Forensic Science International: Genetics 30 (2017) 10–17

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Research paper

Body fluid prediction from microbial patterns for forensic application

Eirik Nataas Hanssen[a,b,*], Ekaterina Avershina[c], Knut Rudi[c], Peter Gill[a,b], Lars Snipen[c,**]

[a] Department of Forensic Biology, Oslo University Hospital, Oslo, Norway
[b] Department of Forensic Medicine, University of Oslo, Oslo, Norway
[c] IKBM, Norwegian University of Life Sciences, Aas, Norway

ARTICLE INFO

Article history:
Received 3 April 2017
Received in revised form 28 May 2017
Accepted 29 May 2017
Available online 1 June 2017

MSC:
00-01
99-00***

Keywords:
Body fluid prediction
Microbiome
Massively parallel sequencing
Principal component analysis
Linear discriminant analysis

ABSTRACT

The association of a DNA profile with a certain body fluid can be of essential importance in the evaluation of biological evidence. Several alternative methods for body fluid prediction have been proposed to improve the currently used presumptive tests. Most of them measure gene expression. Here we present a novel approach based on microbial taxonomic profiles obtained by standard 16S rRNA gene sequencing. We used saliva deposited on skin as a forensically relevant study model, but the same principle can be applied for predicting other bacteria rich body fluids. For classification we used standard pattern recognition based on principal component analysis in combination with linear discriminant analysis. A cross-validation of the experimental data shows that the new method is able to successfully classify samples from saliva deposited on skin and samples from pure skin in 94% of the cases. We found that there is a person-effect influencing the result, especially from skin, indicating that a reference sample of pure skin microbiota from the same person could improve accuracy. In addition the pattern recognition methods could be further optimized. Although there is room for improvement, this study shows the potential of microbial profiles as a new forensic tool for body fluid prediction.

1. Introduction

The short tandem repeat DNA-profile is used to identify the person from which a biological trace originates. However, the DNA-profile provides no information on how the trace was deposited. In this context, information on type of body fluid can be of crucial importance when evaluating biological evidence.

Classical presumptive tests are still preferred methods for body fluid prediction in many laboratories [1]. Generally, a body fluid specific enzyme catalyzes a chemical reaction, and the result is visually detected, often as a color change. The tests are fast and easy to use, but have high error rates as the target enzyme is also present in low quantities in other body fluids [2,3]. In addition, common household items and chemicals can give false results [4]. A few immunochromatographic lateral flow strip tests are commercially available as an alternative [5,6], but even if these tests are more specific, the presence of enzymes in other body fluids still give false positive results for some tests. Traditionally, no probabilistic statements have been associated with a presumptive test result.

Lately several alternative detection technologies have been reported to be applicable for body fluid prediction [7,8]. Most of these measure gene expression using mRNA, miRNA or epigenetic markers. The European DNA Profiling Group (EDNAP) has performed collaborative studies on mRNA tests for blood [9], saliva and semen [10] menstrual blood and vaginal secretion [11] and finally on skin [12]. The chemically more stable miRNA markers can also be used to differentiate between body fluids [13–15]. However, both RNA methods currently lack a reliable quantification method. In addition, there is often a large difference in abundance between the RNA makers within a sample. This is especially challenging for minor components in mixed samples where it can be difficult to separate between real and background signal. Another approach is to measure degree of methylation at GpC islands. Although this has been promising [13,14], it is not yet ready for implementation in casework as methylation levels can differ between individuals, tissues and exhibit age or environmental dependency [8]. Since none of the aforementioned methods will detect and separate all body fluids, it has been proposed to combine their use, but even then some mixtures might be challenging [7].

* Corresponding author at: Department of Forensic Biology, Oslo University Hospital, Oslo, Norway.
** Principal corresponding author.
E-mail addresses: kjeirik@gmail.com (E.N. Hanssen), lars.snipen@nmbu.no (L. Snipen).

http://dx.doi.org/10.1016/j.fsigen.2017.05.009
1872-4973/

As an alternative to gene expression measurements, microbial markers have been proposed as a way to discriminate between various body fluids [7,8]. The main idea is to look for the taxonomic composition of bacteria in the various body fluids, and recognize them based on specific patterns in this composition. The standard genetic marker for taxonomic profiling of microbial communities is the small subunit ribosomal RNA gene, also known as the 16S gene. Large data repositories specifically devoted to 16S rRNA gene data exist, e.g. the Silva database (https://www.arb-silva.de/), the Ribosomal Database Project (RDP, https://rdp.cme.msu.edu/) and the Greengenes database (http://greengenes.lbl.gov/). Microbiota-based body fluid recognition is most likely best suited for bacteria-rich body fluids such as saliva, vaginal secretion, feces and menstrual blood, while sterile or nearly sterile body fluids such as blood, semen and tears are probably more problematic to recognize [16]. Other limitations may be geographical variation [17] and drug use [18], but for a large proportion of cases such limitations can be ruled out.

Since inception of the Human Microbiome Project (HMP) [19], many efforts have been made to study the human microbiota by amplicon sequencing of the 16S rRNA gene. While most such studies have been health related and targeted the human gut, there are also some studies with a forensic focus. In [20,21] a search for body fluid-specific taxonomic markers was conducted, but with a negative result. Although it would be convenient to have unique markers to identify a specific body fluid, this may be unrealistic. The *taxonomic profile* of a given body fluid is a vector of quantitative values describing the bacterial composition. Provided that there is sufficient specificity, body fluids can be identified. A large variety of multivariate pattern recognition approaches are already available for the data analysis part of this problem. Such methods have already been used to separate microbiota from phones and shoes [22] and could even potentially be used to identify persons based on skin samples [23].

In microbiome sequencing two major sources of bias have been thoroughly discussed in the literature. First, different DNA extraction method may have an impact on microbial community profiling [24–28]. However there seems to be consensus on that a bead-beating step increases the yield, and that the same extraction protocol should be used throughout a study to ensure reproducibility. The other source of bias is PCR amplification which can result in artifacts such as chimeras [29] and skewed fragment distributions [30–32]. Chimeras form when short aborted extension products function as primers in later PCR cycles to create full length artificial fragments. Chimeras and other PCR artifacts are problematic for bacteria rich samples, but little is known about artifact formation in samples with low levels of bacterial DNA. Digital droplet PCR (ddPCR) use micro droplets as reaction chambers with just one or a few fragments in each droplet. This results in unbiased amplification (see Droplet Digital PCR Applications Guide at www.bio-rad.com).

Health related microbiota studies have investigated pure body fluids sampled directly from the human body. In a forensic context the conditions will be different and care should be taken when adopting standardized lab protocols and bioinformatics workflows. Biological traces are typically collected with cotton swabs [33] and stored in dry state until analysis [34]. Most trace samples will have relatively low bacterial levels and require highly sensitive methods [35] and appropriate routines to prevent contamination [31]. Low bacterial levels might also enhance different biases e.g. in the sampling [36,37] and PCR amplification [29,30,32] steps. Trace samples are rarely single source, but often mixtures of different body fluids. In addition the data interpretation should not be exploratory as in many health studies, but based on pattern recognition.

In this paper we present a study where we have investigated potential effects of sampling and lab-protocols on the detection and recognition of saliva deposited on human skin. This is a typical example of a biological trace from a crime scene, and to our knowledge the first study to demonstrate the identification of body fluids from microbiota data in this context.

2. Materials and methods

2.1. Experimental setup

Six healthy persons participated in this study. They were told not to eat or wash hands during a period of 1 h before the experiment. Traces of both pure and diluted saliva were deposited between the base of the fingers on the back of each participants hands. The liquid was smeared in the sampling areas using the pipette tip and then dried for 10 min before sampling. The experiment was designed so that each of the six individuals had saliva donated from another participant deposited onto their hands (one donation per participant). All experiments were performed on the same day.

The following samples were collected from each participant: (1) Pure saliva sampled directly from the mouth (to be deposited on another participant), (2) the trace consisting of 20 μL pure saliva deposited between fingers, (3) the trace consisting of 20 μL saliva diluted in PCR water (1:10) deposited between fingers and (4) a sample from pure skin between fingers.

Initially three different sampling techniques were evaluated. 20 μL saliva were applied onto cotton swabs (Medical Wire), synthetic swab (DNA Genotek) and tape (Scenesafe) and processed in parallel with 20 μL pure saliva. Bacterial DNA extraction was performed as described below, and recovery was measured for all three techniques. The use of the cotton swab was discontinued based on the results. Tape was used when sampling the left hand and synthetic swabs when sampling the right hand. Diluted saliva was only collected by tape from the left hand. Thus, we define 6 different types of samples:

1. Pure saliva from mouth.
2. Saliva deposited on skin, collected with tape.
3. Saliva deposited on skin, collected with swab.
4. Diluted saliva on skin, collected with tape.
5. Pure skin, collected with tape.
6. Pure skin, collected with swab.

One droplet of PCR grade water was added to the swab before sampling to mimic standard procedure [38].

2.2. Soaking, extraction and quantification

The samples were first soaked to release the sample material. The tape was cut with a sterile razor before being transferred to a 1.5 mL Eppendorf tube with 200 μL S.T.A.R. buffer (Roche Diagnostics). The synthetic swab was placed in the associated tube containing 1 mL soaking solution. The Eppendorf tubes were put on a horizontal shaker at 1400 rpm and 56 °C for 30 min while the tubes with the synthetic swab was briefly vortexed according to producers recommendations. For each sample 150 μL soaking solution was transferred to a 2 mL conical tube (Sarstedt) with approximate 0.24 g acid-washed glass beads (<106 μm; Sigma Aldrich). The samples were homogenized at 1800 rpm for 2 Ô 30 s using FastPrep96 (MPBio) and then centrifuged at 13,000 *g* for 5 min. DNA was extracted using LGC mag midi kit (LGC Genomic) following the manufacturer's recommendations. The resulting DNA extracts were quantified by digital droplet PCR (Bio-Rad