A method for determining the animals of origin in crude drugs using MinION, a compact next-generation sequencer

Hiroaki Nakanishi (hnakani@juntendo.ac.jp)  
Juntendo University School of Medicine

Aya Takada  
Saitama Medical University

Katsumi Yoneyama  
Saitama Medical University

Masaaki Hara  
Saitama Medical University

Kentaro Sakai  
Tokyo Medical Examiner’s Office

Kazuyuki Saito  
Juntendo University School of Medicine

Article

Keywords:

Posted Date: March 30th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1499627/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

We evaluated whether MinION, an inexpensive, portable sequencer, can be applied for identifying the animals of origin in crude drugs. Standard and nonstandard crude drugs with different species of origin were examined. In addition, the standards mixed with nonstandard samples were used. As a target gene, cytochrome c oxidase I was amplified and sequenced. The Fast mode results had a slightly lower match ratio than High-accuracy mode, but the animals of origin were correctly determined by BLAST for all samples. For antler velvet derived from Rangifer tarandus, even the sequences were aligned based on Cervus elaphus, the animal of origin was determined correctly. Minor contents could be detected from mixtures of two animals, if the mixtures contained at least 19:1 mtDNA when the coverage allele-fraction threshold was 0.05. By contrast, in Fast mode, two sequences could not be separated due to the low accuracy of the base-calling in each read. For field work, the species of origin of crude drugs could be identified, by only simple DNA extraction and library preparation. Therefore, MinION appears to be convenient tool for identifying the animal of origin in crude drugs.

Introduction

In Japan, the compositions of crude drugs listed in the Japanese Pharmacopoeia are regulated. However, nonstandard or counterfeit crude drugs are often made using different plants or animals. Quality control of crude drugs is to confirm that the crude drugs have a certain level of bioactivity. The bioactivity of crude drugs is depended on the chemical components contained in the origin species of crude drugs. Therefore, in the quality control for crude drugs, it is important to identify whether original species is conformed to regulations. However, it is difficult to tell by the naked eye in a crude drug that is chopped or powdered. To solve this problem, molecular techniques have been used to identify the origins of crude drugs. For animal crude drugs, polymerase chain reaction (PCR) methods using species-specific primers have been applied. Direct sequencing methods using universal primers can target cytochrome b, cytochrome c oxidase I (COI), and 12S and 16S ribosome RNA genes. The development of molecular techniques has enabled identification of the origin of samples containing a mix of standard and nonstandard drugs. We reported a method for estimating the animal species included in mixed crude drugs using next-generation sequencing (NGS). Although this method is useful, it might not be used widely because NGS is expensive. If a device is inexpensive with accuracy, molecular examination techniques will likely be used widely.

Recently, Oxford Nanopore Technologies developed a compact inexpensive NGS device called MinION. However, the accuracy of MinION base calling is lower than that of conventional NGS. If MinION could be used to analyze animal crude drugs, instead of conventional and expensive NGS, more advanced analysis would be used more widely. Therefore, to use MinION for identifying the animals in crude drugs, it is necessary to evaluate the accuracy of MinION base-calling. In this study, we estimated the animals of origin in crude drugs using MinION, including samples containing mixtures of standard and nonstandard drugs, to evaluate whether MinION can be applied for identifying the animals of origin.
of crude drugs. Moreover, to identify animal of origin in crude drugs on-site, application for field work was evaluated.

**Results**

**Analysis of single-source samples**

Table 1 summarizes the results of Fast and High-accuracy mode analyses of each sample. Supplemental Figure 1 shows the differences in base-calling accuracy between Fast and High-accuracy modes of Integrative Genomics Viewer (IGV). The results in Fast mode had slightly lower match ratios than those in High-accuracy mode, but the animals in all samples were determined correctly by BLAST. For antler velvet derived from *Rangifer tarandus*, even the sequences were aligned based on *Cervus elaphus*, the animal was determined correctly. Similarly, for chopped cicada sloughs, although the sequences were aligned based on *Oncotympana maculaticollis*, the animal was correctly determined for chopped cicada sloughs derived from *Graptoptsaltria nigrofuscata*. However, the origin of snake gallbladder from *Varanus salvator*, which was used as a nonstandard crude drug, could not be determined because mapping was impossible when sequences were aligned based on *Python reticulatus*, which is a standard crude drug.

**Analysis of mixed samples**

Figure 1 shows the base-calling results of the analyses of mixed samples in High-accuracy mode. Our results indicate that minor contents could be detected from mixtures of two animals if the mixtures contained at least 19:1 mtDNA when the coverage allele-fraction threshold was 0.05. By contrast, for Fast mode base-calling, the two sequences could not be separated because the target base could not be defined due to low base-calling accuracy.

**Application to field work**

To apply origin identification by MinION to field work, DNA extraction, library preparation, and sequencing were simplified. In this study, standard and nonstandard antler velvets were used. Identification of the origin of antler velvet was successful by the simple DNA extraction by incubation after adding the reagents. The concordance rates of standard (*Cervus elaphus*) and nonstandard (*Rangifer tarandus*) deer velvet were 99% and 95%, respectively (data not shown).

**Discussion**

We evaluated whether MinION can be applied for identifying the origin of animal crude drugs. When using either Fast or High-accuracy mode for base-calling, the species of origin in all samples were estimated using consensus sequences. Although the accuracy of MinION read is lower than with conventional NGS, the consensus sequence has sufficient accuracy to identify species in not only High-accuracy mode but also Fast mode. Vasiljevic *et al.* obtained consensus sequences in MinION using the clustering-based bioinformatics pipeline NGSpeciesID, and reported that they were useful for forensic species
identification. As described, obtaining consensus sequences is important for species identification in MinION. To obtain consensus sequences with this method, the alignment process is essential. The expected animal species must be decided in advance. If the animal of origin of a nonstandard drug is completely different from that of the standard drug, the MinION reads may not align. For example, examining snake gallbladder in this study, nonstandard *Varanus salvator* could not be aligned with the standard *Python reticulatus*. In contrast, regardless of whether the drug is standard or nonstandard, if the species of origin are similar, the sequences can be aligned, as seen for antler velvet and chopped cicada sloughs. Regarding the origin of antler velvet, the nonstandard drug could be identified, as could mixtures of the standard and nonstandard drugs. The *Cervus elaphus* and *Rangifer tarandus* reads could be separated by sorting based on different bases in IGV. When the coverage allele-fraction threshold in IGV was set at 0.05, the origin of minor components could be determined in mixture ratios of up to 19:1. The accuracy of the R10.3 flow cell exceeded 99% (99.83%) for human leukocyte antigen (HLA) typing, which was higher compared with Sanger sequencing for fungal identification. Moreover, the maximum incorrect base calls in mixed bases was only 4% in the MinION consensus sequence in human mitochondrial DNA analysis. Therefore, it was considered that threshold value in this study, 0.05, was suitable for distinguishing and identifying animals from a mixture sample. However, two conditions are necessary for mixture analysis: the expected nonstandard species must be similar to the standard species and High-accuracy mode must be used for base-calling. In Fast mode, IGV could not decide on base calls because there were many variant bases due to its low accuracy. Base-calling in High-accuracy mode is time-consuming. If it is necessary to confirm only nonstandard contamination, conventional NGS or other methods may be more suitable than MinION. Nevertheless, MinION is useful for identifying the animal of origin of crude drugs because it is portable and inexpensive.

In field work, MinION is commonly used for sequencing. Therefore, we evaluated whether MinION is optimal for examining the origin of crude drugs in field work by using a Flongle flow cell. The Flongle flow cell is disposable and R9.4.1, which has sufficient sequencing accuracy due to excellent consensus Q-scores. Even though simple DNA extraction and library preparation were used, the species of origin could be identified correctly. Both the library preparation time and the run time were short, around 1 and 2 hour(s), respectively. In the field, the use of instruments is sometimes restricted. This method includes PCR, so a thermal cycler is essential. Moreover, the library preparation kit used here fragments DNA, so cannot be applied to mixed samples. If a sample has a single source and a thermal cycler is available, this method can be used for identification of animal of origin in crude drugs in field.

With the evolution of nanopore technology, the base-call accuracy of MinION has been improving. When the base-call accuracy of MinION equals that of conventional NGS, MinION will prove indispensable for identifying the origin of animal crude drugs.

**Methods**

**Sample preparation**
Antler velvet (Cervi Parvum Cornu) derived from *Cervus elaphus*, bear bile (Fel Ursi) derived from *Ursus thibetanus*, oriental bezoar (Bezoar Bovis) derived from *Bos taurus*, musk (Moschus) derived from *Moschus chrysogaster* and snake gallbladder derived from *Python reticulatus* were supplied by Tochimoto Tenkaido (Osaka, Japan) or purchased from a pharmacy in Japan. As nonstandard crude drugs, we purchased antler velvet derived from *Rangifer tarandus*, and snake gallbladder derived from *Varanus salvator*, from a pharmacy in Japan. Two chopped cicada sloughs with different origins were derived from *Oncotympana maculaticollis* supplied by Tochimoto Tenkaido, and *Graptopsisaltria nigrofuscata* purchased from a pharmacy in Japan.

DNA was extracted from the crude drugs (10–200 mg) following a previously described protocol combining a QIAamp DNA Mini Kit (QIAGEN, Venlo, The Netherlands) with phenol treatment.\(^8\),\(^11\) Mixed antler velvet samples were prepared in 1:1, 2:1, 4:1, 9:1, 19:1 and 39:1 ratios of *Cervus elaphus* to *Rangifer tarandus* after normalizing the copy number based on the results of real-time PCR, which was performed in 20-µL reaction mixtures containing 10 µL TB Green Premix Ex Taq II (Takara Bio, Otsu, Japan), 1.6 µL each of 10-µM forward (5'-CTGTACTAGCAGCCGGAATTA-3') and reverse (5'-CAGACTATTCCTATGTACCCAAATG-3') oligonucleotide primers targeting the COI gene, and 2 µL of template DNA. PCR amplifications were performed using the StepOnePlus Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA, USA) at 95°C for 30 s, followed by 40 cycles of 95°C for 5 s and 60°C for 30 s.

**Library preparation and MinION**

To amplify a partial COI region, PCR was performed in 20-µL reaction mixtures containing 10 µL of KOD One PCR Master Mix (Toyobo, Osaka, Japan), 1 µL each of 10 µM forward (5'-GGTCAACAAATCATAAAGATATTGG-3') and reverse (5'-TAAACTTCAGGGTGACCAAAAAATCA-3') oligonucleotide primers,\(^23\) and 1 µL of prepared DNA sample. PCR amplification was performed using a SimpliAmp Thermal Cycler (Thermo Fisher Scientific) with 35 cycles at 98°C for 10 s, 50°C for 5 s, and 68°C for 1 s. The amplified product was quantified using the Qubit dsDNA HS Assay Kit with Qubit 3.0 (Thermo Fisher Scientific). Following PCR cleanup using Sera-Mag Select (Cytiva, Sheffield, UK), a library was prepared with Native Barcoding and Ligation Sequencing kits (Oxford Nanopore Technologies, Oxford, UK) according to the manufacturer's protocols. A MinION Mk1B device with an R10.3 flow cell (Oxford Nanopore Technologies) was used for nanopore sequencing to obtain approximately 40–60 × 10\(^3\) reads per sample. Using MinKNOW software (Oxford Nanopore Technologies), the raw data (FAST5 files) were converted into FASTQ files and aligned (producing BAM files) using Fast or High-accuracy mode.

**Estimation of origin animal species**

The BAM files produced by MinKNOW were imported into the CLC Genomics Workbench 20 (QIAGEN) and exported as BAM files after re-alignment. These BAM files were viewed using IGV.\(^24\) The consensus sequence in IGV was compared with all available sequences using BLAST.
(www.ncbi.nlm.nih.gov/BLAST). The animal species with the top BLAST score was defined as the species of origin for a given crude drug.

To analyze mixed samples, the coverage allele-fraction threshold was set depending on the mixture ratio (threshold = 0.05 or 0.2). To separate the sequences, a base that differed in *Cervus elaphus* and *Rangifer tarandus* was defined as a target base, and then the sequences could be separated by sorting at the target base. The origin of each consensus sequence was confirmed by BLAST.

**Application for field work**

Ten milligrams of powdered sample was used, and DNA was extracted using the Kaneka Easy DNA Extraction Kit (ver. 2; Kaneka, Tokyo, Japan) according to the manufacturer’s protocol. Libraries were prepared using the Rapid Sequencing Kit (Oxford Nanopore Technologies) according to the manufacturer’s protocol. Sequencing was performed using a MinION Mk1B device with a Flongle flow cell (R9.4.1; Oxford Nanopore Technologies). Amplicon DNA that amplified the COI region was examined. The raw data (FAST5 files) were converted into FASTQ files and aligned (thus producing BAM files) in Fast mode using MinKNOW software. The alignment was based on *Cervus elaphus*. The estimation of animal species was the same as above.

**References**


**Declarations**

**Acknowledgments**

This study was supported by JSPS KAKENHI grant number JP19K07151. The authors thank Dr. Yutaka Yamamoto (Tochimoto Tenkaido) for providing some of the animal-derived crude drugs.

**Conflicts of interest**

The authors declare that they have no conflicts of interest.

**Table**

Table 1 is available in the Supplementary Files section

**Figures**
Figure 1

Results of analyses of mixed samples using base-calling in High-accuracy mode. The base in the red frame is the target base used to separate reads. The coverage allele-fraction threshold was 0.2 (mixture ratio = 1:1~9:1) or 0.05 (mixture ratio = 9:1~39:1). The upper reads originated from *Cervus elaphus* and the lower reads from *Rangifer tarandus*. 
Supplementary Files

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

- Table1.xlsx
- SupplementalFigure1.pdf