frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 28 March 2023
DOI 10.3389/fmicb.2023.1154664

OPEN ACCESS

EDITED BY
Huizhou Fan,
Rutgers Robert Wood Johnson Medical School, United States

REVIEWED BY
Margaret Hammerschlag,
Downstate Health Sciences University, United States
Stephen Allen Morse,
IHRC, Inc., Headquarters, United States

*CORRESPONDENCE
Deborah Dean
✉ deborah.dean@ucsf.edu
Timothy D. Read
✉ tread@emory.edu

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Infectious Agents and Disease, a section of the journal Frontiers in Microbiology

RECEIVED 30 January 2023
ACCEPTED 06 March 2023
PUBLISHED 28 March 2023

CITATION
Joseph SJ, Bommana S, Ziklo N, Kama M, Dean D and Read TD (2023) Patterns of within-host spread of *Chlamydia trachomatis* between vagina, endocervix and rectum revealed by comparative genomic analysis. *Front. Microbiol.* 14:1154664. doi: 10.3389/fmicb.2023.1154664

Patterns of within-host spread of *Chlamydia trachomatis* between vagina, endocervix and rectum revealed by comparative genomic analysis

Sandeep J. Joseph [1], Sankhya Bommana [2], Noa Ziklo[2], Mike Kama[3], Deborah Dean [2,4,5,6,7,8*†] and Timothy D. Read [9*†]

[1]Division of STD Prevention, Centers for Disease Control and Prevention, Atlanta, GA, United States, [2]Department of Pediatrics, University of California, San Francisco, Oakland, CA, United States, [3]Ministry of Health and Medical Services, Suva, Fiji, [4]Department of Medicine, University of California, San Francisco, San Francisco, CA, United States, [5]Department of Bioengineering, Joint Graduate Program, University of California, San Francisco, San Francisco, CA, United States, [6]Department of Bioengineering, Joint Graduate Program, University of California, Berkeley, Berkeley, CA, United States, [7]Bixby Center for Global Reproductive Health, University of California, San Francisco, San Francisco, CA, United States, [8]Benioff Center for Microbiome Medicine, University of California, San Francisco, San Francisco, CA, United States, [9]Division of Infectious Diseases, Department of Medicine, Emory University School of Medicine, Atlanta, GA, United States

Introduction: *Chlamydia trachomatis*, a gram-negative obligate intracellular bacterium, commonly causes sexually transmitted infections (STIs). Little is known about *C. trachomatis* transmission within the host, which is important for understanding disease epidemiology and progression.

Methods: We used RNA-bait enrichment and whole-genome sequencing to compare rectal, vaginal and endocervical samples collected at the same time from 26 study participants who attended Fijian Ministry of Health and Medical Services clinics and tested positive for *C. trachomatis* at each anatomic site.

Results: The 78 *C. trachomatis* genomes from participants resolved into two major clades of the *C. trachomatis* phylogeny (the "prevalent urogenital and anorectal" clade and "non-prevalent urogenital and anorectal" clade). For 21 participants, genome sequences were almost identical in each anatomic site. For the other five participants, two distinct *C. trachomatis* strains were present in different sites; in two cases, the vaginal sample was a mixture of strains.

Discussion: The absence of large numbers of fixed SNPs between *C. trachomatis* genomes within many of the participants could indicate recent acquisition of infection prior to the clinic visit without sufficient time to accumulate significant genetic variation in different body sites. This model suggests that many *C. trachomatis* infections may be resolved relatively quickly in the Fijian population, possibly reflecting common prescription or over-the-counter antibiotics usage.

KEYWORDS

***Chlamydia trachomatis*, single nucleotide polymorphisms, single variable polymorphisms, sexually transmitted diseases, chlamydiae**

Highlights

- *Chlamydia trachomatis* is a bacterial pathogen that causes millions of sexually transmitted infections (STIs) annually across the globe. Because *C. trachomatis* lives inside human cells, it has historically been hard to study. We know little about how the bacterium spreads between body sites. Here, samples from 26 study participants who were simultaneously postive in their vagina, rectum and endocervix were genetically analyzed using an improved method to extract *C. trachomatis* DNA directly from clinical samples for genome sequencing. By analyzing patterns of mutations in the genomes, we found that 21 participants shared very similar *C. trachomatis* strains in all three anatomic sites, suggesting recent infection and spread. For five participants, two *C. trachomatis* strains were evident, indicating multiple infections. This study is significant in that improved enrichment methods for genome sequencing provides robust data to genetically trace patterns of *C. trachomatis* infection and transmission within an individual for epidemiologic and pathogenesis interrogations.

Introduction

The obligate intracellular bacterium *Chlamydia trachomatis* is the most common worldwide cause of bacterial sexually transmitted infections (STIs) with over 129 million annual cases in 2020 (World Health Organization, 2020). In 2019, 1.8 million cases were reported in the United States alone, representing a 19% increase since 2015 (Centers for Disease Control Prevention and Department of Health Human Services, 2021). Approximately 80% of female and 50% of male *C. trachomatis* STIs are asymptomatic (Batteiger, 2020), increasing the risk of transmission and complications at a yearly cost of billions of dollars (Satterwhite et al., 2013).

The endocervix is considered the most common initial site of chlamydial sexually transmitted, non-lymphogranuloma venereum (LGV) infections. Sloughed *C. trachomatis* infected cells and the organism itself can be secreted into the vagina but neither are known to infect the squamous epithelium of that organ (Batteiger, 2020). Cervicitis, an inflammation of the uterine endocervix, is a strong predictor of upper genital tract inflammation and disease (Peipert et al., 2000), including pelvic inflammatory disease, tubal-factor infertility, ectopic pregnancy and poor pregnancy outcomes (Haggerty et al., 2010). The rectum is another site of infection. A growing number of studies now show that *C. trachomatis* rectal infections are more common than previously thought, ranging from 2 to 77% of women seen in clinical settings (Chan et al., 2016). In one study, over 70% of women with urogenital *C. trachomatis* also had rectal *C. trachomatis* infection (van Liere et al., 2015). Of the 24 studies reporting on both urogenital and rectal infections in the same women, six showed a higher prevalence of *C. trachomatis* in the rectum (Chan et al., 2016). These data suggest that, while the rectum is known to be a common site of infection with LGV strains among men who have sex with men (Stoner and Cohen, 2015), it may also be a more frequent primary site of non-LGV strain infections among women. However, no studies to date have evaluated this issue.

There are several hypotheses for *C. trachomatis* transmission between sexual partners and within anatomic sites of the same individual given our fragmentary knowledge of the genetic structure of *C. trachomatis* populations in natural human infections. The ascertainment of *C. trachomatis* infection in females could be affected by rectal infections persisting longer than endocervical infections and/or increased transmissibility during receptive anal intercourse. However, a recent study found no association between receptive anal intercourse and rectal *C. trachomatis* infections (Chandra et al., 2018) and another found that screening given a history of receptive anal intercourse did not significantly influence the rate of detection of *C. trachomatis* infections in the rectum (van Liere et al., 2015). We also know that women may develop urinary tract infections from enteric bacteria that are transferred from the perineum or anorectal area during sex (van Liere et al., 2015). It is therefore possible that rectal *C. trachomatis* infections could similarly be spread to the endocervix and urethra and possibly vice versa from the endocervix and urethra to the rectum. The concern here is that single dose treatment that is effective for uncomplicated urogenital tract infections is inadequate for rectal infections, as has been shown in recent studies (Drummond et al., 2011; Khosropour et al., 2014; Lanjouw et al., 2016; van Liere et al., 2017; Foschi et al., 2018). Indeed, a study that followed cervicovaginal and anorectal *C. trachomatis* loads following treatment with one gram of azithromycin found consistently higher loads in the anorectal site at 16 days after therapy with increasing loads up to 51 days when the study was terminated (Dukers-Muijrers et al., 2013). Due to the requirement to treat non-LGV *C. trachomatis* infections of the rectum for seven days and LGV strains for 21 days, adherence to treatment and/or treatment failure, as a result of lack of adherence, are also concerns (Kong et al., 2014). These studies show that rectal infection, if not treated appropriately, could have a significant effect on persistence and within-host transmission and disease. Therefore, it is important to understand the pathways of transmission between anatomic sites.

To understand the dynamics and pathobiology of within-host transmission of *C. trachomatis*, we explored the relationships among *C. trachomatis* genomes sequenced using DNA purified directly from endocervical, vaginal and rectal swabs from the same women. Our cohort comprised a population of Fijian women that have an unusually high prevalence of *C. trachomatis* STIs (Svigals et al., 2020). We sought to reveal evidence of within-host dissemination that may promote maintenance of infection in the rectum and increase transmission both within the host and to sexual partners in addition to providing data to select optimal anatomic sites for diagnostic screening, appropriate treatment and duration of therapy.

Results

Direct enrichment and sequencing of *Chlamydia trachomatis* genomes and comparison of bacterial loads between anatomic sites

Clinical endocervical, rectal, and vaginal swab samples collected from 26 women who attended the Fijian Ministry of Health and Medical Services clinics and tested positive for *C. trachomatis* at each anatomic site simultaneously were supplied

de-identified from an ongoing parent study (Svigals et al., 2020). The women ranged in age from 19 to 38 years (mean 26.69 years) of whom 13 (34.6%) had signs and/or symptoms of STIs. There was no association of age (i.e., <25 years vs ≥25 years) with signs or symptoms, *omp*A genotype or plasmid strain (**Supplementary Table 1**).

We successfully extracted DNA from clinical swabs and used our recently redesigned Agilent RNA bait library (Bowden et al., 2021) to enrich *C. trachomatis* genomic sequences from Illumina sequencing libraries (see Methods). We defined a threshold for a "good quality" genome of at least 10× average *C. trachomatis* genome read redundancy ("coverage") post-enrichment and at least 5 reads mapped to > 900,000 bases of the 1,042,519 bp *C. trachomatis* reference D/UW-3/CX chromosome (**Supplementary Table 1**). The median coverage of these 78 samples was 127× with an average of 308×; only three samples were lower than 20×. The RNA bait method was therefore able to enrich *C. trachomatis* genomic DNA even though the samples from the three anatomic sites likely contain high levels of other viral and bacterial organisms. These data are supported by our previous study using the same methodology that successfully generated genomes derived from DNA purified directly from clinical vaginal-rectal pairs from Fijian participants (Bowden et al., 2021).

Using qPCR with conserved *omp*A primers, the chromosomal yield for 25/26 women with *C. trachomatis* successfully sequenced from each body site ranged from 69 to 9,600,000 copies/μL. Given the obligate intracellular nature of *C. trachomatis*, and to normalize the sample against the number of human cells collected in the sample, the ratio of the *C. trachomatis* genomic copy number—based on the single copy *omp*A gene in the *C. trachomatis* genome—to the human beta-actin (a single gene in the human genome) copy number was calculated as an estimated relative load of the organism in each anatomic site. In comparing the vaginal with the rectal site for each woman using a paired t-test, there was a statistically significant higher load in the rectum than the vagina (P = 0.0124; **Supplementary Figure 1**). However, there were no statistically significant differences between rectum/endocervix and vagina/endocervix sites. When comparing body sites from the same person, 21 of the 26 women had a higher load in the rectum compared to the vagina (**Figure 1**). However, the differences in qPCR loads across body sites were not reflected in the redundancy of genome coverage. Within the 78 genomes, there was a significantly higher coverage in the endocervical samples compared to rectal (T-test; P = 0.031) and vaginal (P = 0.0016) samples (**Supplementary Figure 2**).

Fiji sample genomes in the context of the global *Chlamydia trachomatis* phylogeny

We investigated the phylogenetic distribution of assembled genomes from this study and selected chlamydial reference and other clinical genomes representing known global *C. trachomatis* clades corresponding to four major *C. trachomatis* clades: LGV, ocular, "prevalent urogenital and anorectal (P-UA) and "non-prevalent urogenital and anorectal" (NP-UA) (Smelov et al., 2017; **Figure 2** and **Table 1**). The reference genome D/UW-3/CX was in the NP-UA clade. All Fiji genomes were in the NP-UA and P-UA clades, forming two subclades of NP-UA and one in P-UA, suggesting that the Fiji genomes were derived from at least two independent introductions in NP-UA and one in P-UA (**Figure 2**). Based on sequencing of the *omp*A gene, referred to as the *omp*A genotype, 32 genomes in NP-UA had *omp*A genotype D (4), F (3), G (23) and Ja (1) plus one that was not possible to determine, and the 46 genomes in P-UA had E (21), G (2), and Ja (23) *omp*A genotypes. Twenty-four Fiji samples dominated a sub-lineage of NP-UA (*omp*A genotypes G and F) that included one publicly submitted genome sequence G/11222 (BioSample: SAMN02603694, Assembly NC_017430.1) (Jeffrey et al., 2010), which was a cervical sample but with no notation of geographic source. This Fijian subclade may represent a local endemic clone. We also found genomes with *omp*A genotype Ja with an E backbone that we had previously described in the Fiji population (Bowden et al., 2021) and also a plasmid genotype E.

Numerous studies have shown that *omp*A alleles recombine frequently between *C. trachomatis* genomic backbones (Gomes et al., 2004; Somboonna et al., 2011; Joseph et al., 2012; Hadfield et al., 2017; Seth-Smith et al., 2021). While the association of *omp*A genotypes with clades in Fiji strains was broadly consistent with patterns found in the Hadfield et al. study (Hadfield et al., 2017), there were some combinations of genomic clade and *omp*A in this work not previously reported: G in P-UA and F in NP-UA (**Figure 2**). fastGEAR (Mostowy et al., 2017) inferred recombination events in ancestors of the global P-UA clade (primarily from NP-UA into P-UA) as well as recent recombinational exchange of DNA within the branches of the tree containing Fiji strains (**Supplementary Figure 3**). Recent inferred events included donors from all clades, including a small number of importation events from LGV and ocular clades, respectively, at recombination hotspots in the chromosome (**Supplementary Table 2**).

Participants with samples from three anatomic body sites fell into two groups based on levels of *Chlamydia trachomatis* genome diversity

Of the 26 study participants, there was good quality genome sequence data across the three anatomic sites, and 21 had the same *omp*A genotype strain consistent with the rest of its genome that formed a monophyletic clade on the global *C. trachomatis* phylogenetic tree (**Figure 2** and **Supplementary Table 3**). We inferred these strains shared a recent common ancestor. We termed these 21 participants "Group A". Five participants ("Group B") had three samples that appeared not to derive from a single recent infection event. For participant #1078, the rectal sample and vaginal/endocervical samples were different *omp*A genotypes/genomes from different clades (E in P-UA and D in NP-UA, respectively). For participant #564, all samples were in P-UA but the vaginal and rectal samples were both E while the endocervical sample was Ja and more distantly related on the core genome phylogeny than the other two (**Figure 2** and **Supplementary Figure 3**). The rectal and endocervical samples of participant #1176 were both Ja in P-UA, but the vaginal sample was a G in NP-UA. In participant #32, all of the strains were *omp*A

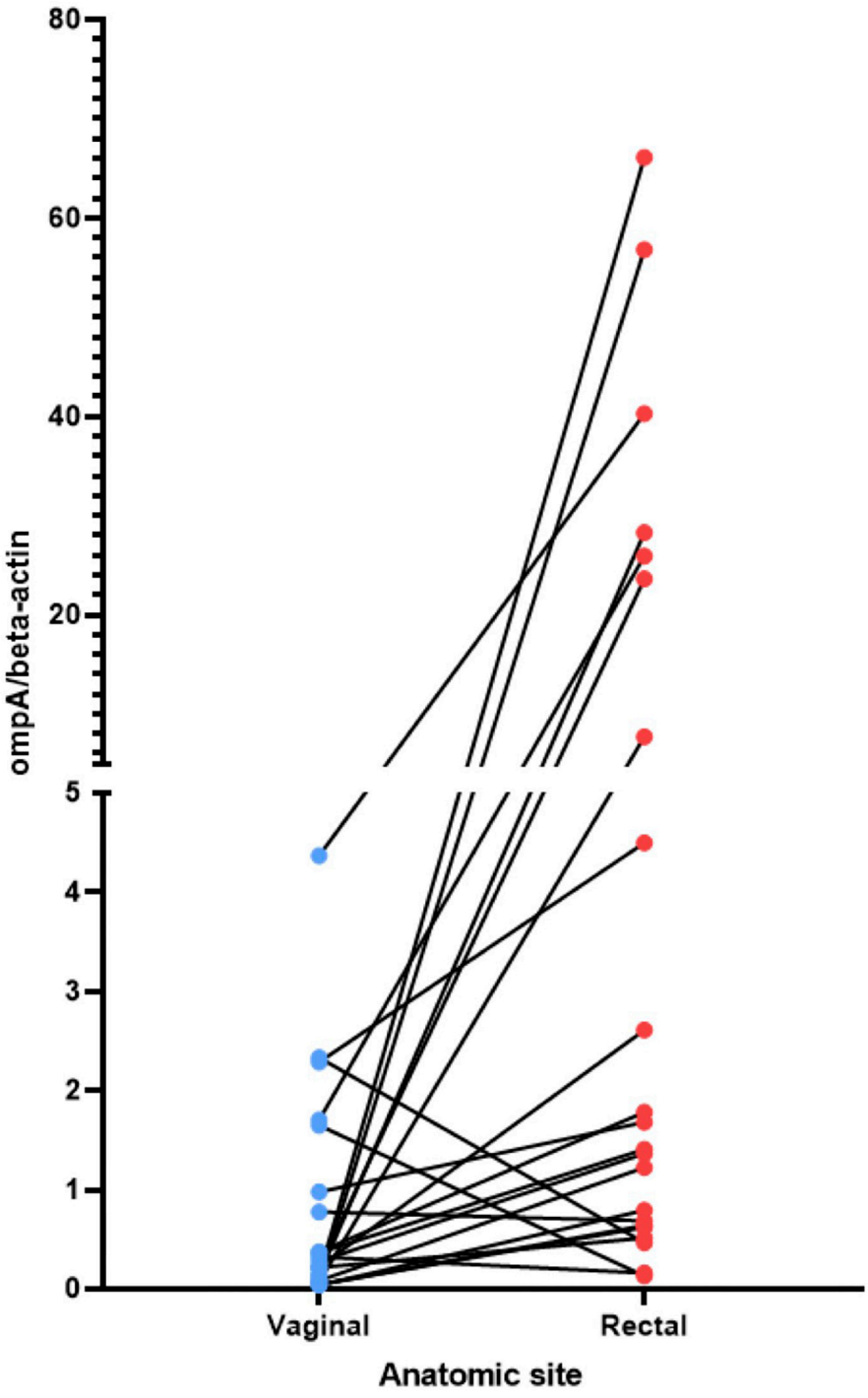


FIGURE 1
Relative load of *C. trachomatis* in the vagina and rectum estimated by qPCR. The non-transformed ratio of the *C. trachomatis ompA* genome copy number to the beta-actin genome copy number is shown (see section "Materials and methods"). C, endocervix; R, rectum; V, vagina. The lines connect the *C. trachomatis* load value for the vagina to the load value for the rectum for the same woman.

genotype G. However, the endocervical and rectal genomes were closely related in the P-UA clade while the vaginal strain was in NP-UA. For participant #1182, all strains were *ompA* genotype Ja but in this case, while the vaginal and endocervical genomes were closely related in P-UA, the rectal genome was in NP-UA. The *ompA* genotype E is rare in NP-UA strains (**Figure 2**), suggesting the rectal strain acquired the allele through recent recombination.

The differences in *C. trachomatis* strains between the vagina and endocervix of the same individual confirm that these sites can be effectively sampled without cross-contamination. In addition, shotgun metagenomics from some of the same samples as in this study also revealed related but diverged communities at each site (Bommana et al., 2022). Further, while the endocervix is the site of infection and secretions along with the infected cells flow into the

vagina, the vaginal environment may promote unique pressures on the genomes that are then detected as noted above.

The *C. trachomatis* ~7 kb virulence plasmid was amplified and sequenced in 66/78 samples. For each participant, the genotype based on comparison with reference strain plasmid sequences was identical across the anatomic sites (**Supplementary Table 1**). All plasmids were either E genotype or "D/G" genotype (D and G plasmids had identical sequences in our typing scheme). Plasmid genotype E was linked to P-UA genomes (36 out of 37 samples with data) and D/G linked to NP-UA (26/29 samples with data). The strong association between chromosome and plasmid genotype suggested that vertical transmission was the dominant mode for plasmid inheritance (Hadfield et al., 2017). Only in Group B patients were incongruent combinations seen (plasmid genotype E-P-UA for 32V, 1176V, and 1182R samples and plasmid D/G-NP-UA for 1078R). These samples likely have had plasmid replacement events, with the donor strain containing the transferred plasmid infecting another anatomic site.

Patterns of shared fixed single nucleotide polymorphisms (SNPs) and single nucleotide variants (SNVs) in *Chlamydia trachomatis* from anatomic body sites of the same participant are different in group A and group B participants

We looked first at the Group A participants to see what the patterns of SNPs revealed about the relationships between the body sites. We defined "fixed" SNPs to mean nucleotide positions on the reference genome where 10% or less of the mapped sequence read coverage matched the reference base. The number of fixed SNPs in all three body sites was 512-1944 for NP-UA samples and 2169-5229 SNPs for P-UA (**Supplementary Table 3**). The higher number for P-UA was because the reference strain D/UW-3/CX was in the NP-UA clade. This pattern was consistent with these SNPs being shared by the common ancestor of the sample that infected the three body sites of each participant.

Fixed SNPs found in only one or two body site samples were rare in Group A participants. The presence of these SNPs would be suggestive of independently evolving populations at different sites. Only five Group A participants had a rectal sample with a fixed SNP, one had a fixed SNP in the endocervix but zero had fixed SNPs unique to the vaginal sample (**Figure 3** and **Supplementary Table 3**). One participant had a SNP shared by rectal and vaginal samples and one shared between rectal and endocervical samples. There were three SNPs shared between endocervical and vaginal pairs that were fixed in one of the sites but intermediate frequency in the other (see below) (**Figure 3**). Since these mutations probably occurred within the host, these data point to a recent common ancestor of the bacterium in each body site of the Group A participants.

Next, we looked at intermediate frequency single nucleotide variants (SNVs), which we defined as having reference allele frequencies in the 10–90% range. Reference alleles over 90% were inferred to be the same as the reference while less than 10% were defined as fixed, as in the paragraph above. Across the 78 high-quality Fiji genomes, we found 8,694 SNVs. Most SNVs were found in a small number of samples, with 2,039 (23.5%) found in only one. Of particular note were the 3,818 "rare SNVs" that were only found in one or more anatomic sites of the same participant. The remaining "common" SNVs (found in 4 + of the 78 samples) appeared to be frequently occurring polymorphic sites within *C. trachomatis* populations. They were distributed across the genome but there was a peak in regions around the highly recombinogenic *omp*A gene. SNVs could be generated by genetic drift and/or sharing of populations between body sites. Alternatively, they could be artifacts of random sequencing error. Artifacts would be more likely to occur where there was lower coverage, as one or two miscalled bases could put the position in the 10–90% range for SNV calling. Some Group A participants with lower coverage had as many as 500 SNVs in only one of the body site samples but on inspection we found that SNVs at these positions were close to the 90% reference threshold, suggesting that they were likely to be false positives generated by sequencing error. Positions that had SNPs that were either fixed in two sites and SNV in the other, or fixed in one and SNV in the other two also were likely artifacts. In this case the SNVs were found at the 10% threshold and probably represented false positive SNPs that were fixed in all three sites. However, positions that were fixed SNPs in one body site but SNV at one other would be expected to be generated infrequently by sequence error. This pattern only occurred in three participants where, in each case, the body site that shared the mutations were the endocervix and vagina (**Figure 3**).

To help understand patterns of sharing within individuals we identified 5,520 genome positions that differentiated NP-UA and P-UA Fiji strains (see Methods). Because of pervasive recombination in *C. trachomatis* every strain had some alleles assigned to both clades but were overrepresented in alleles common in their own clade. In Group A samples, these clonal SNP sites (CSS) segregated across the chromosome as fixed differences (i.e., either mostly > 90% or < 10% reference allele frequency). The pattern seen in participant #1201 (**Figure 4**) is representative of the simple relationships seen in Group A. In this case, CSSs were dominated by NP-UA alleles (>90% reads aligning to reference bases) with few intermediate frequency SNVs. In Group A participants where the dominant strain was from the P-NP clade, the majority of CSS alleles were different from the reference genome (<10% reads aligning).

We saw more complex patterns of SNPs and SNVs in Group B participants compared to Group A. The simplest Group B participant was #564 where all genomes were in P-UA: the rectal and vaginal genomes were genotype E while the endocervix was genotype Ja. Therefore, the CSS showed all three sites exhibited a pattern typical of P-UA but the rectal and vaginal samples shared a large number of fixed SNPs (179 SNPs) not found in the endocervical sample. Conversely, the endocervical sample had unique fixed SNPs (231 SNPs) not found in the other two body sites (**Supplementary Table 3** and **Supplementary Figure 4**). Approximately 50% of these unique SNPs were found within blocks predicted by fastGEAR, suggesting that recombination was a major contributor to genetic differences between the two strains. A simple explanation of these patterns was that participant #564 contained multiple strains: caused by a P-UA Ja strain coinfecting the

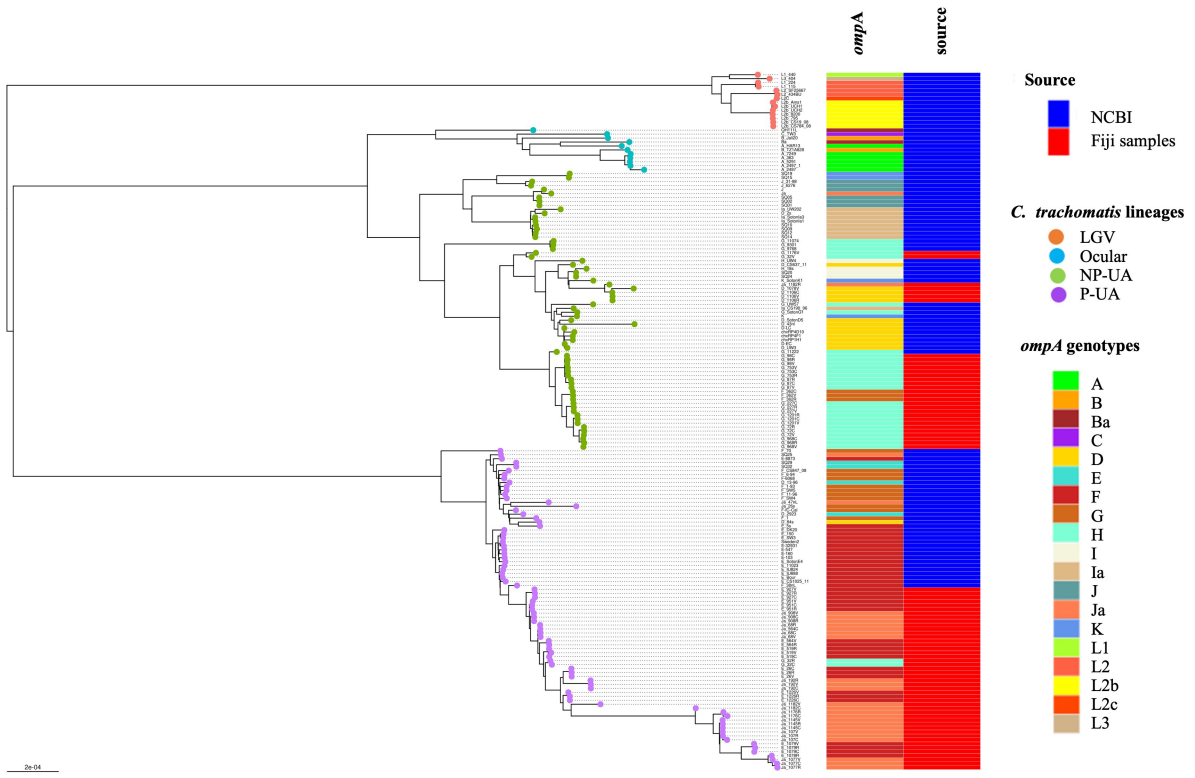


FIGURE 2
Global phylogeny with clade designations. The global phylogeny of high-quality *C. trachomatis* Fiji genomes plus selected complete *C. trachomatis* reference and clinical genomes representing global diversity from the National Center for Biotechnology Information (NCBI). Sample names are < *omp*A genotype > - < participant ID > - < body site code, where C = endocervix, R = rectum and V = vagina >. The round tips are colored by the 4 clade designations [LGV, ocular, prevalent- urogenital and anorectal (P-UA), non-prevalent urogenital and anorectal (NP-UA)]. The first column to the right of the tree denotes the *omp*A genotype with code at the lower right; the second column represents the source of the genomes from NCBI or the Fijian samples.

endocervix after another P-UAE strain had previously infected the rectum and vagina; the reverse order, with E strains coinfecting was also possible. The recombination events between genotypes could have occurred pre- or post-coinfection as natural transformation only requires that chlamydial DNA from a prior, non-viable infection or co-occurring infection be present that can be taken up by a newly infected cell.

In participants #32 and #1176, CSS patterns clearly showed strain mixing in the vaginal genome (**Figures 5**, **6**). While the endocervical and rectal genomes were dominated by alleles typical of P-UA strains, the vaginal genomes, located in the NP-UAs clade, had intermediate allele frequency across the length of the chromosome. Our interpretation of this pattern is that the vaginal samples contain a mixture of strains with P-UA and NP-UA chromosomes.

In participant #1078 interpretation was complicated by the lower data quality of endocervical and rectal samples: (only 249,618 and 875,018 bases with > 10× redundancy, respectively) (**Supplementary Figure 5**). There were many fixed SNPs in the rectal E genome, indicating it was from a different clade to the vaginal D genome. The pattern of CSS suggested some mixing of P-UA and NP-UA backgrounds in the vaginal genome. In #1078 all samples had the same plasmid subtype "D" despite differences in chromosome backgrounds suggesting possible plasmid transmission.

Discussion

The Fijian genomes sequenced in this study represent a sampling from the globally distributed P-UA and NP-UA clades. Although the phylogeny suggested multiple introductions of *C. trachomatis* strains from outside Fiji, there was also evidence for clonal expansion in both clades, presumably due to endemic local transmission (Svigals et al., 2020). There was evidence of recent DNA exchange between P-UA and NP-UA clades and possible local introductions of DNA from LGV and ocular clades into both clades of Fijian strains. This suggests that LGV and ocular strains might be present locally in Fiji but not common in the cohort we sampled, which would be expected as LGV is more common in Men who have Sex with Men, and the ocular strains are associated with the non-STI disease trachoma. However, trachoma is endemic to the Pacific Islands of the Western Pacific Region, and previous research has shown that both ocular and urogenital strains can be present at the same time in trachoma patients (Dean et al., 2008), which would provide an opportunity for genetic exchange.

This work is centered around sequencing *C. trachomatis* genomes directly from clinical samples using Agilent RNA bait libraries. This approach has been used to sequence bacterial species such as *C. trachomatis* and *Treponema pallidum* that are difficult to

TABLE 1 Terms used specific to this work.

Term	Explanation
CSS	"Clonal SNP sites". A set of 5,520 SNPs that were used to differentiate Fijian NP-UA from P-UA chromosomal backgrounds. They were defined at positions where 90% NP-UA had one allele and 90% P-UA had the other (reference C. trachomatis D/UW-3/CX is in the NP-UA clade).
"fixed SNP"	Single Nucleotide Polymorphism is defined here as a position with an allele frequency of less than 0.1 compared to the reference C. trachomatis D/UW-3/CX chromosome. For example, if the reference nucleotide at a position is "A", a fixed SNP would have > 90% sequencing reads as either "G", "C" or "T" aligning to that position.
LGV	"Lymphogranuloma venereum".
NP-UA	""non-prevalent urogenital and anorectal" clade in the *C. trachomatis* species phylogeny
P-UA	"Prevalent urogenital and anorectal" clade.
SNV	"Single nucleotide variant". Defined here as an allele frequency of > 0.1–0.9 < compared to the reference.
STI	"Sexually transmitted infection"

culture and are present in only small fractions of the metagenome (Seth-Smith et al., 2013, 2021; Hadfield et al., 2017; Pickering et al., 2020; Beale et al., 2021; Bowden et al., 2021). Here, we showed that the newly redesigned bait library (Bowden et al., 2021) could be used efficiently to produce high-quality genome sequences from samples with low yields of *C. trachomatis,* as measured by qPCR. Some samples had a high proportion of human DNA even after enrichment, leaving lower-coverage regions in the *C. trachomatis* genomes. However, we achieved good sequence data from 78 samples representing all three anatomic sites from 26 study participants.

There were complexities in the bioinformatic interpretation of the data, which arose because what was being sequenced was actually a within-species pool of strains rather than the pure cultures normally used in bacterial genomics projects. We showed that SNVs, which we defined here as having an allele frequency of 10-90%, were common across all samples but unevenly distributed, with some having many thousands more than the average due to random sequence errors in low-coverage regions. Non-artifactual SNVs could theoretically come from two sources: (1) the presence of more than one *C. trachomatis* strain through mixed infections; (2) mutation accumulation over time through population growth. Interpretation of the sequence pools in the absence of being able to culture pure cell lines is complex as multiple processes may be occurring, especially allowing for the possibility of recombination between subpopulations of strains within the sample pool (Somboonna et al., 2011). We also found that SNVs complicated analysis based on calling consensus nucleotide positions (e.g., *de novo assembly* or reference mapping using tools such as SNIPPY). While these methods worked well for placing samples on a phylogenetic tree, detailed analysis can be confused if SNVs are around the 50% consensus line. Consensus base calling means that distinct subpopulations are not recognized if they are distributed at significantly less than 50% frequency or alternatively, if over 50%, they are incorporated into the consensus.

To our knowledge, this is the first study to use genomics to assess within-host transmission dynamics for *C. trachomatis* STIs. Our analysis revealed two strikingly different patterns within participants: "Group A" (n = 21) had three anatomic samples with similar genetic background and *omp*A genotype or known

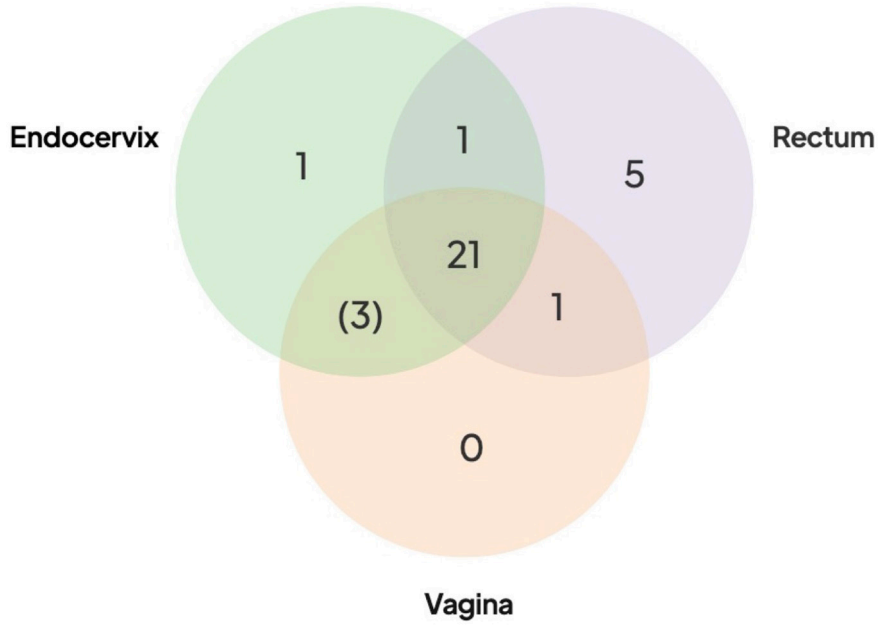


FIGURE 3
Distribution of shared SNPs by anatomic site in the 21 Group A participants. Venn diagram shows the number of participants with fixed SNPs (or fixed in one site with intermediate frequency in the other site in brackets) compared to the reference genome. All 21 participants had shared fixed SNPs in three body sites compared to the reference (center of the Venn diagram). More extensive breakdown of numbers of SNPs by participants are shown in **Supplementary Table 3**.

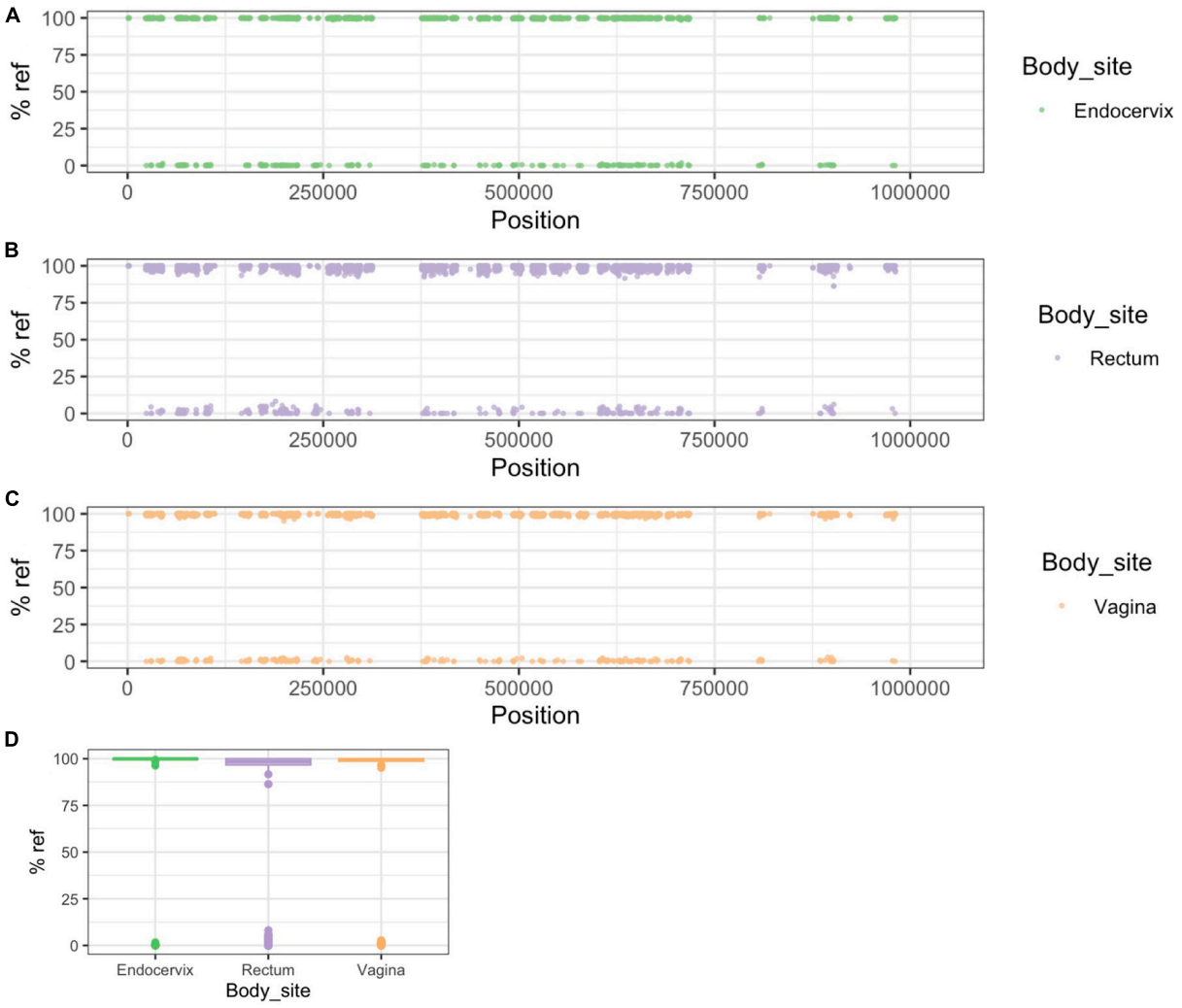


FIGURE 4
Patterns of SNP and SNV frequency across anatomic sites for representative Group A participant #1201. **(A–C)** Percent reference scores versus position on reference genome for "clonal SNPs" (CSSs) by body site. The set of 5,520 CSSs were chosen to differentiate NP-UA and P-UA genetic backgrounds. Each point shows the percentage of reads that mapped with the reference allele at each CSS position. The strains from #1201 are from the NP-UA clade and therefore, at most, CSSs are close to 100% match to the reference **(D)** allele, which is also in the NP-UA clade. The gaps in the distribution of CSSs across the chromosome are where there were regions of low variation or high recombination. **(D)** Box plot of distribution of% reference for clonal SNPs by body site. The minority of the CSSs with alternative alleles (<10% of reference genome) were likely the product of recombination events that have occurred since the divergence of the strains. Notably there is an intermediate frequency of SNVs.

recombinant (i.e., Ja *omp*A but E backbone), while "Group B" (n = 5) had one sample with a different background, implying a coinfection event. In the case of Group A, it was notable that only a minority of participants had samples with any fixed SNP differences and, if present, the modal number of SNPs was one (**Figure 3** and **Supplementary Table 3**). We argue that positions that were SNPs or rare SNVs shared between two samples from different anatomic sites were likely to be real. These, too, were rare in Group A women (**Figure 3**). As the mutation rate of *C. trachomatis* inferred from dated whole genome comparison is ~0.2 SNPs per genome per year (Hadfield et al., 2017; Seth-Smith et al., 2021), the most likely implication of these patterns is that there has been recent acquisition and transmission between anatomic sites in these participants. The simplest explanation is that these infections are quite transient and resolve before there has been time to accumulate significant variation between sites. This resolution may be due to recent infection and prescribed treatment proximal to a clinic visit or self-treatment with antibiotics that are available over-the-counter, limiting the longitudinal acquisition of SNPs. Either of these scenarios could result from symptomatic infection and health care seeking behavior—as would be presumed for the nine women (34.6%) in the study who were symptomatic—or asymptomatic infection with concern over sexual exposure to someone with an STI. These patterns could also be explained by more complex alternative models, for example, population contractions across all body sites followed by rapid re-seeding from one site with a small bottleneck.

The patterns of mutation might reveal pathways of transfer of *C. trachomatis* between anatomic sites, although care must be taken to not over-interpret the findings as the number of participants

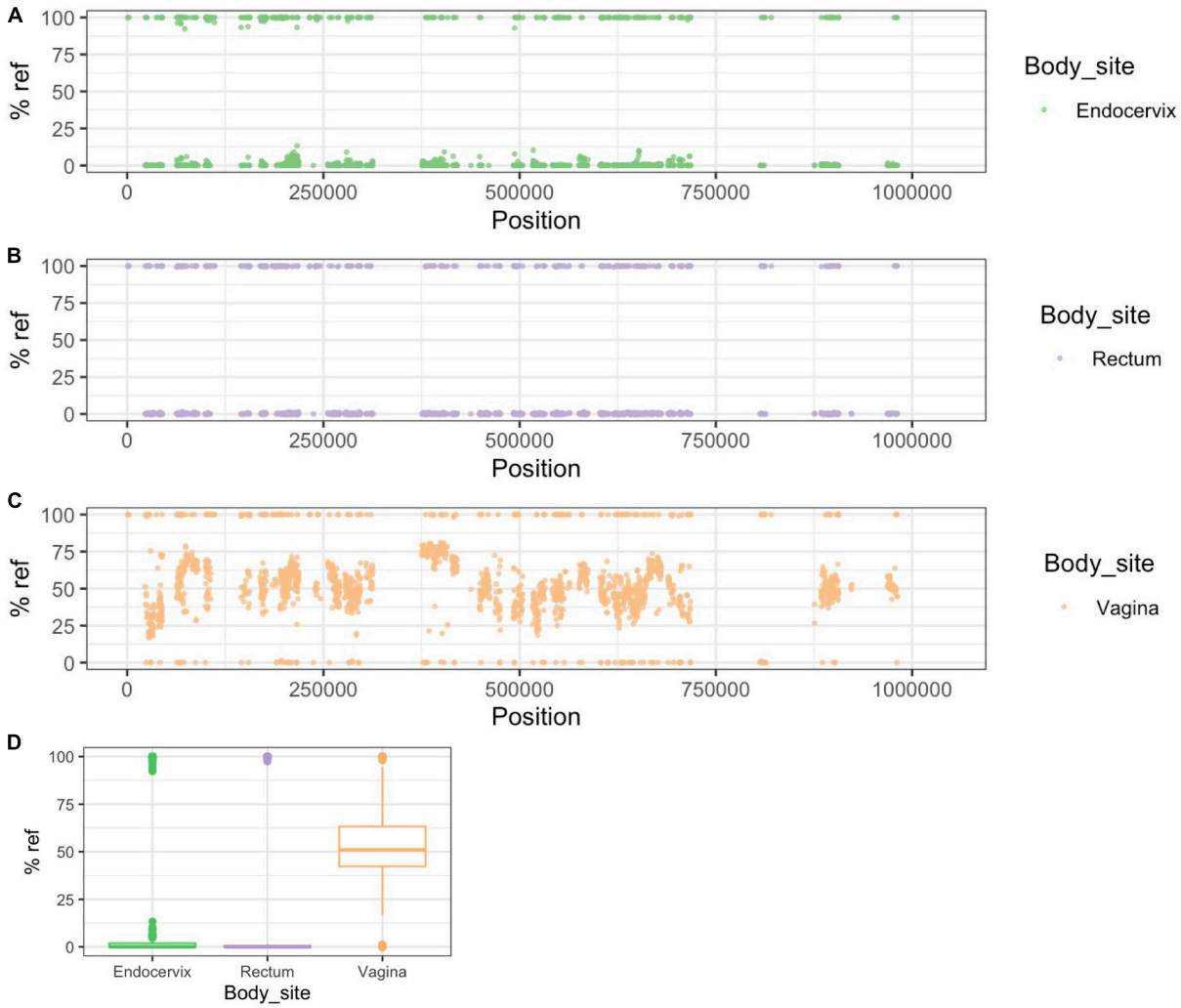


FIGURE 5
Patterns of SNP and SNV frequency across anatomic sites for Group B participant #32. **(A–C)** Percent reference scores versus position on reference genome for "clonal SNPs" (CSSs) by body site. **(D)** Box plot of distribution of % reference for clonal SNPs by body site. The endocervical and rectal strains were in the P-UA clade and therefore the majority of the CSSs had an alternative allele (<10% reference genome). The vaginal genome showed intermediate allele frequency across the chromosome, which was evidence of mixture between P-UA and NP-UA strains.

in this pilot study was small. SNPs and SNVs have been used to infer transmission between individuals (Worby et al., 2017) and in theory could also be used for potential events occurring between body sites of the same individual. It is possibly a sign of the biases in transmission between sites that unique fixed SNPs in Group A participants were more common in rectal samples, and that vaginal and endocervical samples more often had shared fixed and SNVs (**Figure 3**). The accumulation of SNPs in one site could be seen as a sign of population stratification caused by anatomy: The vaginal and endocervical *C. trachomatis* populations transmit between each other more frequently, given their proximity, than *C. trachomatis* in the rectum.

The Group B participant samples had much greater numbers of fixed and intermediate SNPs in pairwise comparisons than Group A. The simplest explanation for these is the coinfection of one anatomic site. The site with the divergent strain was not constant: in two cases it was the rectum (participants #1078R and #1082R), in two cases the vagina (participants #32V and #1176V) and in one case the endocervix (#564C). In four of these samples (#32V, #1078R, #1176V, #1182R) there was evidence of mixtures between *C. trachomatis* strains from different clades. These data show that sequencing of enriched *C. trachomatis genomes* directly from DNA of clinical samples can be used to identify co-infections, which are necessary for inter-strain recombination events to occur. The harmonization of plasmid genotypes in women containing *C. trachomatis* from different clades suggested that the process of plasmid replacement can be rapid. However, the caveat is that plasmid sequences in this study are based on PCR amplification and Sanger sequencing rather than Agilent bait pulldown, so it may not be possible to identify minor plasmid subpopulations.

This study revealed the intricacies of *C. trachomatis* within-host diversity and transmission during natural human infections and suggested that further investigation will yield information that

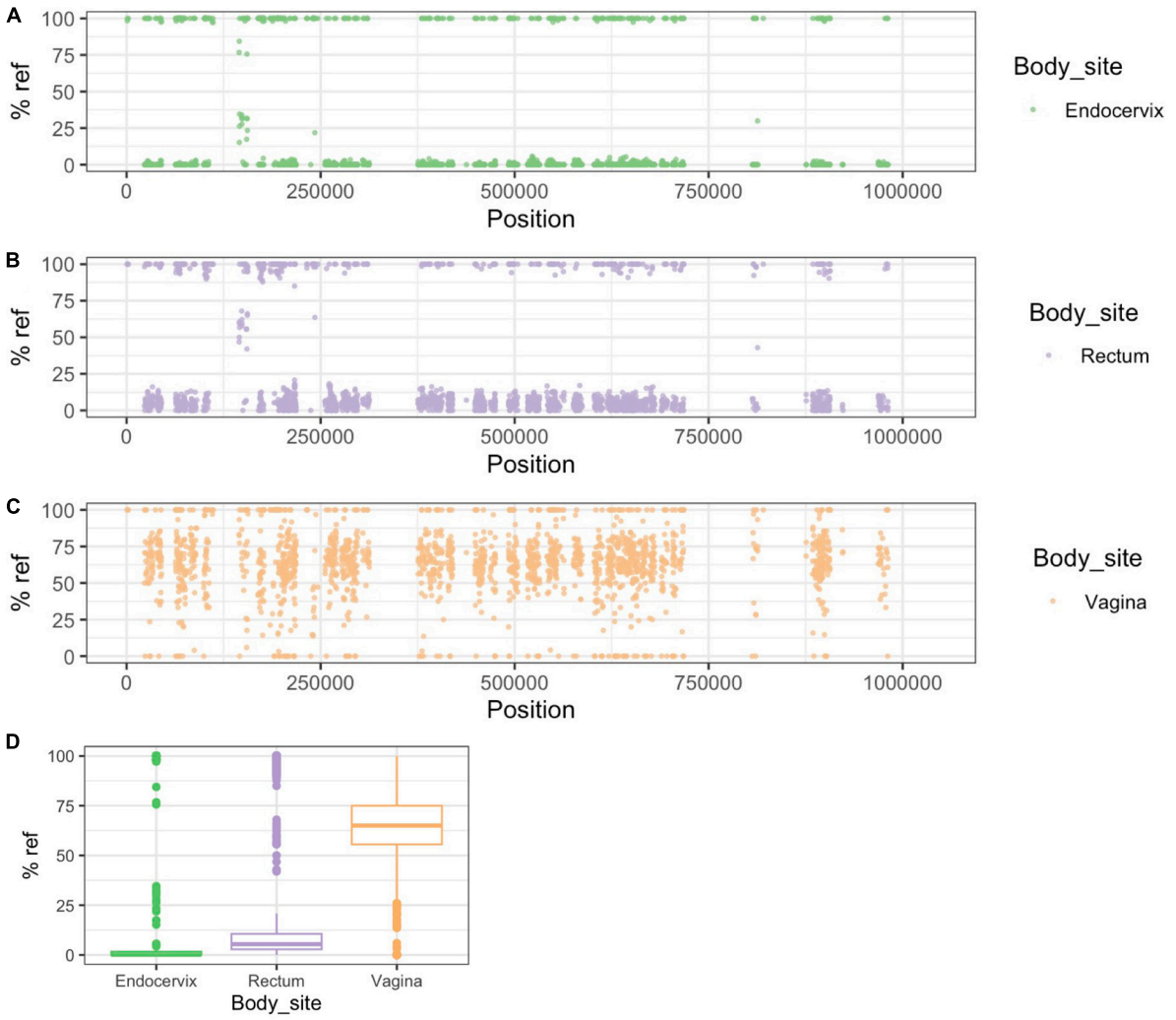


FIGURE 6
Patterns of SNP and SNV frequency across anatomic sites for Group B participant #1176. **(A–C)** Percent reference scores versus position on reference genome for "clonal SNPs" (CSSs) by body site. **(D)** Box plot of distribution of % reference for clonal SNPs by body site. The patterns in this participant are similar to **Figure 5** in showing evidence for the vaginal strain being a mixture of P-UA and NP-UA strains.

will help understand infection spread and disease processes. More samples are needed from a global sample set to know if these results can be extrapolated across human populations. Integration with additional bio-behavioral data will also be important to fully understand causes and direction of *C. trachomatis* transmission, although partner and sexual history are not always reliable. It would be ideal to expand individual datasets by conducting longitudinal studies to help resolve the dynamics of recombination and determine if multiple cycles of cross-infection occur between sites, but this would not be ethical as identification of infection requires treatment to eradicate *C. trachomatis*. Genomic approaches that resolve the potential subpopulations, such as single-cell sequencing (Zheng et al., 2022) and Hi-C (Kent et al., 2020), are hampered by *C. trachomatis* being only a minor component of the DNA in the clinical metagenomic sample. It may be possible to dissect recombination by isolating clonal *C. trachomatis* populations from individual samples and sequencing them independently. The technique commonly used for this is the plaque assay that is labor-intensive and not always guaranteed to completely separate out subpopulations (Somboonna et al., 2008). The most productive near-term strategy may be to continue to build up our picture of *C. trachomatis* natural infection by taking more "snapshots" of populations at single time points across multiple anatomic sites from a larger sample sizes of participants across Fiji, using the efficient RNA-bait methodology, to see if the patterns hold or diverge across a more global population, especially as tourism is a major part of the economy in Fiji.

Materials and methods

Study design and sample collection

The parent study was cross-sectional in design, enrolling women 18 years of age and older attending Ministry of Health and Medical Services (MoHMS) Health Centers in Fiji following

written informed consent as described (Svigals et al., 2020). Appropriate IRB approval had been obtained from UCSF (21-33864) and the Fijian MoHMS (FNHRERC 2015.100.MC) prior to commencement of the parent study. Women attended the Health Centers for various reasons including family planning, contraception, annual pap test, infertilty, a general check-up or interest in STI testing. Only women who did not use antibiotics within the prior month were consented and enrolled.

The current study was supplied with *C. trachomatis* positive endocervical, vaginal and rectal swab samples that had all been collected at the same time during the parent study to determine whether *C. trachomatis* was present in all three sites, especially the rectum as reported in other countries. All samples were de-identified with a unique ID number. All endocervical samples were collected by trained clinicians after cleaning the exocervix with a large cotton swab prior to inserting the collection swab directly into the endocervix, avoiding contact with the exocervix, vaginal wall or speculum. In addition, data on age and signs and/or symptoms of STIs were provided at the time of sample collection. The women were of iTauki ethnicity. None of the women engaged in transactional sex or reported anal intercourse or had a history of anal intercourse or any extragenital sex. All women were heterosexual and had one reported sexual partner who was also reported to be heterosexual.

Paired vaginal and rectal swabs were screened for *C. trachomatis* using the Cepheid Xpert CT/NG assay (Sunnyvale, CA) according to manufacturer's instructions. *C. trachomatis* positive endocervical samples were identified using a *C. trachomatis*-specific in-house qPCR assay as described (Bowden et al., 2021).

DNA extraction and determination of *Chlamydia trachomatis* copy number and load

Genomic (g)DNA was extracted from remnant Xpert CT/NG transport media for vaginal swabs and remnant M4 transport media (Thermo Fisher, South San Francisco, CA) for endocervical and rectal swabs as described previously (Bowden et al., 2021). Briefly, 59 μl consisting of 50 μL lysozyme (10 mg/mL; MilliporeSigma, St. Louis, MO), 3 μl of lysostaphin (4,000 U/mL in sodium acetate; MilliporeSigma) and 6 μl of mutanolysin (25,000 U/mL; MilliporeSigma) was added to 200 μl of remnant transport media and incubated for 1 h at 37°C as described (Bommana et al., 2022). The QIAamp DNA mini kit (Qiagen, California) was then used for DNA extraction, according to manufacturer's instructions. 5 μL of the resulting DNA underwent one or more displacement amplifications using the Repli-G MDA kit (Qiagen), to enrich microbial DNA. DNA concentration was measured using the Qubit dsDNA broad-range assay kit (Invitrogen).

Quantitative PCR (qPCR) was used to determine *C. trachomatis* genomic copy number and *C. trachomatis* load as described (Gomes et al., 2006; Sharma et al., 2018). Primers specific for the *C. trachomatis ompA* gene and for human Beta-Actin were used to generate standard curves of 10-fold serial increases in plasmids containing a single copy of each gene, respectively. Copy number of *C. trachomatis* and Beta-Actin for the clinical sample was determined based on comparison with the standard curve for the respective control plasmid. *C. trachomatis* load was estimated based on the ratio of bacteria (*C. trachomatis* genome copy number) per human cell (Beta-actin genome copy number) for each clinical sample to normalize the data against the host cell.

Chlamydia trachomatis ompA genotyping and plasmid sequencing

The *ompA* genotype was determined for each clinical sample as described previously (Somboonna et al., 2008). PCR was performed using primer pairs that flank the *ompA* gene; the product was sequenced in both directions and aligned using MAFFT v7.45062 to create the consensus sequence, which was then aligned with the 19 known *C. trachomatis* reference sequences to determine the *ompA* genotype. The reference strains were A/HAR-13, B/TW-5/OT, Ba/Apache-2, C/TW-3/OT, D/UW-3/Cx, Da/TW-448, E/Bour, F/IC-Cal-13, G/UW-57/Cx, H/UW-4/Cx, I/UW-12/Ur, Ia/UW-202, J/UW-36/Cx, Ja/UW-92, K/UW-31/Cx, L1/440, L2/434, L2a/UW-396, L2b/UCH-1/proctitis, L2c, and L3/404.

The plasmid for each clinical sample was sequenced as described (Bowden et al., 2021). Five primer pairs that flanked and covered the entire plasmid sequence were used, and the PCR products were sanger sequenced and aligned as above using MAFFT v7.45062 (Katoh and Standley, 2013). Each plasmid sequence was aligned to the 19 reference sequences to determine the plasmid identity.

Enrichment of *Chlamydia trachomatis* sequences from clinical samples using an agilent bait library

We used a methodology for RNA bait capture of *C. trachomatis* described in detail by Bowden et al. (2021). Human gDNA (Promega, San Luis Obispo, CA) was added to the extracted gDNA from the clinical swabs to reach a total input of 3 μg/130 μL for fragmentation and library prep. Samples were sheared on the Covaris LE220 plus (Covaris, Woburn, MA). After shearing and magnetic bead purification, the SureSelectXT Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library (VC2 Dec 2018) and all recommended quality control steps were performed on all gDNA samples. The 2.698 Mbp RNA bait library consisted of 34,795 120-mer probes spanning 85 GenBank *C. trachomatis* reference genomes (Bowden et al., 2021)(Agilent Technologies, INC, Santa Clara, CA, reference: ELID: 3173001). A 16-hour incubation at 65°C was performed for RNA bait library hybridization. Post-capture PCR cycling was set at 12 cycles based on a capture library size > 1.5 Mb. The libraries were paired end sequenced for 150 nt using an Illumina HiSeq instrument. Sequence data from this project was submitted to the NCBI Sequence Read Archive under the BioProject accession ID: PRJNA609714.

Post-sequencing bioinformatic isolation of *Chlamydia trachomatis* sequences

The post-enrichment raw sequencing reads were processed to remove the host genome and *C. trachomatis reads* were extracted and assembled into contigs as described in Bowden et al. (2021). We used an arbitrary threshold for good quality sequence data if the samples had at least 10× average *C. trachomatis* genome coverage post-enrichment and at least 5 reads mapped to > 900,000 bases of the 1,042,519 Mbp *C. trachomatis* reference D/UW-3/CX chromosome. To genotype the patient samples, *de novo* contigs were used to extract and compare the *omp*A genes against a customized BLAST (Altschul et al., 1990) database of the 21 reference *omp*A sequences as we described (Bowden et al., 2021).

Phylogeny and recombination inference

For the global phylogenetic analysis of the main chromosomes (total n = 176), we included all "good quality" genome sequences from the 26 participants (n = 77, with the exception of 1078C, which assembled into too many small contigs); and a collection of diverse C. *trachomatis* chromosomes available in NCBI (n = 99). We used a reference mapping approach with a custom version of *C. trachomatis* D/UW-3/CX by masking the 6 rRNA genes present in the repeated rRNA operons as described in Bowden et al. (2021), and generated a full-length whole genome alignment using snippy v4.3.8[1]. Snippy mapped the *C. trachomatis* reads from each sample to the reference genome using bwa and identified variants using Freebayes v1.0.2 (Garrison and Marth, 2012). The length of the region common to all samples with at least 10× read coverage and 90% read concordance at each site was 699,239 nucleotides with 11,971 polymorphic sites. Regions of increased density of homoplasious SNPs introduced by possible recombination events were predicted iteratively and masked using Gubbins (Croucher et al., 2014). The final maximum-likelihood (ML) global phylogenetic tree on 10,045 polymorphic sites was reconstructed using RAxML v8.2.9 (Kozlov et al., 2019) on the recombination removed (MRE) convergence criterion, along with ascertainment bias corrected using Stamatakis method. Lineage-specific phylogenetic trees were inferred as described above by using only the genomes from Fiji samples from their respective lineages.

fastGEAR (Mostowy et al., 2017) was run on a whole alignment that contained all "good quality" Fiji *C. trachomatis* genomes along with representative reference genomes from the clade on the global phylogenetic tree. This software infers the population structure and detects the "ancestral" and "recent" recombinations between the genomes present in the alignment. FastGEAR was run by clades with 100 iterations and checking for convergence. The statistical significance of the inferred recombination events (changes in SNP density between the two lineages) were assessed based on the natural log of Bayes factor calculated within FastGEAR. To understand the recombination events within group A individuals, we generated individual whole genome alignments from each of the three body sites by reference mapping the *C. trachomatis* reads to *C. trachomatis* D/UW-3/CX genomes using snippy and the within individual recombination events were inferred using Gubbins as described above.

Comparison of SNPs patterns between samples from the same participant

We used samtools mpileup (Li et al., 1000) to process the BAM files created by aligning sample FASTQ files against the reference chromosome to create tables of the numbers of each base (A, C, T, G) mapped to each individual base of reference. For each pair of samples from the same participant, we used R tidyverse tools (R Core Team, 2016; Wickham et al., 2019) to merge the positions with at least 10× read mapping redundancy. Code for analysis of the merged mpileup output was deposited to GitHub.[2]

To create a list of clonal SNP positions (CSSs), we performed Snippy alignment of all contigs from Fiji samples against the reference and identified positions where at least 90% of P-UA strains were identical but different to at least 90% of NP-UA strains. We then filtered out those falling in recombinant regions identified by Gubbins (see section above), leaving 5,520 CSS positions.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA609714, PRJNA609714.

Ethics statement

Appropriate IRB approval had been obtained from UCSF (21-33864) and the Fijian MoHMS (FNHRERC 2015.100.MC) prior to commencement of the parent study. The participants provided their written informed consent to participate in this study.

Author contributions

DD, MK, and NZ: sample collection. TR and SJ: primary bioinformatic analysis. TR, SJ, SB, and DD: writing. DD and TR: study concept and design. All authors contributed to the article and approved the submitted version.

Funding

TR and DD were supported by United States National Institutes of Health awards R21 AI138079 and R01 AI151075.

1 https://github.com/tseemann/snippy

2 https://github.com/Read-Lab-Confederation/Ct_MAP_analysis

Acknowledgments

We thank the parent study for providing the de-identified samples and for this study and Fijian colleagues: Rachel Devi, Kinisimere Nadredre, Mere Kurulo, and Darshika Balak. Thanks to Brian Raphael and Ellen Kersh for reading through the manuscript. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1154664/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Distribution of log-transformed ratio of the *C. trachomatis ompA* genome copy number to the beta-actin genome copy number (y-axis) is shown for each site. The load was significantly higher in the rectum compared to the vagina (P = 0.0124). C, endocervix; R, rectum; V, vagina.

SUPPLEMENTARY FIGURE 2
Comparison of the mean depth of sequencing coverage based on mapping of quality trimmed reads to the reference genome. Significant differences for endocervical depth compared to rectal and vaginal depth are shown. C, endocervix; R, rectum; V, vagina.

SUPPLEMENTARY FIGURE 3
Whole genome phylogenies of strains from this study from clades **(a)** P-UA and **(b)** NP-UA. Two G strains were found in the P-UA clade while three F and four D strains were found in the NP-UA clade. P-UA, prevalent urogenital and anorectal; NP-UA, non-prevalent urogenital and anorectal.

SUPPLEMENTARY FIGURE 4
Patterns of CSS frequency across anatomic sites for participant 1,078. For details of the plots see **Figure 4**. In this case, strains from the endocervix and vagina were in the NP-UA clade, and the rectum in the P-NP clade.

SUPPLEMENTARY FIGURE 5
Patterns of CSS frequency across anatomic sites for participant 564. For details of the plots see **Figure 3**. The high number of SNVs seen in this strain were a result of random errors in low sequence coverage regions.

References

Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Batteiger, B. (2020). "Chlamydia trachomatis," in *Mandell, douglas, and bennett's principles and practice of infectious diseases*, eds J. Bennett, R. Dolin, and M. Blaser (Elsevier).

Beale, M., Marks, M., Cole, M., Lee, M., Pitt, R., Ruis, C., et al. (2021). Global phylogeny of treponema pallidum lineages reveals recent expansion and spread of contemporary syphilis. *Nat. Microbiol.* 6, 1549–1560. doi: 10.1038/s41564-021-01000-z

Bommana, S., Richards, G., Kama, M., Kodimerla, R., Jijakli, K., Read, T., et al. (2022). Metagenomic shotgun sequencing of endocervical, vaginal, and rectal samples among fijian women with and without chlamydia trachomatis reveals disparate microbial populations and function across anatomic sites: a pilot study. *Microbiol. Spectr* 2022:e0010522. doi: 10.1128/spectrum.00105-22

Bowden, K., Joseph, S., Cartee, J., Ziklo, N., Danavall, D., Raphael, B., et al. (2021). Whole-genome enrichment and sequencing of chlamydia trachomatis directly from patient clinical vaginal and rectal swabs. *mSphere* 6, e1302–e1320. doi: 10.1128/mSphere.01302-20

Centers for Disease Control and Prevention and Department of Health and Human Services (2021). *Sexually transmitted diseases surveillance 2019*. Atlanta: CDC. Available online at: https://www.cdc.gov/std/statistics/2019/stdsurveillance-2019.pdf (accessed March 2023).

Chan, P., Robinette, A., Montgomery, M., Almonte, A., Cu-Uvin, S., Lonks, J., et al. (2016). Extragenital infections caused by chlamydia trachomatis and *Neisseria gonorrhoeae*: a review of the literature. *Infect. Dis. Obstet. Gynecol.* 2016:5758387. doi: 10.1155/2016/5758387

Chandra, N., Broad, C., Folkard, K., Town, K., Harding-Esch, E., Woodhall, S., et al. (2018). Detection of chlamydia trachomatis in rectal specimens in women and its association with anal intercourse: a systematic review and meta-analysis. *Sex Trans. Infect.* 2017:53161. doi: 10.1136/sextrans-2017-053161

Croucher, N., Page, A., Connor, T., Delaney, A., Keane, J., Bentley, S., et al. (2014). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* 2014:1196. doi: 10.1093/nar/gku1196

Dean, D., Kandel, R., Adhikari, H., and Hessel, T. (2008). Multiple chlamydiaceae species in trachoma: implications for disease pathogenesis and control. *PLoS Med.* 5:e14. doi: 10.1371/journal.pmed.0050014

Drummond, F., Ryder, N., Wand, H., Guy, R., Read, P., McNulty, A., et al. (2011). Is azithromycin adequate treatment for asymptomatic rectal chlamydia? *Int. J. STD AIDS* 22, 478–480. doi: 10.1258/ijsa.2011.010490

Dukers-Muijrers, N., Speksnijder, A., Morré, S., Wolffs, P., van der Sande, M., Brink, A., et al. (2013). Detection of anorectal and cervicovaginal chlamydia trachomatis infections following azithromycin treatment: prospective cohort study with multiple time-sequential measures of rRNA, DNA, quantitative load and symptoms. *PLoS One* 8:e81236. doi: 10.1371/journal.pone.0081236

Foschi, C., Salvo, M., Cevenini, R., and Marangoni, A. (2018). Chlamydia trachomatis antimicrobial susceptibility in colorectal and endocervical cells. *J. Anti. Chem.* 73, 409–413. doi: 10.1093/jac/dkx392

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv* [preprint]. doi: 10.48550/arXiv.1207.3907

Gomes, J., Borrego, M., Atik, B., Santo, I., Azevedo, J., Brito, A., et al. (2006). Correlating chlamydia trachomatis infectious load with urogenital ecological success and disease pathogenesis. *Microbes Infect.* 8, 16–26. doi: 10.1016/j.micinf.2005.05.014

Gomes, J., Bruno, W., Borrego, M., and Dean, D. (2004). Recombination in the genome of chlamydia trachomatis involving the polymorphic membrane protein C gene relative to OMP A and evidence for horizontal gene transfer. *J. Bacteriol.* 186, 4295–4306. doi: 10.1128/JB.186.13.4295-4306.2004

Hadfield, J., Harris, S., Seth-Smith, H., Parmar, S., Andersson, P., Giffard, P., et al. (2017). Comprehensive global genome dynamics of chlamydia trachomatis show ancient diversification followed by contemporary mixing and recent lineage expansion. *Geno. Res.* 2017:116. doi: 10.1101/gr.212647.116

Haggerty, C., Gottlieb, S., Taylor, B., Low, N., Xu, F., and Ness, R. (2010). Risk of sequelae after chlamydia trachomatis genital infection in women. *J. Infect. Dis.* 201, S134–S155.

Jeffrey, B., Suchland, R., Quinn, K., Davidson, J., Stamm, W., and Rockey, D. (2010). Genome sequencing of recent clinical chlamydia trachomatis strains identifies loci

associated with tissue tropism and regions of apparent recombination. *Infect. Immun.* 78, 2544–2553. doi: 10.1128/IAI.01324-09

Joseph, S., Didelot, X., Rothschild, J., de Vries, H., Morré, S., Read, T., et al. (2012). Population genomics of chlamydia trachomatis: insights on drift, selection, recombination, and population structure. *Mol. Biol. Evol.* 29, 3933–3946. doi: 10.1093/molbev/mss198

Katoh, K., and Standley, D. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kent, A., Vill, A., Shi, Q., Satlin, M., and Brito, I. (2020). Widespread transfer of mobile antibiotic resistance genes within individual gut microbiomes revealed through bacterial Hi-C. *Nat. Commun.* 11:4379. doi: 10.1038/s41467-020-18164-7

Khosropour, C., Dombrowski, J., Barbee, L., Manhart, L., and Golden, M. (2014). Comparing azithromycin and doxycycline for the treatment of rectal chlamydial infection: a retrospective cohort study. *Sex Trans. Dis.* 41, 79–85.

Kong, F., Tabrizi, S., Law, M., Vodstrcil, L., Chen, M., Fairley, C., et al. (2014). Azithromycin versus doxycycline for the treatment of genital chlamydia infection: a meta-analysis of randomized controlled trials. *Clin. Infect. Dis.* 59, 193–205.

Kozlov, A., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455. doi: 10.1093/bioinformatics/btz305

Lanjouw, E., Ouburg, S., de Vries, H., Stary, A., Radcliffe, K., and Unemo, M. (2016). 2015 European guideline on the management of chlamydia trachomatis infections. *Int. J. STD AIDS* 27, 333–348.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (1000). Genome project data processing subgroup. 2009. the sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Mostowy, R., Croucher, N., Andam, C., Corander, J., Hanage, W., and Marttinen, P. (2017). Efficient inference of recent and ancestral recombination within bacterial populations. *Mol. Biol. Evol.* 2017:66. doi: 10.1093/molbev/msx066

Peipert, J., Ness, R., Soper, D., and Bass, D. (2000). Association of lower genital tract inflammation with objective evidence of endometritis. *Infect. Dis. Obstet. Gynecol.* 8, 83–87.

Pickering, H., Chernet, A., Sata, E., Zerihun, M., Williams, C., Breuer, J., et al. (2020). Genomics of ocular chlamydia trachomatis after 5 years of safe interventions for trachoma in amhara, Ethiopia. *J. Infect. Dis.* 2020:138982. doi: 10.1093/infdis/jiaa615

R Core Team. (2016). *R foundation for statistical computing*. Vienna, Austria: R: A language and environment for statistical computing.

Satterwhite, C., Torrone, E., Meites, E., Dunne, E., Mahajan, R., Ocfemia, M., et al. (2013). Sexually transmitted infections among US women and men: prevalence and incidence estimates, 2008. *Sex Tran. Dis.* 40, 187–193.

Seth-Smith, H., Bénard, A., Bruisten, S., Versteeg, B., Herrmann, B., Kok, J., et al. (2021). Ongoing evolution of chlamydia trachomatis lymphogranuloma venereum: exploring the genomic diversity of circulating strains. *Microb. Genom.* 7:000599. doi: 10.1099/mgen.0.000599

Seth-Smith, H., Harris, S., Skilton, R., Radebe, F., Golparian, D., Shipitsyna, E., et al. (2013). Whole-genome sequences of chlamydia trachomatis directly from clinical samples without culture. *Gen. Res.* 23, 855–866.

Sharma, M., Recuero-Checa, M., Fan, F., and Dean, D. (2018). Chlamydia trachomatis regulates growth and development in response to host cell fatty acid availability in the absence of lipid droplets. *Cell Microbiol.* 20:10.1111/cmi.12801. doi: 10.1111/cmi.12801

Smelov, V., Vrbanac, A., van Ess, E., Noz, M., Wan, R., Eklund, C., et al. (2017). Chlamydia trachomatis strain types have diversified regionally and globally with evidence for recombination across geographic divides. *Front. Microbiol.* 8:2195. doi: 10.3389/fmicb.2017.02195

Somboonna, N., Mead, S., Liu, J., and Dean, D. (2008). Discovering and differentiating new and emerging clonal populations of chlamydia trachomatis with a novel shotgun cell culture harvest assay. *Emerg Infect. Dis.* 14, 445–453. doi: 10.3201/eid1403.071071

Somboonna, N., Wan, R., Ojcius, D., Pettengill, M., Joseph, S., Chang, A., et al. (2011). Hypervirulent chlamydia trachomatis clinical strain is a recombinant between lymphogranuloma venereum (L(2)) and D lineages. *MBio* 2:e45. doi: 10.1128/mBio.00045-11

Stoner, B., and Cohen, S. (2015). Lymphogranuloma venereum 2015: clinical presentation, diagnosis, and treatment. *Clin. Infect. Dis.* 61, S865–S873.

Svigals, V., Blair, A., Muller, S., Sahu Khan, A., Faktaufon, D., Kama, M., et al. (2020). Hyperendemic chlamydia trachomatis sexually transmitted infections among females represent a high burden of asymptomatic disease and health disparity among pacific islanders in fiji. *PLoS Negl. Trop. Dis.* 14:e0008022. doi: 10.1371/journal.pntd.0008022

van Liere, G., Dukers-Muijrers, N., Levels, L., and Hoebe, C. (2017). High proportion of anorectal chlamydia trachomatis and *Neisseria gonorrhoeae* after routine universal urogenital and anorectal screening in women visiting the sexually transmitted infection clinic. *Clin. Infect. Dis.* 64, 1705–1710. doi: 10.1093/cid/cix243

van Liere, G., van Rooijen, M., Hoebe, C., Heijman, T., de Vries, H., and Dukers-Muijrers, N. (2015). Prevalence of and factors associated with rectal-only chlamydia and gonorrhoea in women and in men who have sex with men. *PLoS One* 10:e0140297. doi: 10.1371/journal.pone.0140297

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the tidyverse. *J. Open Source Softw* 4:1686.

Worby, C., Lipsitch, M., and Hanage, W. (2017). Shared genomic variants: identification of transmission routes using pathogen deep sequence data. *Am. J. Epid.* 2017:182. doi: 10.1093/aje/kwx182

World Health Organization (2020). *Sexually transmitted infections (STIs) key facts*. Available online at: http://www.who.int/mediacentre/factsheets/fs110/en/ (accessed March 2023).

Zheng, W., Zhao, S., Yin, Y., Zhang, H., Needham, D., Evans, E. D., et al. (2022). High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome. *Science* 376:eabm1483. doi: 10.1126/science.abm1483

Figure S1

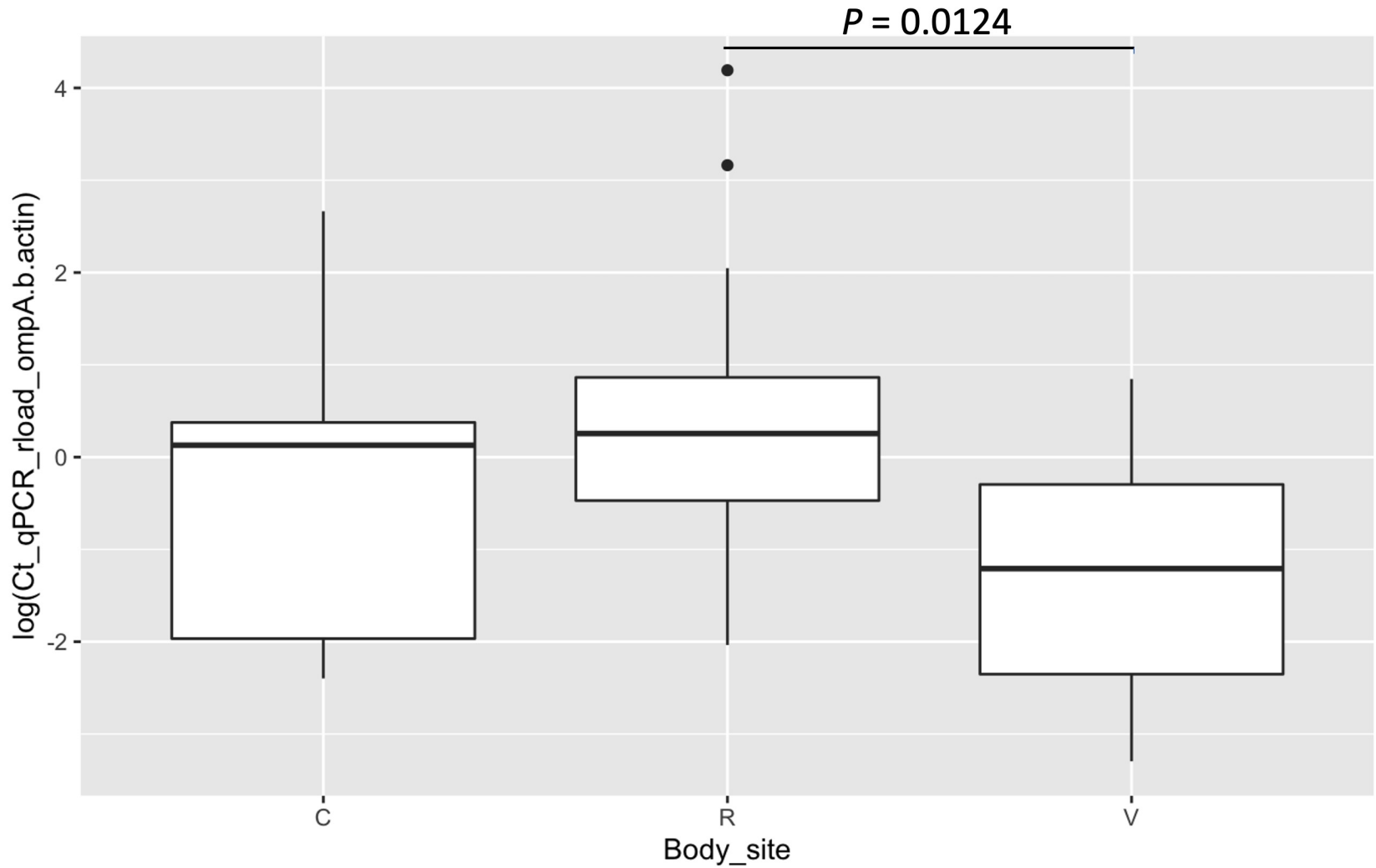


Figure S2

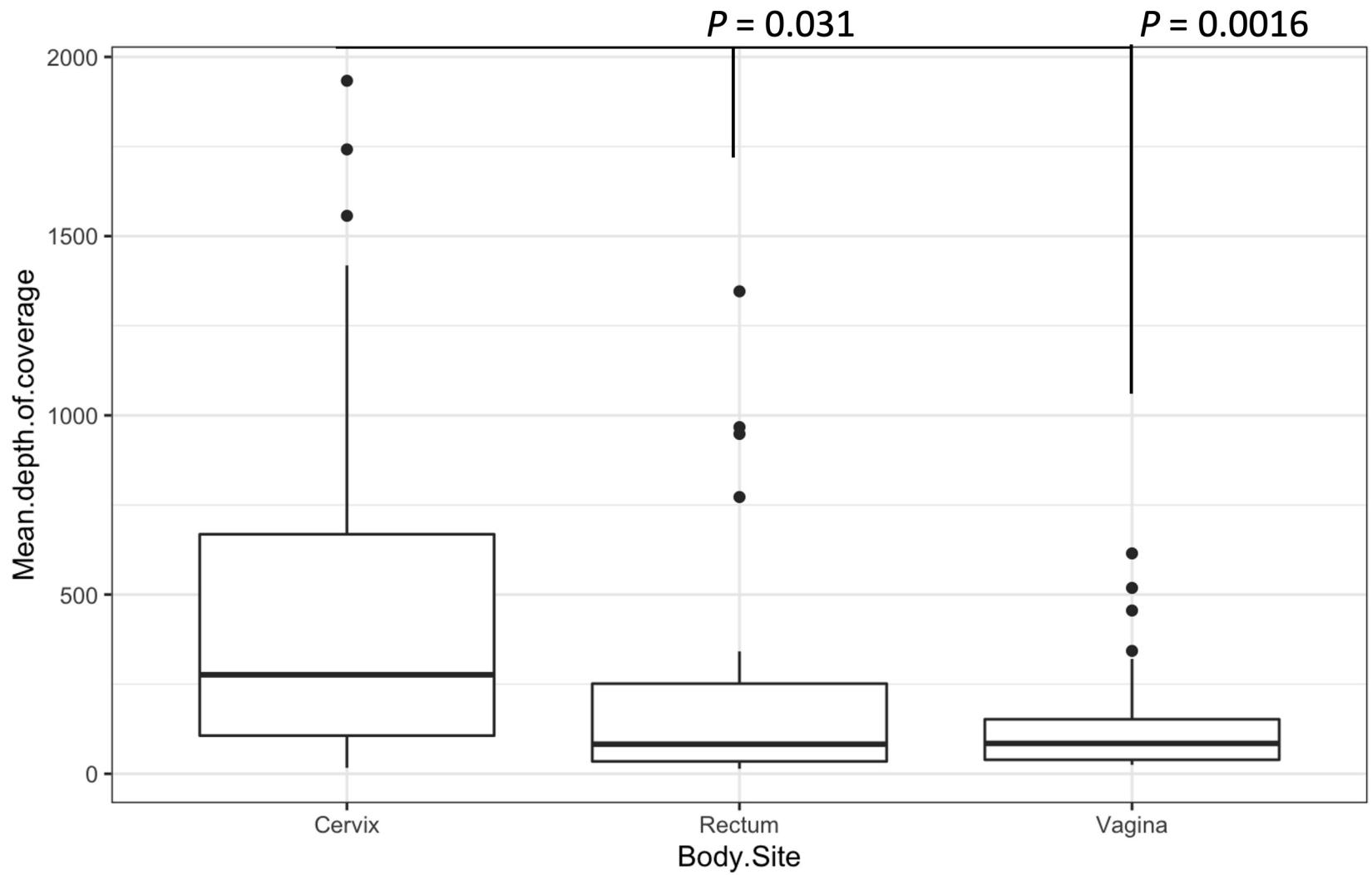
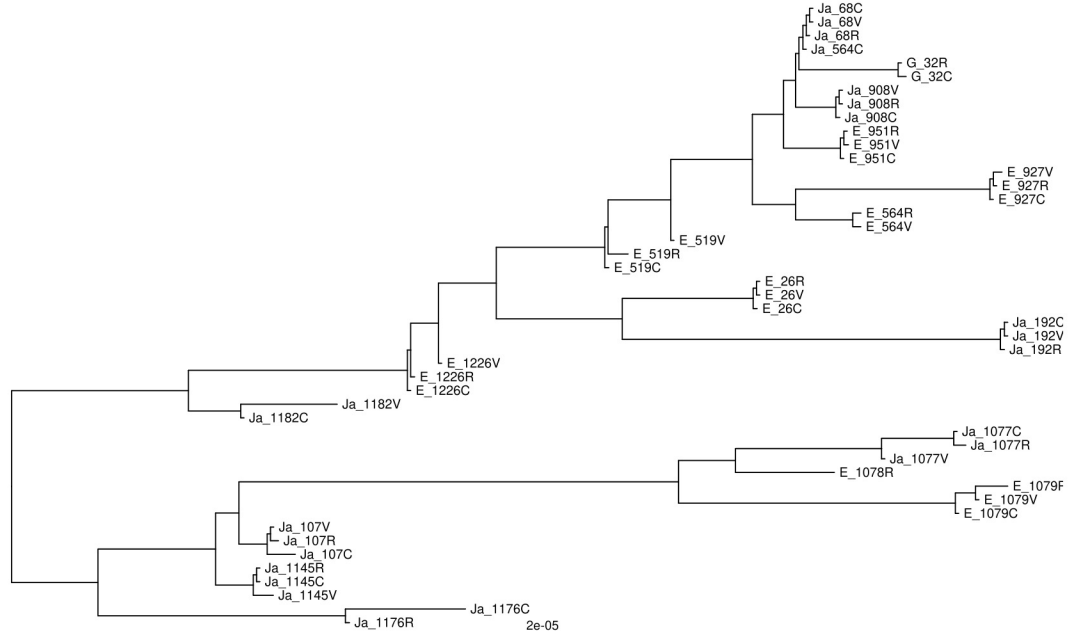

P = 0.031
P = 0.0016
2000
1500
1000
500
0
Mean.depth.of.coverage
Cervix
Rectum
Vagina
Body.Site

Figure S3

a


Ja_68C
Ja_68V
Ja_68R
Ja_564C
G_32R
G_32C
Ja_908V
Ja_908R
Ja_908C
E_951R
E_951V
E_951C
E_927V
E_927R
E_927C
E_564R
E_564V
E_519V
E_519R
E_519C
E_26R
E_26V
E_26C
Ja_192C
Ja_192V
Ja_192R
E_1226V
E_1226R
E_1226C
Ja_1182V
Ja_1182C
Ja_1077C
Ja_1077R
Ja_1077V
E_1078R
E_1079F
E_1079V
E_1079C
Ja_107V
Ja_107R
Ja_107C
Ja_1145R
Ja_1145C
Ja_1145V
Ja_1176C
Ja_1176R
2e-05

b

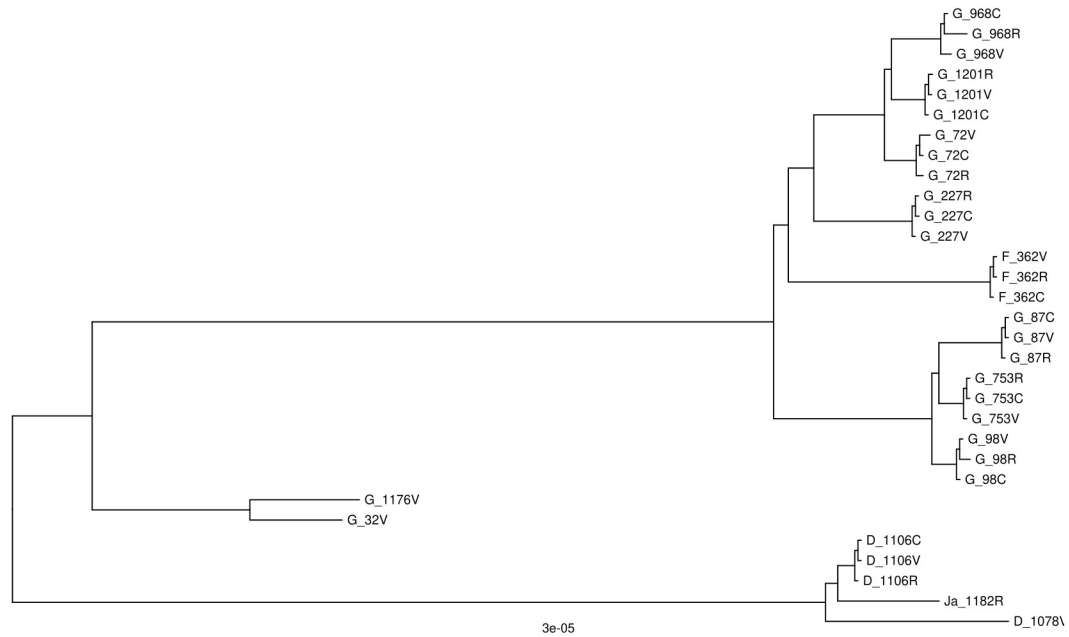

G_968C
G_968R
G_968V
G_1201R
G_1201V
G_1201C
G_72V
G_72C
G_72R
G_227R
G_227C
G_227V
F_362V
F_362R
F_362C
G_87C
G_87V
G_87R
G_753R
G_753C
G_753V
G_98V
G_98R
G_98C
G_1176V
G_32V
D_1106C
D_1106V
D_1106R
Ja_1182R
D_1078\
3e-05

Figure S4

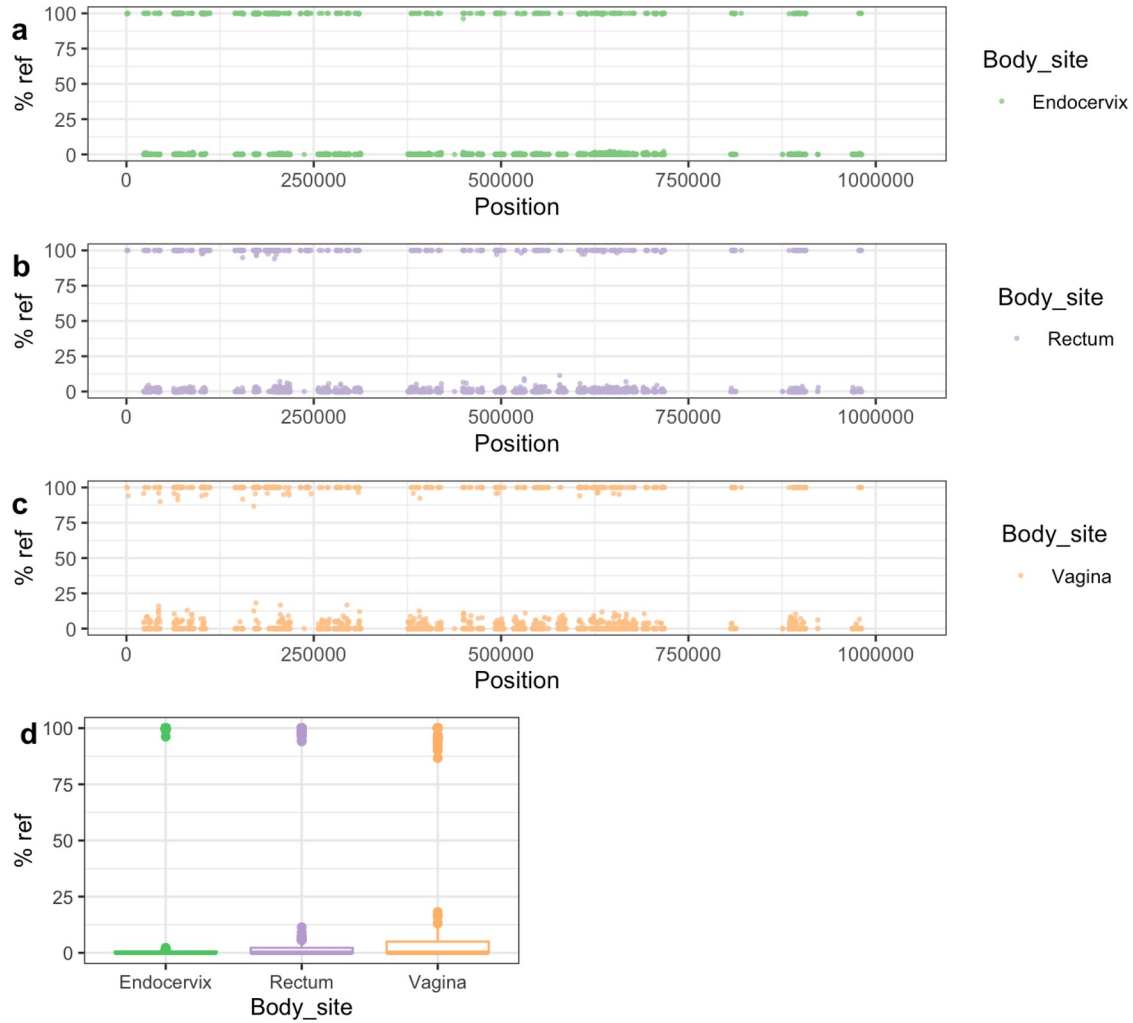
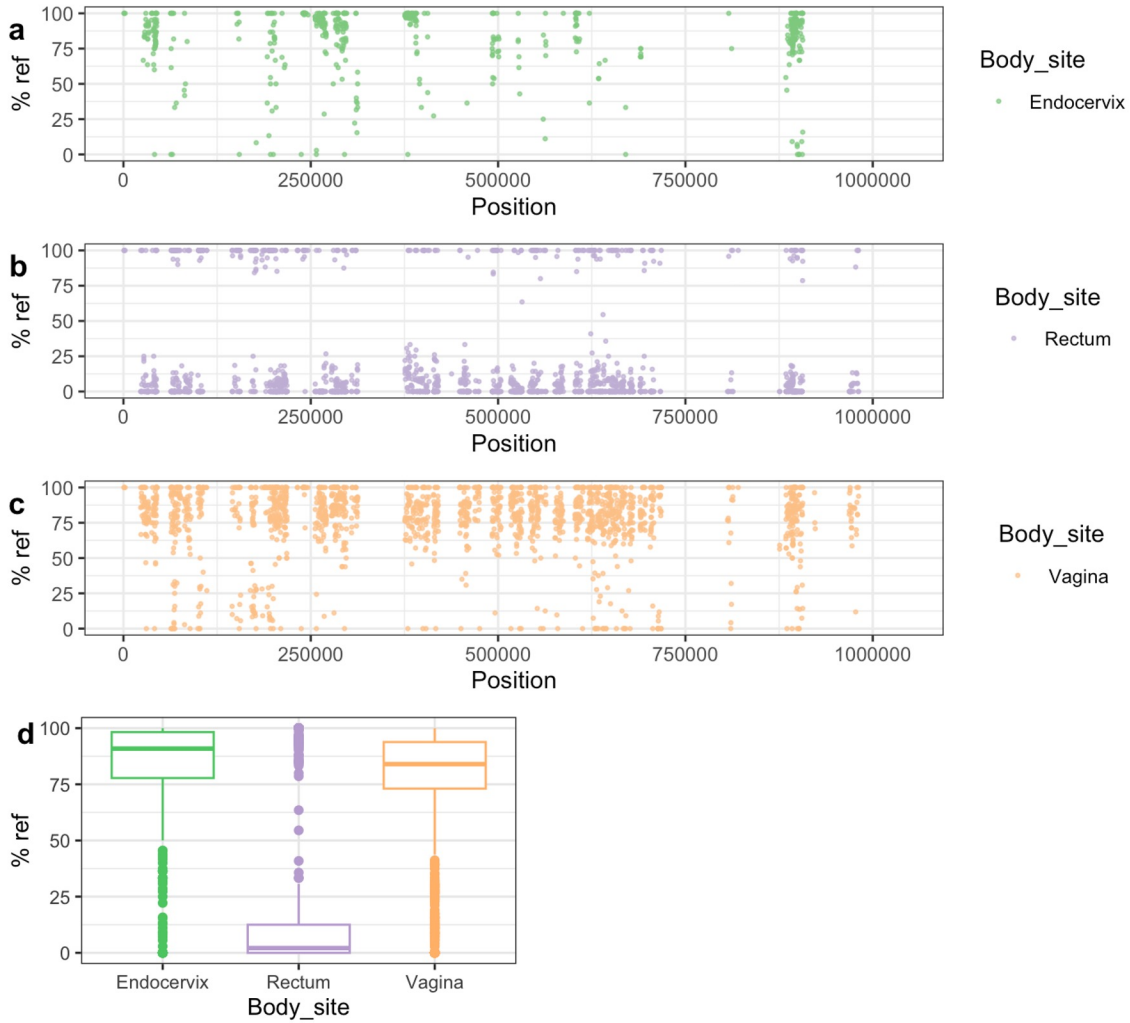

a
b
c
d
% ref
100
75
50
25
0
0
250000
500000
750000
1000000
Position
Body_site
Endocervix
Rectum
Vagina
Endocervix
Rectum
Vagina
Body_site

Figure S5


a
b
c
d
% ref
Position
0
250000
500000
750000
1000000
0
25
50
75
100
Body_site
Endocervix
Rectum
Vagina

Supplemental Table 1. Metadata, typing and genome sequence quality control statistics associated with the samples from 26 women with good quality genomic sequences in three anatomic sites.

Sample_name	Age*	Signs & Symptoms**	Anatomic Site	# of Raw Read pairs	# of read pairs after adaptor trimming and Quality score (>=15)	paired classified as C.	per base on Ct reference	positions covered in the reference with >=10 reads	positions covered in the reference with	nucleotide positions covered in	ompA genotype	Clade***	plasmid genotype	Ct ompA (copies/µl)	Ct b-actin (copies/µl)	Load (ompA/b-actin)
26C	20	yes	Endocervix	3881771	1973633	1128927	309.7515	1030529	1030662	1030986	E	P-UA	No data	534.5	233.672	2.287
26R	20	no	Rectum	4376496	2155100	607714	164.5129	1030451	1030558	1030993	E	P-UA	E	3075.5	684.098	4.496
26V	20	no	Vagina	2538201	875282	360482	98.48688	1029795	1030419	1030997	E	P-UA	E	1645	NA	NA
32C	26	no	Endocervix	2768653	1300007	1047385	284.9352	979976	1011880	1029955	G	P-UA	E	452.382	2103.437	0.215
32R	26	yes	Rectum	8518792	4120862	1249195	341.3122	1030360	1030575	1031049	G	P-UA	No data	828.676	1617.977	0.512
32V	26	yes	Vagina	3493014	1038876	575175	156.9291	1031261	1031464	1031627	G	NP-UA	E	875.107	305.099	2.868
68C	26	no	Endocervix	3255345	1494551	839394	227.504	1030218	1030319	1030782	Ja	P-UA	E	1683.931	1021.489	1.649
68R	26	no	Rectum	3964230	2171942	300439	82.32851	1028335	1029607	1030976	Ja	P-UA	No data	262.311	2008.173	0.131
68V	26	no	Vagina	2399972	825022	232670	63.32957	1029398	1030136	1030693	Ja	P-UA	E	2005.341	1433.92	1.399
72C	32	no	Endocervix	9061129	5449349	5188999	1418.151	1031271	1031313	1031928	G	NP-UA	D/G	2107.48	482.501	4.368
72R	32	no	Rectum	1764699	1429554	53096	13.99784	778741	1002017	1030750	G	NP-UA	D/G	1664	41.292	40.299
72V	32	no	Vagina	1568358	463667	169916	46.18477	1028273	1030168	1031293	G	NP-UA	D/G	349.2	2495.272	0.14
87C	24	no	Endocervix	3379564	858759	584847	156.5297	1030948	1031164	1031720	G	NP-UA	D/G	186.8	2233.7	0.084
87R	24	no	Rectum	4378845	2091933	125227	33.88708	1026633	1030429	1031485	G	NP-UA	D/G	953.3	778.639	1.224
87V	24	yes	Vagina	3583086	639211	142218	38.48257	1027794	1030251	1031481	G	NP-UA	D/G	151.5	1666.666	0.091
98C	22	yes	Endocervix	10110856	6618509	6510165	1742.018	1031308	1031462	1031966	G	NP-UA	D/G	344	3947.599	0.087
98R	22	no	Rectum	1930840	1715197	962268	262.0773	1031377	1031487	1031574	G	NP-UA	D/G	31549	555.293	56.815
98V	22	yes	Vagina	1259385	791372	307678	84.66804	1029099	1030820	1031531	G	NP-UA	D/G	434.4	1809.354	0.24
107C	29	no	Endocervix	33923070	2934946	324782	86.66605	859337	928194	994243	Ja	P-UA	E	2218	1301.548	1.704
107R	29	no	Rectum	2302152	1924399	705294	190.4395	1030324	1030537	1030800	Ja	P-UA	E	18377	708.586	25.935
107V	29	no	Vagina	1347716	942944	175198	47.05494	1028088	1029717	1030586	Ja	P-UA	E	120.18	815.621	0.147
192C	26	no	Endocervix	5476564	2157973	1576686	425.9618	1030173	1030356	1031167	Ja	P-UA	E	328	2383.453	0.138
192R	26	no	Rectum	2175904	1793255	1184836	312.305	1030450	1030549	1030744	Ja	P-UA	E	43905	1550.326	28.32
192V	26	no	Vagina	1390989	940352	331250	90.01676	1030137	1030378	1030584	Ja	P-UA	E	275.1	NA	NA
227C	21	yes	Endocervix	4827086	2972494	2651587	728.0298	1031271	1031343	1031864	G	NP-UA	D/G	1824	2790.564	0.654
227R	21	yes	Rectum	17615509	5210586	86270	22.26701	987852	1026120	1030982	G	NP-UA	D/G	470.5	NA	NA
227V	21	no	Vagina	4115404	1431140	402322	109.5889	1030961	1031102	1031458	G	NP-UA	D/G	3257.49	NA	NA
362C	35	no	Endocervix	8046629	4720587	887492	239.4672	1030593	1031012	1031791	F	NP-UA	D/G	1933.677	829.484	2.331
362R	35	no	Rectum	7028809	3899905	613008	165.2597	1029180	1030515	1031524	F	NP-UA	D/G	576.25	1236.356	0.466
362V	35	no	Vagina	2785894	1277347	297558	80.73862	1030914	1031031	1031456	F	NP-UA	D/G	136.42	1067.452	0.128
519C	19	no	Endocervix	10915045	7121754	7083988	1933.492	1030759	1030904	1031711	E	P-UA	No data	532.044	1918.343	0.277
519R	19	no	Rectum	23015761	3637332	175582	46.22006	927851	995316	1027451	E	P-UA	E	185.797	136.429	1.362
519V	19	yes	Vagina	2521704	1440450	457803	123.0051	1029979	1030347	1031041	E	P-UA	E	63753.539	NA	NA
564C	32	no	Endocervix	4201603	1928478	1393171	379.1702	1030246	1030390	1031396	Ja	P-UA	E	894.99	2781.221	0.322
564R	32	no	Rectum	4643615	1954793	259807	70.08115	1029757	1030218	1030957	E	P-UA	No data	404.064	2515.309	0.161
564V	32	no	Vagina	3677998	1262700	128965	34.80634	1024900	1028851	1030891	E	P-UA	E	2953.966	794.379	3.719
753C	23	no	Endocervix	4811591	2477195	351316	94.55193	1030538	1031074	1031552	G	NP-UA	No data	772.9	3714.723	0.208
753R	23	no	Rectum	3858570	2348430	75702	20.22568	925660	1016985	1030573	G	NP-UA	D/G	121.89	5.158	23.633
753V	23	no	Vagina	2489300	1120281	560065	152.646	1030820	1031105	1031588	G	NP-UA	D/G	570.39	NA	NA
908C	31	no	Endocervix	3044175	587948	351365	95.79307	1028111	1029501	1031150	Ja	P-UA	E	116.754	2771.414	0.042
908R	31	no	Rectum	3595960	2193839	802050	219.5479	1029593	1029970	1030676	Ja	P-UA	E	483.091	779.604	0.62
908V	31	no	Vagina	25289804	2368406	342739	87.13967	1027162	1029070	1030287	Ja	P-UA	E	1477.985	NA	NA
927C	38	no	Endocervix	4153249	2063566	1804757	488.6605	1030526	1030707	1031018	E	P-UA	No data	1295.605	2437.583	0.532
927R	38	yes	Rectum	7173247	4290127	3482170	948.1692	1030640	1030810	1031504	E	P-UA	E	1904.665	NA	NA
927V	38	yes	Vagina	4960468	2505383	1894046	518.5069	1030516	1030630	1031226	E	P-UA	E	2350.484	NA	NA
951C	30	no	Endocervix	3002396	1697833	1591389	434.9692	1030210	1030390	1031095	E	P-UA	No data	840.525	2249.041	0.374
951R	30	no	Rectum	11964131	7658027	4952271	1345.959	1030963	1031735	1032243	E	P-UA	E	1159.613	651.188	1.781
951V	30	no	Vagina	4113765	2118273	555930	151.452	1030100	1030593	1031837	E	P-UA	E	2638.747	NA	NA
968C	22	no	Endocervix	9296911	5961175	5692279	1556.688	1031318	1031470	1032082	G	NP-UA	No data	865.662	6158.418	0.141
968R	22	yes	Rectum	4395902	2364069	65255	17.4249	882607	1009672	1030457	G	NP-UA	No data	1201.895	18.175	66.129
968V	22	yes	Vagina	4098543	1864162	742576	201.6281	1031096	1031313	1031790	G	NP-UA	D/G	15179.26	NA	NA
1077C	30	no	Endocervix	2707271	1379498	973867	265.4235	1030256	1030606	1031899	Ja	P-UA	No data	122.744	2572.965	0.048
1077R	30	no	Rectum	4536604	2112188	146444	36.00623	966434	1021812	1030909	Ja	P-UA	No data	368.139	NA	NA
1077V	30	yes	Vagina	2828731	817391	111963	30.16658	1019869	1027960	1030554	Ja	P-UA	E	1135.017	NA	NA
1078C	26	no	Endocervix	44313968	8314392	62220	18.3445	437082	751250	1008294	na	na	D/G	77.516	1113.549	0.07
1078R	26	no	Rectum	30823554	10871094	104468	30.8532	987934	1025420	1030651	E	P-UA	D/G	123.705	15.964	7.749
1078V	26	yes	Vagina	15664189	2304311	149052	38.30531	1026002	1030627	1031695	D	NP-UA	D/G	118.389	NA	NA
1079C	32	no	Endocervix	3518444	1377695	483959	131.2056	1030045	1030473	1031026	E	P-UA	E	103.751	2800.482	0.037
1079R	32	no	Rectum	3543575	1632939	173076	46.54158	1026667	1030411	1031639	E	P-UA	E	306.39	387.255	0.791
1079V	32	no	Vagina	3446135	795669	264367	71.45122	1029549	1030744	1031682	E	P-UA	E	402.732	NA	NA
1106C	26	yes	Endocervix	7390409	4895814	4867637	1330.369	1031633	1031676	1032115	D	NP-UA	D/G	2362.054	2419.986	0.976
1106R	26	no	Rectum	7878537	4880882	3546987	967.0821	1031652	1031738	1032043	D	NP-UA	D/G	2018.854	1201.252	1.681
1106V	26	yes	Vagina	2673088	1497698	1177170	320.2839	1031568	1031628	1031896	D	NP-UA	D/G	20731.73	1444.637	14.351
1145C	27	no	Endocervix	3245911	605268	342941	93.50092	1029412	1029994	1031149	Ja	P-UA	E	249.978	1119.778	0.223
1145R	27	no	Rectum	6315303	3675922	2839618	772.0722	1030400	1030565	1031347	Ja	P-UA	E	1323.749	507.37	2.609
1145V	27	no	Vagina	4261594	1499177	133698	36.1844	992587	1026948	1031199	Ja	P-UA	E	1771.264	1528.859	1.159
1176C	28	no	Endocervix	1704397	402441	290255	78.88034	1029623	1030448	1031097	Ja	P-UA	E	84.805	1951.951	0.043
1176R	28	no	Rectum	3079956	1095024	304242	82.72563	1030642	1031200	1031893	Ja	P-UA	E	641.326	1003.104	0.639
1176V	28	no	Vagina	3031154	537031	145531	39.68367	1026976	1030621	1031844	G	NP-UA	E	331.935	2495.802	0.133
1182C	25	yes	Endocervix	4248230	2014575	1744982	477.8301	1030301	1030532	1031354	Ja	P-UA	E	2556.347	3294.033	0.776
1182R	25	no	Rectum	2578932	1335796	186473	51.01578	1030165	1031376	1031791	Ja	NP-UA	E	540.352	787.653	0.686
1182V	25	no	Vagina	4025248	1989224	1675784	455.4528	1031233	1031630	1031873	Ja	P-UA	E	351.756	241.598	1.456
1201C	20	yes	Endocervix	5865884	3620178	3577270	979.1686	1031246	1031480	1032125	G	NP-UA	D/G	1029.991	2832.509	0.364
1201R	20	no	Rectum	3035709	1400669	447950	122.1489	1030976	1031407	1031829	G	NP-UA	D/G	413.114	293.635	1.407
1201V	20	no	Vagina	2940741	1331200	1261434	342.8366	1031186	1031311	1031871	G	NP-UA	D/G	1745.979	914.434	1.909
1226C	24	no	Endocervix	2127706	721573	360904	98.11438	1029491	1030279	1030790	E	P-UA	E	2431.069	1919.655	1.266
1226R	24	no	Rectum	4552437	2653408	88464	23.96163	868261	999909	1030018	E	P-UA	E	454.206	NA	NA
1226V	24	no	Vagina	3779246	1735200	518926	141.4167	1028643	1029935	1030975	E	P-UA	E	69.292	60.915	1.138

*Age, participant age when samples were taken.

**Signs & Symptoms, whether or not the anatomic site was showing any signs and/or participant had symptoms suggestive of a sexually transmitted infection.

***Clades, P-UA, prevalent urogenital and anorectal; NP-UA, non prevalent urogenital and anorectal

NOTE: plasmid genotypes D and G are identical

Supplemental Table 2. Coordinates of recent cross-clade recombination events inferred by fastGEAR.

Start	End	Donor Clade	log(BF)	Sample	Fiji *omp* A genotypes
1038741	1038783	LGV	30.8	1077C	Ja
58310	58671	LGV	52	1078V	D
58310	58530	LGV	56.2	1106C	D
746274	746804	LGV	1.4	1106V	D
58310	58686	LGV	51.3	1182R	Ja
732899	733369	LGV	27.2	1182V	Ja
58310	58493	LGV	35.7	1201R	G
1039115	1042519	LGV	1.2	192C	Ja
1039115	1042519	LGV	1.2	192R	Ja
1039115	1042519	LGV	1.2	192V	Ja
1038578	1038783	LGV	97.1	32C	G
1038758	1038783	LGV	19	32R	G
58310	58686	LGV	51.6	32V	G
1038558	1038783	LGV	103.9	519V	E
73592	98346	LGV	39.5	Ct_refDa	
160233	177750	LGV	49.4	Ct_refDa	
195737	220733	LGV	48	Ct_refDa	
233403	237307	LGV	0.7	Ct_refDa	
784059	785445	LGV	0.7	Ct_refDa	
786670	788427	LGV	2	Ct_refDa	
790699	791639	LGV	5.7	Ct_refDa	
942428	954148	LGV	17.9	Ct_refDa	
1026738	1029022	LGV	96.4	Ct_refDa	
1031970	1035741	LGV	3.6	Ct_refDa	
1037401	1042312	LGV	182.5	Ct_refDa	
236999	240719	LGV	0.8	F_70	
255752	256238	LGV	3.7	F_70	
697276	699446	LGV	5.7	F_70	
58310	58748	LGV	48.2	D_43nl	
155911	163894	LGV	5	F_38nL	
58310	58748	LGV	48.2	G	
58310	58748	LGV	47.7	H	
1022818	1026750	LGV	52.6	J	
448581	450045	LGV	0.2	Ja_26s	
473619	474831	LGV	1.6	Ja_26s	
732899	733258	LGV	24	Ja_26s	
58310	58748	LGV	47.9	K	
58310	58748	LGV	48.4	D_UW3	
825804	849082	LGV	102.1	L2C	
1022890	1027524	LGV	140.4	G_9768	
1022890	1027524	LGV	139.9	G_9301	
58310	58748	LGV	48.3	D-EC	
58310	58748	LGV	48.2	D-LC	
1025518	1027561	LGV	85.6	E_150	
1022890	1027524	LGV	140.1	G_11074	
155911	163894	LGV	5.5	E_IU824	
155911	163894	LGV	5.5	E_IU888	
749035	749094	LGV	4.2	L1_115	
58310	58748	LGV	48.2	i_SotonG1	
767990	771123	LGV	16.5	i_SotonG1	
58310	58748	LGV	47.9	)_SotonD5	
155911	163894	LGV	5.6	E_Bour	
58310	58748	LGV	48.3	_CS637_11	
155441	163894	LGV	4.3	:S1025_11	
58310	58748	LGV	48.4	_CS190_96	
166185	170075	LGV	1	E-32931	
255752	256238	LGV	3.8	SQ32	
255752	256238	LGV	3.8	SQ29	
58310	58748	LGV	48.5	SQ20	
58310	58748	LGV	48.5	SQ24	
58406	58748	LGV	21.3	SQ15	
58406	58748	LGV	20.6	SQ19	
58310	58748	LGV	48.3	chxRP4F1	
58310	58748	LGV	48.2	hxRP4D10	
58310	58748	LGV	48.2	chxRP1H1	
57817	58298	NP-UA	57.6	1077C	Ja
747018	747565	NP-UA	124	1077C	Ja
770942	771475	NP-UA	37.9	1077C	Ja
771727	772398	NP-UA	53	1077C	Ja
744966	745205	NP-UA	53.1	1077R	Ja
770942	771475	NP-UA	38	1077R	Ja
771727	772398	NP-UA	53.2	1077R	Ja
770942	771475	NP-UA	38	1077V	Ja
772005	772374	NP-UA	48.6	1077V	Ja
1038531	1038783	NP-UA	109.7	1077V	Ja
719232	732487	NP-UA	39.1	1078R	E
791507	791761	NP-UA	36	1078R	E
768502	772831	NP-UA	108.2	1078V	D
1024613	1034464	NP-UA	26.2	1078V	D
57213	58309	NP-UA	84.1	1079C	E
770942	771475	NP-UA	37.6	1079C	E
772005	772374	NP-UA	48.2	1079C	E
1029706	1033303	NP-UA	97.1	1079C	E

1037321	1042519	NP-UA	176.8	1079C	E
57213	57860	NP-UA	37.3	1079R	E
57861	58748	NP-UA	46.1	1079R	E
744966	745167	NP-UA	43.5	1079R	E
770942	771475	NP-UA	37.8	1079R	E
1029706	1033303	NP-UA	97.7	1079R	E
1037307	1042519	NP-UA	180.1	1079R	E
57213	58598	NP-UA	75.7	1079V	E
58599	58686	NP-UA	7.6	1079V	E
744966	745227	NP-UA	55.4	1079V	E
747018	747362	NP-UA	94.7	1079V	E
770942	771475	NP-UA	37.8	1079V	E
1029706	1033303	NP-UA	97.8	1079V	E
1037307	1042519	NP-UA	180.3	1079V	E
770942	771475	NP-UA	38.4	107C	Ja
770942	771475	NP-UA	38.4	107R	Ja
770942	771475	NP-UA	38.4	107V	Ja
768502	772831	NP-UA	107.8	1106C	D
1024613	1034464	NP-UA	25.9	1106C	D
768502	772831	NP-UA	107.9	1106R	D
1024613	1034464	NP-UA	26	1106R	D
768502	772831	NP-UA	108.2	1106V	D
1024613	1034464	NP-UA	26.2	1106V	D
58083	58298	NP-UA	9.5	1145C	Ja
747285	747565	NP-UA	50.6	1145C	Ja
770942	771475	NP-UA	38.5	1145C	Ja
770942	771475	NP-UA	38.4	1145R	Ja
1038461	1038721	NP-UA	86.5	1145R	Ja
57963	58139	NP-UA	37.8	1145V	Ja
770942	771475	NP-UA	38.5	1145V	Ja
1	20947	NP-UA	10.8	1176C	Ja
57213	58298	NP-UA	73.2	1176C	Ja
111506	120344	NP-UA	1.9	1176C	Ja
770942	771421	NP-UA	40.5	1176C	Ja
1013693	1025311	NP-UA	23.8	1176C	Ja
1026984	1028007	NP-UA	5.1	1176C	Ja
1028570	1029449	NP-UA	0.4	1176C	Ja
1029706	1031027	NP-UA	4.7	1176C	Ja
1037334	1042519	NP-UA	184.7	1176C	Ja
57213	58298	NP-UA	73	1176R	Ja
111814	112517	NP-UA	6	1176R	Ja
744990	745386	NP-UA	40.3	1176R	Ja
770942	771421	NP-UA	40.4	1176R	Ja
1013809	1016084	NP-UA	1.2	1176R	Ja
1018145	1024440	NP-UA	0.3	1176R	Ja
1027274	1029229	NP-UA	7.7	1176R	Ja
1037946	1039376	NP-UA	265.1	1176R	Ja
57020	58493	NP-UA	27.7	1176V	G
770942	772415	NP-UA	87	1182C	Ja
768502	772831	NP-UA	102.5	1182R	Ja
1024613	1034464	NP-UA	24.5	1182R	Ja
25851	26192	NP-UA	21.6	1182V	Ja
57757	58298	NP-UA	89.1	1182V	Ja
161219	161530	NP-UA	16.1	1182V	Ja
321995	322235	NP-UA	8.7	1182V	Ja
348202	348418	NP-UA	27	1182V	Ja
358309	358607	NP-UA	7.2	1182V	Ja
408332	408583	NP-UA	2.6	1182V	Ja
410089	410372	NP-UA	4.7	1182V	Ja
540891	541245	NP-UA	1	1182V	Ja
547014	547456	NP-UA	12.7	1182V	Ja
597494	597625	NP-UA	2.5	1182V	Ja
697276	697564	NP-UA	4.6	1182V	Ja
700241	701180	NP-UA	25.2	1182V	Ja
741420	741711	NP-UA	34.3	1182V	Ja
742134	742367	NP-UA	18.9	1182V	Ja
744189	751321	NP-UA	299.4	1182V	Ja
770531	770914	NP-UA	26.8	1182V	Ja
770915	771726	NP-UA	30.6	1182V	Ja
771727	774463	NP-UA	54.1	1182V	Ja
786960	787168	NP-UA	13	1182V	Ja
791507	791761	NP-UA	36.3	1182V	Ja
796467	796602	NP-UA	5.6	1182V	Ja
895259	895681	NP-UA	0.2	1182V	Ja
966526	966887	NP-UA	8.7	1182V	Ja
1001738	1002840	NP-UA	8.4	1182V	Ja
1026422	1029537	NP-UA	9.4	1182V	Ja
1038181	1039376	NP-UA	240.6	1182V	Ja
57680	58482	NP-UA	31.2	1201C	G
57680	58349	NP-UA	17.2	1201V	G
1030034	1034464	NP-UA	0.2	1226C	E
1030034	1034464	NP-UA	0.2	1226R	E
1030034	1034464	NP-UA	0.2	1226V	E
57213	57962	NP-UA	51.2	192C	Ja
57963	58748	NP-UA	34.4	192C	Ja
91094	92117	NP-UA	5.1	192C	Ja

313206	348342	NP-UA	108	192C	Ja
361477	366640	NP-UA	10.9	192C	Ja
747047	747271	NP-UA	77	192C	Ja
771159	772415	NP-UA	80.9	192C	Ja
57213	57962	NP-UA	51.2	192R	Ja
57963	58748	NP-UA	34.4	192R	Ja
91094	92117	NP-UA	5.1	192R	Ja
313206	348342	NP-UA	107.8	192R	Ja
361477	366640	NP-UA	10.9	192R	Ja
771159	772415	NP-UA	80.9	192R	Ja
57213	57962	NP-UA	51.2	192V	Ja
57963	58748	NP-UA	34.4	192V	Ja
91094	92117	NP-UA	5.1	192V	Ja
313206	348342	NP-UA	107.7	192V	Ja
361477	366640	NP-UA	10.9	192V	Ja
771159	772415	NP-UA	80.8	192V	Ja
125050	140910	NP-UA	6.6	227C	G
126602	140910	NP-UA	7.8	227R	G
126602	140910	NP-UA	7.8	227V	G
1	20947	NP-UA	11.4	26C	E
57213	57962	NP-UA	50.4	26C	E
57963	58748	NP-UA	33.8	26C	E
1037832	1038008	NP-UA	33.3	26C	E
1039130	1042519	NP-UA	4.8	26C	E
1	20947	NP-UA	11.4	26R	E
57213	57962	NP-UA	50.4	26R	E
57963	58748	NP-UA	33.8	26R	E
1037832	1038008	NP-UA	33.3	26R	E
1039130	1042519	NP-UA	4.8	26R	E
1	20947	NP-UA	11.4	26V	E
57213	57962	NP-UA	50.4	26V	E
57963	58748	NP-UA	33.8	26V	E
1037832	1038008	NP-UA	33.3	26V	E
1039130	1042519	NP-UA	4.8	26V	E
770942	772415	NP-UA	85.9	32C	G
770942	772415	NP-UA	85.9	32R	G
125578	131240	NP-UA	1.6	362C	F
125578	131240	NP-UA	1.6	362R	F
125578	131240	NP-UA	1.6	362V	F
774153	774250	NP-UA	14.8	519C	E
1030034	1034464	NP-UA	0.3	519C	E
1038461	1038757	NP-UA	99.9	519C	E
1030034	1034464	NP-UA	0.2	519R	E
1030034	1034464	NP-UA	0.3	519V	E
770942	772415	NP-UA	86.3	564C	Ja
1	7529	NP-UA	0.3	564R	E
1026818	1042519	NP-UA	226.3	564R	E
1	7529	NP-UA	0.3	564V	E
1026818	1042519	NP-UA	226	564V	E
770942	772415	NP-UA	86.3	68C	Ja
1038398	1038445	NP-UA	45.7	68C	Ja
57757	57962	NP-UA	51.9	68R	Ja
770942	772415	NP-UA	86.3	68R	Ja
770942	772415	NP-UA	86.3	68V	Ja
57680	58530	NP-UA	58.7	72C	G
770942	772426	NP-UA	86.2	908C	Ja
57757	57962	NP-UA	52	908R	Ja
770942	772426	NP-UA	86.2	908R	Ja
770942	772426	NP-UA	86.2	908V	Ja
1	20947	NP-UA	13.3	927C	E
111506	120245	NP-UA	0.2	927C	E
1013693	1025125	NP-UA	20.5	927C	E
1026738	1042519	NP-UA	237.4	927C	E
1	20947	NP-UA	13.3	927R	E
111506	120245	NP-UA	0.2	927R	E
1013693	1025125	NP-UA	20.5	927R	E
1026738	1042519	NP-UA	203.7	927R	E
1	20947	NP-UA	13.3	927V	E
111615	120245	NP-UA	0.4	927V	E
1013693	1025125	NP-UA	20.5	927V	E
1026738	1042519	NP-UA	237.4	927V	E
719232	732487	NP-UA	43.2	951C	E
57757	58309	NP-UA	88.8	951R	E
719232	732612	NP-UA	46.2	951R	E
733061	735089	NP-UA	14.8	951R	E
748462	749152	NP-UA	1.9	951R	E
1038268	1038517	NP-UA	111	951R	E
719232	732487	NP-UA	43.2	951V	E
56796	57816	NP-UA	4.9	968V	G
56796	58748	NP-UA	36.1	98C	G
56796	57816	NP-UA	4.9	98R	G
56796	57816	NP-UA	4.9	98V	G
97784	99849	NP-UA	9.2	Ba	
363860	364929	NP-UA	1	Ba	
408332	409929	NP-UA	10.2	Ba	
961126	963202	NP-UA	0.1	Ba	

1030673	1035390	NP-UA	80.7	Ba
774194	775753	NP-UA	42.2	Ct_RefC
1030673	1035390	NP-UA	84.5	Ct_RefC
57213	57962	NP-UA	50.8	Ct_refDa
57963	58748	NP-UA	25.3	Ct_refDa
412356	448486	NP-UA	66	Ct_refDa
448487	454822	NP-UA	21.7	Ct_refDa
473918	475218	NP-UA	1.2	Ct_refDa
581883	582242	NP-UA	3.4	Ct_refDa
632048	632262	NP-UA	0.6	Ct_refDa
642496	674846	NP-UA	23.3	Ct_refDa
831688	834931	NP-UA	2.9	Ct_refDa
1030689	1031469	NP-UA	40.3	Ct_refDa
57213	57962	NP-UA	50.7	D_2923
57963	58748	NP-UA	25.3	D_2923
61351	62117	NP-UA	1.1	D_2923
96678	120155	NP-UA	5.4	D_2923
195090	204801	NP-UA	13	D_2923
412356	442321	NP-UA	49.6	D_2923
532095	534130	NP-UA	3.7	D_2923
1025905	1034464	NP-UA	27.9	D_2923
69317	88069	NP-UA	19.4	F_70
177751	181271	NP-UA	14.5	F_70
183935	192026	NP-UA	6.4	F_70
251546	255658	NP-UA	12.9	F_70
498976	502720	NP-UA	4.4	F_70
511159	513562	NP-UA	9.4	F_70
518973	520574	NP-UA	4.1	F_70
538742	539300	NP-UA	5.1	F_70
628901	629159	NP-UA	2.3	F_70
638304	641356	NP-UA	7.4	F_70
646039	655313	NP-UA	15.7	F_70
719592	734283	NP-UA	54.1	F_70
1012132	1020426	NP-UA	8.1	F_70
1020427	1025348	NP-UA	1.9	F_70
1026756	1028233	NP-UA	12.6	F_70
78692	102624	NP-UA	15.1	D_43nl
140811	160823	NP-UA	2.1	D_43nl
196578	238427	NP-UA	12.3	D_43nl
768360	772831	NP-UA	93.9	D_43nl
57213	57962	NP-UA	50.4	D_84s
57963	58802	NP-UA	24	D_84s
100669	120155	NP-UA	2.7	D_84s
195030	199995	NP-UA	3.3	D_84s
199996	204801	NP-UA	3.8	D_84s
412356	442321	NP-UA	52.1	D_84s
532095	534130	NP-UA	3.6	D_84s
650218	659457	NP-UA	2.4	D_84s
700800	701180	NP-UA	37.1	D_84s
732899	733258	NP-UA	24.1	D_84s
772663	774999	NP-UA	39.5	D_84s
1025905	1034464	NP-UA	27.5	D_84s
57213	57962	NP-UA	50.8	E_5s
57963	58820	NP-UA	23.9	E_5s
96678	120155	NP-UA	7.8	E_5s
195030	199463	NP-UA	3.9	E_5s
199464	204801	NP-UA	4.8	E_5s
412356	442321	NP-UA	48.9	E_5s
532095	534130	NP-UA	3.7	E_5s
700800	701180	NP-UA	37.3	E_5s
732899	733258	NP-UA	24.3	E_5s
1026009	1034464	NP-UA	28.6	E_5s
57213	57962	NP-UA	50.7	-Cal.fasta
57963	58802	NP-UA	24.2	-Cal.fasta
96678	120155	NP-UA	5.5	-Cal.fasta
195090	204801	NP-UA	13.1	-Cal.fasta
412356	442321	NP-UA	50.8	-Cal.fasta
532095	534130	NP-UA	3.7	-Cal.fasta
700800	701180	NP-UA	37.2	-Cal.fasta
732899	733258	NP-UA	24.2	-Cal.fasta
1025905	1034464	NP-UA	28	-Cal.fasta
57213	57962	NP-UA	50.1	F_1
57963	58820	NP-UA	23.4	F_1
96678	100668	NP-UA	0.2	F_1
100669	120155	NP-UA	1.6	F_1
157904	163060	NP-UA	20.4	F_1
195090	204801	NP-UA	12.3	F_1
412356	442321	NP-UA	54.1	F_1
532095	534130	NP-UA	3.4	F_1
700800	701180	NP-UA	36.8	F_1
732899	733258	NP-UA	23.9	F_1
1025905	1034464	NP-UA	26.9	F_1
98026	124913	NP-UA	5.3	F_38nL
700800	701180	NP-UA	37	F_38nL
732899	733258	NP-UA	24	F_38nL
1008427	1013240	NP-UA	1.3	F_38nL

1025905	1034464	NP-UA	27.3	F_38nL
278849	279877	NP-UA	2.4	G
1024613	1034464	NP-UA	25.5	G
181024	196322	NP-UA	5.2	H
768502	773345	NP-UA	102.2	H
920081	920415	NP-UA	3.4	H
1002482	1034464	NP-UA	6.7	H
184815	189032	NP-UA	9.9	H_18s
768502	773345	NP-UA	101.3	H_18s
1024613	1028007	NP-UA	61.7	H_18s
743302	747173	NP-UA	3	J
1026751	1027648	NP-UA	9.9	J
739057	751267	NP-UA	163.1	Ja
1026738	1028233	NP-UA	0.6	Ja
1028234	1042519	NP-UA	24.1	Ja
1	11655	NP-UA	0.6	Ja_26s
57169	57962	NP-UA	52.2	Ja_26s
57963	58820	NP-UA	23.6	Ja_26s
96678	120155	NP-UA	9.4	Ja_26s
195090	198531	NP-UA	1.9	Ja_26s
198532	204801	NP-UA	4.1	Ja_26s
335988	336927	NP-UA	3.5	Ja_26s
412356	439135	NP-UA	44.9	Ja_26s
530832	533191	NP-UA	0.6	Ja_26s
700800	701180	NP-UA	37	Ja_26s
1025905	1042519	NP-UA	162	Ja_26s
1	11655	NP-UA	0.7	Ja_47nL
54059	55173	NP-UA	4.1	Ja_47nL
55460	55844	NP-UA	8.1	Ja_47nL
57253	58309	NP-UA	79.5	Ja_47nL
59266	95314	NP-UA	63.6	Ja_47nL
96678	100874	NP-UA	0.9	Ja_47nL
100875	120155	NP-UA	0.9	Ja_47nL
195090	204976	NP-UA	11.1	Ja_47nL
412356	442321	NP-UA	55.2	Ja_47nL
532095	534130	NP-UA	3.5	Ja_47nL
700800	701180	NP-UA	37	Ja_47nL
732899	733258	NP-UA	24.1	Ja_47nL
772005	772426	NP-UA	45.9	Ja_47nL
1025905	1042519	NP-UA	136.1	Ja_47nL
767854	773506	NP-UA	89.1	K
1024613	1034464	NP-UA	25.1	K
768360	772831	NP-UA	94.6	D_UW3
1024613	1034464	NP-UA	25.9	D_UW3
97053	99849	NP-UA	9	A_HAR13
408332	409912	NP-UA	8	A_HAR13
961126	968208	NP-UA	9.6	A_HAR13
97784	99849	NP-UA	5.3	B_Jali20
292877	303462	NP-UA	0.6	B_Jali20
362659	365145	NP-UA	0.3	B_Jali20
961126	967948	NP-UA	9.2	B_Jali20
786479	791040	NP-UA	60.8	L2C
798093	823642	NP-UA	60.4	L2C
823643	825803	NP-UA	13.3	L2C
849083	856518	NP-UA	3.4	L2C
774194	775877	NP-UA	46.6	A_2497_1
1	49237	NP-UA	10.3	G_9768
181401	183067	NP-UA	5.6	G_9768
191628	230980	NP-UA	3.2	G_9768
259420	305251	NP-UA	12.4	G_9768
56796	57816	NP-UA	4.9	G_11222
125578	131240	NP-UA	1.6	G_11222
1	11655	NP-UA	1.2	E_11023
700800	701180	NP-UA	37.5	E_11023
732899	733258	NP-UA	24.4	E_11023
1025905	1042519	NP-UA	166.5	E_11023
1	49237	NP-UA	9.2	G_9301
181401	183067	NP-UA	5.5	G_9301
191628	230980	NP-UA	3	G_9301
259420	310101	NP-UA	8.8	G_9301
768360	772831	NP-UA	94.2	D-EC
1024613	1034464	NP-UA	25.7	D-EC
768360	772831	NP-UA	94.1	D-LC
1024613	1034464	NP-UA	25.6	D-LC
774194	775873	NP-UA	46.4	A_2497
1	22361	NP-UA	18.5	E_150
57213	57962	NP-UA	51.4	E_150
57963	58820	NP-UA	24.3	E_150
111506	120245	NP-UA	0.4	E_150
700800	701180	NP-UA	37.7	E_150
732899	733258	NP-UA	24.5	E_150
1013693	1025517	NP-UA	20.9	E_150
1027562	1042519	NP-UA	129.1	E_150
1	49237	NP-UA	7.8	G_11074
181401	183067	NP-UA	5.5	G_11074
191628	230980	NP-UA	3.7	G_11074

259420	310101	NP-UA	7.5	G_11074
1	20947	NP-UA	14.3	Sweden2
57213	57962	NP-UA	51.7	Sweden2
57963	58832	NP-UA	24.3	Sweden2
111506	120245	NP-UA	1.5	Sweden2
157904	182713	NP-UA	57.8	Sweden2
700800	701180	NP-UA	37.9	Sweden2
732899	733258	NP-UA	24.7	Sweden2
1013693	1025311	NP-UA	22	Sweden2
1026738	1028233	NP-UA	13.6	Sweden2
1028234	1042519	NP-UA	133.5	Sweden2
1	11655	NP-UA	1.3	F_SW4
57213	57962	NP-UA	51.2	F_SW4
57963	58832	NP-UA	24	F_SW4
96678	120155	NP-UA	6.1	F_SW4
195090	204801	NP-UA	13.7	F_SW4
412356	442321	NP-UA	51.3	F_SW4
532095	534130	NP-UA	3.9	F_SW4
700800	701180	NP-UA	37.6	F_SW4
709781	724809	NP-UA	42.3	F_SW4
732899	733258	NP-UA	24.5	F_SW4
1025905	1042519	NP-UA	166.8	F_SW4
1	20947	NP-UA	13	E_SW3
57213	57962	NP-UA	51.4	E_SW3
57963	58832	NP-UA	24.1	E_SW3
111506	120245	NP-UA	1.3	E_SW3
700800	701180	NP-UA	37.7	E_SW3
732899	733258	NP-UA	24.6	E_SW3
1013693	1025311	NP-UA	21.7	E_SW3
1026738	1028233	NP-UA	13.4	E_SW3
1028234	1042519	NP-UA	132.6	E_SW3
1	11655	NP-UA	1	F_SW5
57213	57962	NP-UA	50.8	F_SW5
57963	58832	NP-UA	23.7	F_SW5
96678	120155	NP-UA	5.6	F_SW5
195090	204801	NP-UA	13.2	F_SW5
412356	442321	NP-UA	50	F_SW5
532095	534130	NP-UA	3.7	F_SW5
700800	701180	NP-UA	37.3	F_SW5
732899	733258	NP-UA	24.3	F_SW5
1025905	1042519	NP-UA	164.7	F_SW5
1	11655	NP-UA	1	E_IU824
700800	701180	NP-UA	37.4	E_IU824
732899	733258	NP-UA	24.3	E_IU824
1025905	1042519	NP-UA	165.4	E_IU824
1	11655	NP-UA	1	E_IU888
700800	701180	NP-UA	37.4	E_IU888
732899	733258	NP-UA	24.3	E_IU888
1025905	1042519	NP-UA	165.5	E_IU888
774194	775877	NP-UA	46.6	A_5291
771124	772831	NP-UA	86.5	i_SotonG1
1024613	1034464	NP-UA	25.5	i_SotonG1
403117	408696	NP-UA	4	)_SotonD5
768360	772831	NP-UA	93	)_SotonD5
1024613	1034464	NP-UA	25	)_SotonD5
774194	775885	NP-UA	46.3	A_7249
55268	56443	NP-UA	5.8	K_SotonK1
758824	772831	NP-UA	63.7	K_SotonK1
1024613	1034464	NP-UA	24.4	K_SotonK1
774194	775870	NP-UA	46.5	A_363
1	20947	NP-UA	13.8	E_SotonE4
111506	120245	NP-UA	1.3	E_SotonE4
700800	701180	NP-UA	37.7	E_SotonE4
732899	733258	NP-UA	24.6	E_SotonE4
1013693	1025311	NP-UA	21.6	E_SotonE4
1026738	1028233	NP-UA	13.4	E_SotonE4
1028234	1042519	NP-UA	132.3	E_SotonE4
1	11655	NP-UA	1.1	E_Bour
700800	701180	NP-UA	37.4	E_Bour
732899	733258	NP-UA	24.4	E_Bour
1025905	1042519	NP-UA	165.9	E_Bour
58176	58349	NP-UA	42.8	L1_224
58191	58349	NP-UA	39.3	L3_404
1	11655	NP-UA	1.1	F_11-96
57154	58349	NP-UA	82.5	F_11-96
96678	120155	NP-UA	5.9	F_11-96
195090	204801	NP-UA	13.5	F_11-96
218058	226446	NP-UA	6.8	F_11-96
412356	442321	NP-UA	50.7	F_11-96
532095	534130	NP-UA	3.8	F_11-96
700800	701180	NP-UA	37.5	F_11-96
732899	733258	NP-UA	24.4	F_11-96
1025905	1042519	NP-UA	165.8	F_11-96
1	11655	NP-UA	1.1	F_1-93
57154	58349	NP-UA	82.4	F_1-93
96678	120155	NP-UA	5.8	F_1-93

195090	204801	NP-UA	13.4	F_1-93
412356	442321	NP-UA	50.6	F_1-93
532095	534130	NP-UA	5.4	F_1-93
700800	701180	NP-UA	37.4	F_1-93
732899	733258	NP-UA	24.4	F_1-93
1025905	1042519	NP-UA	167.5	F_1-93
1	11655	NP-UA	0.5	F_6-94
57154	58349	NP-UA	81.2	F_6-94
96678	120155	NP-UA	4.8	F_6-94
157904	163060	NP-UA	19	F_6-94
195030	195868	NP-UA	2.4	F_6-94
196837	204801	NP-UA	6.1	F_6-94
412356	442321	NP-UA	51.1	F_6-94
532095	534130	NP-UA	3.5	F_6-94
700800	701180	NP-UA	36.9	F_6-94
732899	733258	NP-UA	24	F_6-94
995741	1015714	NP-UA	44.1	F_6-94
1018057	1042519	NP-UA	172.7	F_6-94
1	11655	NP-UA	1	D_13-96
57154	58349	NP-UA	82.1	D_13-96
61311	62117	NP-UA	0.8	D_13-96
96678	120155	NP-UA	5.5	D_13-96
195090	204801	NP-UA	13.2	D_13-96
412356	442321	NP-UA	49.9	D_13-96
532095	534130	NP-UA	3.7	D_13-96
700800	701180	NP-UA	37.3	D_13-96
732899	733258	NP-UA	24.3	D_13-96
1025905	1042519	NP-UA	164.5	D_13-96
58176	58298	NP-UA	28.9	C_TW3
293124	302319	NP-UA	1.1	C_TW3
774194	775870	NP-UA	46.5	C_TW3
55268	56443	NP-UA	6.1	CS637_11
251912	258094	NP-UA	7.3	CS637_11
768309	772831	NP-UA	102.2	CS637_11
1024613	1034464	NP-UA	25.8	CS637_11
58191	58349	NP-UA	39.2	CS19_08
58191	58349	NP-UA	39.1	CS784_08
1	11655	NP-UA	1	CS1025_11
700800	701180	NP-UA	37.3	CS1025_11
732899	733258	NP-UA	24.3	CS1025_11
1025905	1042519	NP-UA	164.9	CS1025_11
1	11655	NP-UA	1	CS847_08
57169	58349	NP-UA	82.7	CS847_08
412356	442321	NP-UA	50	CS847_08
532095	534130	NP-UA	5.3	CS847_08
644228	654822	NP-UA	20.3	CS847_08
700800	701180	NP-UA	37.3	CS847_08
732899	733258	NP-UA	24.3	CS847_08
1025905	1042519	NP-UA	164.8	CS847_08
768066	773118	NP-UA	88.4	CS190_96
1024613	1034464	NP-UA	25.9	CS190_96
1	20947	NP-UA	11.9	E-103
57169	58349	NP-UA	83.4	E-103
111506	120245	NP-UA	0.3	E-103
700800	701180	NP-UA	37.6	E-103
732899	733258	NP-UA	24.5	E-103
1013693	1026008	NP-UA	21.8	E-103
1026009	1042519	NP-UA	168.4	E-103
1	20947	NP-UA	12.5	E-160
57169	58349	NP-UA	83.2	E-160
111506	120245	NP-UA	1.1	E-160
700800	701180	NP-UA	37.5	E-160
732899	733258	NP-UA	24.4	E-160
1013693	1025311	NP-UA	21.3	E-160
1026738	1028233	NP-UA	13.3	E-160
1028234	1042519	NP-UA	131.3	E-160
1	20947	NP-UA	12.9	E-547
57169	58349	NP-UA	83.6	E-547
111506	120245	NP-UA	1.3	E-547
700800	701180	NP-UA	37.7	E-547
732899	733306	NP-UA	23	E-547
1013693	1025311	NP-UA	21.6	E-547
1026738	1028233	NP-UA	13.4	E-547
1028234	1042519	NP-UA	132.3	E-547
1	11655	NP-UA	1.2	E-8873
57154	58349	NP-UA	82.6	E-8873
96678	120116	NP-UA	6.7	E-8873
195090	204801	NP-UA	13.5	E-8873
700800	701180	NP-UA	37.5	E-8873
732899	733306	NP-UA	22.8	E-8873
1025905	1042519	NP-UA	166.2	E-8873
1	20947	NP-UA	12.9	E-32931
57154	58349	NP-UA	83	E-32931
111506	120245	NP-UA	0.4	E-32931
164296	166184	NP-UA	4.7	E-32931
184768	194529	NP-UA	4.9	E-32931

700800	701180	NP-UA	37.7	E-32931
732899	733306	NP-UA	23	E-32931
1013693	1025311	NP-UA	21.6	E-32931
1026738	1028233	NP-UA	13.4	E-32931
1028234	1042519	NP-UA	132.2	E-32931
1	20947	NP-UA	12.3	E_DK20
121260	123640	NP-UA	8.5	E_DK20
179423	179834	NP-UA	0.4	E_DK20
247629	258599	NP-UA	2.3	E_DK20
700800	701180	NP-UA	37.4	E_DK20
732899	733306	NP-UA	22.8	E_DK20
1015715	1025311	NP-UA	11.4	E_DK20
1026716	1028233	NP-UA	15.1	E_DK20
1028234	1042519	NP-UA	130.6	E_DK20
1	11655	NP-UA	0.9	F-6068
57154	58349	NP-UA	81.9	F-6068
96678	120116	NP-UA	6.9	F-6068
157904	163060	NP-UA	19.5	F-6068
195090	204801	NP-UA	13	F-6068
412356	442321	NP-UA	54	F-6068
532095	534130	NP-UA	3.6	F-6068
700800	701180	NP-UA	37.2	F-6068
732899	733306	NP-UA	22.6	F-6068
1025876	1042519	NP-UA	163.6	F-6068
1	11655	NP-UA	0.4	SQ32
57143	58349	NP-UA	80.6	SQ32
96678	120155	NP-UA	4.6	SQ32
179423	179834	NP-UA	0.2	SQ32
247390	255658	NP-UA	13.2	SQ32
700800	701180	NP-UA	36.8	SQ32
732899	733306	NP-UA	22.3	SQ32
1026422	1042519	NP-UA	153.4	SQ32
1	11655	NP-UA	0.4	SQ29
57143	58349	NP-UA	80.6	SQ29
96678	120155	NP-UA	4.6	SQ29
179423	179834	NP-UA	0.2	SQ29
247390	255658	NP-UA	13.1	SQ29
700800	701180	NP-UA	36.8	SQ29
732899	733306	NP-UA	22.3	SQ29
1026422	1042519	NP-UA	153.3	SQ29
55290	56443	NP-UA	6.3	SQ20
184369	189727	NP-UA	11.5	SQ20
768502	773345	NP-UA	101.9	SQ20
1024613	1034464	NP-UA	26.1	SQ20
55290	56443	NP-UA	6.3	SQ24
184369	189727	NP-UA	11.5	SQ24
768502	773345	NP-UA	101.9	SQ24
1024613	1034464	NP-UA	26.1	SQ24
742947	747173	NP-UA	0.5	SQ01
1	11655	NP-UA	1.1	SQ25
57143	58349	NP-UA	82.2	SQ25
96678	120155	NP-UA	5.9	SQ25
160877	161749	NP-UA	13.4	SQ25
700800	701180	NP-UA	37.4	SQ25
732899	733306	NP-UA	22.8	SQ25
1025905	1042519	NP-UA	166	SQ25
1	52385	NP-UA	0	SQ15
55460	56443	NP-UA	8.4	SQ15
57118	58405	NP-UA	13.7	SQ15
99774	133032	NP-UA	0.8	SQ15
161146	164916	NP-UA	7.5	SQ15
219432	219821	NP-UA	6.2	SQ15
553069	558653	NP-UA	2.7	SQ15
55460	56443	NP-UA	7.8	SQ19
57118	58405	NP-UA	13	SQ19
161146	164916	NP-UA	6.6	SQ19
219432	219821	NP-UA	5.9	SQ19
553048	558653	NP-UA	1.9	SQ19
23957	40177	NP-UA	8.4	QH111L
41442	49237	NP-UA	7	QH111L
57487	57816	NP-UA	18.2	QH111L
58176	58275	NP-UA	30.2	QH111L
66045	67698	NP-UA	3.2	QH111L
72139	76886	NP-UA	7.3	QH111L
179656	191627	NP-UA	22.6	QH111L
240720	243829	NP-UA	20.2	QH111L
246875	257350	NP-UA	63.3	QH111L
258952	271304	NP-UA	4.7	QH111L
277645	280981	NP-UA	7.8	QH111L
350094	385284	NP-UA	0.1	QH111L
429029	438039	NP-UA	9.1	QH111L
444859	454372	NP-UA	39.2	QH111L
481583	482294	NP-UA	5.7	QH111L
526117	528445	NP-UA	5.3	QH111L
635952	645097	NP-UA	2	QH111L
757409	771312	NP-UA	7.2	QH111L

783800	786478	NP-UA	0.7	QH111L	
981635	1019101	NP-UA	23.3	QH111L	
768090	772831	NP-UA	91.3	chxRP4F1	
1024613	1034464	NP-UA	25.7	chxRP4F1	
768090	772831	NP-UA	91.1	hxRP4D10	
1024613	1034464	NP-UA	25.6	hxRP4D10	
768090	772831	NP-UA	91.1	chxRP1H1	
1024613	1034464	NP-UA	25.6	chxRP1H1	
56342	57312	P-UA	48	1078V	D
183216	189371	P-UA	7.6	1078V	D
444988	448392	P-UA	17.6	1078V	D
56138	57312	P-UA	43.8	1106C	D
183216	189371	P-UA	7.5	1106C	D
444988	448392	P-UA	17.5	1106C	D
56138	57312	P-UA	43.8	1106R	D
183216	189371	P-UA	7.5	1106R	D
444988	448392	P-UA	17.5	1106R	D
56138	57312	P-UA	43.9	1106V	D
183216	189371	P-UA	7.6	1106V	D
444988	448392	P-UA	17.6	1106V	D
54417	57019	P-UA	107.7	1176V	G
181826	190352	P-UA	0.7	1176V	G
1026793	1027561	P-UA	7.2	1176V	G
54294	55173	P-UA	32.8	1182R	Ja
55624	57312	P-UA	39.2	1182R	Ja
183216	189371	P-UA	7.6	1182R	Ja
444988	448392	P-UA	15.5	1182R	Ja
1026984	1027561	P-UA	12.2	227C	G
1026984	1027561	P-UA	12.2	227R	G
1026984	1027561	P-UA	12.2	227V	G
54294	57019	P-UA	58.9	32V	G
183068	189274	P-UA	6.2	32V	G
760234	761754	P-UA	6.4	32V	G
790344	797331	P-UA	87.9	32V	G
799963	819898	P-UA	7.8	32V	G
825301	826002	P-UA	2.5	32V	G
827105	871020	P-UA	99	32V	G
133033	134933	P-UA	6.3	362C	F
773993	775893	P-UA	135.9	362C	F
783109	883485	P-UA	253.5	362C	F
923775	941098	P-UA	39.8	362C	F
942211	962496	P-UA	29.7	362C	F
981310	985396	P-UA	10.6	362C	F
990878	1007315	P-UA	59.1	362C	F
133033	134933	P-UA	6.3	362R	F
773993	775893	P-UA	135.9	362R	F
783109	883485	P-UA	253.5	362R	F
923775	941098	P-UA	39.8	362R	F
942211	962496	P-UA	29.7	362R	F
981310	985396	P-UA	10.6	362R	F
990878	1007315	P-UA	59.1	362R	F
133033	134933	P-UA	6.3	362V	F
773993	775893	P-UA	135.9	362V	F
783109	883485	P-UA	253.5	362V	F
923775	941098	P-UA	39.8	362V	F
942319	962496	P-UA	29.9	362V	F
981310	985396	P-UA	10.6	362V	F
990878	1007315	P-UA	59.1	362V	F
760234	761621	P-UA	5	72C	G
790344	797331	P-UA	90.8	72C	G
799963	819898	P-UA	11.2	72C	G
825301	826002	P-UA	2.6	72C	G
827105	870833	P-UA	104.7	72C	G
760234	761621	P-UA	4.9	72R	G
790344	797331	P-UA	90.3	72R	G
799963	819898	P-UA	10.9	72R	G
825301	826002	P-UA	2.5	72R	G
827105	870700	P-UA	104	72R	G
760234	761621	P-UA	4.9	72V	G
790344	797331	P-UA	90.3	72V	G
799963	819898	P-UA	10.9	72V	G
825301	826002	P-UA	2.5	72V	G
827105	870700	P-UA	104	72V	G
159603	163454	P-UA	8.1	Ba	
160316	163454	P-UA	15.5	Ct_RefC	
774014	774193	P-UA	9.2	Ct_RefC	
4437	7334	P-UA	8.2	Ct_refL3	
774014	774116	P-UA	22.2	Ct_refL3	
775665	775885	P-UA	61.6	Ct_refL3	
56607	57019	P-UA	37.2	D_43nl	
183216	189371	P-UA	8.8	D_43nl	
56607	57019	P-UA	37.2	G	
183216	189371	P-UA	7.3	G	
56607	57019	P-UA	36.8	H	
444988	448392	P-UA	14.7	H	
56607	57312	P-UA	47.1	H_18s	

556834	558381	P-UA	5.3	J	
56607	57019	P-UA	36.9	K	
183216	189371	P-UA	10.2	K	
56607	57019	P-UA	37.3	D_UW3	
183216	189371	P-UA	7.5	D_UW3	
158321	164185	P-UA	3.4	A_HAR13	
56607	56749	P-UA	62.3	L2_434BU	
160618	163454	P-UA	10.8	B_Jali20	
56607	56749	P-UA	62.3	L2C	
56607	57312	P-UA	44	G_9768	
158498	161953	P-UA	5.9	G_9768	
183068	186806	P-UA	48.8	G_9768	
230981	259419	P-UA	12.5	G_9768	
133033	134933	P-UA	6.3	G_11222	
56607	57312	P-UA	43.8	G_9301	
158498	161953	P-UA	5.8	G_9301	
183068	186806	P-UA	48.5	G_9301	
230981	259419	P-UA	12.2	G_9301	
56607	57019	P-UA	37.2	D-EC	
183216	189371	P-UA	7.4	D-EC	
56607	57019	P-UA	37.2	D-LC	
183216	189371	P-UA	7.3	D-LC	
56607	57312	P-UA	43.9	G_11074	
158498	161953	P-UA	5.8	G_11074	
183068	186806	P-UA	48.6	G_11074	
230981	259419	P-UA	12.3	G_11074	
56607	56749	P-UA	62.3	!_SF25667	
56607	56749	P-UA	62.9	L1_440	
56607	57019	P-UA	37.1	i_SotonG1	
183216	189371	P-UA	7.3	i_SotonG1	
56607	57019	P-UA	36.9	)_SotonD5	
183216	189371	P-UA	7	)_SotonD5	
56444	57312	P-UA	38.8	K_SotonK1	
183216	189371	P-UA	6.8	K_SotonK1	
759982	768192	P-UA	32.3	J_31-98	
160349	163454	P-UA	13.8	C_TW3	
56444	57312	P-UA	39.6	CS637_11	
183216	189371	P-UA	7.4	CS637_11	
56607	57019	P-UA	37.3	CS190_96	
65929	72014	P-UA	31.1	CS190_96	
183216	189371	P-UA	7.5	CS190_96	
56444	57312	P-UA	39.8	SQ20	
56444	57312	P-UA	39.8	SQ24	
886192	887789	P-UA	2.6	SQ09	
556834	558441	P-UA	5	SQ01	
556834	558441	P-UA	4.9	SQ02	
556834	558441	P-UA	4.9	SQ05	
56444	57117	P-UA	31.5	SQ15	
56444	57117	P-UA	30.5	SQ19	
56444	57486	P-UA	28.4	QH111L	
155631	163454	P-UA	7	QH111L	
56607	57019	P-UA	37.2	chxRP4F1	
183216	189371	P-UA	7.4	chxRP4F1	
56607	57019	P-UA	37.2	hxRP4D10	
183216	189371	P-UA	7.3	hxRP4D10	
56607	57019	P-UA	37.2	chxRP1H1	
183216	189371	P-UA	7.3	chxRP1H1	
547176	547381	Trachoma	12.9	1079R	E
25451	25733	Trachoma	10.6	1182V	Ja
55340	55372	Trachoma	28.1	927V	E
782882	784058	Trachoma	5.7	Ct_refDa	
788428	788439	Trachoma	1.7	Ct_refDa	
111229	111548	Trachoma	1.3	F_70	
54204	55451	Trachoma	65.7	D_43nl	
54204	55451	Trachoma	65.7	G	
690319	690410	Trachoma	18.8	G	
55174	55451	Trachoma	38.4	H	
55174	55459	Trachoma	30.6	Ja_47nL	
782093	783896	Trachoma	1.6	Ja_47nL	
54204	55451	Trachoma	65.3	K	
54204	55451	Trachoma	66	D_UW3	
54204	55451	Trachoma	65.8	D-EC	
54204	55451	Trachoma	65.7	D-LC	
54204	55451	Trachoma	65.6	i_SotonG1	
54204	55451	Trachoma	65.1	)_SotonD5	
869427	893862	Trachoma	77	J_6276	
54204	55451	Trachoma	66	CS190_96	
111229	111548	Trachoma	1.5	E_DK20	
113026	114074	Trachoma	1	E_DK20	
52386	55459	Trachoma	25.7	SQ15	
52386	55459	Trachoma	24.3	SQ19	
55309	55459	Trachoma	43.9	QH111L	
54204	55451	Trachoma	65.8	chxRP4F1	
54204	55451	Trachoma	65.7	hxRP4D10	
54204	55451	Trachoma	65.7	chxRP1H1	
757409	759981	Unknown	0.1	1078V	D

775878	782794	Unknown	0.1	1079C	E
368170	375328	Unknown	0.2	1078R	E
228878	230623	Unknown	0.2	F_70	
368170	375328	Unknown	0.3	1079R	E
368170	375328	Unknown	0.3	1079V	E
368170	375328	Unknown	0.4	1079C	E
546715	547318	Unknown	0.5	Ja_26s	
228878	230560	Unknown	0.6	SQ32	
228878	230560	Unknown	0.6	SQ29	
234361	234931	Unknown	0.9	E_DK20	
564829	572854	Unknown	1.1	1077V	Ja
997249	997589	Unknown	1.1	1077V	Ja
997249	997589	Unknown	1.1	1078R	E
771476	772004	Unknown	1.1	1079C	E
564829	572854	Unknown	1.2	1077C	Ja
564829	572854	Unknown	1.2	1077R	Ja
771476	772004	Unknown	1.2	1077V	Ja
564829	572854	Unknown	1.2	1078R	E
564829	572854	Unknown	1.3	1079R	E
564829	572854	Unknown	1.3	1079V	E
564829	572854	Unknown	1.4	1079C	E
240720	242279	Unknown	1.5	E_DK20	
179835	180770	Unknown	1.6	SQ32	
179835	180770	Unknown	1.6	SQ29	
908055	909649	Unknown	1.7	1078V	D
183776	183934	Unknown	1.7	F_70	
234320	234786	Unknown	1.7	F_70	
179835	180770	Unknown	1.8	E_DK20	
234320	234786	Unknown	1.8	SQ32	
234320	234786	Unknown	1.8	SQ29	
410939	412132	Unknown	1.9	1176R	Ja
410939	412132	Unknown	1.9	1182C	Ja
113026	116013	Unknown	2	F_70	
670547	673508	Unknown	2.2	1079C	E
670547	673508	Unknown	2.3	1079R	E
670547	673508	Unknown	2.3	1079V	E
1	4436	Unknown	14.7	Ct_refL3	
54539	56341	Unknown	208.6	1078V	D
54539	56137	Unknown	216.6	1106C	D
54539	56137	Unknown	216.8	1106R	D
54539	56137	Unknown	217.1	1106V	D
55174	55623	Unknown	82.3	1182R	Ja
105937	111228	Unknown	5	E_DK20	
124692	125577	Unknown	8.5	362C	F
124692	125577	Unknown	8.5	362R	F
124692	125577	Unknown	8.5	362V	F
124692	125577	Unknown	8.5	G_11222	
183068	185466	Unknown	33.7	SQ32	
183068	185466	Unknown	33.7	SQ29	
183068	185466	Unknown	34.8	E_DK20	
187879	191434	Unknown	6.3	Ct_RefC	
187879	191434	Unknown	9.3	Ba	
191628	192722	Unknown	9.1	QH111L	
192027	192606	Unknown	2.6	F_70	
192027	192606	Unknown	2.7	SQ32	
192027	192606	Unknown	2.7	SQ29	
192027	192606	Unknown	2.9	E_DK20	
194157	194897	Unknown	7.6	SQ32	
194157	194897	Unknown	7.6	SQ29	
194299	194897	Unknown	9.4	F_70	
194299	194897	Unknown	9.9	E_DK20	
240720	243742	Unknown	23.3	F_70	
240720	243742	Unknown	23.6	SQ32	
240720	243742	Unknown	23.6	SQ29	
248933	251545	Unknown	24.8	F_70	
255659	255751	Unknown	14.9	F_70	
255659	255751	Unknown	15	SQ32	
255659	255751	Unknown	15	SQ29	
256239	259175	Unknown	5	F_70	
256239	259175	Unknown	5.1	SQ32	
256239	259175	Unknown	5.1	SQ29	
327249	327324	Unknown	12.4	-Cal.fasta	
363708	363751	Unknown	20.5	Ja_47nL	
410939	412132	Unknown	3.7	1176C	Ja
410939	412132	Unknown	11.3	1079C	E
410939	412132	Unknown	11.5	1077C	Ja
410939	412132	Unknown	11.5	1079R	E
410939	412132	Unknown	11.5	1079V	E
410939	412132	Unknown	11.6	1077R	Ja
410939	412132	Unknown	11.6	1077V	Ja
410939	412132	Unknown	11.6	1078R	E
410939	412132	Unknown	11.9	107C	Ja
410939	412132	Unknown	11.9	107R	Ja
410939	412132	Unknown	11.9	107V	Ja
410939	412132	Unknown	11.9	1145C	Ja
410939	412132	Unknown	11.9	1145R	Ja

410939	412132	Unknown	11.9	1145V	Ja
475704	477304	Unknown	107.3	Ja_26s	
533192	534985	Unknown	12.2	Ja_26s	
552194	553288	Unknown	6.2	Ja_26s	
560155	564195	Unknown	11.7	F_70	
562294	563151	Unknown	4.2	Ja_26s	
607013	610302	Unknown	11.3	Ja_26s	
670547	673508	Unknown	2.4	1077C	Ja
670547	673508	Unknown	2.4	1077R	Ja
670547	673508	Unknown	2.4	1077V	Ja
670547	673508	Unknown	3.8	1078R	E
699525	700240	Unknown	3.7	1182V	Ja
700608	701180	Unknown	132.4	Ct_refL3	
708134	709277	Unknown	61.6	1078V	D
708134	708338	Unknown	87.5	1106C	D
708134	708338	Unknown	87.5	1106R	D
708134	708338	Unknown	87.7	1106V	D
708707	709277	Unknown	5.2	1078R	E
708707	708980	Unknown	10.7	1077C	Ja
708707	708980	Unknown	10.7	1077R	Ja
708707	708980	Unknown	10.7	1077V	Ja
732899	733258	Unknown	39.5	Ct_refL3	
747566	747913	Unknown	6.1	1077C	Ja
747639	747913	Unknown	7.6	1078R	E
747639	747913	Unknown	7.7	1077R	Ja
747639	747913	Unknown	7.7	1077V	Ja
752032	755143	Unknown	23.7	1078V	D
770865	771562	Unknown	13.2	1078R	E
771476	772426	Unknown	44.4	1079R	E
771476	772415	Unknown	44.8	1079V	E
771476	772426	Unknown	45.2	107C	Ja
771476	772426	Unknown	45.3	107R	Ja
771476	772426	Unknown	45.3	107V	Ja
771476	772426	Unknown	45.3	1145R	Ja
771476	772426	Unknown	45.3	1145V	Ja
771476	772426	Unknown	45.4	1145C	Ja
805980	806131	Unknown	4.9	Ja_47nL	
847368	847574	Unknown	16.3	Ja_26s	
896817	901645	Unknown	2.8	SQ25	
1026738	1026792	Unknown	33.7	1176V	G
1028008	1028569	Unknown	10.5	1176C	Ja
1028570	1042519	Unknown	763	Ct_refL3	
1028881	1029705	Unknown	16.1	1079C	E
1028881	1029705	Unknown	16.2	1079R	E
1028881	1029705	Unknown	16.2	1079V	E
1029023	1030688	Unknown	115.8	Ct_refDa	
1029050	1030672	Unknown	137.4	Ct_RefC	
1029050	1030672	Unknown	139.7	Ba	
1031470	1031969	Unknown	60.8	Ct_refDa	
1035391	1037320	Unknown	169.4	Ba	
1035391	1037333	Unknown	174.2	Ct_RefC	
1035742	1036028	Unknown	42.1	Ct_refDa	
1036354	1037400	Unknown	104.4	Ct_refDa	
1038578	1038595	Unknown	21.4	68C	Ja

Start & End, coordinates of the putative recombination region on the reference chromosome
P-UA, prevalent urogenital and anorectal; NP-UA, non-prevalent urogenital and anorectal (as described in **Figure 2**)
Unknown, In some cases, the donor is unknown or uncertain, probably representing unsampled lineages of *C. trachomatis*
logBF, log of the Bayes Factor score
Fiji genome names, as in **Supplemental Table 1**
omp A genotypes, based on complete sequence of the *omp* A gene and provided for each Fiji genome
Blocks of shared recombination events with "unknown donor" assignments are highlighted in different colors.

Supplemental Table 3. Sample information from Group A and B participants

Participant	Group	Endocervical clade/ompA genotype	Rectal clade/ompA genotype	Vaginal clade/ompA genotype	Fixed SNPs found in all 3 sites	Rare Fixed SNPs in 3 sites	Rare fixed SNPs found only in rectal sample	Rare iSNPs found only in rectal sample	Rare fixed SNPs found only in endocervical sample	Rare iSNPs found only in endocervical sample	Rare fixed SNPs found only in vaginalsample	Rare iSNPs found only in vaginal sample	Rare fixed SNPs found in endocervical and rectal but not vaginal sample	Rare iSNPs found in endocervical and rectal but not vaginal sample	Rare fixed SNPs found in vaginal and rectal but not endocervical sample	Rare iSNPs found in vaginal and rectal but not endocervical sample	Rare fixed SNPs found in vaginal and endocervical but not rectal sample	Rare iSNPs found in vaginal and endocervical but not rectal sample	Fixed SNP in rectal sample rare iSNP in vagina	Fixed SNP in rectal sample rare iSNP in endocervix	Fixed SNP in rectal sample rare iSNP in others	Fixed SNP in endocervix sample rare iSNP in vagina	Fixed SNP in endocervix sample rare iSNP in rectum	Fixed SNP in endocervix sample rare iSNP in others	Fixed SNP in vagina sample rare iSNP in endocervix	Fixed SNP in vagina sample rare iSNP in rectum	Fixed SNP in vagina sample rare iSNP in others
927	A	P-UA/E	P-UA/E	P-UA/E	5464	1388	0	27	0	13	0	22	0	11	0	8	0	6	0	0	0	0	0	0	0	0	0	1
26	A	P-UA/E	P-UA/E	P-UA/E	5316	1264	0	91	0	16	0	50	0	10	0	2	0	6	0	0	0	0	0	3	0	0	0	1
951	A	P-UA/E	P-UA/E	P-UA/E	5291	1192	0	23	1	15	0	88	0	13	0	4	0	6	0	0	0	0	0	7	0	0	0	1
192	A	P-UA/Ja	P-UA/Ja	P-UA/Ja	5245	1330	2	25	0	81	0	90	0	5	1	11	0	2	0	0	0	0	0	0	2	0	0	2
68	A	P-UA/Ja	P-UA/Ja	P-UA/Ja	5241	1085	0	53	0	18	0	57	0	7	0	3	0	2	0	0	0	0	0	0	0	0	0	0
1226	A	P-UA/E	P-UA/E	P-UA/E	5187	1123	1	183	0	35	0	72	0	3	0	5	0	3	0	0	2	0	0	1	0	0	0	0
908	A	P-UA/Ja	P-UA/Ja	P-UA/Ja	4911	1010	0	51	0	120	0	228	0	4	0	4	0	1	0	0	1	0	0	0	0	0	0	3
564	B	P-UA/Ja	P-UA/E	P-UA/E	4750	949	0	105	72	43	0	128	0	4	107	16	0	3	0	0	1	0	0	0	0	0	0	0
1145	A	P-UA/Ja	P-UA/Ja	P-UA/Ja	4242	841	1	17	0	47	0	244	0	3	0	4	0	2	0	0	0	0	0	0	0	0	0	1
519	A	P-UA/E	P-UA/E	P-UA/E	4066	860	0	250	0	2	0	200	0	3	0	1	0	3	0	0	1	0	0	0	0	0	0	0
107	A	P-UA/Ja	P-UA/Ja	P-UA/Ja	3600	729	2	30	0	274	0	74	0	2	0	5	0	1	0	0	0	0	0	1	0	0	0	0
1077	A	P-UA/Ja	P-UA/Ja	P-UA/Ja	3172	670	0	226	0	9	0	79	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1079	A	P-UA/E	P-UA/E	P-UA/E	2189	917	0	329	0	31	0	78	0	5	0	11	0	5	0	0	1	0	0	16	0	0	0	1
362	A	NP-UA/F	NP-UA/F	NP-UA/F	1986	939	0	195	0	124	0	76	0	9	0	3	0	2	0	0	2	0	0	2	0	0	0	1
98	A	NP-UA/G	NP-UA/G	NP-UA/G	1247	880	0	27	0	30	0	69	0	3	0	7	0	2	0	0	0	0	0	0	0	0	0	1
1201	A	NP-UA/G	NP-UA/G	NP-UA/G	1139	788	1	58	0	6	0	6	0	1	0	3	0	5	0	0	0	0	0	8	0	0	0	0
87	A	NP-UA/G	NP-UA/G	NP-UA/G	1006	721	0	126	0	149	0	110	1	0	0	0	0	3	0	0	0	1	0	0	0	0	0	0
227	A	NP-UA/G	NP-UA/G	NP-UA/G	825	582	0	192	0	1	0	35	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
32	B	P-UA/G	P-UA/G	NP-UA/G	759	547	0	18	0	251	0	415	14	14	0	1	0	5	0	0	1	0	0	1	0	0	0	0
1106	A	NP-UA/D	NP-UA/D	NP-UA/D	695	452	0	17	0	3	0	18	0	4	0	1	0	0	0	0	0	1	0	0	0	0	0	2
753	A	NP-UA/G	NP-UA/G	NP-UA/G	633	451	0	176	0	26	0	25	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
968	A	NP-UA/G	NP-UA/G	NP-UA/G	605	436	0	208	0	1	0	9	0	0	0	0	0	2	0	0	0	1	0	1	0	0	0	1
72	A	NP-UA/G	NP-UA/G	NP-UA/G	554	341	0	65	0	13	0	40	0	2	0	3	0	5	0	0	1	0	0	0	0	0	0	0
1176	B	P-UA/Ja	P-UA/Ja	NP-UA/G	385	378	0	78	1	54	1	402	25	117	0	12	0	4	0	0	0	0	9	8	0	0	0	52
1182	B	P-UA/Ja	NP-UA/Ja	P-UA/Ja	250	231	13	252	6	61	0	21	0	2	0	13	218	93	1	0	0	125	0	37	0	0	0	0
1078	B	nd	P-UA/E	NP-UA/D	32	32	41	119	0	118	1	44	0	3	0	12	2	13	53	57	63	9	0	0	5	0	0	0

NOTE: "Fixed" SNPs are defined as < 10% reference allele frequency and not in fastGEAR defined recombination blocks. SNVs have 10-90% reference allele frequency. “Rare” SNPs or SNVs are only found in <= 3 samples, which in almost all cases means that they only appeared in samples isolated from one study participant.