

Vaginal microbial flora analysis by next generation sequencing and microarrays; can microbes indicate vaginal origin in a forensic context?

Corina C. G. Benschop · Frederike C. A. Quaak ·
Mathilde E. Boon · Titia Sijen · Irene Kuiper

Received: 21 October 2011 / Accepted: 14 December 2011 / Published online: 27 January 2012
© Springer-Verlag 2012

Abstract Forensic analysis of biological traces generally encompasses the investigation of both the person who contributed to the trace and the body site(s) from which the trace originates. For instance, for sexual assault cases, it can be beneficial to distinguish vaginal samples from skin or saliva samples. In this study, we explored the use of microbial flora to indicate vaginal origin. First, we explored the vaginal microbiome for a large set of clinical vaginal samples ($n=240$) by next generation sequencing ($n=338,184$ sequence reads) and found 1,619 different sequences. Next, we selected 389 candidate probes targeting genera or species and designed a microarray, with which we analysed a diverse set of samples; 43 DNA extracts from vaginal samples

and 25 DNA extracts from samples from other body sites, including sites in close proximity of or in contact with the vagina. Finally, we used the microarray results and next generation sequencing dataset to assess the potential for a future approach that uses microbial markers to indicate vaginal origin. Since no candidate genera/species were found to positively identify all vaginal DNA extracts on their own, while excluding all non-vaginal DNA extracts, we deduce that a reliable statement about the cellular origin of a biological trace should be based on the detection of multiple species within various genera. Microarray analysis of a sample will then render a microbial flora pattern that is probably best analysed in a probabilistic approach.

Electronic supplementary material The online version of this article (doi:10.1007/s00414-011-0660-8) contains supplementary material, which is available to authorized users.

Both C. C. G. Benschop and F. Quaak contributed equally. M. Boon is the pathologist responsible for the morphologic evaluation of the cervical brush samples.

C. C. G. Benschop · F. C. A. Quaak · T. Sijen (✉) · I. Kuiper
Netherlands Forensic Institute,
Laan van Ypenburg 6,
2497GB The Hague, The Netherlands
e-mail: t.sijen@nfi.minvenj.nl

C. C. G. Benschop
e-mail: c.benschop@nfi.minvenj.nl

F. C. A. Quaak
e-mail: f.quaak@nfi.minvenj.nl

I. Kuiper
e-mail: i.kuiper@nfi.minvenj.nl

M. E. Boon
Leiden Cytology and Pathology Laboratory,
P.O. Box 16084, 2301 GB Leiden, The Netherlands
e-mail: M.E.boon@lcpl.nl

Keywords Forensic science · Vaginal swabs · Microbial flora · Next generation sequencing · Microarray · Body site identification

Introduction

In forensic analysis, information on the presence of someone's cell material in an evidentiary trace does not always suffice as recurrently the cellular origin of a sample is questioned. From a criminalistic point of view, the first is regarded as reporting at source level (which donor is the source of the biological material?), the latter relates to an expert opinion at activity level (what action has led to deposition of the biological material?). For instance, indications whether the cell material of a female donor is of buccal, skin or vaginal origin may lead to a different evaluation at the activity level in a sexual assault case. In some cases, the cellular origin of a sample is determined using microscopic analysis, which can be assisted by histological or immunological staining to detect sperm or epithelial cells [1–4].

In addition to microscopic analysis, presumptive tests based on the detection of enzymes or antigens can be used to indicate the presence of semen, blood and saliva. Furthermore, alternative approaches for body fluid identification have been published, such as DNA methylation [5], mRNA markers or microRNA markers [6–8] or the use of Raman spectroscopy [9, 10]. Notwithstanding these methods, it is hard to unambiguously discriminate vaginal epithelial cells from other epithelial cells as the discriminative power of the various markers used is not absolute.

This study investigates whether the presence of certain microbial flora can indicate a vaginal origin of the sample of interest. The human body is colonised by a wide variety of microbes [11, 12]. Different body sites harbour different populations of microbial flora. The detection of these different populations may be indicative of the sampled body site. Various studies have examined the vaginal microbiome, via microbiological culturing techniques and recently also via deep sequencing approaches [13–21]. A drawback of the culturing method is that a large part of the microbial flora present in a sample may be lost, due to the selectivity of the culturing media and/or the inability to culture a particular microorganism. Extracting DNA directly from a sample, however, allows the analysis of a large part of the present microbiome without selection. By using next generation sequencing techniques such as 454 sequencing, the DNA extract containing DNA from a large proportion of the microbiome can be analysed in one run. In this study, next generation sequencing was performed on a large set of clinical vaginal samples to assay the vaginal microbial flora. Based on this sequencing data set, candidate probes for vaginal flora identification were selected. These probes were spotted on microarrays together with control probes and candidate probes for species or genera known to be present in saliva, faeces and/or on skin. The microarrays were used to (1) evaluate the effect of different DNA extraction methods on isolating DNA from the different microbial species found in the vaginal samples, (2) determine whether species can be found that are present in all or a majority of the vaginal samples and (3) infer what approach has the highest potential for identifying the vaginal origin of a sample based on microbial flora.

Materials and methods

Next generation sequencing and development of a (vaginal) microbial flora microarray

Next generation 454 sequencing (Roche, Branford, USA) of 240 clinical cervical brush samples, collected by general practitioners and stored in the coagulant fixative BoonFix [22], was performed by TNO Quality of Life (Zeist, The Netherlands [23]). These samples were not subjected to

human DNA profiling for which an informed consent would have been needed. Sample diversity was obtained by using vaginal samples from different donors with ages between 15 and 84 and with a healthy vaginal community or with bacterial overgrowth. Eighteen pools were prepared after DNA extraction, for which, grouping predominantly was based on the age of the donor. DNA extraction was done by phenol bead-beating followed by silica column extraction [24]. A specific sequence tag was added for each of the 18 pools during amplification of the 16S rDNA V5 and V6 region. The universal PCR primers used for these amplifications are presented in Supplementary Table 1. The raw sequence data were processed by removing the tag and primer sequences and microbial taxonomy was determined using the Ribosomal Database Project (RDP), which is a 16S rDNA database, that is continuously updated, as described in [25, 26]. A set of 220 oligonucleotide probes, targeting families or (groups of) genera or species, was designed using the next generation sequencing data from the vaginal samples. In addition, 169 probes targeting families of (groups of) genera or species known to be common in saliva, in faeces or on skin were selected [24]. These probes were spotted on microarrays [24] together with 16 general bacterial spots (positive controls) and ten buffer spots (negative controls) to reach a total number of 415 probe and control spots.

Samples used for microarray analysis

In order to gain a representative forensic dataset, 12 female volunteers donated 36 self-collected vaginal swabs. According to Forney et al., self-collected vaginal swabs reveal the same microbial diversity as vaginal swabs collected by a physician [27]. We aimed to obtain an overview of the microbial flora in vaginal samples including factors that can influence the composition of these communities [16, 27, 28]. Therefore, several variables were covered in our dataset, i.e. different periods in menstrual cycle or menopause, variable time between intercourse and sampling and condom or lubricant use. Furthermore, vaginal samples were collected using different swab types, as the characteristics of the swab type may influence the uptake and release of (bacterial) cells [29]. Eight volunteers donated multiple vaginal samples (up to six swabs), some collected at different time points and some collected at the same time point but using different swab types.

In addition to the vaginal swabs, 25 samples from several other forensically relevant body sites were tested. These samples comprised nine double swabs [30] from skin [four hands (three females, one male), three female groins and two penis (two males)], eight saliva samples (five females, three males), three semen samples (one fertile man, two non-fertile men), three urine samples (two males, one

female), one faecal sample (female) and one blood sample (female, obtained by fingertip puncture).

DNA extraction, quantification and STR profiling

Twenty-nine of the 36 vaginal swabs were subjected to differential extraction (DE) resulting in a non-sperm fraction (NF) and a sperm fraction (SF) [29], of which 22 NFs and 16 SFs were further analysed. Five of the 36 vaginal swabs and all non-vaginal samples were processed by the QIAamp DNA mini kit as described by the manufacturer (Qiagen, Venlo, The Netherlands). For 2 of the 36 vaginal swabs, the FastDNA[®] SPIN Kit for Soil (FastPrep) (MP Biomedicals, USA) was used for the extraction of bacterial DNA. DNA was extracted according to the manufacturer, with the adaptation that bead-beating was performed for 30 s at a speed setting of 5.5 which was followed by 1 min centrifugation.

Bacterial DNA quantification was performed using qPCR with universal bacterial primers Eub338 and Eub518 (Supplementary Table 1) [31]. Amplifications were carried out in 25 µl reactions containing 1x iQ[™] SYBR[®] Green Supermix (Bio-Rad, Veenendaal, The Netherlands), 10 pmol of each primer, 3 µl template and distilled water. The following PCR programme was used on a MiniOpticon[™] Real-Time PCR system (Bio-Rad, Veenendaal, The Netherlands): 4 min 95°C, 40 cycles of 30 s 94°C, 40 s 52°C and 40 s 72°C. After each cycle, a plate read was performed. DNA concentrations were calculated using a dilution series (0.3–54 ng/µl) of a *Lactobacillus casei* DNA extract as standard.

Additionally, most extracts were used for human DNA profiling (using the AmpF/STR[®] SGM Plus[®] kit and/or the AmpF/STR[®] Identifiler[®] kit, Applied Biosystems, Nieuwerkerk aan de IJssel, The Netherlands) to confirm the presence of DNA corresponding to the donor. All donors gave an informed consent for this short tandem repeat (STR) profiling, and for all donors, the STR genotypes were known from a reference sample.

16S rDNA PCR and microarray analysis

16S rDNA PCR and microarray analysis were performed according to [24]. In brief, approximately 1 ng of bacterial DNA was amplified using universal primers for 16S rDNA (Supplementary Table 1). After exonuclease treatment, single-stranded PCR products were hybridised to a (vaginal) microbial flora microarray for 4 h at 37°C. After hybridization, four wash steps were performed, and the slides were dried. Fluorescent signals were scanned using a ScanArray Express 4000 scanner (Packard Bioscience, MA, USA). For each spot, the fluorescent signal and background were measured and the signal/noise ratio was calculated [24].

Spots with a signal-to-noise ratio above 5 were regarded as positive spots.

Results and discussion

Next generation sequencing data and microarray development

Eighteen pools of DNA extracts obtained from 240 clinical cervical brush samples were used to analyse the microbiome by next generation sequencing. After data processing, a total of 338,184 useful sequence reads, distributed equally over the 18 pools [representing different age groups of the donors (Supplementary Fig. 1)] remained. Thereby, a representative dataset is obtained indicative of the vaginal microbial flora of women between 15 and 84 years of age. The sequencing reads correspond to 1,619 different sequences, of which 265 occurred with a frequency greater than 0.01% and 56 sequences had a frequency greater than 0.1%. The sequences were assigned to species, genera or groups within a genus using RDP. Reads corresponding to 88 different genera were found with percentages of sequence reads varying from 59% to single reads (Supplementary Fig. 2 and Supplementary Table 2). More genera are found when a larger next generation sequencing dataset encompassing more women of various ethnic origins is analysed, but all these represent low abundance genera [18]. The next generation sequencing dataset gives an average of the abundance for all 240 women and does not address the variation between women [18], as reflected by different abundances of the genera for the 18 pools (Supplementary Fig. 1). Most sequence reads (59%) correspond to species within the genus *Lactobacillus* (Supplementary Fig. 2 and Supplementary Table 2), which are known to be common inhabitants of the human vagina [21, 27]. The genus *Lactobacillus* is known to encompass more than 125 different species/subspecies. Within our dataset, 22 different *Lactobacillus* species were found with read percentages ranging from 48% (of the 199,433 *Lactobacillus* reads) to single reads (Supplementary Fig. 3 and Supplementary Table 3). Apparently, also at the species level, there are large differences in average abundance, which is in agreement with earlier reports [13–15, 32]. Likewise in our dataset, the two most abundant *Lactobacillus* species were *Lactobacillus iners* and *Lactobacillus crispatus* [18]. Next to the genus *Lactobacillus*, a predominant genus in the next generation sequencing dataset was *Gardnerella* (21%). *Gardnerella vaginalis* (*G. vaginalis*) is commonly found in women with bacterial vaginosis [13, 14, 17, 19, 20]. The presence of *G. vaginalis* is consistent with the fact that the sample set used for the next generation sequencing contained cervical brush samples of women with *Gardnerella*

morphotypes, which was identified in stained cytological slides (data not shown).

To develop the (vaginal) microbial flora microarray, 220 oligonucleotide probes that aim to detect a family, a group of genera, a specific genus, a species or a set of species were designed using the obtained next generation sequencing results. In addition, probes that aimed to detect species known to be common in saliva, in faeces or on skin were added. The targets of the probes match to the data in RDP (August 2011). In total, 389 oligonucleotide 16S rDNA probes (covering 101 genera) were spotted on microarrays to analyse the microbial species in DNA extracts from forensic samples of interest.

Microbial DNA analysis using “human DNA extracts”

In the context of forensic casework, inferring the type(s) of cell material present in an evidentiary sample is only of value when accompanied by information regarding the possible donor of the cells. Consequently, evidentiary traces will be subjected to DNA extraction using methods that comply with human DNA typing. We therefore decided to assess whether these “human DNA extracts” can be used for microbial analysis. Bacterial DNA quantification was performed on 43 vaginal DNA extracts which were obtained using two human DNA extraction methods—DE (resulting in NF and SF) and QIAamp extraction. For comparison, two vaginal samples were subjected to a bacterial DNA extraction method (FastPrep). Table 1a shows that the yield of bacterial DNA is similar for QIAamp extracts, NFs and FastPrep extracts. The SFs on the other hand show a substantially lower yield of bacterial DNA. Similar results were obtained when comparing NFs and SFs originating from the same swab ($n=9$) (Table 1b). These NFs and SFs are obtained during sequential lysis steps within the DE. It appears that most bacterial DNA is extracted in the first

Table 1 Bacterial DNA quantification results for all vaginal swab samples subjected to different DNA extraction methods (A); bacterial DNA quantification results for NF and SF extracts obtained from the same swab (B)

DNA extraction method	Number of samples	Average yield of bacterial DNA (ng) ^a
A		
Differential extraction (NF)	22	238
Differential extraction (SF)	16	46
QIA-amp	5	363
FastPrep	2	295
B		
Differential extraction (NF)	9	295
Differential extraction (SF)	9	44

^a The standard deviation is high

NF non-sperm fraction, SF sperm fraction

(mild) lysis step (the NF). In addition, 25 non-vaginal DNA samples were assayed for the amount of bacterial DNA, and a large variation in DNA yield was observed between body sites. Low yields were found for blood and urine which are sterile body fluids under healthy conditions, although transfer of microbes from the surrounding epithelial layers may have occurred during collection of these samples. High bacterial DNA yields were obtained from faeces, which is concordant with the high mass percentage of bacteria in faeces. Next, most vaginal and all non-vaginal DNA extracts were subjected to human DNA profiling, and 64 of the 68 DNA extracts resulted in full or partial DNA profiles that were concordant with the donor (data not shown). Two skin and two urine DNA extracts did not generate DNA profiles. This was probably due to an insufficient amount of human cellular material, which is common for these types of samples.

Since we found bacterial DNA to be co-extracted when applying human DNA extraction procedures, we proceeded to test whether a wide range of bacterial species had been isolated for the vaginal samples. The 43 vaginal DNA extracts (22 NF, 16 SF and 5 QIAamp extracts) were hybridised to the microbial flora arrays. When all 43 microarray profiles were compiled, 121 of the 389 probes were detected (Supplementary Table 4). On average, 26 probes were detected per DNA extract, with a maximum of 51 probes and a minimum of 11 probes. Although less bacterial DNA was obtained in the SF (Table 1), the average number of detected probes was slightly higher (28 ± 10) than with NF (25 ± 10) or QIAamp extracts (21 ± 9). Similar trends were obtained for the microbial species in NF and SF extracts obtained from the same swab ($n=9$) (Supplementary Fig. 4). The distribution of Gram-positive and Gram-negative species/genera is quite similar for both NF and SF extracts (Supplementary Fig. 4). Thus, a diversity of the species occurs in these vaginal DNA extracts which does not depend on the total amount of microbial DNA that is extracted but appears to differ for extraction conditions.

In summary, DNA extracts obtained by two commonly used human DNA extraction methods contained both human and microbial DNA and seem suited not only for human DNA profiling but also for microbial flora analysis. It may even be possible to use stored DNA extracts from old or cold cases for microbial analyses, although this has not been tested. This may be a benefit over the use of mRNA markers for body fluid identification [6, 7], since it is unlikely that mRNA is present in stored DNA extracts as no measures were taken to extract or preserve the mRNA.

Microbial species as potential identifiers for vaginal origin

Ideally, in order to use microbial species as identifiers for vaginal origin, the corresponding probes should be detected

in all vaginal DNA extracts. We assessed the sensitivity of the microarrays by comparing microarray results to the next generation sequencing dataset. Both sets were obtained with DNA extracts from vaginal samples but differed very much in the methodology (regarding both DNA extraction and analysis method). *Lactobacillus* species are abundant in the next generation sequencing dataset: 59% of the reads correspond to this genus. *Corynebacterium* species on the other hand are much less abundant in the next generation sequencing dataset, with only 0.06% of the reads corresponding to this genus (Supplementary Table 2). When using the microarray, positive signals were obtained for probes targeting species for both the genus *Lactobacillus* and *Corynebacterium*, suggesting that the microarray is able to detect both high and low abundant microbes.

When compiling all vaginal DNA extracts, 121 of the 389 probes were detected corresponding to 39 different genera of bacteria. Sixty-five of these 121 probes were derived from the next generation dataset. Only two of these probes were detected in all vaginal swab extracts and 15 were detected in at least 22 of the 43 DNA extracts (Table 2). When only the DNA extract type with the largest diversity of microbes (the SF) was taken into account, four probes were detected in all SF extracts (Table 2).

Fleming et al. (2010) describe *L. crispatus* and *L. gasseri* as vaginal-specific bacteria which can be used for the identification of vaginal secretions. They used an end-point RT-PCR approach, and the fluorescently labelled amplicons that correspond to the 16S-23S intergenic spacer region were detected by capillary electrophoresis. In that study, *L. crispatus* and *L. gasseri* were detected in all vaginal samples ($n=14$) without being detected in blood, saliva and semen

samples [33]. As shown in Supplementary Fig. 3 and Supplementary Table 3, *L. crispatus* covers 39% of the next generation sequencing reads for *Lactobacillus* species, while *L. gasseri* covers only 0.44% of the reads. The low percentage of reads for *L. gasseri* means either that this species has a low abundance in all women or that only some women carry this species, as was reported before [18]. Also, *L. gasseri* was not detected in 4 of the 18 pools in the next generation sequencing dataset (data not shown), indicating absence in some women. The designed array contains three probes for *L. crispatus/kefirano-faciens* and three probes for *L. gasseri/johnsonii*. For 2 of our 12 donors (corresponding to 3 of the 43 samples), none of these six probes were detected in the microarray profiles. Lactic acid-producing bacteria, like these *Lactobacillus* species, are known to provide a healthy vaginal environment with a low pH. For some women, *Lactobacillus* species can be replaced by other lactic acid-producing bacteria [13, 15, 19, 34, 35], and for the two donors mentioned above, probes for some of these other lactic acid-producing bacteria and also for other *Lactobacillus* species were detected on the microarray. This presents a biological explanation why for some women vaginal identification based on microarray signals for *L. crispatus* and *L. gasseri* may fail. Alternatively, lactic acid-producing bacteria may also be low or absent due to bacterial vaginosis.

Comparing vaginal microbial flora to the flora of other body sites

In addition to the vaginal samples, we analysed 25 samples of other body sites, namely saliva, skin, semen, urine, faeces and blood. The skin samples included body sites which are

Table 2 Probes detected in all or the majority of vaginal DNA extracts or in all SF extracts. The number between brackets reflects the number of probes detected per target

Probe target(s)	All vaginal DNA extracts ($n=43$)	Majority of the vaginal DNA extracts (at least 22)	All SF extracts ($n=16$)
<i>Anaerococcus prevotii/tetradium/lactolyticus</i> (1)		X	
<i>Bacteroides</i> group (1)		X	
<i>Corynebacterium genitalium/imitans/capitovis/appendicis</i> (1)		X	X
<i>Corynebacterium</i> group (1)			X
<i>Gemella</i> group (1)		X	
<i>Lactobacillus crispatus/kefirano-faciens</i> (3)		X	
<i>Lactobacillus iners/uncultured</i> (1)		X	
<i>Lactobacillus salivarius</i> (1)		X	
<i>Lactobacillus gasseri/johnsonii</i> (2)		X	
<i>Lactobacillus panis/pontis/vaginalis/psittaci</i> (1)		X	
<i>Lactobacillus plantarum/paraplantarum/coryniformis/pentosus</i> (1)	X	X	X
<i>Prevotella disiens</i> (1)		X	
<i>Streptococcus</i> group (1)	X	X	X

SF sperm fraction

in close proximity to or can be in contact with the vagina, like (female) groin and penis, thereby challenging the specificity tests. Even though the sample sizes for these body sites are small, some insight regarding body site specificity of the probes will be obtained. In our dataset, no species were detected in all or a high percentage of the vaginal samples, while they were not detected in samples from other body sites. In agreement with Fleming et al. [33], we did not detect *L. crispatus* and *L. gasseri* in DNA extracts isolated from saliva, blood and semen samples, but we did detect *Lactobacillus* species (including *L. crispatus* and *L. gasseri*) in skin samples (from hand, female groin and penis) and in the female urine sample (Table 3). Also, like many other *Lactobacillus* species, *L. crispatus* and *L. gasseri* are known to be commonly isolated from stool samples [36]. There are different explanations for the presence of these species at other body sites; (1) the overlap in the microbial populations inhabiting different body sites [11], (2) the close proximity of and (3) the contact between body sites.

The microarray results illustrate the complexity of establishing a single microbial marker that identifies vaginal origin for all donors among a wide range of body sites, as no single 16S rDNA probe was able to include all vaginal samples and at the same time exclude all samples from other body sites. Consequently, alternative strategies are required when using vaginal microflora markers in a forensic context. These strategies may include: (1) the use of a larger number of vaginal markers of which a subset needs to be detected; (2) the addition of other microbial species that indicate different body sites; and/or (3) evaluation of

microbial flora data in a probabilistic approach, for example resulting in support for hypothesis A or B (body site A versus body site B). We assayed the feasibility to base the microflora analysis on a larger number of genera/species in which a subset of the selected species need to be detected for a positive identification. Within our dataset, 21 probes (corresponding to 11 genera) were detected exclusively in vaginal samples and not in samples from other body sites. These 21 probes together would mark the vaginal origin in 34 of the 43 vaginal DNA extracts, thereby giving a false exclusion for 19% of the vaginal samples. One of the 21 probes, targeting four *Lactobacillus* species (*Lactobacillus panis/pontis/vaginalis/psittaci*), was detected in 51% of the vaginal DNA extracts that corresponded to 8 of the 12 donors (Tables 2 and 3). The probe with a sequence determined as “unclassified *Lachnospiraceae*” was detected in 12% of the vaginal extracts, corresponding to 4 of the 12 donors. Both probes were not always detected in all of the DNA extracts of a donor even when the DNA extracts were isolated with the same method which suggests that the corresponding microbial species may not be sufficiently abundant for robust detection on the microarray or that other factors, e.g. swab type used or sampling time, influence the diversity seen. Thirteen of the 21 probes were detected in just 1 of the 43 vaginal DNA extracts (Supplementary Table 5). Clearly, more analyses are needed to establish true vaginal specificity for these probes.

Another approach to discriminate vaginal and non-vaginal samples could involve probes that can exclude vaginal origin. We have used 25 DNA extracts from non-vaginal body sites.

Table 3 Percentage of DNA extracts per body site in which a specific *Lactobacillus* probe was detected. The number between brackets reflects the number of probes per target detected on the microarray. A sample is regarded positive if at least one of the probes responds

Probe target	% of DNA extracts per body site for which a <i>Lactobacillus</i> probe was detected								
	Vagina n=43	Saliva n=8	Hand n=4	Groin n=3	Penis n=2	Semen n=3	Urine n=3	Faeces n=1	Blood n=1
<i>Lactobacillus brevis</i> (1)	9	75	50	0	0	0	33	0	0
<i>Lactobacillus crispatus/kefiranofaciens</i> (3)	86	0	50	67	100	0	33	0	0
<i>Lactobacillus fermentum</i> (1)	5	0	0	0	0	0	0	0	0
<i>Lactobacillus frumenti/reuteri/secaliphilus/coleohominis</i> (1)	2	0	0	0	0	0	0	0	0
<i>Lactobacillus gasseri/johnsonii</i> (3)	81	0	25	67	50	0	33	0	0
<i>Lactobacillus iners/uncultured</i> (3)	58	0	25	33	50	33	33	0	100
<i>Lactobacillus jensenii/saerimneri/fornicalis/salivarius</i> (1)	42	0	0	67	0	0	33	0	0
<i>Lactobacillus oris/antri</i> (1)	7	0	0	0	0	0	0	0	0
<i>Lactobacillus panis/pontis/vaginalis/nsittaci</i> (1)	51	0	0	0	0	0	0	0	0
<i>Lactobacillus plantarum/paraplantarum/coryniformis/pentosus</i> (1)	100	100	75	100	50	33	100	100	100
<i>Lactobacillus salivarius</i> (1)	67	75	0	33	50	100	0	0	100
<i>Lactobacillus</i> species (1)	0	25	0	0	0	0	0	0	0
<i>Lactobacillus vaginalis</i> (1)	5	0	0	0	0	0	0	0	0

This sample size is too small to realistically study which probes can indeed exclude vaginal origin. Nonetheless, 64 probes were detected in non-vaginal samples only. These 64 probes correspond to a total of 38 genera of which 25 genera of bacteria were detected in non-vaginal samples only (Supplementary Table 4). Due to the small sample size, these probes were not studied any further. However, we did examine the microarray results for *Prevotella* and *Streptococcus* species (like *S. salivarius* [30, 31]), as these microbes are known to colonise the oral cavity [37–39]. Nine probes targeting *Streptococcus* species/groups (including one probe for *S. salivarius*) and 16 probes targeting *Prevotella* species/groups gave a positive result on the (vaginal) microbial flora microarray. Three *Streptococcus* group probes, one *Prevotella* group probe and one probe targeting *Prevotella melanogenica/veroralis*, were detected in all saliva samples. Unfortunately, these probes were also detected in vaginal and other samples, such as skin DNA extracts. One *Streptococcus* group probe even gave a positive result for all of the 43 vaginal samples. In the next generation sequencing data, 0.36% and 3.5% of the reads were assigned to *Streptococcus* and *Prevotella*, respectively, confirming the presence of these genera in the vaginal samples (Supplementary Table 2).

Concluding remarks

This study gives insight in the human vaginal microbial flora and the possibilities of using microbial flora in forensic investigations. A total number of 338,184 next generation sequencing reads were obtained from a set of 240 clinical cervical brush samples. The next generation sequencing dataset consists of 1,619 different sequences representing 88 different genera. The abundance of the various microbial genera shows extreme variation and ranges from <0.01% to 59%. We examined whether microbial flora analysis could be used to indicate vaginal origin in a forensic context. Cell type analysis addresses questions on activity level, which is ideally combined with analysis on source level (human DNA typing) using the same DNA extract. Most often, human DNA profiling precedes the cell type analysis and determines the DNA extraction method that is applied. Therefore, we subjected vaginal samples to two “human DNA extraction” methods and analysed the diversity of the microbes that were co-extracted. To be able to analyse a wide range of microbial species, we used a microarray approach. The designed array carried 389 assessment probes (not counting positive and negative controls), of which, 220 were selected from the vaginal next generation sequencing dataset; the remaining probes corresponded to species known to be present in saliva, in faeces and/or on skin. Both vaginal and non-vaginal samples were analysed, although the number of each type of non-vaginal sample was limited.

The aim of the analysis was to deduce the most promising approach for vaginal identification via microbes: if only few species would suffice to mark vaginal origin, methods like multiplex (RT)-PCR could be applied; if tens (or hundreds) of species are needed, methods like microarray analysis would be more appropriate.

In the next generation sequencing data, most reads corresponded to the genus *Lactobacillus* but on the microarray platform, single probes targeting *Lactobacillus* species could not mark all vaginal samples. Moreover, the species were also detected in non-vaginal samples, especially when residing from body sites that are proximate or can be in contact with the vagina. Neither probes for microbial species other than *Lactobacillus* were found to suffice, which is in agreement with other studies assessing the vaginal microbiome [18, 19]. We infer that a larger set of microbes is required, for which microarrays appear a suitable analysis platform as species with both a high and a low percentage of next generation sequence reads can be detected. Our study suggests that for a future microarray-based assay attempting to determine the vaginal origin of a sample, three types of probes should be included: (1) probes for species detected in vaginal DNA extracts only (but not necessarily in all vaginal DNA extracts), (2) probes for species detected in all or the majority of vaginal DNA extracts and (3) probes for species less common in vaginal DNA extracts but frequently found on other body sites. Microarray analysis of a sample will render a microbial flora pattern that is probably best analysed in a probabilistic approach resulting in support for hypothesis A or B (body site A versus B). Other assays examining sample origin like microscopic analysis, presumptive tests and mRNA profiling may be added to assist and supplement evidence evaluation.

Acknowledgements We thank Frank Schuren, Anita Ouwens and Michel Ossendrijver (TNO Quality of Life, The Netherlands) for the development of the (vaginal) microbial flora microarray and Hans Korporaal for retrieving the 240 clinical samples from the archives of the Leiden Cytology and Pathology Laboratory (The Netherlands). We are thankful to the donors who provided the various types of samples and we thank Rolla Voorhamme for critically reading the manuscript.

References

1. French CE, Jensen CG, Vintiner SK, Elliot DA, McGlashan SR (2008) A novel histological technique for distinguishing between epithelial cells in forensic casework. *Forensic Sci Int* 178:1–6
2. Paterson SK, Jensen CG, Vintiner SK, McGlashan SR (2006) Immunohistochemical staining as a potential method for the identification of vaginal epithelial cells in forensic casework. *J Forensic Sci* 51:1138–1143
3. Virkler K, Lednev IK (2009) Analysis of body fluids for forensic purposes: from laboratory testing to non-destructive rapid confirmatory identification at a crime scene. *Forensic Sci Int* 188:1–17

4. Schulz MM, Buschner MG, Leidig R, Wehner HD, Fritz P, Habig K, Bonin M, Schutz M, Shiozawa T, Wehner F (2010) A new approach to the investigation of sexual offenses-cytoskeleton analysis reveals the origin of cells found on forensic swabs. *J Forensic Sci* 55:492–498
5. Lee HY, Park MJ, Choi A, An JH, Yang WI, Shin KJ (2011) Potential forensic application of DNA methylation profiling to body fluid identification. *Int J Legal Med*. doi:10.1007/s00414-011-0569-2
6. Haas C, Klessner B, Maake C, Bar W, Kratzer A (2009) mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR. *Forensic Sci Int Genet* 3:80–88
7. Fleming RI, Harbison S (2010) The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids. *Forensic Sci Int Genet* 4:244–256
8. Vennemann M, Koppelkamm A (2010) Postmortem mRNA profiling II: practical considerations. *Forensic Sci Int* 203:76–82
9. Virkler K, Lednev IK (2009) Raman spectroscopic signature of semen and its potential application to forensic body fluid identification. *Forensic Sci Int* 193:56–62
10. Sikirzhyskaya A, Sikirzhyski V, Lednev IK (2011) Raman spectroscopic signature of vaginal fluid and its potential application in forensic body fluid identification. *Forensic Sci Int* 193:56–62. doi:10.1016/j.fosciint.2011.08.15, in press
11. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R (2009) Bacterial community variation in human body habitats across space and time. *Science* 326:1694–1697
12. Kuczynski J, Costello EK, Nemergut DR, Zaneveld J, Lauber CL, Knights D, Koren O, Fierer N, Kelley ST, Ley RE, Gordon JI, Knight R (2010) Direct sequencing of the human microbiome readily reveals community differences. *Genome Biol* 11:210
13. Hyman RW, Fukushima M, Diamond L, Kumm J, Giudice LC, Davis RW (2005) Microbes on the human vaginal epithelium. *Proc Natl Acad Sci U S A* 102:7952–7957
14. Zhou X, Bent SJ, Schneider MG, Davis CC, Islam MR, Forney LJ (2004) Characterization of vaginal microbial communities in adult healthy women using cultivation-independent methods. *Microbiology* 150:2565–2573
15. Zhou X, Brown CJ, Abdo Z, Davis CC, Hansmann MA, Joyce P, Foster JA, Forney LJ (2007) Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women. *ISME J*:121–133
16. Thies FL, Konig W, Konig B (2007) Rapid characterization of the normal and disturbed vaginal microbiota by application of 16S rRNA gene terminal RFLP fingerprinting. *J Med Microbiol* 56:755–761
17. Hummelen R, Fernandes AD, Macklaim JM, Dickson RJ, Changalucha J, Gloor GB, Reid G (2010) Deep sequencing of the vaginal microbiota of women with HIV. *PLoS One* 5:e12078
18. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ (2011) Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A* 108(Suppl 1):4680–4687
19. White BA, Creedon DJ, Nelson KE, Wilson BA (2011) The vaginal microbiome in health and disease. *Trends Endocrinol Metab* 22:389–393
20. Hummelen R, Macklaim JM, Bisanz JE, Hammond JA, McMillan A, Vongsa R, Koenig D, Gloor GB, Reid G (2011) Vaginal microbiome and epithelial gene array in post-menopausal women with moderate to severe dryness. *PLoS One* 6:e26602
21. Brenig B, Beck J, Schutz E (2010) Shotgun metagenomics of biological stains using ultra-deep DNA sequencing. *Forensic Sci Int Genet* 4:228–231
22. Boon ME, Kok LP (2008) Theory and practice of combining coagulant fixation and microwave histoprocessing. *Biotech Histochem* 83:261–277
23. www.tno.nl. Accessed 6 January 2010
24. Dols JA, Smit PW, Kort R, Reid G, Schuren FH, Tempelman H, Romke BT, Korporaal H, Boon ME (2011) Microarray-based identification of clinically relevant vaginal bacteria in relation to bacterial vaginosis. *Am J Obstet Gynecol* 204(4):305–307
25. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 37:D141–D145
26. Nocker A, Richter-Heitmann T, Montijn R, Schuren F, Kort R (2010) Discrimination between live and dead cells in bacterial communities from environmental water samples analyzed by 454 pyrosequencing. *Int Microbiol* 13:59–65
27. Forney LJ, Gajer P, Williams CJ, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Brotman RM, Davis CC, Ault K, Ravel J (2010) Comparison of self-collected and physician-collected vaginal swabs for microbiome analysis. *J Clin Microbiol* 48:1741–1748
28. Eschenbach DA, Patton DL, Hooton TM, Meier AS, Stapleton A, Aura J, Agnew K (2001) Effects of vaginal intercourse with and without a condom on vaginal flora and vaginal epithelium. *J Infect Dis* 183:913–918
29. Benschop CC, Wiebosch DC, Kloosterman AD, Sijen T (2010) Post-coital vaginal sampling with nylon flocked swabs improves DNA typing. *Forensic Sci Int Genet* 4:115–121
30. Sweet D, Lorente M, Lorente JA, Valenzuela A, Villanueva E (1997) An improved method to recover saliva from human skin: the double swab technique. *J Forensic Sci* 42:320–322
31. Fierer N, Jackson JA, Vilgalys R, Jackson RB (2005) Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. *Appl Environ Microbiol* 71:4117–4120
32. Verhelst R, Verstraelen H, Claeys G, Verschraegen G, Delanghe J, Van Simaey SL, De Ganck GC, Temmerman M, Vaneechoutte M (2004) Cloning of 16S rRNA genes amplified from normal and disturbed vaginal microflora suggests a strong association between *Atopobium vaginae*, *Gardnerella vaginalis* and bacterial vaginosis. *BMC Microbiol* 4:16
33. Fleming RI, Harbison S (2010) The use of bacteria for the identification of vaginal secretions. *Forensic Sci Int Genet* 4:311–315
34. Witkin SS, Linhares IM, Giraldo P (2007) Bacterial flora of the female genital tract: function and immune regulation. *Best Pract Res Clin Obstet Gynaecol* 21:347–354
35. Yamamoto T, Zhou X, Williams CJ, Hochwalt A, Forney LJ (2009) Bacterial populations in the vaginas of healthy adolescent women. *J Pediatr Adolesc Gynecol* 22:11–18
36. Song Y, Kato N, Liu C, Matsumiya Y, Kato H, Watanabe K (2000) Rapid identification of 11 human intestinal *Lactobacillus* species by multiplex PCR assays using group- and species-specific primers derived from the 16S-23S rRNA intergenic spacer region and its flanking 23S rRNA. *FEMS Microbiol Lett* 187:167–173
37. Nakanishi H, Kido A, Ohmori T, Takada A, Hara M, Adachi N, Saito K (2009) A novel method for the identification of saliva by detecting oral streptococci using PCR. *Forensic Sci Int* 183:20–23
38. Donaldson AE, Taylor MC, Cordiner SJ, Lamont IL (2010) Using oral microbial DNA analysis to identify expired bloodspatter. *Int J Legal Med* 124:569–576
39. Power DA, Cordiner SJ, Kieser JA, Tompkins GR, Horswell J (2010) PCR-based detection of salivary bacteria as a marker of expired blood. *Sci Justice* 50:59–63