Journal of Medical Microbiology (2007), **56**, 755–761 DOI 10.1099/jmm.0.46562-0

Rapid characterization of the normal and disturbed vaginal microbiota by application of 16S rRNA gene terminal RFLP fingerprinting

Frank L. Thies, Wolfgang König and Brigitte König

Institute of Medical Microbiology, Otto-von-Guericke-University Magdeburg, 39120 Magdeburg, Germany

Correspondence
Frank L. Thies
frank.thies@medizin.uni-magdeburg.de

Received 8 February 2006
Accepted 31 January 2007

Bacterial vaginosis (BV) is a prevalent infection in women of reproductive age associated with numerous sequelae, including preterm delivery, amniotic fluid infections and an increased risk of acquiring sexually transmitted diseases. The vaginal microbiota in BV patients is characterized by a shift from lactobacilli to a diverse spectrum of mostly anaerobic bacteria. In this study, terminal restriction fragment length polymorphism (T-RFLP) was used to characterize the vaginal bacterial communities from 50 women with BV and 20 healthy subjects. In the BV samples, 23 species or phylotypes from 17 genera could be identified, including *Atopobium vaginae*, *Megasphaera* sp., *Lactobacillus iners*, *Gardnerella vaginalis* and three recently described phylotypes from the order *Clostridiales*. The number of detected species or phylotypes was on average 6.3 per sample (range 2–14). In contrast, in normal samples, only *Lactobacillus* species could be identified. In conclusion, T-RFLP provides a rapid and reliable technique to investigate the diversity of the predominant vaginal microbiota and allows differentiation of the flora of BV and healthy women. As such, T-RFLP may be helpful both in the diagnosis of BV from vaginal fluids and in a better understanding of the bacterial succession involved in the aetiology of BV.

INTRODUCTION

The composition of the human vaginal microbiota is affected by several host factors, including, among others, age, menarche, sexual activity, pregnancy and the use of contraceptives or spermicides, as well as individual habits such as douching (Wilson, 2005). At menarche, lactobacilli become the dominating organisms, but numerous other microbes, such as anaerobic bacteria and *Candida* spp., may be detectable in the vagina of the healthy woman. The various *Lactobacillus* spp. inhabiting the vagina are assumed to confer protection against the overgrowth of potentially pathogenic bacteria.

Bacterial vaginosis (BV) is a common disease with a reported prevalence ranging from 4 to 40 % in diverse populations and is associated with a number of severe gynaecological and obstetric complications (Sobel, 2000). From the microbiological point of view, BV is a polymicrobial syndrome characterized by disturbance of the normal vaginal microbiota and appears to be the result of an aetiologically unknown process in which the physiological microbial flora is replaced by a still ill-defined set of mostly anaerobic bacteria.

Based on conventional culture identification, an association has been found between BV and the detection of *Gardnerella vaginalis*, *Mobiluncus mulieris* and anaerobic Gram-negative bacteria, especially *Prevotella* spp. Recent findings, mostly based on molecular methods, have emphasized the association between *Atopobium vaginae* and BV (Ferris *et al.*, 2004). However, the alleged pivotal role of these bacteria has been questioned, as they were also found in the vaginal fluid of healthy women. Therefore, recent research has focused on the development of culture-independent methods for in-depth analysis of vaginal microbiota. Detailed information on the complex microbial communities in BV as well as under normal conditions has been obtained by cloning and sequencing of 16S rRNA gene libraries (Verhelst *et al.*, 2004; Zhou *et al.*, 2004; Hyman *et al.*, 2005; Fredricks *et al.*, 2005). As a result of this work, in addition to *Atopobium vaginae*, relatively unknown bacterial species such as '*Leptotrichia amnionii*' (Shukla *et al.*, 2002), *Eggerthella* sp., *Sneathia sanguinegens* and *Megasphaera* sp. have been implicated in the pathogenesis of BV.

Diagnosis of BV is based on the presence of clinical signs as well as microbiological findings. Among the many

Abbreviations: BV, bacterial vaginosis; BVAB, bacterial vaginosis-associated bacterium; 6-FAM, 6-carboxyfluorescein; T-RF, terminal restriction fragment; T-RFLP, terminal restriction fragment length polymorphism.

laboratory methods used for the diagnosis of BV, the Gram-stain criteria as defined by Nugent *et al.* (1991) are currently regarded as the standard procedure. Although this method has been repeatedly refined (Pereira *et al.*, 2005; Verhelst *et al.*, 2005), morphological assessment alone may not reflect the microbiological variations among BV individuals and should be extended by molecular methods (Fredricks & Marrazzo, 2005).

Terminal RFLP (T-RFLP) is a promising molecular approach for the analysis of microbial ecosystems that is increasingly being used by microbiologists for the study of complex human microbiota in health and disease (Rogers *et al.*, 2004; Sakamoto *et al.*, 2004). Unlike competing methods used to analyse microbial communities, such as denaturing gradient gel electrophoresis and single-strand conformation polymorphism, T-RFLP offers the advantage that appropriate computer software can generate *in silico* predictions of terminal restriction fragments (T-RFs), thereby permitting straightforward identification of bacterial species from fingerprinting data. As analysis of clone libraries is time-consuming as well as cost-intensive, it cannot be applied in microbiological laboratories for routine purposes. Therefore, we attempted to establish T-RFLP profiling as a rapid and relatively inexpensive method for community analysis of the vaginal fluid. This will allow the application of T-RFLP as a molecular tool whenever an in-depth analysis of the vaginal flora is desirable, for example within a clinical studies framework.

METHODS

Sample collection and grading. Vaginal swabs were obtained from women of reproductive age attending the Obstetrics and Gynecology Clinic of the University of Magdeburg, Germany. The presence or absence of clinical criteria for BV were not known to the investigators. BV status was assessed microscopically after Gram staining according to Nugent's scoring system (Nugent *et al.*, 1991). Only swabs from women diagnosed as BV-positive (Nugent's score >6) or BV-negative (Nugent's score <4) were processed further. The first 50 BV-positive and the first 20 BV-negative samples were included in this study.

DNA isolation. For DNA isolation, a QIAamp DNA mini kit (Qiagen) was used according to the manufacturer's instructions, with minor modifications. Swab specimens were swirled in 180 µl lysis buffer containing 1 % (w/v) Triton X-100, 0.5 % (w/v) Tween 20 and 1 mmol EDTA in 10 mM Tris/HCl (pH 8.0). Mutanolysin (2 µl containing 50 U; Sigma) was added to the lysate and incubated for 30 min at 37 °C. After mixing samples with 200 µl buffer AL (Qiagen) and 20 µl proteinase K (20 mg ml^{-1}), samples were incubated for 30 min at 56 °C followed by 15 min at 95 °C. After adding 200 µl ethanol, samples were loaded onto a QIAmp spin column, washed and eluted in 200 µl buffer AE (Qiagen), as described by the manufacturer. DNA samples were stored at –20 °C until further use.

T-RFLP analysis

(i) PCR and digestion. Universal bacterial primers 27F (5′-AGAGTTTGATCCTGGCTCAG-3′) and 926R [5′-CCGTCAATTC(A/C)TTT(A/G)AGTTT-3′] were used to amplify internal fragments of 16S rRNA genes in the genomic DNA obtained from samples. Primers 27F und 926R were labelled at their 5′ ends with the dyes 6-carboxyfluorescein (6-FAM) and 4,7,2′,4′,5′,7′-hexachloro-6-carboxyfluorescein (HEX), respectively. All primers were obtained from MWG-Biotech. PCR mixtures comprised PCR buffer, 1.5 mM $MgCl_2$, each deoxynucleoside triphosphate at a concentration of 0.2 mM, each primer at a concentration of 0.2 µM and 1 U *Taq* polymerase (Qiagen) in a final volume of 50 µl. The cycling program was performed using a Perkin-Elmer 2400 thermocycler with the following conditions: an initial denaturation step at 95 °C for 5 min, followed by 35 cycles of denaturation for 30 s at 95 °C, annealing for 30 s at 50 °C and elongation for 1 min at 72 °C, with a final additional elongation step for 7 min at 72 °C. PCR products were analysed on 1 % agarose gels stained with ethidium bromide. After purification using the QIAquick PCR purification kit, PCR products were digested with *Hin*fI, *Hha*I and *Msp*I (New England Biolabs). Digestion reactions were performed separately for 4 h at 37 °C, followed by treatment for 20 min at 65 °C for enzyme inactivation.

(ii) T-RF length analysis. The lengths of T-RFs were determined by electrophoresis with a model 3100 automated sequencer (Applied Biosystems Instruments), as described previously (Trotha *et al.*, 2002). In brief, samples were prepared by combining 1 µl restriction digestion product, 21 µl Hi-Di formamide solution (Applied Biosystems), 0.5 µl Genescan 500 ROX size standard (Applied Biosystems) and 0.5 µl of a 780 bp ROX-labelled DNA fragment (12 ng μl^{-1}). After electrophoresis, the lengths of T-RFs were determined by comparison with the internal standard, using the local Southern algorithm as the size-calling method (Southern, 1979). Only peaks with heights exceeding 50 fluorescence units were evaluated.

(iii) *In silico* analysis. An in-house software program (JOparin; unpublished) permitted microbial identification from T-RF data. The program performed an *in silico* digestion of all 16S rRNA gene sequences available in the 16S rRNA gene database (release 9.24) downloaded from the RDP-II website (http://rdp.cme.msu.edu/misc/resources.jsp) (Maidak *et al.*, 2001). The original dataset contained 119 821 sequences, but all entries without unambiguous identification of the bacterial source were skipped, leaving 52 725 rRNA gene sequences. To this well-annotated database generated so far, we added 439 rRNA gene sequences from Hyman *et al.* (2005) that have been deposited in GenBank (GenBank accession nos AY958774–AY959212). Sample T-RF patterns were compared with all *in silico* patterns and a matching score (RF score) was assigned to each database entry. The RF score was computed from the size-weighted sum of differences between observed and predicted T-RF lengths.

T-RF sequencing. To prove experimentally the correct *in silico* identification of T-RFs, the 16S rRNA gene was amplified from genomic DNA with unlabelled primers 27F and 926R using the cycling conditions as described above. The PCR product was column purified and digested with an appropriate restriction enzyme (*Hin*fI, *Hha*I or *Msp*I). The restriction fragments were separated in a 2.5 % agarose gel and the fragment corresponding to the T-RF was cut from the gel and purified (QIAquick gel extraction kit; Qiagen). Cycle sequencing of this fragment with 27F or 926R as the sequencing primer was performed with the Big Dye terminator kit (Applied Biosystems) as recommended by the manufacturer. Sequences were analysed using the BIBI software tool (http://pbil.univ-lyon1.fr/bibi/query.php) (Devulder *et al.*, 2003).

Statistical analysis. Basic summary statistics and hierarchical cluster analysis were carried out using the free software package R (R Development Core Team, 2005). For cluster analysis, we constructed a binary matrix with bacterial species (or respective taxonomic units) as rows and BV samples as columns considering only the presence or absence of species. Similarity, i.e. frequency of co-occurrence, between species was measured with Jaccard's coefficient, which is equal to the number of samples in which both species

occur, divided by the total number of samples with either species present (Riley, 2004). Hierarchical clustering was performed by the unweighted pair group method using arithmetic averages (UPGMA).

RESULTS AND DISCUSSION

Vaginal samples from 70 women were used in our study: 50 BV-positive (score >6) and 20 BV-negative (score <4), according to Nugent's criteria. Genomic DNA was isolated from all 70 specimens and T-RFLP patterns were generated.

Reproducibility and accuracy of the T-RFLP assay

Initially, we investigated the technical reproducibility of T-RFLP patterns. PCR, enzyme digestions and T-RFLP analysis were performed on three vaginal samples. After 3 weeks, the procedures were repeated using the same DNA, which had been stored at –20 °C. Comparison of corresponding T-RFLP patterns demonstrated excellent reproducibility, confirming the observations of Osborn *et al.* (2000). The mean difference between T-RFs was <1 bp. The peak heights displayed a higher level of variation; however, this had no impact on species identification (data not shown).

Differences between observed and predicted T-RF sizes usually were within –2 to +2 bp of the predicted T-RF size. However, the differences grew larger for T-RFs >500 bp (data not shown). This T-RF drift has been reported previously and appears to be affected by subtle differences in molecular mass, from either purine content or dye label (Kaplan & Kitts, 2003). We observed 109 distinct T-RFs from 50 to 500 bp, which could be attributed to a total of 23 identifiable T-RF patterns. The mean difference between maximum and minimum values of all T-RFs was 1.29 bp, demonstrating the very low variability among observed T-RF sizes corresponding to distinct species or phylotypes from different samples (as a representative example, see Table 1 for a detailed analysis of *Msp*I-generated T-RFs). These data suggested that T-RF length variation due to strain-to-strain differences was negligible and that overall reproducibility, defined as technical and biological reproducibility combined, could be regarded as sufficiently high.

Microbial identification by T-RF profiling

Depending on the size of the database used for prediction, multiple species or phylotypes may be assigned to a single T-RF. Therefore, it is helpful for definite species identification to analyse digests produced using multiple restriction enzymes (Dunbar *et al.*, 2001). Furthermore, reliable *in silico* prediction is dependent on high-quality databases. In this respect, the almost full-length 16S rRNA gene sequences of bacteria from vaginal communities made available by Hyman *et al.* (2005) were particularly useful. In our experimental approach, a typical T-RF pattern consisted of a maximum of six T-RFs and represented an 'operational taxonomic unit' (Felske & Osborn, 2005), which could be a phylotype, a valid species, a proposed species ('*Leptotrichia amnionii*'), a genus, a subgroup within a genus, or, as in the case of the *Enterobacteriaceae*, a family of bacteria. The term 'phylotype' is used for classifying microbes solely on the basis of their 16S rRNA gene sequence, if no other taxonomic information is available.

Table 1. Predicted and experimentally observed T-RF lengths for bacterial organisms frequently (≥10%) detected in BV samples

T-RFs were generated with 6-FAM-labelled primer and digested with *Msp*I. The mean absolute value of the range was 0.65 and the mean difference between observed and predicted lengths was 1.05 (weighted by number of samples, *n*).

Organism	*n*	Observed length (bp)		Predicted length (bp)
		Mean	Range	
Atopobium vaginae	48	161.8	161.3–162.1	163
Megasphaera sp.	34	152.0	151.7–152.2	152
Lactobacillus iners	32	187.3	186.7–187.5	189
Gardnerella vaginalis	32	281.9	281.3–282.1	281
Eggerthella sp.	22	280.5	279.9–281.1	280
Clostridium-like sp. (GenBank no. AY958888)	18	228.8	228.5–229.0	230
Prevotella bivia/*P. buccalis*	26	96.4	95.9–97.1	97
Peptostreptococcus micros	17	225.2	225.1–225.4	226
Uncultured *Prevotella* (GenBank no. AY959212)	15	491.8	491.4–492.2	496
Aerococcus christensenii	12	146.2	145.8–146.5	148
'*Leptotrichia amnionii*'	9	278.7	278.6–278.8	278
Peptoniphilus sp.	9	168.6	168.2–168.8	169
Dialister sp.	8	304.0	303.5–304.3	305
Clostridium-like sp. (GenBank no. AY959097)	8	146.8	146.7–146.9	147
Mycoplasma hominis	5	154.1	154.0–154.1	153

Based on the concordance between observed and predicted fragment sizes, a ranked list of bacterial species was generated. The bacterial species with the highest RF score was regarded as the most probable source of the T-RF pattern. As noticed by others (Kaplan & Kitts, 2003), for some T-RF patterns there was a considerable difference between observed and predicted T-RF sizes, making *in silico* assignments problematic. In all cases with an equivocal assignment, which represented approximately 40 % of all patterns, we sequenced at least one T-RF from the pattern in question. T-RF sequencing provided DNA sequences that were mostly short (300–500 bp), but of sufficient quality. Together with the T-RF pattern, this usually permitted bacterial identification at least to the genus level. Otherwise, further T-RFs were sequenced. Thereafter, due to the high reproducibility of the assay, the assignment was considered accurate if the same T-RF pattern occurred in further samples.

Depending on the T-RF pattern, it was not always possible to obtain a conclusive identification at the species level. For example, *Lactobacillus iners* could be identified unambiguously, whereas *Lactobacillus crispatus* could not be differentiated from *Lactobacillus suntoryeus*, *Lactobacillus amylovorus*, *Lactobacillus kalixensis* or *Lactobacillus ultunensis* (further designated the *Lactobacillus crispatus* group). Similarly, *Mobiluncus mulieris* was indistinguishable from *Mobiluncus curtisii*, leading to our diagnosis as *Mobiluncus* sp. Assigning valid species names was particularly difficult for bacteria from the genus *Prevotella*. At least three distinct T-RF patterns were detected, which corresponded to *Prevotella bivia*, *Prevotella buccalis* and an uncultured *Prevotella* sp. (most similar to GenBank accession no. AY959212). However, T-RF sequencing showed that the first two T-RF patterns comprised more than a single phylotype (mostly uncultured *Prevotella* spp.) and were therefore designated *P. bivia* group and *P. buccalis* group. Due to the low variation among the 16S rRNA genes of bacteria from the family *Enterobacteriaceae*, it was not possible to achieve identification at the genus level with reasonable certainty. Therefore, bacteria accounting for such T-RF patterns were designated 'enterobacteria'.

Not all T-RFs could be assigned to a bacterial taxon. In such cases, no corresponding database entry may exist, or pseudo-T-RFs may be produced by incomplete digestion (Osborn *et al.*, 2000) or by a malfunctioning of the PCR process (Egert & Friedrich, 2003). In our experience, incomplete digestion did occur, especially if a species was abundant in a bacterial community (see, for example, the peaks corresponding to *Atopobium vaginae* in Fig. 1a). However, the computer software we used was capable of including such information so that species identification might be further supported by partially digested terminal

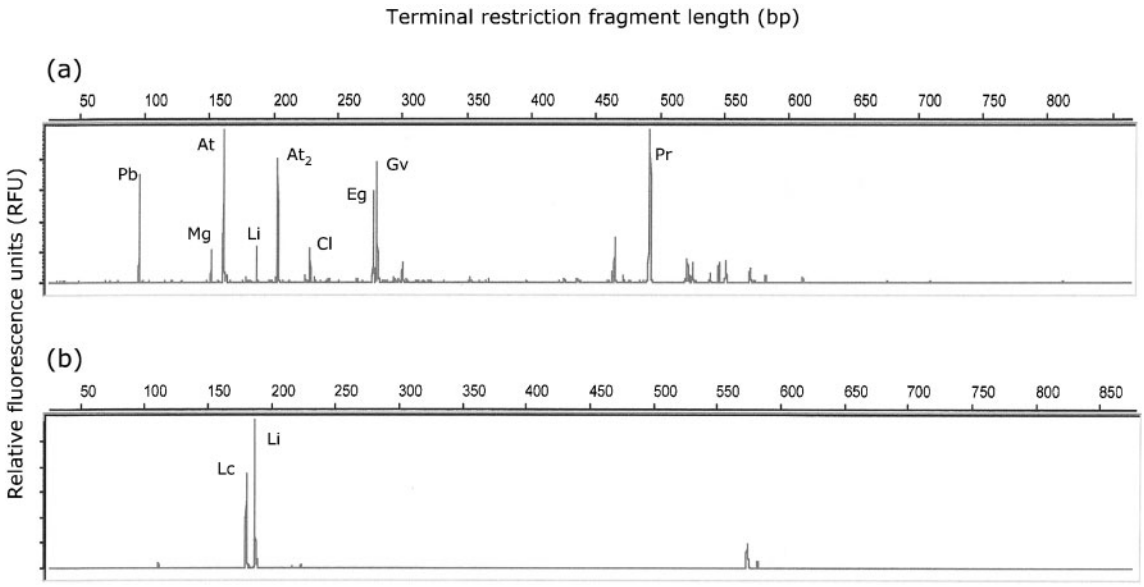

Fig. 1. Examples of T-RFLP patterns from vaginal smears. The 16S rRNA gene was amplified with primers 27F and 926R and amplicons were digested with *Msp*I. Only 6-FAM-labelled T-RFs are displayed. (a) Bacterial vaginosis sample; (b) normal sample. Pb, *P. bivia*; Mg, *Megasphaera* sp.; At, *Atopobium vaginae*; Li, *Lactobacillus iners*; At_2, *Atopobium vaginae* (incomplete digestion); Cl, *Clostridium*-like sp. (GenBank accession no. AY958888); Eg, *Eggerthella* sp.; Gv, *G. vaginalis*; Pr, uncultured *Prevotella* sp. (GenBank accession no. AY959212); Lc, *Lactobacillus crispatus* group. The range of values of the ordinate of T-RFLP patterns was 0–800 RFU (a) and 0–2040 RFU (b). Note that some minor peaks could not be assigned to a bacterial taxon.

fragments. However, in some rare cases involving co-occurrence of very closely related species or strains, partial digests made a definite diagnosis extremely difficult.

By imposing a subtle amplification bias, selection of universal primers may influence the spectrum of bacterial taxa detected by the T-RFLP technique. Based on simulation studies, primer pair 27F and 926R was favoured by Liu *et al.* (1997) and utilized, for example, by Zhou *et al.* (2004) for creating 16S rRNA gene clone libraries from vaginal fluid samples. Several other groups used different broad-range 16S rRNA gene PCR primers for analysis of vaginal microbiota (Fredricks & Marrazzo, 2005). However, as the experimental settings of these studies differed substantially, the impact of primer choice is difficult to determine.

Analysis of the BV-associated bacterial community

In the 50 BV samples, T-RFLP displayed a high level of bacterial diversity. Twenty-three different T-RF patterns corresponding to 'operational taxonomic units' (further named bacterial species or phylotypes) were identified unambiguously (Table 2). On average, in each BV sample, 6.3 T-RF patterns were found (range 2–14). Fig. 1(a) shows a typical T-RFLP pattern. The frequency at which each of these 23 species or phylotypes occurred in the BV sample set ranged from 96 % (*Atopobium vaginae*, the dominating bacterium in our BV samples) to 2 % (*Lactobacillus crispatus*, present in a single sample). An uncultured *Megasphaera* sp., *Lactobacillus iners* and *G. vaginalis* were frequently detected (68, 64 and 64 %, respectively). The 16S rRNA gene sequence with the most significant RF score for the *Megasphaera* T-RF pattern was GenBank accession number AY959093, which showed highest similarity (95 %), among all unambiguously annotated sequences, to the *Megasphaera elsdenii* 16S rRNA gene. Several bacterial species or phylotypes occurred with moderate relative frequencies from 44 to 10 %: *Eggerthella* sp., several *Clostridium*-like and *Prevotella* spp., *Peptostreptococcus* (*Micromonas*) *micros*, *Aerococcus christensenii*, '*Leptotrichia amnionii*', *Peptoniphilus* sp., *Dialister* sp., *Mycoplasma hominis* and bacteria from the family *Enterobacteriaceae*. The three clostridial T-RFs showed highest concordance with the sequences of GenBank accession numbers AY958888, AY959097 and AY995273, respectively. The corresponding bacteria were designated bacterial vaginosis-associated bacterium 2 (BVAB2), BVAB1 and BVAB3 by Fredricks *et al.* (2005), as they were only distantly related to known clostridia (16S rRNA gene identity ~90 %). In our BV sample set, BVAB2 occurred with highest frequency (36 %), followed by BVAB1 (18 %). BVAB3 was only rarely detected (4 %). In addition to the aforementioned bacteria, five further species were identified in <10 % of all specimens: *Sneathia sanguinegens*, *Anaerococcus tetradius*, *Mobiluncus* sp., *Finegoldia magna* and *Lactobacillus crispatus*. Although high concentrations of *Mobiluncus*-like morphotypes, visible on a Gram stain, are commonly regarded as indicative of BV (with Nugent scores of 9 or 10), we could identify *Mobiluncus* sp. in only two samples. This surprising result is in concordance with data from Fredricks *et al.* (2005), who detected *Mobiluncus mulieris* in only one expanded rRNA gene library (out of nine patients with BV) when 420 clones of this library were sequenced. It is further noteworthy that in the study of Hyman *et al.* (2005), in which approximately 1000 16S rRNA gene library clones from the vaginal fluids of 20 premenopausal women were sequenced, no *Mobiluncus*-typical sequences were found. It may be speculated that in older studies, which relied partially on microscopical examination of vaginal smears, the frequency of *Mobiluncus* sp. was overestimated. Fluorescent *in situ* hybridization technology has demonstrated that at least one uncultured BV-associated bacterium (BVAB1) has a curved rod morphology very similar to that of the *Mobiluncus* morphotype (Fredricks & Marrazzo, 2005).

We defined microbial abundance as the number of complete T-RF patterns that could be correlated unambiguously

Table 2. Number of vaginal samples harbouring bacterial species or phylotypes

Phylotype/species	BV samples (n=50)	Samples from healthy subjects (n=20)
Atopobium vaginae	48	
Megasphaera sp. (GenBank no. AY959093)	34	
Lactobacillus iners	32	11
Gardnerella vaginalis	32	
Eggerthella sp.	22	
Clostridium-like sp. (GenBank no. AY958888)	18	
Prevotella bivia group	18	
Peptostreptococcus micros	17	
Prevotella sp. (GenBank no. AY959212)	15	
Aerococcus christensenii	12	
'*Leptotrichia amnionii*'	9	
Peptoniphilus sp.	9	
Bacterium from the order *Clostridiales* (GenBank no. AY959097)	8	
Dialister sp.	8	
Prevotella buccalis group	8	
'Enterobacteria'	5	
Mycoplasma hominis	5	
Sneathia sanguinegens	3	
Anaerococcus tetradius	3	
Clostridium-like sp. (GenBank no. AY995273)	2	
Mobiluncus sp.	2	
Finegoldia magna	2	
Lactobacillus crispatus group	1	14
Lactobacillus gasseri group		2

with a bacterial taxon, i.e. as the number of identified taxa in the sample. This has the advantage that incomplete digestion and other processes leading to spurious T-RFs did not bias the estimation of microbial abundance. However, the drawback of this approach is that the estimate may be too conservative, and sometimes bacteria corresponding to indistinct T-RFs may be missed. Compared with the results of Fredricks *et al.* (2005), who reported a mean of 12.6 phylotypes per clone library, the mean number of T-RF patterns observed per sample in our study (6.3) was relatively low. However, this was not surprising, as sequencing of clones is a more discriminatory assay than T-RFLP profiling, especially when closely related species or clones are concerned. For example, Fredricks *et al.* (2005) could distinguish among nine different *Prevotella* spp., whereas T-RFLP only afforded discrimination of three separate groups of *Prevotella* phylotypes. As well as employing a different genetic-based assay for examining the vaginal microbiota, the selection of subjects differed between our study and that of Fredricks *et al.* (2005), which may further account for the discrepancies in microbial diversity seen between these studies.

In order to search systematically for the co-occurrence of phylotypes within the BV microbiota, we used hierarchical cluster analysis as an explanatory data analysis tool. The resulting dendrogram displayed one cluster formed by *Atopobium vaginae*, *Lactobacillus iners*, *G. vaginalis* and *Megasphaera* sp. (Fig. 2). However, this may simply reflect the abundance of these species in the BV samples. In 22 % of the BV-positive samples, all four species were detected by T-RFLP, whilst at least three were seen in 64 % of BV samples. Zhou *et al.* (2004) reported that *Megasphaera* sp. and *Atopobium vaginae* were highly coincident in vaginal 16S rRNA gene libraries, with one being found only when the other was also present, and Verhelst *et al.* (2004) demonstrated the co-occurrence of *Atopobium vaginae* and *G. vaginalis* in BV samples based on species-specific PCR analysis. In the current study, we observed that *Megasphaera* sp. was only found in BV-positive samples when *Atopobium vaginae* was present, although this may simply reflect the abundance of *Atopobium vaginae* in our BV-positive samples. However, we did not observe a strict co-existence, as *Atopobium vaginae* was detected without concomitant *Megasphaera*-typical T-RFs in 14 BV-positive samples (Table 2). This was similar for the co-existence of *Atopobium vaginae* and *G. vaginalis*, with 16 BV-positive samples shown to contain *Atopobium vaginae* but not *G. vaginalis*. These discrepancies between the current data and earlier findings may reflect differences in the sensitivity of T-RFLP and other molecular-based assays.

In one BV sample, *Lactobacillus crispatus* was identified alongside *Megasphaera* sp., *Atopobium vaginae*, *G. vaginalis*, '*Leptotrichia amnionii*' and *Dialister* sp. Identification was confirmed by T-RF sequencing. This is an interesting observation, as *Lactobacillus crispatus* is commonly considered to be an indicator species for the healthy vaginal microbiota (Antonio *et al.*, 1999).

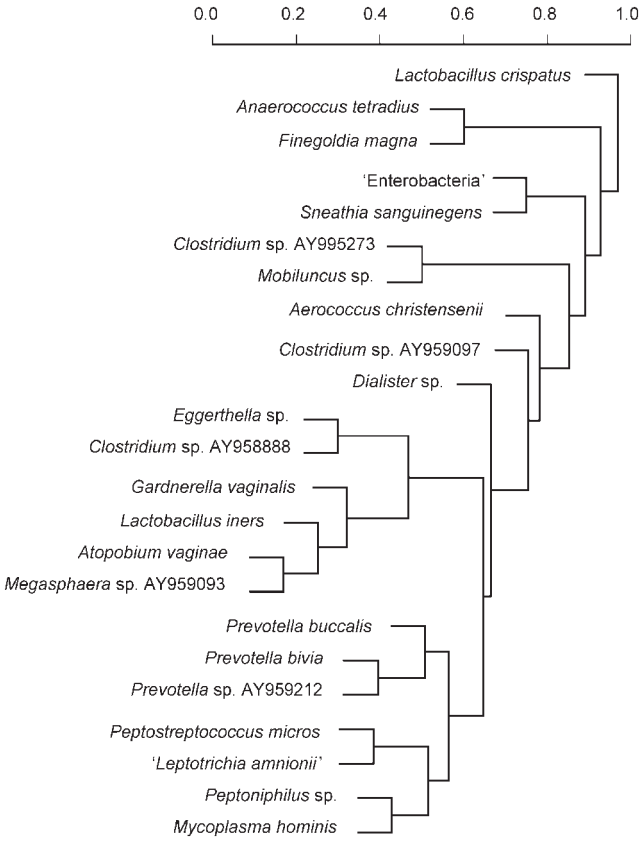

Fig. 2. Hierarchical clustering showing the co-occurrence of bacteria in 50 BV samples. Jaccard's coefficient was used as a similarity measure and UPGMA as an agglomeration method. The scale bar indicates dissimilarity.

Analysis of the vaginal bacterial community not associated with BV

In 20 samples with normal vaginal microbiota, only *Lactobacillus* spp. could be detected using our T-RFLP approach (see Fig. 1b for a typical T-RFLP pattern). Lactobacilli from the *Lactobacillus crispatus* group were detected in 70 % of BV-negative samples, whilst *Lactobacillus iners* was seen in 55 % and bacteria from the *Lactobacillus gasseri* group (comprising *Lactobacillus gasseri*, *Lactobacillus acidophilus* and *Lactobacillus johnsonii*) in 10 % of non-BV samples. The fact that only *Lactobacillus* spp. were observed in the BV-negative samples provided a clear distinction from the BV-positive samples. The discrepancy of this observation (only lactobacilli seen in non-BV samples) compared with other studies is probably due to the competitive nature of the T-RFLP amplification technique. If a mixed bacterial community comprises highly abundant (or predominant) species, subdominant species or phylotypes may be below the detection limit of T-RFLP. However, such subdominant bacterial groups may readily be detectable by specific PCR or culturing on selective media.

Conclusions

We have established T-RFLP profiling as a molecular tool for the routine analysis of the vaginal microbiota. T-RFLP profiling accurately distinguished between all BV-positive and BV-negative samples, as assessed by Nugent's Gram-stain criteria. As BV is a polymicrobial disease, it is important to consider the microbial community as a whole, rather than focusing on single presumptive 'key organisms'. T-RFLP profiling may prove particularly valuable in an attempt to characterize subgroups of women who are more prone to certain complications of BV, such as preterm birth or amniotic fluid infection. Thus T-RFLP profiling may contribute to a better understanding of the aetiology of BV and its complications.

REFERENCES

Antonio, M. A., Hawes, S. E. & Hillier, S. L. (1999). The identification of vaginal *Lactobacillus* species and the demographic and microbiologic characteristics of women colonized by these species. *J Infect Dis* **180**, 1950–1956.

Devulder, G., Perrière, G., Baty, F. & Flandrois, J. P. (2003). BIBI, a bioinformatics bacterial identification tool. *J Clin Microbiol* **41**, 1785–1787.

Dunbar, J., Ticknor, L. O. & Kuske, C. R. (2001). Phylogenetic specificity and reproducibility and new method for analysis of terminal restriction fragment profiles of 16S rRNA genes from bacterial communities. *Appl Environ Microbiol* **67**, 190–197.

Egert, M. & Friedrich, M. W. (2003). Formation of pseudo-terminal restriction fragments, a PCR-related bias affecting terminal restriction fragment length polymorphism analysis of microbial community structure. *Appl Environ Microbiol* **69**, 2555–2562.

Felske, A. & Osborn, A. M. (2005). DNA fingerprinting of microbial communities. In *Molecular Microbial Ecology*, pp. 65–90. Edited by A. M. Osborn & C. J. Smith. New York: Taylor & Francis.

Ferris, M. J., Masztal, A., Aldridge, K. E., Fortenberry, J. D., Fidel, P. L., Jr & Martin, D. H. (2004). Association of *Atopobium vaginae*, a recently described metronidazole resistant anaerobe, with bacterial vaginosis. *BMC Infect Dis* **4**, 5–13.

Fredricks, D. N. & Marrazzo, J. M. (2005). Molecular methodology in determining vaginal flora in health and disease: its time has come. *Curr Infect Dis Rep* **7**, 463–470.

Fredricks, D. N., Fiedler, T. L. & Marrazzo, J. M. (2005). Molecular identification of bacteria associated with bacterial vaginosis. *N Engl J Med* **353**, 1899–1911.

Hyman, R. W., Fukushima, M., Diamond, L., Kumm, J., Giudice, L. C. & Davis, R. W. (2005). Microbes on the human vaginal epithelium. *Proc Natl Acad Sci U S A* **102**, 7952–7957.

Kaplan, C. W. & Kitts, C. L. (2003). Variation between observed and true Terminal Restriction Fragment length is dependent on true TRF length and purine content. *J Microbiol Methods* **54**, 121–125.

Liu, W. T., Marsh, T. L., Cheng, H. & Forney, L. J. (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* **63**, 4516–4522.

Maidak, B. L., Cole, J. R., Lilburn, T. B., Parker, C. T., Jr, Saxman, P. R., Farris, R. J., Garrity, G. M., Olsen, G. J., Schmidt, T. M. & Tiedje, J. M. (2001). The RDP-II (Ribosomal Database Project). *Nucleic Acids Res* **29**, 173–174.

Nugent, R. P., Krohn, M. A. & Hillier, S. L. (1991). Reliability of diagnosing bacterial vaginosis is improved by a standardized method of Gram stain interpretation. *J Clin Microbiol* **29**, 297–301.

Osborn, A. M., Moore, E. R. & Timmis, K. N. (2000). An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. *Environ Microbiol* **2**, 39–50.

Pereira, L., Culhane, J., McCollum, K., Agnew, K. & Nyirjesy, P. (2005). Variation in microbiologic profiles among pregnant women with bacterial vaginosis. *Am J Obstet Gynecol* **193**, 746–751.

R Development Core Team (2005). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Riley, L. W. (editor) (2004). Analysis of similarity and relatedness in molecular epidemiology. In *Molecular Epidemiology of Infectious Diseases*, pp. 91–124. Washington, DC: American Society for Microbiology.

Rogers, G. B., Carroll, M. P., Serisier, D. J., Hockey, P. M., Jones, G. & Bruce, K. D. (2004). Characterization of bacterial community diversity in cystic fibrosis lung infections by use of 16S ribosomal DNA terminal restriction fragment length polymorphism profiling. *J Clin Microbiol* **42**, 5176–5183.

Sakamoto, M., Huang, Y., Ohnishi, M., Umeda, M., Ishikawa, I. & Benno, Y. (2004). Changes in oral microbial profiles after periodontal treatment as determined by molecular analysis of 16S rRNA genes. *J Med Microbiol* **53**, 563–571.

Shukla, S. K., Meier, P. R., Mitchell, P. D., Frank, D. N. & Reed, K. D. (2002). *Leptotrichia amnionii* sp. nov., a novel bacterium isolated from the amniotic fluid of a woman after intrauterine fetal demise. *J Clin Microbiol* **40**, 3346–3349.

Sobel, J. D. (2000). Bacterial vaginosis. *Annu Rev Med* **51**, 349–356.

Southern, E. M. (1979). Measurement of DNA length by gel electrophoresis. *Anal Biochem* **100**, 319–323.

Trotha, R., Reichl, U., Thies, F. L., Sperling, D., König, W. & König, B. (2002). Adaption of a fragment analysis technique to an automated high-throughput multicapillary electrophoresis device for the precise qualitative and quantitative characterization of microbial communities. *Electrophoresis* **23**, 1070–1079.

Verhelst, R., Verstraelen, H., Claeys, G., Verschraegen, G., Delanghe, J., Van Simaey, L., De Ganck, C., Temmerman, M. & Vaneechoutte, M. (2004). Cloning of 16S rRNA genes amplified from normal and disturbed vaginal microbiota suggests a strong association between *Atopobium vaginae*, *Gardnerella vaginalis* and bacterial vaginosis. *BMC Microbiol* **4**, 16–25.

Verhelst, R., Verstraelen, H., Claeys, G., Verschraegen, G., Van Simaey, L., De Ganck, C., De Backer, E., Temmerman, M. & Vaneechoutte, M. (2005). Comparison between Gram stain and culture for the characterization of vaginal microbiota: definition of a distinct grade that resembles grade I microbiota and revised categorization of grade I microbiota. *BMC Microbiol* **5**, 61–71.

Wilson, M. (editor) (2005). The reproductive system and its indigenous microbiota. In *Microbial Inhabitants of Humans: their Ecology and Role in Health and Disease*, pp. 206–250. Cambridge, UK: Cambridge University Press.

Zhou, X., Bent, S. J., Schneider, M. G., Davis, C. C., Islam, M. R. & Forney, L. J. (2004). Characterization of vaginal microbial communities in adult healthy women using cultivation-independent methods. *Microbiology* **150**, 2565–2573.