Y.; Way, L.; et al. Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. Forensic Sci. Int.: Genet. 2017, 28, 52–70.

(34) Reis, M. D.; Gunnell, G. F.; Barba-Montoya, J.; Wilkins, A.; Yang, Z.; Yoder, A. D. Using Phylogenomic Data to Explore the Effects of Relaxed Clocks and Calibration Strategies on Divergence Time Estimation: Primates as a Test Case. Syst. Biol. 2018, 67, 594–615.

(35) Malhis, N.; Jones, S. J. M. High quality SNP calling using Illumina data at shallow coverage. Bioinformatics 2010, 26, 1029–1035.

(36) Marino, P.; Touzani, R.; Perrier, L.; Rouleau, E.; Kossi, D. S.; et al. Cost of cancer diagnosis using next-generation sequencing targeted gene panels in routine practice: a nationwide French study. Eur. J. Hum. Genet. 2018, 26, 314–323.

(37) Dissing, J.; Rudbeck, L.; Marcher, H. In A Five Minute Procedure for Extraction of Genomic DNA from Whole Blood, semen and Forensic Stains for PCR, 16th Congress of the International Society for Forensic Haemogenetics (Internationale Gesellschaft für forensische Hämogenetik eV), Santiago de Compostela, Sep 12–16, 1995 Springer: Berlin, Heidelberg, 1996; pp 269–271.

(38) Butler, J. M. Advanced Topics in Forensic DNA Typing: Methodology; Academic Press, 2011; pp 569–585.

(39) Drmanac, R.; Sparks, A. B.; Callow, M. J.; Halpern, A. L.; Burns, N. L.; Kermani, B. G.; Carnevali, P.; Nazarenko, I.; Nilsen, G. B.; Yeung, G.; et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 2010, 327, 78–81.

(40) Tao, R.; Qi, W.; Chen, C.; Zhang, J.; Yang, Z.; Song, W.; et al. Pilot study for forensic evaluations of the Precision ID GlobalFiler NGS STR Panel v2 with the Ion SS system. Forensic Sci. Int.: Genet. 2019, 43, No. 102147.

(41) Teoh, W. L.; Khoo, M. B. C.; Castagliola, P.; Yeong, W. C.; Teh, S. Y. Run-sum control charts for monitoring the coefficient of variation. Eur. J. Oper. Res. 2017, 257, 144–158.

(42) Kozich, J. J.; Westcott, S. L.; Baxter, N. T.; Highlander, S. K.; Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. 2013, 79, 5112–5120.

(43) Fadross, D. W.; Ma, B.; Gajer, P.; Sengamalay, N.; Ott, S.; Brotman, R. M.; et al. An improved dual-indexing approach for multiplexed 16s rRNA gene sequencing on the Illumina MiSeq platform. Microbiome 2014, 2, 6.

(44) Herbold, C. W.; Pelikan, C.; Kuzyk, O.; Hausmann, B.; Angel, R.; Berry, D.; Loy, A. A flexible and economical barcoding approach for highly multiplexed amplicon sequencing of diverse target genes. Front. Microbiol. 2015, 6, No. 731.

(45) Xu, Y.; Lin, Z.; Tang, C.; Tang, Y.; Cai, Y.; Zhong, H.; et al. A new massively parallel nanoball sequencing platform for whole exome research. BMC Bioinf. 2019, 20, No. 153.

(46) Catanesi, G.; Catanesi, G.; Infante, C.; Manchado, M.; Catanesi, G.; Infante, C.; Manchado, M. Complete mitochondrial DNA sequences of the frigate tunaAuxis thazard and the bullet tunaAuxis rochei. DNA Sequence 2008, 19, 159–166.

(47) Alvarado Bremer, J. R.; Viñas, J.; Mejuto, J.; Ely, B.; Pla, C. Comparative phylogeography of Atlantic bluefin tuna and swordfish: the combined effects of vicariance, secondary contact, introgression, and population expansion on the regional phylogenies of two highly migratory pelagic fishes. Mol. Phylogenet. Evol. 2005, 36, 169–187.

(48) Teletchea, F.; Maudet, C.; Hänni, C. Food and forensic molecular identification: update and challenges. Trends Biotechnol. 2005, 23, 359–366.