Optimization of Illumina® Nextera® XT Library Preparation for Analysis of Complete Mitochondrial Genomes from Human Reference Samples

Viktorija Sukser

Forensic Science Centre Ivan Vucetic: Centar za forenzična ispitivanja istraživanja i vještacenja Ivan Vucetic

Ivana Račić

Forensic Science Centre Ivan Vucetic: Centar za forenzična ispitivanja istraživanja i vještacenja Ivan Vucetic

Sara Rožić

Forensic Science Centre Ivan Vucetic: Centar za forenzična ispitivanja istraživanja i vještacenja Ivan Vucetic

Lucija Barbarić

Forensic Science Centre Ivan Vucetic: Centar za forenzična ispitivanja istraživanja i vještacenja Ivan Vucetic

Marina Korolija (✉ mkorolija@mup.hr)

Forensic Science Centre Ivan Vucetic: Centar za forenzična ispitivanja istraživanja i vještacenja Ivan Vucetic <https://orcid.org/0000-0003-4892-0204>

Short Report

Keywords: MiSeq, mitochondrial DNA, Nextera XT, optimization, small genome sequencing

Posted Date: October 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1004103/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Optimized and efficient library preparation workflow is one of the most important prerequisites for obtaining high quality and quantity of results in massively parallel sequencing (MPS). Our aim was to assess and optimize different steps of Illumina® Nextera® XT assay for analysis of whole mitochondrial genomes.

Methods and Results: Among the three long-range high-fidelity DNA polymerases tested here, PrimeSTAR® GXL performed best in aspects of specificity and yield for mitochondrial DNA (mtDNA) enrichment. Furthermore, library quantification combined with individual library-by-library dilution outperformed bead-based normalization in terms of more equal distribution of reads per library, reduced hands-on time and simplified workflow. Increasing the number of amplification cycles in the index-adapters-adding PCR step had no adverse effect on the level of sequencing noise, which remained low both in negative controls and in samples.

Conclusions: Optimizations described herein provide beneficial insights for laboratories aiming at implementation and/or advancement of similar MPS workflows (e.g. small genomes, PCR amplicons and plasmids).

1. Introduction

Library preparation for massively parallel sequencing (MPS) is a sensitive process, usually consisting of numerous steps. To ensure high quality of sequencing results, it is important to optimize library prep, or rather tailor it to the characteristics of a particular target molecule. There are several critical steps, such as initial enrichment of the target molecule (achieved, for example, by long-range PCR), the PCR step for adding index-adaptor oligonucleotides and library amplification, and library normalization method prior to pooling for sequencing. Our aim was to test and optimize the following aspects of Illumina® assay for whole mitochondrial DNA (mtDNA) analysis on MiSeq® instrument [1]: target enrichment by long-range PCR, limited-cycle PCR, and library normalization methods. Various optimizations and evaluations have been previously published [2-6], but, to our knowledge, none assessed the impact of varying the cycling conditions of the index-adapters-adding PCR step (i.e. limited-cycle PCR) onto sequencing results. The latter, however, was a necessity since we observed that the recommended cycling conditions produced libraries of low quantities, thus risking low read depth of mtDNA genomes and potential loss of sequencing information. Therefore, it was essential to ensure that increasing the number of cycles in the step of limited-cycle PCR would not adversely affect the sequencing results.

2. Materials And Methods

Samples were collected from volunteers who gave detailed informed consent. From each person, two types of samples were collected: buccal swabs (collected on Whatman™ Sterile Omniswab, GE

Healthcare, UK) and blood (collected on Whatman™ FTA™ Classic Cards, GE Healthcare, UK). DNA was extracted from buccal swabs using the EZ1® DNA Investigator® Kit on EZ1® Advanced XL instrument (Qiagen, Hilden, Germany), and from dried blood using QIAamp® DNA Micro Kit (Qiagen), according to the manufacturer's instructions for both kits. Extracts were quantified on Qubit™ 3.0 Fluorometer using Qubit™ dsDNA High Sensitivity kit (Thermo Fisher Scientific, Waltham, MA, USA).

In Illumina® whole mtDNA assay [1] target molecules of mtDNA are initially amplified in two large fragments of sizes 9.1 kb and 11.2 kb (Fig. S1), which is achieved by employing long-range PCR. Three long-range, high-fidelity DNA polymerases were tested for target enrichment step: Platinum™ PCR SuperMix High Fidelity (Thermo Fisher Scientific), LA Taq® Hot Start (TaKaRa, Kusatsu, Japan) and PrimeSTAR® GXL (TaKaRa). Layout of polymerase testing experiments is given in Table S1. Since PrimeSTAR® GXL exhibited better performance of the three, thermal cycling conditions were optimized further only for this polymerase, for both mtDNA amplicons (9.1 kb and 11.2 kb) and for both sample types (buccal epithelia and blood). PCR products were evaluated via agarose gel electrophoresis at every step (Table S2), and quantified with Qubit™ dsDNA High Sensitivity kit.

After establishing optimal thermal cycling conditions for the chosen DNA polymerase, these were applied subsequently to all samples. For input into library prep, 9.1 kb and 11.2 kb mtDNA amplicons were first normalized to 0.2 ng/μL, then equal volumes of each amplicon were pooled for each sample, and 5 μL of pooled amplicons (1 ng of DNA in total) was input into library prep. Negative controls were introduced at three stages of workflow: DNA extraction (NC-EX), long-range PCR mtDNA enrichment (NC-PCR), and library preparation (NC-LIB). Libraries were prepared using Nextera® XT Library Prep Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions [1], until the step of index-adapters-adding PCR (termed "limited-cycle PCR" in the manufacturer's protocol), where we tested the impact of increasing the number of amplification cycles from 12 (as per protocol) to 15 cycles. Libraries were purified with Agencourt AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA), whereupon followed either bead-based normalization (reagents provided in the library prep kit), or quantification of libraries using LabChip® DNA High Sensitivity Assay on LabChip® GX Touch HT (PerkinElmer, Waltham, MA, USA) plus individual dilution of libraries. Normalized libraries were pooled, denatured and diluted as described in Illumina® protocol [7], with a 5% spike-in of PhiX Sequencing Control v3 (Illumina). Paired-end sequencing was performed on Illumina® MiSeq FGx™ instrument using MiSeq® Reagent Kit v2, 300 cycles (2x151 bp).

Sequencing quality metrics were reviewed in Illumina® Sequencing Analysis Viewer v.1.11.1 software. FASTQ files generated on instrument by MiSeq® Reporter were extracted and uploaded to Illumina® BaseSpace® Sequence Hub online platform, where they were analysed by BaseSpace® mtDNA Variant Processor v1.0.0 App [8], using rCRS genome [9, 10] for alignment and analysis thresholds previously established by internal evaluation [11]. Sequence variants and read depth profiles were visually inspected in BaseSpace® mtDNA Variant Analyzer v1.0.0 App, as well as in IGV browser [12, 13]. Variant reports

were exported from BaseSpace® in Excel format for further review, variant confirmation and comparison of results.

3. Results And Discussion

3.1. Optimization of long-range PCR for mtDNA enrichment

Quality of amplicons was assessed by observing the definition of bands on agarose gel: clearly defined bands of expected size, without smears and/or non-specific products, were aimed at. At the initial stage, Platinum PCR SuperMix yielded either poor bands or large smears (Fig. S2a), while LA Taq Hot Start and PrimeSTAR GXL both yielded distinguishable bands of expected size with primers for longer, more challenging, 11.2 kb fragment (Fig. S2b and S2c, respectively). Thus, LA Taq Hot Start and PrimeSTAR GXL entered stage II, where optimal DNA input and number of amplification cycles were tested (Fig. S3). PrimeSTAR GXL performed better than LA Taq Hot Start for 11.2 kb amplicon, by producing clearer and better-defined bands, as well as optimal yield required for downstream library preparation (Fig. S3a and S3b). Therefore, decision was made to proceed only with PrimeSTAR GXL to stages III and IV, in order to establish optimal conditions for 9.1 kb amplicon. As shown in Fig. S4 and S5, optimal genomic DNA input amounted to 1 ng (in 12.5 µL reaction volume), while 25 amplification cycles provided good balance between yield and specificity for both buccal epithelia and blood sample types. Increasing annealing temperature to 60°C improved specificity, thus producing visually better results when compared to 55°C annealing temperature (Fig. S5 and S4, respectively). To confirm optimal PCR conditions, we amplified both mtDNA fragments in different samples and sample types (Fig. S6). Annealing temperature of 60°C was retained for 9.1 kb amplicon in 3-step PCR, while 2-step PCR conditions were applied for 11.2 kb amplicon. Final optimized long-range PCR conditions for PrimeSTAR GXL DNA polymerase are shown in Table S1. PrimeSTAR GXL has already been reported as the best-performing long-range DNA polymerase in comparison to five other DNA polymerases, and specifically for the purpose of obtaining long PCR products for sequencing on MiSeq instrument [14]. Even though the previous publication used different targets for long-range PCR, this study corroborates the best performance of PrimeSTAR GXL, while also expanding its application to long-range PCR of mtDNA amplicons for MPS. Additionally, PrimeSTAR GXL provided accurate, repeatable and reproducible results in our previous study [11], thus confirming its reliability for long-range PCR purposes.

3.2. Evaluation of limited-cycle PCR step

We noticed that, while quality metrics of sequencing runs were quite satisfactory and within respective parameter's range given in manufacturer's specifications, the yield of generated data did not achieve its full capacity. Moreover, libraries that produced low quantity electropherograms on LabChip also produced lower cluster density (hence less clusters and subsequently less data) in sequencing, regardless of the loading concentration. Nextera XT assay has been known to produce uneven read depth profiles [2-4, 6, 15], which risks that some regions receive very low read depth. Thus, in order to maximize the yield of

data – primarily to ensure that each sample gets sufficient read depth on all positions of mtDNA genome – increment to 15 cycles of amplification was introduced to limited-cycle PCR step, wherein index adapters are added and libraries are amplified. Molar concentrations of libraries amplified with 15 cycles showed substantial increase as opposed to libraries that underwent 12-cycle PCR (Table S3), which was expected, and LabChip electropherograms showed improved, larger quantities of library fragments (Fig. 1). All variant calls were concordant between 12-cycle and 15-cycle libraries of corresponding samples, including occurrences of point heteroplasmy, as well as insertions and deletions (Table S4).

However, it was necessary to exclude the possibility that prolonged amplification of indexed libraries affected sequencing results in any way (e.g. elevating the level of noise or introducing sequence errors). For that reason, negative controls (NC-EX, NC-PCR and NC-LIB) were analysed as previously described [11], from runs containing libraries prepared with 12 and 15 cycles, to assess the level of noise and exogenous signals detected in sequencing. Cumulatively, an average of 7,370 positions with reads were detected in 15-cycles NCs, which is higher than the average of 6,543 detected in 12-cycles NCs (Table S5). Nonetheless, average read depth was only slightly elevated (6 reads and 5 reads for 15-cycles and 12-cycles NCs, respectively), while maximum read depth (Table S5) detected in any NC was well below the established minimum read depth threshold of 220 reads [11]. Noise in samples was also evaluated by analysing signals of alternative bases (different than haplotypes, excluding positions with point heteroplasmy): average read depth of alternative signals was 44 reads for 12-cycle libraries, and 52 reads for 15-cycle libraries. While in both cases signals exceeding the threshold of 220 reads were detected, these would not be of concern to impact variant calling and interpretation since all such signals were either below the 3% analysis threshold [11], or exhibited poor strand balance and/or displayed low quality score; thusly, they would be excluded from final variant calling. Increasing amplification cycles produced larger quantity of libraries, which, ultimately, improved the yield of sequencing data by enabling maximal usage of sequencing chemistry capacity, while still maintaining good quality of run metrics (Table S6). Since the modified PCR conditions did not affect the level of noise nor variant calling at the established analysis and interpretation thresholds, they were deemed safe and so applied in subsequent sequencing runs.

3.3. Comparison of library normalization methods

Sequencing metrics parameters such as cluster density, clusters passing filter and quality of bases, are mostly dependent on the loading concentration of pooled libraries (which is, in turn, most influenced by the accuracy of library quantification [16], and are thus not directly dependent on normalization method. However, the chosen method of library normalization may greatly impact the proportion of reads for a sample (expressed as “% reads identified”), in the sense of better or worse uniformity in representation of samples. Naturally, greater uniformity between samples means more even distribution of reads per sample, and consequently achieving sufficient read depth across the sequenced targets. The greatest risk of low proportion of reads in a sample is losing valuable information from regions that would potentially receive very few or no reads, leading to increasingly difficult detection and interpretation of variants in

those regions, and eventually requiring repeated library preparation and sequencing – with additional costs of reagents and consumables.

Two library normalization methods were compared: magnetic beads-based normalization against “standard” normalization (i.e. quantification of libraries followed by individual normalization). From the standard deviation of % reads identified, with corresponding coefficients of variation (Table S7), it is evident that magnetic beads normalization introduced greater variation to the distribution of reads per library (Fig. 2). This observation is concordant to previously reported [4], and is likely caused by the sensitivity of magnetic beads to numerous handling steps (dependent on accuracy, precision, speed and dexterity of particular analyst). Even though normalization beads are included in Nextera XT library preparation kit and require no additional expenses, LabChip quantification and individual, library-by-library normalization allow faster processing of larger sample batches. The latter method also requires less hands-on time (thereby reducing cross-contamination), and produces both concentration/molarity and fragment distribution information for each library, so the risk of (costly) repeated sequencing is greatly diminished. Additionally, it enables more flexible regulation of run plexity, which is particularly relevant for achieving the desired read depth of samples for certain applications (e.g. detection of novel variants will require higher read depth than population studies, which means sequencing less samples per run).

4. Conclusions

In conclusion, we optimized Illumina® whole mtDNA MPS assay on MiSeq FGx™ instrument. High-fidelity long-range DNA polymerase TaKaRa PrimeSTAR® GXL performed best in aspects of specificity and yield for mtDNA enrichment from reference samples of buccal swabs and blood samples. Increment of amplification cycles in the limited-cycle PCR (indexing and amplification) is recommended, as it maximizes the usage of sequencing chemistry capacity as well as the yield of sequencing results, while not affecting the sequencing noise in any way that would interfere with variant calling and interpretation. Individual library normalization, following quantification on LabChip®, enables fast and efficient processing of large number of samples, simultaneously allowing visualization of library fragment distribution, which is a particularly useful step for quality assurance prior to sequencing. In comparison to less-precise bead-based normalization, LabChip®-based approach provides more uniform distribution of reads per library, which is a crucial feature towards ensuring sufficient read depth for all samples. Since the latter feature is of paramount importance for sample comparison within and between sequencing runs, we propose the LabChip®-based normalization as the preferred method, especially in studies involving detection of minor variants (e.g. mitochondrial heteroplasmy). Altogether, optimizations reported herein surpass the solely intra-laboratory usage, and can be employed by other groups pursuing broad variety of PCR-based MPS workflows.

Declarations

Funding:

This work was supported by Ministry of the Interior of the Republic of Croatia.

Authors' contributions (CRediT):

Conceptualization [Viktorija Sukser, Marina Korolija]; Investigation [Viktorija Sukser, Ivana Račić, Sara Rožić, Lucija Barbarić, Marina Korolija]; Data Curation [Viktorija Sukser, Ivana Račić, Sara Rožić]; Formal Analysis, Methodology, Writing – Original Draft [Viktorija Sukser]; Supervision, Writing – Review and Editing [Marina Korolija].

Conflicts of interest:

The authors declare that they have no conflict of interest.

Ethics approval:

This study involved samples collected from human participants. All procedures performed in the study were in accordance with the institutional and national ethical standards.

Consent to participate:

Informed consent was obtained from all individual participants included in this study.

Availability of data and material:

The datasets generated and analysed during this study are available from the corresponding author on reasonable request.

Code availability:

Not applicable.

Consent for publication:

Not applicable.

References

1. Illumina (2016) Protocol: Human mtDNA Genome for the Illumina Sequencing Platform.
https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_legacy/human-mtdna-genome-guide-15037958-01.pdf
2. King JL, LaRue BL, Novroski NM, Stoljarova M, Seo SB, Zeng X, Warshauer DH, Davis CP, Parson W, Sajantila A, Budowle B (2014) High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. *Forensic Science International: Genetics* 12:128-135. <https://doi.org/10.1016/j.fsigen.2014.06.001>
3. McElhoe JA, Holland MM, Makova KD, Su MS, Paul IM, Baker CH, Faith SA, Young B (2014) Development and assessment of an optimized next-generation DNA sequencing approach for the mitogenome using the Illumina MiSeq. *Forensic Science International: Genetics* 13:20-29. <https://dx.doi.org/10.1016%2Fj.fsigen.2014.05.007>
4. Peck MA, Brandhagen MD, Marshall C, Diegoli TM, Irwin JA, Sturk-Andreaggi K (2016) Concordance and reproducibility of a next generation mtGenome sequencing method for high-quality samples using the Illumina MiSeq. *Forensic Science International: Genetics* 24:103-111. <https://doi.org/10.1016/j.fsigen.2016.06.003>
5. Peck MA, Sturk-Andreaggi K, Thomas JT, Oliver RS, Barritt-Ross S, Marshall C (2018) Developmental validation of a Nextera XT mitogenome Illumina MiSeq sequencing method for high-quality samples. *Forensic Science International: Genetics* 34:25-36. <https://doi.org/10.1016/j.fsigen.2018.01.004>
6. Ring JD, Sturk-Andreaggi K, Peck MA, Marshall C (2017) A performance evaluation of Nextera XT and KAPA HyperPlus for rapid Illumina library preparation of long-range mitogenome amplicons. *Forensic Science International: Genetics* 29:174-180. <https://doi.org/10.1016/j.fsigen.2017.04.003>
7. Illumina (2019) MiSeq® System Denature and Dilute Libraries Guide, https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-denature-dilute-libraries-guide-15039740-10.pdf
8. Illumina (2016) mtDNA Variant Processor v1.0 BaseSpace App Guide, https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/basespace/basespace-mtdna-variant-processor-v1-app-guide-1000000007931-00.pdf
9. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290(5806):457-465. <https://doi.org/10.1038/290457a0>
10. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics* 23(2):147. <https://doi.org/10.1038/13779>
11. Sukser V, Rokić F, Barbarić L, Korolija M (2021) Assessment of Illumina® Human mtDNA Genome assay: workflow evaluation with development of analysis and interpretation guidelines. *International*

12. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative Genomics Viewer. *Nature Biotechnology* 29(1):24-26. <https://doi.org/10.1038/nbt.1754>
13. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP (2017) Variant review with the Integrative Genomics Viewer (IGV). *Cancer Research* 77(21) 31-34. <https://doi.org/10.1158/0008-5472.can-17-0337>
14. Jia H, Guo Y, Zhao W, Wang K (2014) Long-range PCR in next-generation sequencing: comparison of six enzymes and evaluation on the MiSeq sequencer. *Scientific Reports* 4:5737. <https://doi.org/10.1038/srep05737>
15. Parson W, Huber G, Moreno L, Madel MB, Brandhagen MD, Nagl S, Xavier C, Eduardoff M, Callaghan TC, Irwin JA (2015) Massively parallel sequencing of complete mitochondrial genomes from hair shaft samples. *Forensic Science International: Genetics* 15:8-15. <https://doi.org/10.1016/j.fsigen.2014.11.009>
16. Hussing C, Kampmann ML, Smidt Mogensen H, Børsting C, Morling N (2018) Quantification of massively parallel sequencing libraries – a comparative study of eight methods. *Scientific Reports* 8:1110. <https://doi.org/10.1038/s41598-018-19574-w>

Figures

Figure 1

Comparison of fragments' distribution for a library amplified with 12 cycles (a) and 15 cycles (b) in limited-cycle PCR step of Nextera® XT library preparation. Electropherograms produced by LabChip® DNA High Sensitivity Assay on LabChip® GX Touch HT

Figure 2

Distribution of the percentage of reads identified per sample for sequencing runs. Libraries in runs 1-4 underwent bead-based normalization (magnetic beads included in Illumina® Nextera® XT Library Prep Kit), while libraries in runs 5-8 were normalized individually according to molarities obtained by LabChip® DNA High Sensitivity Assay

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial.pdf](#)