ABSTRACT  
Gardnerella is a frequent member of the urogenital microbiota. Given the association between Gardnerella vaginalis and bacterial vaginosis (BV), significant efforts have been focused on characterizing this species in the vaginal microbiota. However, Gardnerella also is a frequent member of the urinary microbiota. In an effort to characterize the bacterial species of the urinary microbiota, we present here 10 genomes of urinary Gardnerella isolates from women with and without lower urinary tract symptoms. These genomes complement those of 22 urinary Gardnerella strains previously isolated and sequenced by our team. We included these genomes in a comparative genome analysis of all publicly available Gardnerella genomes, which include 33 urinary isolates, 78 vaginal isolates, and 2 other isolates. While once this genus was thought to consist of a single species, recent comparative genome analyses have revealed 3 new species and an additional 9 groups within Gardnerella. Based upon our analysis, we suggest a new group for the species. We also find that distinction between these Gardnerella species/groups is possible only when considering the core or whole-genome sequence, as neither the sialidase nor vaginolysin genes are sufficient for distinguishing between species/groups despite their clinical importance. In contrast to the vaginal microbiota, we found that only five Gardnerella species/groups have been detected within the lower urinary tract. Although we found no association between a particular Gardnerella species/group(s) and urinary symptoms, further sequencing of urinary Gardnerella isolates is needed for both comprehensive taxonomic characterization and etiological classification of Gardnerella in the urinary tract.

IMPORTANCE  
Prior research into the bacterium Gardnerella vaginalis has largely focused on its association with bacterial vaginosis (BV). However, G. vaginalis is also frequently found within the urinary microbiota of women with and without lower urinary tract symptoms as well as individuals with chronic kidney disease, interstitial cystitis, and BV. This prompted our investigation into Gardnerella from the urinary microbiota and all publicly available Gardnerella genomes from the urogenital tract. Our work suggests that while some Gardnerella species can survive in both the urinary tract and vagina, others likely cannot. This study provides the foundation for future studies of Gardnerella within the urinary tract and its possible contribution to lower urinary tract symptoms.

KEYWORDS  
Gardnerella, urinary microbiome, phylogenomics, lower urinary tract symptoms

Bacterial vaginosis (BV) is defined as a dysbiosis of the vaginal microbiota that leads to irritation of the vaginal tract and is associated with increased risk of preterm birth and sexually transmitted infections (1, 2). While the cause of BV is unknown,
many suspect *Gardnerella vaginalis* to be either the cause or a biomarker of this dysbiosis. However, *G. vaginalis* also is frequently found within the vaginal and urinary microbiotas of women without BV (3, 4). Therefore, research into *G. vaginalis* pathogenicity has been inconclusive (5).

Whole-genome sequencing of *G. vaginalis* isolates from the urogenital microbiota has incited reevaluation of the genus and species. *Gardnerella* genomes have been separated into variants (6), genotypes (7, 8), genovars (9), or ecotypes (10), but all contain 16S rRNA gene sequences with >98% similarity. The use of other marker genes (e.g., *cpn60*) has led to phylogenies distinct from that determined by the use of the 16S rRNA gene (11, 12). The most recent whole-genome analysis identified at least 13 separate species/groups within the genus, including the description of three new species in addition to *G. vaginalis*: *G. leopoldii*, *G. piotii*, and *G. swidsinskii* (13). These 13 species/groups, which fall into eight major clades, can be differentiated by both their allelic variation within the core genes and the gene content of their accessory genomes (14).

Given the definition of these new species/groups, a new hypothesis is posed: only some *Gardnerella* species/groups are associated with BV, while others can be considered part of the normal vaginal microbiota (4, 12, 15). Prior studies have found that multiple *Gardnerella* species/groups are frequently present within an individual’s vaginal microbiota (12), including the vaginal microbiotas of women with BV (16–18). Furthermore, recent work suggests that BV may be a polymicrobial infection that can include multiple different *Gardnerella* species/groups (19–23).

While the majority of research into *Gardnerella* has focused on its prevalence in the vaginal microbiota, it is also frequently present within the bladders of women with and without lower urinary tract symptoms (24–31). Moreover, *Gardnerella* has been regularly detected within midstream voided urine samples of individuals with chronic kidney disease (32), interstitial cystitis (33, 34), and BV (35). Prior research also has suggested associations between BV and urinary tract infections (UTI) (see the work of Morrill et al. (23) and references therein), and it has been shown that uropathogens can associate with and enhance *G. vaginalis* biofilms (36, 37). Regarding the bladder, periurethral, and urethral microbiotas, *Gardnerella* is more common within younger, premenopausal women (30, 38). As all of the aforementioned investigations relied on 16S rRNA sequencing to detect *Gardnerella* within the microbiome, it is not possible to determine the *Gardnerella* species/group(s) present. Thus, associations between urinary tract symptoms and the newly defined species/groups is unknown.

In an effort to characterize the bacterial species of the urinary microbiota, we previously isolated and sequenced 22 *Gardnerella* strains (28, 39). These strains supplement publicly available genomes, most of which belong to isolates from vaginal samples. Here, we present an additional 10 genomes of urinary *Gardnerella* isolates. With these new genomes available, we conducted a comparative genome analysis of all publicly available *Gardnerella* genomes. In total, 113 genomes were examined. Our analysis includes a pangenome and core genome investigation, complemented with average nucleotide identity (ANI) analysis. Examination of genes associated with pathogenicity provides insight into taxonomic associations with clinical symptoms.

**RESULTS**

**Genome analyses support existence of distinct *Gardnerella* species.** To determine if the matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) assignments are indicative of genetic differences, we purified and sequenced the genomes of 10 clinical isolates (6 identified by MALDI-TOF MS as *G. vaginalis* and 4 identified by MALDI-TOF MS as *Gardnerella* species) (see Materials and Methods). These isolates were collected from 10 different women with overactive bladder (OAB) symptoms (*n* = 7), stress urinary incontinence (SUI) (*n* = 1), and diabetes (*n* = 1) and from a kidney stone (*n* = 1) (see Table S1 in the supplemental material). Statistics regarding these 10 new genomes can be found in Table S1. We then compared the 10 new genome assemblies to 103 publicly available *Gardnerella* assemblies
for a total of 113 genomes (Table S2). Of the 103 publicly available Gardnerella assemblies, 19 were published subsequent to the work of Vaneechoutte et al. (13) and 4 of these strains were deposited as Gardnerella species; these were omitted from the ANI analysis of Vaneechoutte et al. (13). These 23 genomes, in addition to the 10 new genomes produced for this study, have not been previously examined. The 10 new genomes presented here, as well as 22 of the 103 publicly available assemblies included in this analysis, come from our own collection. These 32, plus 1 not from our collection, are the only sequenced urinary isolates for the genus; most isolates come from vaginal samples \( (n = 78) \) (Table S2).

The pan-genome of the 113 Gardnerella assemblies includes 4,542 unique genes, with 1,399 of these genes unique to a single strain assembly. The core genome includes 138 single-copy genes present in all 113 assemblies. The functionalities of these genes were assessed via blastp queries to the complete nr database (Table S3). All of these sequences returned significant hits to annotated Gardnerella sequences as expected, and many also exhibited significant homologies to annotated protein records from other species within the family Bifidobacteriaceae. The majority (87%) of these core sequences exhibited homology to annotated functions. Using these core gene sequences, we derived a phylogenomic tree (Fig. 1). The clade structure of this tree mirrored that of prior work (13, 14). The same 13 groups as identified by Vaneechoutte et al. (13) were identified in this study. However, our phylogenomic tree also provided insight into classification of new strains, including those published after (or omitted) from the earlier work of Vaneechoutte et al. (13) (Fig. 1, gray) and the 10 new genomes published as part of this study (Fig. 1, red).

Next, we conducted ANI analysis for the 113 genomes (Fig. 2). The ANI analysis also provides insight into the classification of the previously unanalyzed genomes (Fig. 1, gray) and the newly sequenced genomes (Fig. 1, red). The combined phylogenomic and ANI analyses identify a new group: group 14, which is represented by the single member, NR010 \( (\text{GCA}_003408845) \). While the core phylogeny shows that this strain is most closely related to the group 7 strains, the ANI analysis shows that it is distinctly different (Fig. S1). ANI comparisons within group 7 strains range between 98.87 and 99.96%. In contrast, NR010 versus group 7 strain ANI values range between 85.20 and 85.33%; this is on par with ANI values between groups. Thus, we designate Gardnerella strain NR010, isolated from the vaginal mucus from a woman in Kenya (11), as the sole representative of group 14.

**Comparison of pathogenicity genes between the Gardnerella species/groups.**

Previous work has suggested that mucin degradation, sialidase activity, and vaginolysin activity contribute to Gardnerella pathogenicity. The mucin degradation pathway consists of 6 enzymes (Table 1) and includes sialidase activity. Only G. vaginalis \( (n = 43/47) \) and group 2 \( (n=5/5) \) strains contained genes that encode the complete mucin degradation pathway; all other species or groups were missing part or all of the pathway (Table 1; Table S4). The genes that encode sialidase A and O-sialoglycoprotein endopeptidase were present in G. vaginalis, G. piontii, and group 2, group 3, and group 11 strains (Table 1). Groups 7 \( (n = 3) \), 12 \( (n = 2) \), and 13 \( (n = 1) \) are omitted from Table 1, as none of the genomes contained recognizable homologs of these 6 gene sequences.

Previously, it was suggested that the presence of the sialidase A gene can be associated with G. vaginalis virulence in the vagina (40), and increased sialidase activity within the vaginal fluid of BV+ individuals has been observed (41). Genome sequences from 9 of the 14 groups include this gene. A phylogenetic tree of all detected sialidase genes is shown in Fig. 3. While genes from G. vaginalis and group 2 were highly related, there remained distinct differences within the G. vaginalis group, with one clade exhibiting greater similarity to group 2 sequences than to the other G. vaginalis strains. The remaining 7 groups exhibited greater similarity to each other than to the G. vaginalis and G. vaginalis/group 2 clades.

Vaginolysin (encoded by vly) is a cholesterol-dependent cytolysin (CDC) (42). It binds to cholesterol receptors on the surface of vaginal epithelial cells, inducing lysis. It
FIG 1 Core genome phylogeny for Gardnerella strains. Accession numbers are listed in parentheses. The color bar indicates the genomic species/group per the study by Vaneechoutte et al. (13) or N/A (gray; not included in the study).

(Continued on next page)
is even hypothesized that the cholesterol level in African American women is what pre-

disposes them to BV (43). Of the 113 genomes examined, the vly coding region was
detected in 95 genomes (Table S4). Only the single group 11 strain’s genome assembly
did not contain vly; further sequencing of isolates belonging to this group will provide
insight into whether this is characteristic for the group. All other species and groups
included strains containing vly. As shown in Fig. 4, the vly sequence is not always con-
gruent with the core genome phylogeny. Species/groups are not monophyletic in this
tree. For instance, group 3 strains (lime green) clade with both G. vaginalis and group 7
strains.

**Gardnerella species/groups and associations with urinary symptoms.** MALDI-
TOF MS analysis separated members the genus Gardnerella into two categories:
Gardnerella vaginalis and Gardnerella species. Isolates of both types were found in
catheterized samples from women with and without OAB, and neither G. vaginalis nor
Gardnerella species are associated with either continence or OAB with statistical signi-
ficance ($\chi^2; P = 0.161$) (Table 2). Thirteen of the 32 sequenced urinary isolates from our
collection were identified as Gardnerella species via MALDI-TOF MS (Table S5). Based
upon our genome analyses, these strains are representative of G. swidsinskii ($n = 6$), G.
leopoldii ($n = 6$), or group 8 ($n = 1$). The remaining 19 sequenced urinary isolates from our
collection were identified as G. vaginalis via MALDI-TOF MS; genomic analysis
determined that 3 were in fact group 3 strains, while the others were G. vaginalis
strains (Table S5).

Given the species/group designations based upon ANI, we investigated the associa-
tion of the 32 sequenced strains from our collection with urinary symptoms. All 32 of
these strains were isolated from urine samples from women and include representa-
tives of G. vaginalis ($n = 16$), G. swidsinskii ($n = 6$), G. leopoldii ($n = 6$), group 3 ($n = 3$), and
group 8 ($n = 1$) (Table S5). These 32 strains grouped with strains isolated from the vagi-
nal microbiota (Fig. 1); thus, these species/groups are not unique to the urinary micro-
bioti. As detailed in Table 3, G. vaginalis, G. swidsinskii, G. leopoldii, and group 3 strains
were isolated from asymptomatic patients, as well as individuals with urinary symp-
toms. There was no significant association between Gardnerella species/group and uri-
nary symptoms.

**FIG 1 Legend (Continued)**

by Vaneechoutte et al. [13] or new genomes produced in this study (red). Strain names and accession numbers
produced as part of this study are listed in red. Strains deposited as “Gardnerella species” are listed in bold. Tree scale
refers to evolutionary distance based upon the FastTree’s approximate maximum-likelihood method.
DISCUSSION

Several previous studies have demonstrated broad genetic diversity within the *Gardnerella* genus (8–11, 13, 14, 44). Whole-genome sequencing of isolates and comparative studies have been pivotal in describing this diversity (8–10, 14, 44). Genome analysis suggests that these species/groups are in fact reproductively isolated (14). We similarly detected broad genetic diversity, identifying a large accessory genome for the genus. Once only a single species represented this genus; now 13 groups, which include three new *Gardnerella* species, have been described (13). Based upon our core genome analysis and ANI comparisons, we suggest a 14th group.

In contrast to recent investigations of the *Gardnerella* core genome (8, 14), we identified a smaller core gene set: 138 genes. Bohr et al. (14) included 608 genes common among 106 assemblies, and Tarracchini et al. (8) included 334 genes common among 72 assemblies. Our smaller core gene set can be contributed to our inclusion of only single-copy-number genes and the threshold for similarity used. Similar to the aforementioned studies, our core was also identified from primarily draft assemblies, i.e., multiple contigs. We did exclude three publicly available genomes from our analysis, as they did not meet the threshold of completeness (see Materials and Methods). Even though a smaller set of genes was considered in this study, our core genome phylogeny is equivalent to these other recent core genome phylogenies. This concurrence further calls for additional formal delineation of species within the genus.

Classification of isolates by the new *Gardnerella* species/groups may help to resolve the debate over the role of *Gardnerella* in BV progression (5). However, this cannot be accomplished via 16S rRNA-based sequence surveys given the >98.5% similarity across the genus (13). One alternative is using another marker gene, e.g., *cpn60* (11, 12) or the sialidase gene, given its association with BV symptoms (4, 41). Despite its putative clinical relevance, our study found that neither the presence (Table 1) nor sequence (Fig. 2) of sialidase can distinguish between *Gardnerella* species/groups. Moreover, a recent study found that presence of the sialidase A gene does not correspond with BV symptoms (18). Similarly, vaginolysin presence or sequence is insufficient (Fig. 3; Table S4). These findings advocate for whole-genome analysis for taxonomic and etiological classification. Recently, Tarracchini et al. (8) showed the viability of this approach; individual *Gardnerella* species/groups were identified in vaginal microbiome data.

Urinary isolates previously identified as *Gardnerella* species by MALDI-TOF MS are members of the newly recognized species *G. swidsinskii* and *G. leopoldii*. While our collection of sequenced *G. vaginalis* strains includes 12 isolates from women with lower urinary symptoms (OAB, n = 9; urge urinary incontinence [UUI], n = 2; SUI, n = 1), it also includes 3 isolates from asymptomatic women. Since these three isolates’ genomes include annotations for virulence factors, including both sialidase and vaginolysin, we cannot speculate that they are commensal strains. As Table 3 shows, no single species/

### TABLE 1 Presence of mucin degradation genes

<table>
<thead>
<tr>
<th>Species or group</th>
<th>n</th>
<th>Sialidase A</th>
<th>O-Sialoglycoprotein endopeptidase</th>
<th>Beta-galactosidase</th>
<th>Alpha-L-fucosidase</th>
<th>M22 family glycoprotease</th>
<th>Alpha-mannosidase</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>G. leopoldii</em></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>G. piotti</em></td>
<td>9</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>G. swidsinskii</em></td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>G. vaginalis</em></td>
<td>47</td>
<td>43</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>Group 2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 8</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 9</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 10</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 11</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 14</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Classified as *G. leptoldii* in BV progression (5). However, this cannot be accomplished via 16S rRNA-based sequence surveys given the >98.5% similarity across the genus (13). One alternative is using another marker gene, e.g., *cpn60* (11, 12) or the sialidase gene, given its association with BV symptoms (4, 41). Despite its putative clinical relevance, our study found that neither the presence (Table 1) nor sequence (Fig. 2) of sialidase can distinguish between *Gardnerella* species/groups. Moreover, a recent study found that presence of the sialidase A gene does not correspond with BV symptoms (18). Similarly, vaginolysin presence or sequence is insufficient (Fig. 3; Table S4). These findings advocate for whole-genome analysis for taxonomic and etiological classification. Recently, Tarracchini et al. (8) showed the viability of this approach; individual *Gardnerella* species/groups were identified in vaginal microbiome data.

Urinary isolates previously identified as *Gardnerella* species by MALDI-TOF MS are members of the newly recognized species *G. swidsinskii* and *G. leopoldii*. While our collection of sequenced *G. vaginalis* strains includes 12 isolates from women with lower urinary symptoms (OAB, n = 9; urge urinary incontinence [UUI], n = 2; SUI, n = 1), it also includes 3 isolates from asymptomatic women. Since these three isolates’ genomes include annotations for virulence factors, including both sialidase and vaginolysin, we cannot speculate that they are commensal strains. As Table 3 shows, no single species/
A group could be associated with a urinary symptom cohort. Just as others hypothesize that some *Gardnerella* species/groups may contribute to the development (rather than be the sole cause) of BV (see reviews in references 45 and 46), *Gardnerella* species/groups also may contribute to lower urinary tract symptoms. Other factors likely contribute to symptom status, including overall urinary microbiota composition. Further isolation and sequencing of urinary *Gardnerella* strains are needed to determine if this is true or if age, race, or other demographics contribute to symptom status.

In our prior study, we found that many of the species that inhabit the bladder are also found within the vagina, suggesting that the bladder and vaginal microbiotas are
FIG 4 Phylogenetic tree of the vaginolysin coding sequences from the *Gardnerella* assemblies. Tree scale refers to evolutionary distance based upon the FastTree’s approximate maximum-likelihood method.
interlinked (39). However, in comparison to the vaginal microbiota, only five species/groups have been identified within the lower urinary tract. While some of the 14 species/groups are represented by a single isolate or just a few isolates, we were intrigued to find that *G. pioi* has yet to be identified in the urinary tract; the nine assemblies examined in this study were all collected from the vagina (Table S2). Indeed, current evidence suggests that *G. pioi* is a commensal of the vagina (12). The urinary genomes examined in this study represent only 6 of the 14 species/groups: *G. leopoldii*, *G. swidsinskii*, *G. vaginalis*, group 3, group 7, and group 8. Future efforts to isolate and sequence *Gardnerella* species/groups from the urinary tract will provide insight into our preliminary observations that some species/groups are not (or are infrequent) members of the urinary microbiota.

**MATERIALS AND METHODS**

**Culture and identification.** The 10 *Gardnerella* strains sequenced as part of this study were isolated from urine samples collected and processed as part of prior institutional review board (IRB)-approved studies (LU206449, LU207152, LU207102, LUC204195, and LU204133) (24–26, 30, 31, 39, 47, 48). Isolation was performed using the enhanced quantitative urine culture (EQUC) protocol, in which all morphologically distinct colonies are purified and identified using MALDI-TOF mass spectrometry as reported previously (24). MALDI-TOF MS identification of *Gardnerella* isolates was conducted prior to the definition of the 3 new *Gardnerella* species. MALDI-TOF MS peaks have been reassessed subsequent to our analysis such that the four *Gardnerella* species can be distinguished (13). Following isolation, samples were stored down in brucella broth (Hardy Diagnostics, Santa Maria, CA) at −80°C for future use.

**Whole-genome sequencing.** For sequencing, the isolates were struck onto BD BBL CDC anaerobe 5% sheep blood agar and grown anaerobically for 48 h, whereupon the colonies were scraped off the plates, resuspended in phosphate-buffered saline (PBS), and pelleted. Genomic DNA was extracted from pelleted cells using a phenol-chloroform method (49). DNA was prepared and sequenced using the Illumina Hi-Seq platform with library fragment sizes of 200 to 300 bp and a read length of 100 bp at the Wellcome Trust Sanger Institute, as previously described (50). Annotated assemblies were produced using the pipeline described previously (51). Briefly, sequence reads were used to create multiple assemblies using Velvet v1.2 (52) and VelvetOptimiser v2.2.5 (https://github.com/tseemann/VelvetOptimiser). An assembly improvement step was applied to the assembly with the best N₅₀, and contigs were scaffolded using SSPACE (53) and sequence gaps filled using GapFiller (54). Automated annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v4.11 (55) upon submission to GenBank’s Assembly database.

**Core genome and average nucleotide identity analysis.** Publicly available *Gardnerella* genomes were identified from the NCBI (Table S1). Genome assemblies were evaluated for their completeness using CheckM (56). As a result, three publicly available sequences were excluded from evaluation. It is worth noting that the final genome assemblies examined include three different sequencing projects for strain ATCC 14018 and two different sequencing projects for strain ATCC 49145. The pangenome was identified using anvï‘o (57). Homologous genes were identified using a Markov clustering (MCL)
inflation of 6. Single-copy core genes were identified by anvi'o, and the functionality of these genes was determined via blastp queries (using the core gene amino acid sequence representative of each core gene from the G. vaginalis 409-05 annotation) against the nr database. The phylogenomic tree was derived based upon the alignment of the concatenation of the single-copy core gene amino acid sequences using anvi'o (57), FastTree (58), and iTOL (59). ANI was calculated by performing pairwise comparisons of complete genome assemblies using fastANI (60).

Gene-specific analyses. Genes that encode mucin degradation, sialidase activity, and vaginolysin were identified for each strain from the PGAP annotations and confirmed by local blastn queries. Gene sequences were aligned with MAFFT v7.388 (61). FastTree (58) and iTOL (59) were used to derive and visualize the phylogenetic relationship for vaginolysin and sialidase.

Data availability. Raw reads and genomes have been deposited in NCBI's SRA and Assembly databases, respectively. Accession numbers for genome assemblies are listed in Table S1 and include GCA_013315005, GCA_013315025, GCA_013315045, GCA_013315075, GCA_013315085, GCA_013315115, GCA_013315125, GCA_013315135, GCA_013315145, and GCA_013315215.

SUPPLEMENTAL MATERIAL
Supplemental material is available online only.

FIG S1, EPS file, 1.8 MB.
TABLE S1, PDF file, 0.1 MB.
TABLE S2, PDF file, 0.1 MB.
TABLE S3, PDF file, 0.1 MB.
TABLE S4, PDF file, 0.2 MB.
TABLE S5, PDF file, 0.1 MB.

ACKNOWLEDGMENTS
We acknowledge Trevor Lawley and his group at the Wellcome Trust Sanger Institute, in particular Nitin Kumar. We also thank Swarnali Banerjee for discussion on statistical analyses.

We acknowledge funding from NIH (R01 DK104718 awarded to A.J.W.) and NSF (1661357 awarded to C.P.). E.C. is supported by the Mulcahy Research Fellowship (Loyola University Chicago).

C. Putonti discloses research support from NSF. A. J. Wolfe discloses research support from NIH, the DOD, and the Kimberly Clark Corporation and membership on the scientific advisory boards of Pathnostics and Urobiome Therapeutics.

REFERENCES


