Breed-specific Reference Sequence Optimizes Mapping Accuracy of NGS Analyses for Pigs

Dan Wang
Shandong Agricultural University
Liu Yang
China Agricultural University
Chao Ning
Shandong Agricultural University
Jian-Feng Liu
China Agricultural University
Xingbo Zhao (zhxbcau@126.com)
China Agricultural University https://orcid.org/0000-0002-5576-349X

Research

Keywords: mitochondrial genome, mapping, reference sequence, SNP calling, pig

Posted Date: August 11th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-54147/v1

Abstract

Background: The reference sequence plays a key role in next-generation sequencing (NGS), which impacts the mapping quality during genome analyses. Especially for mitochondrion which contains plentiful amounts of innate DNA, the optimal reference sequence makes mitochondrial genome (mitogenome) alignment accurate and efficient. In this study, different mapping reference sequences, the commonly used reference sequence (CU-ref), the breed-specific reference sequence (BS-ref) and the sample-specific reference sequence (SS-ref) were compared to test the accuracy of mapping quality in the NGS analyses of pigs.

Results: Four pigs from three breeds were high-throughput sequenced and subsequently mapped using three different reference mentioned above, which indicated that the BS-ref produced the largest number of mappable reads and coverages at acceptable run-times. After that, the SNP calling accuracy was evaluated by 18 detection strategies with three tools SAMtools, VarScan and GATK with different parameters under the BS-ref mapping strategy. Results showed that nine detection strategies achieved the same best specificity and sensitivity, which suggested a high accuracy of mitogenome alignment by the BS-ref alignment strategy, with a low requirement for SNP calling tools and parameter choices.

Conclusions: Overall, using the breed-specific reference sequences in NGS analyses optimized the mapping quality and the SNP calling accuracy. This study indicates that the different reference sequences which represent different genetic distances between reference sequences and samples in mitogenome alignments influence alignment quality.

Background

Next-generation sequencing (NGS) technology is characterized by the ability to read millions of DNA molecules at a time, with high efficiency and low costs, compared to Sanger sequencing [1]. The mitochondrion contains hundreds of innate DNAs [2], which permits high read depths and makes mitogenome alignment from NGS data indeed feasible. With NGS data, the complete mitogenomes of many mammals have been obtained, including pigs [3–6], chickens [7–9] and cattle [10–13]. It is reported that reference sequences affected the accuracy in genome mapping [14–16]. Despite the existence of relevant studies, there are few studies on the effect of reference sequences on genome mapping and the quality of NGS in mammals or agricultural animals.

In this study, we explored whether genetic distances between reference and sample sequences impact the accuracy of mitogenome mapping for pigs. Three breeds of pigs, including one Asian wild pig, two unrelated Diannan small-ear pigs and one Tibetan pig, were high-throughput sequenced and analyzed in genome alignment and SNP calling. We put forward three kinds of reference sequences on behalf of different genetic distances for mitogenome alignment, including the commonly used reference sequence (CU-ref), breed-specific reference sequences (BS-ref) and sample-specific reference sequences (SS-ref). In

addition, the *de novo* assembly strategies were carried out based on three levels of NGS read sets to explore the best *de novo* strategy and to generate the SS-ref.

Results

Performances of *de novo* assembly strategies

In order to produce the optimal SS-ref, three *de novo* assembly strategies were carried out. The BLAST_denovo strategy yielded similar N50 contig size, consensus length and genome coverage to the Denovo strategy. The BLAST_denovo strategy had no polymorphic site, while the Denovo strategy had a large number of polymorphisms (193 sites), which might be caused by the interference from nuclear mitochondrial sequences (NUMTs). The BWA_denovo strategy got bad results of the least N50 contig size and some polymorphic sites. Therefore, the *de novo* assembly by homologous sequences filtered from NGS data through BLAST was used for constitution of the SS-ref. NGS data information were listed in Additional file 1: Table S1, and the assembly quality of each *de novo* strategy was shown in Table 1.

Alignment qualities by different reference sequences

We compared the mitogenome mapping quality using different mapping references. Alignment against BS-ref yielded higher mapping ratio and average coverage than that against CU-ref or SS-ref, and the latter two got similarly higher mapping ratios and average coverages. In the aspect of time consumption in alignment, mapping to SS-ref cost much more run time than mapping to BS-ref or CU-ref, and the latter two had similar time-cost (Figure 1). In addition, mapping against BS-ref showed a more uniform coverage across mitogenome than mapping against CU-ref or SS-ref (Figure 2).

Performances of SNP calling strategies on mitogenome diversity

As the gold standard, the Sanger sequencing data were aligned against the reference sequence KP765605.1 for the sample A1, KM044240.1 for D1 and D2, and KM073256.1 for T1, and the number of SNP was 12 for A1, 7 for D1, 9 for D2, and 10 for T1, which were detailed in Additional file 2: Table S2. Totally 18 SNP calling strategies were performed on bam files resulted from alignments against BS-ref. The concordance between NGS data and Sanger data was analyzed, and all the 18 strategies detected all true SNPs, zero false positives and zero false negatives, which was listed in Additional file 3: Table S3.

Discussion

Optimized highly similar sequences have always been an important selection of reference sequences during the homologous sequences alignment to assist in piecing reads together. In order to accurately construct a genome sequence from NGS data, many studies turn to exploring the optimal sequencing

methods and alignment strategies [14–20]. The *de novo* and reference-based approaches were used to assemble the complete mitogenome of *Clarias batrachus* from NGS data, respectively, which resulted in different consensus sequences in length [14]. Moreover, different reference sequences also might mean different performances on mapping. Liu *et al*. found when the sample-specific sequence, a sequence with the same genotype sequence as the sample, was used as the mapping reference in NGS analyses of HBV(Hepatitis B Virus), the mapping accuracy and variation calling were optimized, compared to when the other four commonly used HBV genome sequences from GenBank database were respectively used as the mapping references [15].The present study provides an important recommendation for the selection of reference sequences for homology analysis. We compared three references (CU-ref, BS-ref and SS-ref) with different genetic distances between the reference sequences and the samples to assess the accuracy of mitogenome mapping for pigs. Through a comprehensive comparison, the BS-ref used in the mitogenome alignment showed the brilliant performance than the other two references. This result was consistent with the previous studies' [18, 21]. Lee *et al*. investigated whether the choice of reference genome would influence the detection of mitochondrial SNPs and epidemiological inferences of *Mycobacterium tuberculosis* transmission by aligning sequencing reads from 162 closely related lineage4 (Euro-American) isolates to seven different references, in result, when CDC1551 and H37Rv both from lineage 4 and with close genetic distance from *Mycobacterium tuberculosis* were respectively acted as the alignment reference, the highest proportion of reads were successfully aligned [21].

Annotation of mitogenome from NGS data relies crucially on SNP calling. Totally 18 different SNP calling strategies using three software programs with different options were compared, and led to same SNP results in terms of true SNPs, false positives and false negatives. The results suggested the mitogenome obtained by the alignment strategy had a low requirement for variation calling tools and parameter choices, which indicated that mapping with the breed-specific reference sequence contributed to an accurate mitogenome.

To sum up, references with close genetic distance, such as breed-specific sequences, were suggested to be chosen for a high reliability and accuracy of mitogenome alignment, and to later apply in sequences variation determination in the germlines of individuals, families, or populations and ascertain disease-causing mutations, which would reduce potential mistakes. When specific reference sequences were used, the erroneous base incorporations can be efficiently avoided, which is important for NGS data of low depth [22], for example in the study of ancient DNA (aDNA). Therefore, breed-specific references are suggested in the case.

In addition, three *de novo* assembly strategies were compared and showed that the homologous reads sets filtered by BLAST from clean data were suitable for *de novo* assembly. BLAST algorithm [23], which was proposed by Altshul *et al* in 1990, is now the most widely used search tool for homologous sequences in nucleotide databases. BLAST can tolerate more mismatches and gaps for a higher recruitment rate than BWA [24], which, only identifies extremely stringent similarities [25].

Conclusions

In this study, the mitogenome alignments against different reference sequences which represent different genetic distances between the reference sequences and the samples revealed that the breed-specific sequence is the optimal mapping reference due to the high mapping quality and SNP calling accuracy. In addition, for an accurate *de novo* assembly, it is recommended first to filter the clean sequences by BLAST, and then to *de novo* assemble. In summary, this study highlighted the importance of reference sequence choose in mitogenome alignment.

Methods

Animals

Ear tissues from four pigs of three breeds were collected, including an Asian wild pig (A1), two unrelated Diannan small-ear pigs (D1 and D2) and a Tibetan pig (T1). Total DNA was extracted by QIAamp DNA Investigator kit (QIAGEN, Hilden, Germany) following the manufacturer's instruction. DNA quality was evaluated by spectrophotometry and agarose gel electrophoresis. The guidelines of the experimental animal management of China Agricultural University (CAU) were followed, and the experimental protocols were approved by the Experimental Animal Care and Use Committee of CAU.

DNA sequencing

DNA templates were ultrasonically sheared using a Covaris E220 (Covaris, Woburn, USA), and were prepared for DNA libraries following the NEBNext Ultra DNA Library preparation protocol. Multiple Ampure Bead XP cleanups (Beckman Coulter, Brea, CA, USA) were conducted to remove any adapter dimer that might have developed. The quality and concentration of libraries were determined on an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). Subsequently, the quality-controlled genomic library for each sample was PE100 sequenced using the Illumina HiSeq 2000 sequencing system.

The traditional sequencing approach, Sanger sequencing, was also performed on the samples to represent a gold standard in variation detection. The mitochondrial DNA was PCR-amplified by the 16 primer pairs used in the previous study [26]. Amplicons were bi-directionally sequenced using BigDye Terminator version 3.1 technology on an ABI 3730 system (Applied Biosystems, Foster City, CA). Mitogenomes were analyzed by the softwares of MEGA6 [27] and DnaSP v5 [28].

Quality control

Read quality was assessed using the FastQC software focusing on base quality scores and sequence length. To ensure quality, a quality control of NGS data was conducted. Adapters and low-quality bases were removed by the Clip&Merge software. Reads were filtered to exclude those of a nucleotide length of shorter than 35bp and a Phred quality score of lower than 20. And then forward and reverse reads were

merged into single sequences if they overlapped by at least 8 bp. These tools were integrated into the EAGER-pipeline [29] .

Alignment to different reference sequences

Mapping was performed by BWA [30] with default “aln” and “samse” parameters, except considering the overall number of mismatches tolerated in the alignment by setting the expected fraction of misalignments to 0.04 (-n). Three reference sequences on behalf of different genetic distances between the reference sequences and the sample were used as follows.

(1) The commonly used reference sequence (CU-ref), which referred to a frequently-used sequence from RefSeq project at NCBI [31] database, and here was NC_000845.1 used.

(2) The breed-specific reference sequence (BS-ref), which referred to a sequence of the same breed as the sample downloaded from NCBI database. For the Asian wild pig, KP765605.1 was used as BS-ref, which is from a Changbai mountains wild boar and 16720 bp in length; for the Diannan small-ear pigs, KM044240.1 was used, which is a complete mitogenome of 16720 bp obtained from a Diannan small-ear pig in Yunnan Province; and for the Tibetan pig, KM073256.1 was used, which is a Tibetan complete mitogenome of 16710 bp.

(3) The sample-specific reference sequence (SS-ref), which referred to the consensus sequence obtained from the NGS reads of the sample through *de novo* assembly.

The produced BAM files from BWA were filtered for sequences with a mapping quality of at least 30. Duplicate removal was carried out on those reads that showed identical start and end coordinates only by the DeDup software. The tools were integrated into the EAGER-pipeline [32].

The qualities of mitogenome mapping were measured by the mapping ratio, average coverage and run time. The mapping ratio refers to the ratio of reads mappable to the mitochondrial reference. The average coverage means the number of times the mitogenome is sequenced. The runtime counts the consumed CPU time consumption of mapping processing, instead of using elapsed time, which includes for example, waiting for input/output operations or entering low-power mode.

De novo assembly for SS-ref

In order to produce the optimal SS-ref, three modified *de novo* assembly strategies were compared based on the NGS data from A1. They were different in the NGS read sets, including all clean read sets, homologous read sets filtered by BLAST or by BWA mapping. *De novo* assembly was performed using SOAPdenove2 [33] by default parameters with the best k-mer size estimated by KmerGennie [34]. The detailed assembly information was as follows.

(1) “Denovo”: the *de novo* assembly directly by all clean [22, 34]. All clean data from NGS were put into SOAPdenovo2, and produced contigs. Then contigs were aligned to NC_000845.1 using MEGA6, and produced a consensus [22].

(2) “BLAST_denovo”: the *de novo* assembly by homologous read sets filtered from clean data by BLAST. Clean data were filtered against a reference panel composed of all complete *Sus Scrofa* mitochondrial genome sequences (219) downloaded from NCBI database by the BLAST tool with the blastn command, and then these sets were put into SOAPdenovo2 for *de novo* assembly.

(3) “BWA_denovo”: the *de novo* assembly by homologous read sets filtered from clean data by BWA. Clean reads were mapped against the above-mentioned reference panel to filter homologous sequences of each sample by BWA, and then these sequences were assembled by SOAPdenovo2.

To assess the three *de novo* assembly strategies, the indicators including the N50, consensus length, coverage and sequence polymorphism resulted from each strategy were measured.

SNP calling of mitochondrial genomes

Three variation callers, *i.e,* SAMtools 1.3.1 [35], VarScan 2.3.9 [36] and GATK 3.7 [37], were respectively applied with different parameter combinations detailed in Table 2 to bam files resulting from mitogenome alignments. These parameters were selected to ensure comparability among different callers. The minimum base quality required to consider a base for calling was set to be 30.

The performance of SNP calling was evaluated in the overall genotype concordance by comparing the NGS results with the Sanger data, with the assumption that the Sanger sequencing gave the correct calling [38-40]. Only the positions where a Sanger sequence were available were kept, and the concordance SNPs between Sanger and NGS data for each individual were considered as true SNPs, while the discrepancies were considered as errors. When NGS data identified an alternate homozygote not observed by Sanger sequence, it was considered as a false positive. Accordingly, when NGS data did not see an alternate homozygote found with Sanger sequence, it was considered as a false negative. The number of true SNPs, false positives and false negatives were analyzed.

Abbreviations

NGS: Next-generation sequencing; mitogenome: mitochondrial genome; SNP: Single nucleotide polymorphism; NUMTS: Nuclear mitochondrial sequences; HBV: Hepatitis B virus; aDNA: Ancient DNA

Declarations

Acknowledgements

Not applicable

Authors’ contributions

X.Z. designed the study. D.W. and J.L. collected the samples and provided data. D.W., L.Y. and C.N. analyzed data, and X.Z., D.W., L.Y., C.N. and J.L. wrote the manuscript.

Funding

This work was funded by programs of National Natural Science Foundation of China-Deutsche Forschungsgemeinschaft (31961133031) and the National Natural Science Foundation of China (31672379).

Ethics approval and consent to participate

The guidelines of the experimental animal management of China Agricultural University (CAU) were followed, and the experimental protocols were approved by the Experimental Animal Care and Use Committee of CAU.

Availability of data and materials

The sequences used in this study were already published in [41].

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]National Engineering Laboratory for Animal Breeding; Ministry of Agricultural Key Laboratory of Animal Genetics, Breeding and Reproduction; College of Animal Science and Technology, China Agricultural University, Beijing, China.[2]College of Animal Science and Technology, Shandong Agricultural University, Tai’an, China.

References

1. Hasan MS, Wu X, Zhang L. Performance evaluation of indel calling tools using real short-read data. Human genomics. 2015;9(1):20.
2. Foote AD, Morin PA, Durban JW, Pitman RL, Wade P, Willerslev E, et al. Positive selection on the killer whale mitogenome. Biol Lett. 2011;7(1):116–8.
3. Tan Y, Shi K-z, Wang J, Du C-l, Zhao C-p, Zhang X, et al. The complete mitochondrial genome of Kele pig (Sus scrofa) using next-generation deep sequencing. Conservation genetics resources. 2018;10(2):195–9.
4. Singh AP, Jadav KK, Kumar D, Rajput N, Srivastav AB, Sarkhel BC. Complete mitochondrial genome sequencing of central Indian domestic pig. Mitochondrial DNA Part B. 2016;1(1):949–50.
5. Lord E, Collins C, deFrance S, LeFebvre MJ, Matisoo-Smith E. Complete mitogenomes of ancient Caribbean Guinea pigs (Cavia porcellus). Journal of Archaeological Science: Reports. 2018;17:678–88.
6. Mao H, Zhao G, Guan Y, Guo X, Lamaocao Z. The complete mitochondrial genome of Juema pig (Suina: Suidae). Conservation genetics resources. 2018:1–3.
7. Kwak W, Song K-D, Oh J-D, Heo K-N, Lee J-H, Lee WK, et al. Uncovering genomic features and maternal origin of Korean native chicken by whole genome sequencing. PloS one. 2014;9(12):e114763.
8. Suwannapoom C, Wu Y-J, Chen X, Adeola AC, Chen J, Wang W-Z. Complete mitochondrial genome of the Thai Red Junglefowl (Gallus gallus) and phylogenetic analysis. Zoological research. 2018;39(2):127.
9. Huang X-H, Li G-M, Chen X, Wu Y-J, Li W-N, Zhong F-S, et al. Identification of a novel mtDNA lineage B3 in chicken (Gallus gallus domesticus). Zoological research. 2017;38(4):208.
10. Guo X, Ding X, Wu X, Bao P, Xiong L, Yan P, et al. Complete mitochondrial genome of Anxi cattle (Bos taurus). Conservation genetics resources. 2018;10(3):393–5.
11. Guo X, Pei J, Xiong L, Bao P, Zhu Y, Wangdui B, et al. The complete mitochondrial genome of Shigaste humped cattle (Bos taurus). Conservation genetics resources. 2018;10(4):789–91.
12. Pramod RK, Velayutham D, Zachariah A, Zachariah A, Dhinoth Kumar B, Iype S, et al. Complete mitogenome reveals genetic divergence and phylogenetic relationships among Indian cattle (Bos indicus) breeds. Animal biotechnology. 2019:1–14.
13. SUJITH S
SEQUENCING OF BOVINE SPECIES
SUJITH S. WHOLE MITOCHONDRIAL GENOME. SEQUENCING OF BOVINE SPECIES.
14. Kushwaha B, Kumar R, Agarwal S, Pandey M, Nagpure NS, Singh M, et al. Assembly and variation analyses of Clarias batrachus mitogenome retrieved from WGS data and its phylogenetic relationship with other catfishes. Meta Gene. 2015;5:105–14.
15. Liu WC, Lin CP, Cheng CP, Ho CH, Lan KL, Cheng JH, et al. Aligning to the sample-specific reference sequence to optimize the accuracy of next-generation sequencing analysis for hepatitis B virus. Hep Intl. 2016;10(1):147–57.

16. Shuai Y, Qin Z, editors. Read-mapping using personalized diploid reference genome for RNA sequencing data reduced bias for detecting allele-specific expression. IEEE International Conference on Bioinformatics & Biomedicine Workshops; 2013.
17. Kushwaha B, Kumar R, Agarwal S, Pandey M, Nagpure N, Singh M, et al. Assembly and variation analyses of Clarias batrachus mitogenome retrieved from WGS data and its phylogenetic relationship with other catfishes. Meta gene. 2015;5:105–14.
18. Machado D, Lyra M, Grant T. Mitogenome assembly from genomic multiplex libraries: comparison of strategies and novel mitogenomes for five species of frogs. Molecular Ecology Resources. 2016;16(3):686–93.
19. Dubchak I, Poliakov A, Kislyuk A, Brudno M. Multiple whole-genome alignments without a reference organism. Genome research. 2009;19(4):682–9.
20. Angiuoli SV, Salzberg SL. Mugsy: fast multiple alignment of closely related whole genomes. Bioinformatics. 2016;27(3):334.
21. Lee RS, Behr MA. Does choice matter? Reference-based alignment for molecular epidemiology of tuberculosis. J Clin Microbiol. 2016;54(7):1891–5.
22. Cronn R, Liston A, Parks M, Gernandt DS, Shen R, Mockler T. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic acids research. 2008;36(19):e122-e.
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology. 1990;215(3):403–10.
24. Li H, Durbin R. Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. Bioinformatics. 2009;25(14):1754–60.
25. Niu B, Zhu Z, Fu L, Wu S, Li W. FR-HIT, a very fast program to recruit metagenomic reads to homologous reference genomes. Bioinformatics. 2011;27(12):1704–5.
26. Wang D, Ning C, Xiang H, Zheng X, Kong M, Yin T, et al. Polymorphism of mitochondrial tRNA genes associated with the number of pigs born alive. J Anim Sci Biotechnol. 2018;9(1):86.
27. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular biology evolution. 2013;30(12):2725–9.
28. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451–2.
29. Peltzer A, Jäger G, Herbig A, Seitz A, Kniep C, Krause J, et al. EAGER: efficient ancient genome reconstruction. Genome biology. 2016;17(1):60.
30. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
31. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic acids research. 2015;44(D1):D733-D45.

32. Peltzer A, Jager G, Herbig A, Seitz A, Kniep C, Krause J, et al. EAGER: efficient ancient genome reconstruction. Genome biology. 2016;17:60.
33. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012;1(1):18-.
34. Chikhi R, Medvedev P. Informed and automated k-mer size selection for genome assembly. Bioinformatics. 2014;30(1):31–7.
35. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987–93.
36. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome research. 2012;22(3):568–76.
37. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome research. 2010;20(9):1297–303.
38. Scarcelli N, Mariac C, Couvreur TLP, Faye A, Richard D, Sabot F, et al. Intra-individual polymorphism in chloroplasts from NGS data: where does it come from and how to handle it. Mol Ecol Resour. 2016;16(2):434–45.
39. Alexander A, Steel D, Slikas B, Hoekzema K, Carraher C, Parks M, et al. Low Diversity in the Mitogenome of Sperm Whales Revealed by Next-Generation Sequencing. Genome Biol Evol. 2013;5(1):113–29.
40. Elsensohn MH, Leblay N, Dimassi S, Campan-Fournier A, Labalme A, Roucher-Boulez F, et al. Statistical method to compare massive parallel sequencing pipelines. BMC Bioinform. 2017;18(1):139.
41. Wang D, Xiang H, Ning C, Liu H, Liu J-F, Zhao X. Mitochondrial DNA enrichment reduced NUMT contamination in porcine NGS analyses. Brief Bioinform. 2020;21(4):1368–77.

Tables

Due to technical limitations, table 1 and Table 2 are only available as a download in the Supplemental Files section.

Figures

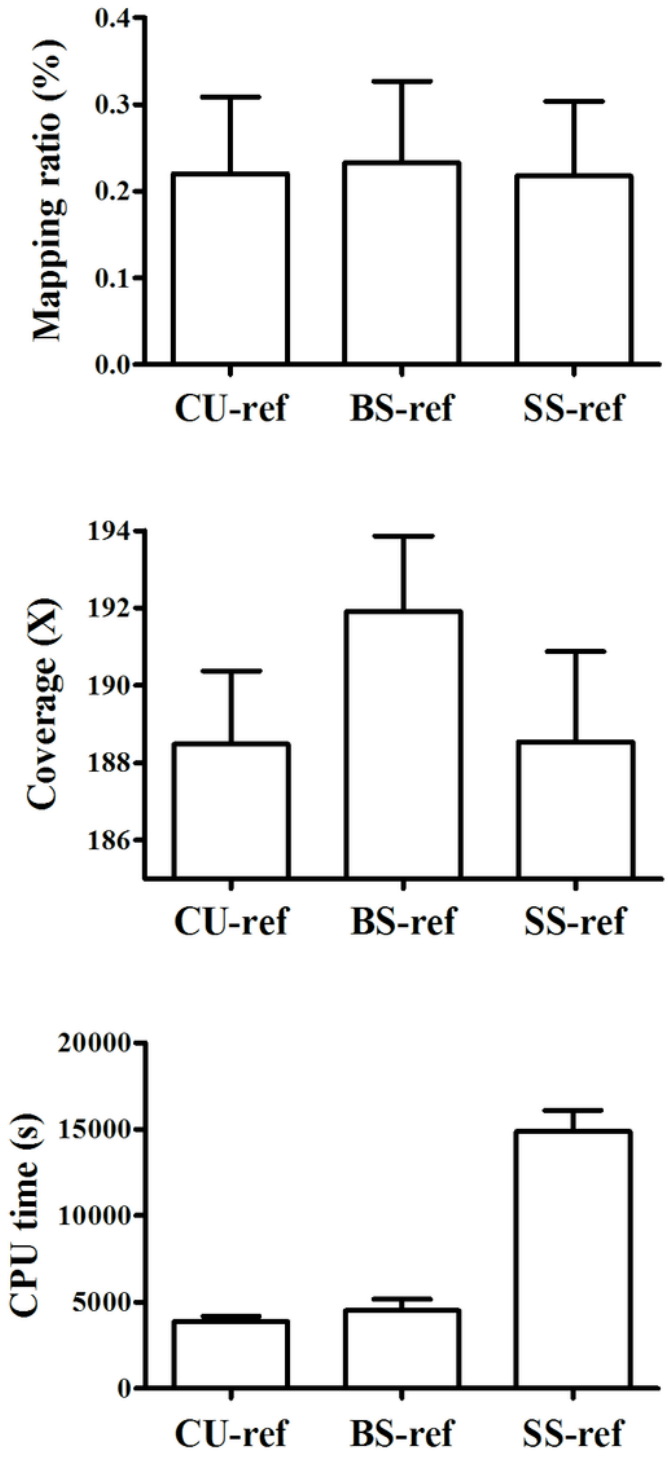

Figure 1

mapping to SS-ref cost much more run time than mapping to BS-ref or CU-ref, and the latter two had similar time-cost

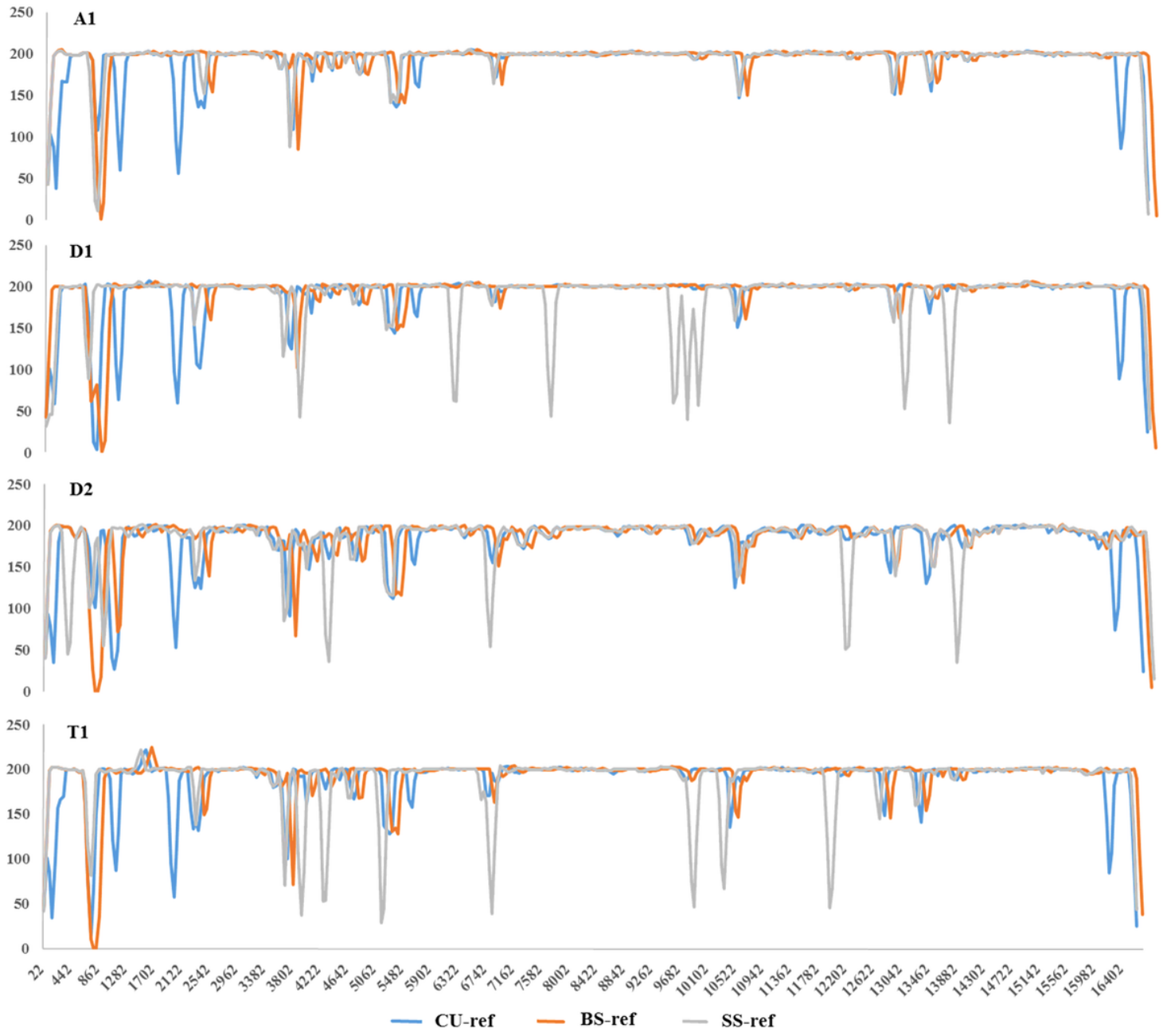

Figure 2

mapping against BS-ref showed a more uniform coverage across mitogenome than mapping against CU-ref or SS-ref

Supplementary Files

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

- Table2.xlsx
- Table1.xlsx

- Additionalfile3TableS3.xlsx
- Additionalfile2TableS2.xlsx
- Additionalfile1TableS1.xlsx