MATERIALS AND METHODS

Bacterial strains and culture. Reference bacterial species in this study are listed in Table S3. Species are from either the ATCC, or were clinical isolates from the Detroit Medical Center Microbiology Laboratory; all were confirmed in our laboratory by sequencing the 16S rRNA gene. DNA was extracted from cells from their original agar plates, or in some cases after being transferred to Luria-Bertani (LB), Brain-Heart Infusion (BHI), Rogosa, or Blood agar plates (Difco, Detroit, MI). Cells were stored suspended in 0.5 ml of 15% glycerol-BHI at -80°C or sterile 13% nonfat dry milk, 0.5% yeast extract, 15% glycerol.

Subject and sample collection. The subject was enrolled as part of a study at the Detroit Medical Center (Detroit, MI). The protocol was approved by the Institutional Review Board of Wayne State University. At enrollment, written informed consent was obtained, and assurance of confidentiality was given.

DNA extraction. Bacteria were lysed from vaginal lavages by adding sodium hydroxide and sodium dodecyl sulfate to final concentrations of 133 mM and 10% respectively, followed by vigorous reciprocal shaking at room temperature for 30 m. This lysate was then extracted with phenol:chloroform:isoamyl alcohol (25:24:1) followed by chloroform extraction. The DNA was precipitated with an equal volume of 100% isopropanol, washed once with 70% isopropanol, and dissolved in 100 µl TE (10 mM Tris pH 8, 1 mM EDTA) (1). DNA was stored at -20°C until further use. Control experiments indicated that this method extracted DNA efficiently (detection of at least 40 cells, range 1 – 40 cells) from Gram-positive, Gram-negative, and acid-fast bacteria, using both fluorescent DNA and PCR-based assays. DNA was amplified from extracts of reference

bacterial species as single colonies, using the above protocol or QuickExtract™ Plant DNA Extraction Solution (Epicentre, Madison, WI), modified by heating at 65°C for 30 m, 95°C for 1 m.

PCR primers, blockers, and amplification. Standard PCR was performed on RoboCycler Gradient 96 (Stratagene, La Jolla, CA) in 10 µL reactions overlaid with 15 µL mineral oil. Real-time PCR was performed using Cepheid SmartCycler (Sunnyvale, CA) or LightCycler 480 II (Roche Applied Sciences, Indianapolis IN) thermal cyclers in 25 µl or 20 µl reaction, respectively. The reaction consisted of individually optimized amounts of each primer, 250 µM dNTPs (Invitrogen Corporation, Carlsbad, CA or GenScript, Piscataway, NJ), 20 mM Tris pH 8.3, 3 – 3.5 mM $MgCl_2$, 50 mM KCl, and *Taq* DNA polymerase, 5 units Taq from either GeneChoice (Frederick, MD) or New England Biolabs (Ipswich, MA), 2.5 units GenScript Green *Taq* (Piscataway, NJ), or 1 unit of glycerol-free *Taq* DNA polymerase from APEX (Genesee Scientific, San Diego CA) per 100 µl reaction volume, including either 0.073X SYBR Green I Nucleic Acid Stain (FMC Bio Products, Rockland, ME) or 75 nM SYTO9 (Invitrogen Corporation, Carlsbad, CA) for qPCR reactions.

Primers (Fig. 1, Tables S1 and S2) were designed using BioEdit software (Tom Hall, Ibis Biosciences, Carlsbad, CA) and the Ribosomal Database Project II, Releases 9 and 10 (RDPII, Michigan State University, East Lansing, MI, http://rdp.cme.msu.edu/index.jsp). Candidate sequences were identified by inspection of selected aligned sequences and verified for inclusiveness and specificity using Probe Match on the RDPII website, and for compatibility and lack of internal base pairing using the tools on the Operon website (http://www.operon.com/tools/oligo-analysis-

tool.aspx?). PB primer sequences were chosen to include as many species as possible in the target phylogenetic branch while excluding or minimizing non-targets. Programs for each primer were optimized for annealing temperature and extension times (Table S2) on pure templates (genomic DNA or sequenced lavage amplicon) from one or several species and tested for detection of template from DNA equivalent to at least 1-10 cells. Primer specificity was then validated using reference species (Table S3) and ultimately by sequencing of the amplicons they generated from vaginal samples. Negative controls included mock samples processed at the same time as real samples, and reactions with no added template. Performance of the broad-spectrum primers was compared to published sequences and found to be as good to superior in inclusiveness or sensitivity (Table S1).

Lavage DNA was tested for the presence of inhibitors assaying for a shift in Cq values (Cq = cycle of quantification, the cycle at which product fluorescence is first detected above background) of a spiked DNA control. The spike DNA was amplified from a cloned 542 bp fragment of a modified firefly (*Photinus pyralis*) luciferase gene from pGEM-luc cloned into pGEM-T Easy (both vectors from Promega, Madison, WI) and amplified with pUC universal (5'-CCCAGTCACGACGTTGTAAAACG) and reverse (5'-AGCGGATAACAATTTCACACAGG) primers with the cycling parameters 1X 1 m @ 95°C; 40X (30s 30s 30s @ 94°C 45°C 72°C); 1X 5m @ 72°C. An aliquot of a 10^{-6} dilution of the amplicon was added to each sample for qPCR using luciferase-specific primers LucF3 (5'- GCTTACTGGGACGAAGACGAA) and LucR3 (5'-GCGGTTGTTACTTGACTGGC), and amplified using cycling parameters 1X 1 m @ 95°C; 40X (30s 30s 30s @ 94°C 64°C 72°C); 1X 5m @ 72°C. Our undiluted lysed

lavage gDNAs showed negligible Cq shifts (-0.08 and +0.14, <1 standard deviation from the mean) compared to spike alone (Cq = 17.68 +/- 0.17 over 14 reactions), indicating that no inhibitors in the sample influenced the efficiency of the qPCR reactions. Reproducibility of the qPCR reactions was high; for example, among 44 duplicate reactions, the average Cq range was 0.35, standard deviation 0.25.

Lactobacillus blocking oligomers (LB-blockers) were designed with the criteria that they partially overlap binding sites for broad-spectrum BU4F+ / BU6R+ primer combination (Table S1, S2), and that their melting temperatures (Tm) were at least 68°C, which are higher than that of BU4F+ / BU6R+. First generation blockers used phosphorothioate nucleotide on both ends to prevent digestion by exonuclease, and mismatched bases at the 3' ends were used to prevent their extension by polymerase. Second generation blockers did not use phosphorothioate nucleotide , but retained the mismatches and added a 3' phosphate to block extension.

Amplicon cloning. Broad-spectrum PCR targets were amplified for varying cycle numbers, typically Cq plus 1 or 2 cycles to achieve a "just visible" product on an agarose ethidium-bromide gel. Amplicons derived from PB primers were sequenced directly with the rationale that any species present at 5 to 10-fold over others amplified by the same PB primer would generate a single, readable sequence; amplicons from PB primers that were unreadable as uncloned amplicons were cloned after nested PCR. They were amplified to maximum fluorescence or maximum cycles (typically 40 cycles) for that primer. Products were cleaned by ultrafiltration with the Montage PCR kit UFC7PCR50 (Millipore Corporation, Bedford, MA) or QIAquick PCR purification kit (QIAGEN, Valencia, CA). 1-3 µl of the 20 µl purified product was ligated into pGEM-T

Easy vector using the pGEM-T Easy Ligation Kit (Promega, Madison, WI) according to the manufacturer's specifications. The ligation reaction was carried out 4 hours to overnight at 4°C. 40 µl of *E. coli* cells (strain XL1-Blue, Stratagene, CA) were electroporated using a BioRad GenePulser and Pulse Controller (set at 2.5 kV, 25 µFD capacitance, and 200 Ω resistance) with 1 µl of the ligation, allowed to recover in LB broth 1 h at 37°C, then plated onto LB agar plus 100 µg/ml ampicillin with 0.1 mM IPTG and 20 µg/ml X-gal and grown overnight at 37°C. White colonies and a few blue control colonies were picked from the plates into 100 µl of 15% glycerol or TE, incubated at 95°C for 5 m; 2 µl of this was amplified in 8 µl of reaction buffer as described, using pUC universal and reverse primers that flank the cloning site in the vector (as described earlier) except extension times for cloned amplicons from PB primers were adjusted as appropriate for each amplicon size according to Table S2. An aliquot of each amplicon was tested by agarose gel electrophoresis.

Sequencing and analysis. Full length amplicons (5 µl) were enzymatically cleaned using a 5 µl solution of 0.1 units of NTPhos thermolabile phosphatase and 1 unit of Exonuclease I (Epicentre Biotechnologies, Madison, WI), in 20 mM Tris pH 8.3, 50 mM KCl and 10 mM $MgCl_2$ (1X PCR reaction buffer, Invitrogen Corporation, Carlsbad, CA), incubated at 37°C for 15 m, then inactivated at 80°C for 15 m, and used for sequencing. Mini-libraries were constructed from amplicons with co-dominant species (which generate mixed template reads when uncloned); 17 mini-libraries each averaging 11 sequences were sequenced. Enzymatically cleaned amplicons were sequenced with the same primers used for amplification or T7 for cloned amplicons at Functional Biosciences, Inc. (http://www.functionalbio.com/index.htm). Sequences were uploaded

to Ribosomal Database Project II using the Pipeline function in MyRDP (https://rdp.cme.msu.edu/login/myrdp/ overview.spr) for trimming, quality scoring, alignment, and identification. Only base calls with PHRED quality scores (Q) above 20 were considered. A small number of high quality sequences were rejected by the website as non-ribosomal and were not considered further in this analysis. Aligned sequences, their closest hits via SeqMatch, and manually selected reference species, were downloaded from myRDP, manually trimmed to common 5' and 3' ends, and analyzed with Molecular Evolutionary Genetics Analysis (MEGA4 and MEGA5) software (2, 3). From this, trees were constructed using the bootstrapped Neighbor-Joining method (4) or the Maximum Composite Likelihood method (5), and used for species identification.

Degree of LB-blocking. In this test, the sample was amplified with broad-spectrum primers with and without LB-blocker, then assayed for the presence of *Lactobacillus* amplicon with qPCR with nested *Lactobacillus*-specific primers. *Lactobacillus* amplicon was not generated if blocked, since the amplicon at a 1:1000 dilution was negative by nested qPCR with *Lactobacillus* primers. In contrast, unblocked initial reactions generated product even when diluted one million-fold. The Cq value increased after blocking by 12 cycles; at its amplification efficiency of 1.85, this indicates that blocking decreased the effective titer by a factor of at least ~1.6 million (1000×1.85^{12}).

To validate their use, known concentrations of amplicons of 15 vaginal bacterial species were mixed in proportions that resembled both acute BV and healthy vaginal samples. PB-qPCR of these mock samples showed that the observed titers agreed with the input, typically with an observed / input ratio of 0.6 (data not shown).

Calculation of titers and percent compositions. Cq values from reactions in LightCycler 480 II Real-Time PCR System (Roche Applied Science, Indianapolis, IN) were converted to molecules by comparison to Cq values of a standard curve from the same run derived from four to eight 10-fold serial dilutions of an amplicon quantified using the Quant-iT assay (Invitrogen, Carlsbad, CA). Molecules per µl were converted to cells per 5 ml lavage, assuming an average of 5 ribosomal genes per cell (6) and proportioned to the ratio of the volume of the lavage used in the DNA prep and the its final volume.

If the titer of the sample was below the limit of detection for the PB assay, nested PCR on a 24-cycle amplicon from a longer broad-spectrum 16S rRNA gene amplicon was performed (Fig. 1, Tables S1 and S2). If the nested PCR was negative, the titer of the original sample was assumed to be 0; if the nested PCR was positive, the titer was reported as below the concentration of most dilute but still positive sample in each specific run. Amplicons were sequenced to verify that targeted species were amplified. Mixed amplicons were cloned into pGEM-T easy (Promega, Madison, WI), and ~10 of the resultant clones sequenced. Percentages were calculated by dividing the titer from each PB primer by the sum of the titers from all the PB primers.

Compositions determined by 16S-C&S were simply tallies of sequences in each taxa, defined here as those with >97% identity. These are depicted graphically in Figs. 2 and 3. These tallies were adjusted in libraries from LB-blocked PCR with a correction factor. To determine this factor, the titers of *Gardnerella* and *Lactobacillus* were independently determined with and without blocking by PB-qPCR. Their ratio in the unblocked sample ($7x10^{-6}$) was increased in the blocked sample to 0.09, indicating an

enrichment of $0.09/7x10^{-6}$ = 12,600. To allow for better comparison of qPCR to 16S-C&S results, we converted percent composition to titers by multiplying the fraction of the broad spectrum sequences in the targeted taxonomic group by the total number of bacteria in the sample as determined by summed titers of all PB-qPCR targets. We justify this summation across different primers by noting that titers for each primer were determined from concurrent standard curves on target sequences, which normalizes for differences in primer efficiencies.

SUPPLEMENT S1 Materials and Methods

SUPPLEMENTAL TABLES:

TABLE S1 Percentage of RDP database entries with perfect complementarity to broad-spectrum primers used in vaginal studies

TABLE S2 Summary of primers, PCR programs, and *in silico* performance

TABLE S3 Bacterial reference species used in this study and their responses to LB-blocker

TABLE S4 Richness and diversity parameters

TABLE S5 Summary of titers and compositions using broad-spectrum primers +/- LB-Blocker and PB-qPCR

SUPPLEMENT REFERENCES

1. **Sambrook J, Fritsch EF, Maniatis T.** 1989. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
2. **Kumar S, Nei M, Dudley J, Tamura K.** 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform. **9:**299-306.
3. **Tamura K, Dudley J, Nei M, Kumar S.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596-1599.
4. **Saitou N, Nei M.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406-425.

5. **Tamura K, Nei M, Kumar S.** 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. U. S. A. **101:**11030-11035.

6. **Fogel GB, Collins CR, Li J, Brunk CF.** 1999. Prokaryotic Genome Size and SSU rDNA Copy Number: Estimation of Microbial Relative Abundance from a Mixed Population. Microb. Ecol. **38:**93-113.