ARTICLES
https://doi.org/10.1038/s41564-022-01070-7
nature microbiology

Check for updates

# Cysteine dependence of *Lactobacillus iners* is a potential therapeutic target for vaginal microbiota modulation

Seth M. Bloom[1,2,3], Nomfuneko A. Mafunda[1,17], Benjamin M. Woolston[4,5], Matthew R. Hayward[1,3], Josephine F. Frempong[1,18], Aaron B. Abai[6], Jiawu Xu[1], Alissa J. Mitchell[7,19], Xavier Westergaard[1,7,20], Fatima A. Hussain[1,3], Nondumiso Xulu[8], Mary Dong[1,9], Krista L. Dong[1,2,3], Thandeka Gumbi[10], F. Xolisile Ceasar[10], Justin K. Rice[3,21], Namit Choksi[7,22], Nasreen Ismail[8], Thumbi Ndung'u[1,8,11,12,13], Musie S. Ghebremichael[1,3], David A. Relman[14,15,16], Emily P. Balskus[4], Caroline M. Mitchell[1,3,7] and Douglas S. Kwon[1,2,3] ✉

**Vaginal microbiota composition affects many facets of reproductive health. *Lactobacillus iners*-dominated microbial communities are associated with poorer outcomes, including higher risk of bacterial vaginosis (BV), compared with vaginal microbiota rich in *L. crispatus*. Unfortunately, standard-of-care metronidazole therapy for BV typically results in dominance of *L. iners*, probably contributing to post-treatment relapse. Here we generate an *L. iners* isolate collection comprising 34 previously unreported isolates from 14 South African women with and without BV and 4 previously unreported isolates from 3 US women. We also report an associated genome catalogue comprising 1,218 vaginal *Lactobacillus* isolate genomes and metagenome-assembled genomes from >300 women across 4 continents. We show that, unlike *L. crispatus*, *L. iners* growth is dependent on L-cysteine in vitro and we trace this phenotype to the absence of canonical cysteine biosynthesis pathways and a restricted repertoire of cysteine-related transport mechanisms. We further show that cysteine concentrations in cervicovaginal lavage samples correlate with *Lactobacillus* abundance in vivo and that cystine uptake inhibitors selectively inhibit *L. iners* growth in vitro. Combining an inhibitor with metronidazole promotes *L. crispatus* dominance of defined BV-like communities in vitro by suppressing *L. iners* growth. Our findings enable a better understanding of *L. iners* biology and suggest candidate treatments to modulate the vaginal microbiota to improve reproductive health for women globally.**

Female genital tract (FGT) microbiota composition is linked to important women's health and reproductive outcomes including human immunodeficiency virus (HIV) risk[1,2], preterm birth[3], mucosal inflammation[4–6], human papillomavirus infection[7] and cervical dysplasia[7]. Diverse, anaerobe-dominated bacterial communities are associated with negative sequelae and are a hallmark of bacterial vaginosis (BV)[8,9], a syndrome involving vaginal discharge and mucosal inflammation that affects up to 58% of women worldwide, with disproportionately high impact in sub-Saharan Africa[10]. By contrast, FGT bacterial communities dominated by *Lactobacillus crispatus* and most other *Lactobacillus* species are generally considered optimal for health[9]. One notable exception involves communities dominated by *L. iners*, the most prevalent and abundant FGT species worldwide[11], which are associated with many of the same unfavourable outcomes as BV[1,7,12] and have higher risk of transitioning to BV and BV-like communities[3,13–17]. However factors influencing the balance between *L. iners* and more health-associated lactobacilli are incompletely understood[16,18].

The close associations between vaginal microbiota and disease make microbiota-targeted therapies a key strategy to improve women's health and reproductive outcomes[8]. Microbiota modulation has primarily been studied for BV, where current treatments exhibit low efficacy and high rates of recurrence. Over 50% of women with symptomatic BV who receive standard antibiotic therapy with metronidazole (MTZ) experience recurrence by 12 months, while up to 80% of women with previous history of recurrent BV relapse within

[1]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA. [2]Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA, USA. [3]Harvard Medical School, Boston, MA, USA. [4]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [5]Department of Chemical Engineering, Northeastern University, Boston, MA, USA. [6]Harvard College, Harvard University, Cambridge, MA, USA. [7]Department of Obstetrics and Gynecology, Massachusetts General Hospital, Boston, MA, USA. [8]HIV Pathogenesis Programme (HPP), The Doris Duke Medical Research Institute, University of KwaZulu-Natal, Durban, South Africa. [9]Massachusetts General Hospital, Boston, MA, USA. [10]Health Systems Trust, Durban, South Africa. [11]Africa Health Research Institute (AHRI), Durban, South Africa. [12]Max Planck Institute for Infection Biology, Berlin, Germany. [13]Division of Infection and Immunity, University College London, London, UK. [14]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA. [15]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. [16]Infectious Diseases Section, Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, USA. [17]Present address: Harvard T.H. Chan School of Public Health, Boston, MA, USA. [18]Present address: Medical Scientist Training Program, Washington University School of Medicine, St Louis, MO, USA. [19]Present address: William Carey University College of Osteopathic Medicine, Hattiesburg, MS, USA. [20]Present address: Department of Biological Sciences, Columbia University, New York, NY, USA. [21]Present address: Ronald O. Perelman Department of Emergency Medicine, NYU School of Medicine, New York, NY, USA. [22]Present address: Rishihood University - School of Healthcare, Sonepat, Haryana, India. ✉e-mail: dkwon@mgh.harvard.edu

16 weeks[19,20]. MTZ typically shifts BV-associated microbiota towards *L. iners*-dominated communities[17,21–24], which frequently transition back to BV-like communities[13,15,16]. Novel strategies to promote *L. crispatus* dominance over *L. iners* dominance during BV treatment are thus a key priority[16,25].

Despite its ubiquity in human populations, *L. iners*'s biology remains incompletely characterized[18]. Several features distinguish it from other lactobacilli including a putative virulence factor (inerolysin)[26], lack of D-lactic acid and hydrogen peroxide production[27–29], and a smaller genome size with reduced metabolic potential, suggesting niche adaptation and dependence on exogenous nutrients[28]. However, *L. iners* research has been substantially hampered by its species-defining failure to grow in standard *Lactobacillus* MRS (de Man, Rogosa and Sharpe) media[30], complicating attempts at isolation and in vitro characterization[11,18,31]. Consequently, few experimental strains and genomes exist, with many studies reporting little or no success in isolating *L. iners*, even when culturing from hundreds of clinical samples[32–34].

Here we show that *L. iners*'s unique in vitro growth limitation is due to a restricted capacity to exploit exogenous cysteine (Cys) sources. Through experiments and analysis of a collection of >1,200 diverse *Lactobacillus* genomes, we demonstrate that major vaginal *Lactobacillus* species lack canonical Cys biosynthesis pathways. We find that vaginal Cys concentrations are higher in women without BV and that in vivo abundances of both *L. iners* and *L. crispatus* correlate with Cys availability. However, *L. iners* lacks transport mechanisms for Cys and Cys-containing molecules possessed by other *Lactobacillus* species, rendering its growth uniquely susceptible to inhibition by cystine uptake inhibitors. Combining an inhibitor with MTZ promotes *L. crispatus* dominance of defined BV-like bacterial communities by suppressing *L. iners*, identifying a previously undescribed target for therapeutically modulating vaginal microbiota to promote women's health.

## Results

**L-cysteine supports growth of diverse *L. iners* strains.** We investigated the nutritional dependencies underlying *L. iners*'s failure to grow in *Lactobacillus* MRS media[31] by testing various nutrient additives. MRS broth supplemented with IsoVitaleX (a defined micronutrient mixture) supported robust *L. iners* growth while retaining selectivity against *Gardnerella vaginalis* (Fig. 1a). Testing IsoVitaleX components individually revealed that L-cysteine (L-Cys) was necessary and sufficient for this growth phenotype (Extended Data Fig. 1a–f; reagent details in Supplementary Tables 1 and 2). L-glutamine (L-Gln) was neither necessary nor sufficient, but augmented growth in the presence of L-Cys, so subsequent experiments employed a base of L-Gln-supplemented MRS broth (MRSQ).

We next assessed species-level generalizability of Cys-dependent growth by establishing a large, geographically diverse *L. iners* isolate collection and genome catalogue of strains from women both with and without BV. The isolate collection included 4 previously reported US strains from culture repositories, plus 34 previously unreported isolates from 14 South African women and 4 previously unreported isolates from 3 US women, isolated using L-Cys-supplemented MRS or blood agar. The associated genome catalogue comprises 327 genomes derived from studies of disparate human populations in the US, South Africa, Sweden, China and Italy, including 21 'reference' genomes of previously reported isolates retrieved from RefSeq[35], 49 previously unreported isolate genomes and 257 previously unreported culture-independent metagenome-assembled genomes (MAGs) generated from a combination of both previously published and previously unreported FGT shotgun metagenomic sequencing studies (M.R.H. et al., manuscript in preparation; Supplementary Table 3 lists isolates tested experimentally in this study; see also additional isolate and genome source information in Methods, Extended Data Fig. 2 and Supplementary Tables 4–11). Species-level phylogenetic reconstruction revealed that *L. iners* genomic diversity was largely independent of geographic source, genome type or clinical context (Fig. 1b). Strains broadly representative of species diversity isolated from South African and US women both with and without BV were selected for further experimental characterization (Fig. 1b and Supplementary Table 3). All selected strains grew in MRSQ broth with L-Cys (representative examples in Fig. 1c), confirming Cys-dependent growth as a species-level phenotype.

**FGT *Lactobacilli* lack canonical Cys biosynthesis pathways.** We investigated whether *L. iners*'s in vitro Cys-dependent growth phenotype reflected differences in Cys biosynthetic capacity relative to other common FGT lactobacilli. Bacteria canonically synthesize Cys de novo from serine (Ser) via either the Cys synthase pathway (*cysE*, followed by *cysK*, *cysM* or *cysO*) or reverse transsulfuration pathway (*cbs* or *mccA*, followed by *mccB*; Fig. 2a)[36]. We assessed the presence of these pathways in a previously unreported catalogue of 1,218 FGT *Lactobacillus* genomes and MAGs, including the catalogue of 327 *L. iners* genomes described above, along with similarly constructed, geographically diverse catalogues for *L. crispatus* ($n = 310$ genomes and 198 MAGs), *L. gasseri* ($n = 174$ genomes and 42 MAGs), *L. jensenii* ($n = 80$ genomes and 57 MAGs) and *L. vaginalis* ($n = 28$ genomes and 2 MAGs; see Extended Data Fig. 2 and Supplementary Tables 4–11; M.R.H. et al., manuscript in preparation). These catalogues represent substantial expansion of known genomic diversity for all species, most notably for *L. iners* with a >10-fold increase in the number of genomes and a 73% increase in pan-genome size compared with RefSeq-derived genomes alone (Supplementary Table 4). No genomes from any species were predicted to encode intact Cys synthase or reverse transsulfuration pathways, although interestingly all non-*iners* species had predicted *mccB* orthologues (Fig. 2b and Supplementary Table 12). Thus, all common FGT *Lactobacillus* species, including *L. iners*, are predicted to lack canonical mechanisms for de novo Cys biosynthesis.

To phenotypically test this result, we used isotopic tracing to measure the ability of *L. iners* and *L. crispatus* to convert Ser into Cys in vitro. Representative strains were grown in MRSQ broth supplemented with labelled Ser ($^{13}$C-L-Ser) plus homocysteine or with labelled L-Cys ($^{13}$C-L-Cys; labelling scheme in Fig. 2a), then cellular amino acid isotopic labelling was analysed via liquid chromatography–mass spectrometry (LC–MS). Both *L. iners* and *L. crispatus* took up labelled Ser but failed to convert it to Cys (Fig. 2c,d, and Supplementary Tables 13 and 14), consistent with mechanistic predictions from genomic analysis of the species genome collections (Fig. 2a,b and Supplementary Table 12). By contrast, both strains readily incorporated labelled Cys, supporting the genomic prediction that, despite their different in vitro growth phenotypes, both utilize exogenous Cys.

**Cys levels correlate with *Lactobacillus* abundance in vivo.** Since vaginal lactobacilli lack canonical Cys biosynthesis pathways and utilize exogenous Cys, we investigated in vivo relationships between vaginal Cys concentrations, BV and *Lactobacillus* colonization levels. We measured Cys concentrations in cervicovaginal lavage (CVL) fluid from 142 participants in a cohort of South African women aged 18–23 years[1,4,37], 53 of whom were also evaluated for BV by Nugent scoring[38]. Cys concentrations differed strikingly by BV status ($P = 6.3 \times 10^{-9}$), with significantly higher concentrations among women without BV compared with those with BV (Fig. 3a). We explored FGT microbiota composition in this cohort by bacterial 16S rRNA gene sequencing, classifying individual bacterial communities into four 'cervicotypes' (CTs) on the basis of previously established criteria:[1] CT1 (*L. crispatus*-dominated), CT2 (*L. iners*-dominated), CT3 (*Gardnerella*-dominated) and CT4 (non-*Lactobacillus*-dominated, diverse communities typically

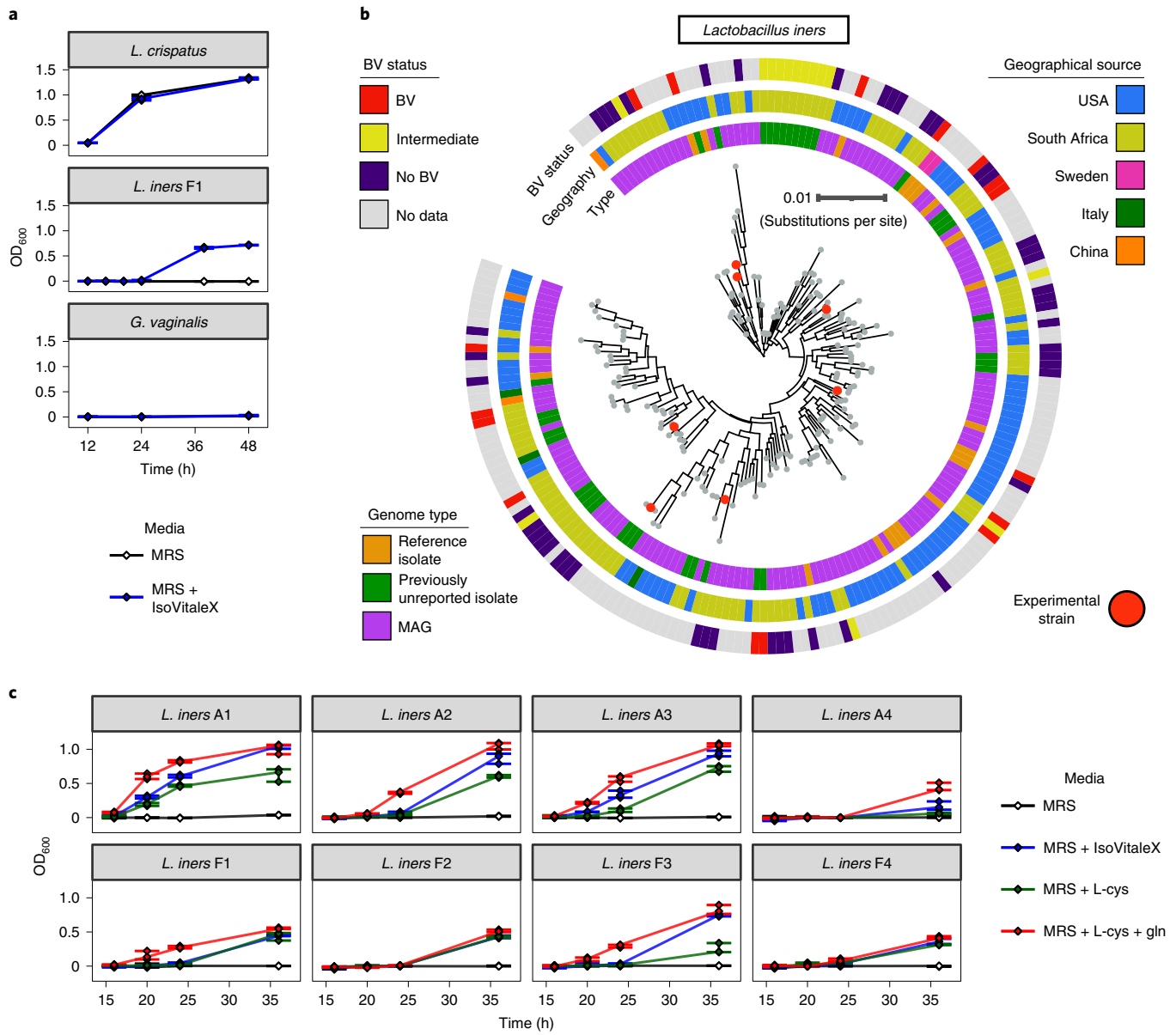



**Fig. 1 | *L. iners* requires L-cysteine supplementation to grow in conditions that support other vaginal lactobacilli. a**, Growth of *L. crispatus*, *L. iners* (strain F1) and *G. vaginalis* in *Lactobacillus* MRS broth ± supplementation with 2% IsoVitaleX. **b**, Unrooted phylogenetic tree of 198 *L. iners* genomes from previously unreported and reference bacterial isolates, and previously undescribed culture-independent MAGs derived from geographically diverse populations. Red dots indicate isolates experimentally studied in this work. Genome type, geographic origin and BV status of the source samples are shown. The tree depicts only genome assemblies exceeding certain quality thresholds to ensure robustness of the phylogenetic reconstruction (Methods); additional strains and genomes were included in other analyses (Supplementary Tables 3–11 and Extended Data Fig. 2). **c**, Growth of 8 representative US (prefix 'A') or South African (prefix 'F') *L. iners* strains cultured in MRS broth supplemented as indicated with IsoVitaleX (2%), L-cysteine (L-Cys, 4 mM) and/or L-glutamine (L-Gln, 1.1 mM). In **a** and **c**, growth was assessed by optical density at 600 nm ($OD_{600}$) and plotted as median ± range for 3 replicates from 1 of ≥2 independent experiments per strain and media condition.

featuring high *Prevotella* abundance) (Fig. 3b; taxonomy assignments in Supplementary Table 15). As expected, CT assignment was strongly related to BV status ($P = 1.902 \times 10^{-11}$, two-sided Fisher's exact test; Extended Data Fig. 3a). Cys concentration differed markedly between CTs ($P = 2.7 \times 10^{-14}$ overall), with significantly higher concentrations in *Lactobacillus*-dominated CTs compared with CT3 and CT4 (Fig. 3c).

We next examined correlations between Cys concentration and abundance of individual taxa within the microbiota. Analysis of CT-associated genera revealed that *Lactobacillus* relative abundance strongly positively correlated with Cys concentration ($\rho = 0.6$, $P = 5.2 \times 10^{-15}$ by Spearman correlation), while *Prevotella* abundance had equivalently strong negative correlation ($\rho = -0.57$, $P = 1.9 \times 10^{-13}$) and *Gardnerella* abundance was weakly negatively correlated (Fig. 3d). *L. crispatus* and *L. iners* each positively correlated with Cys when analysed individually (Extended Data Fig. 3b). To more comprehensively identify taxa associated with Cys, we calculated correlations for each taxon detected in ≥50% of samples at the genus or species level (Extended Data Fig. 3c,d), adjusting for multiple comparisons. *Lactobacillus* (genus level) and *L. crispatus* and *L. iners* (species level) were the only taxa positively correlated with Cys (Fig. 3e,f; full statistical results in Supplementary

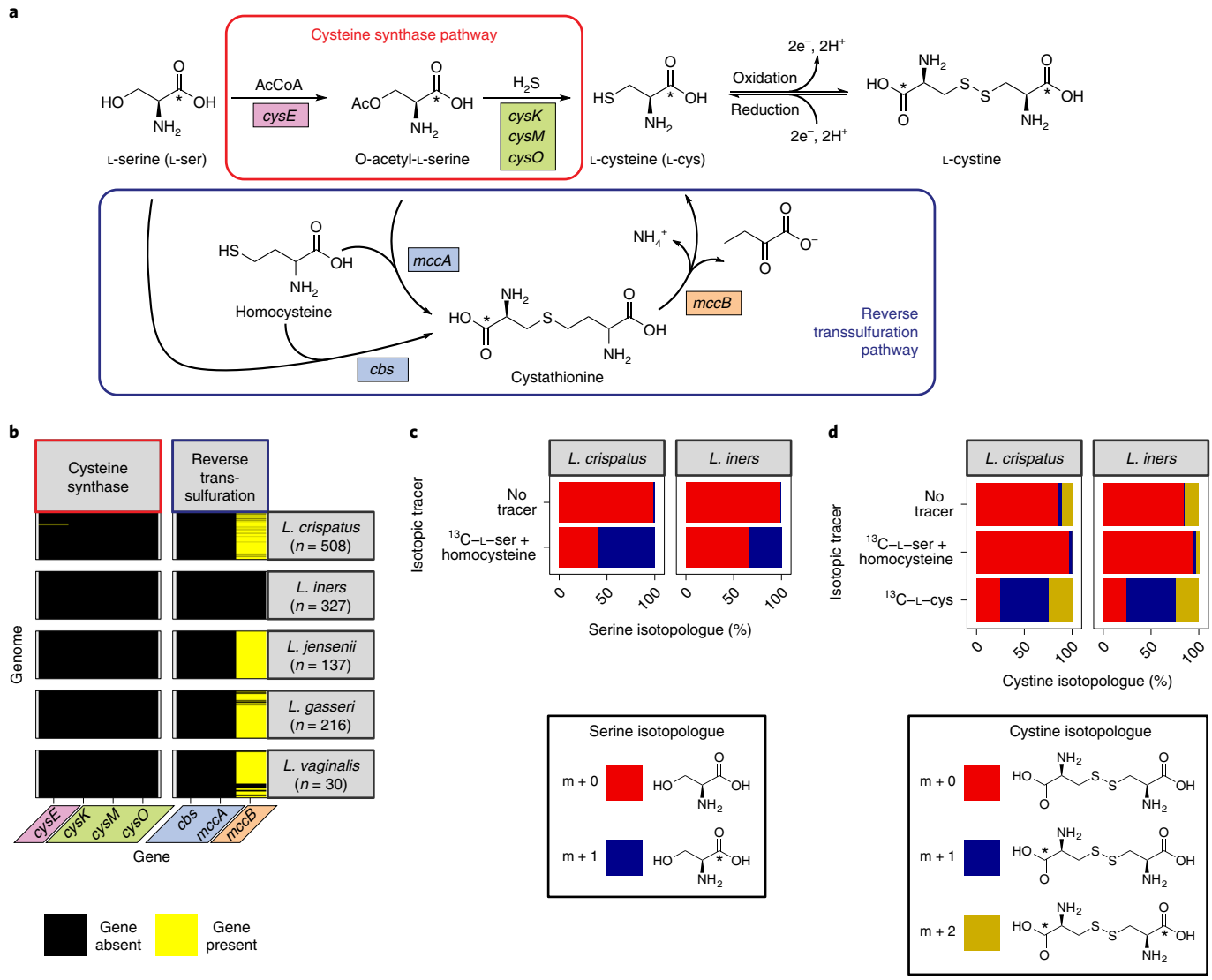



**Fig. 2 | Vaginal lactobacilli lack canonical Cys biosynthesis pathways. a**, Canonical bacterial Cys biosynthesis pathways. Cysteine synthase pathway: serine *O*-acetyltransferase (encoded by *cysE*), then cysteine synthase (*cysK*, *cysM* or *cysO*). Reverse transsulfuration pathway: cystathionine β-synthase (*cbs*) or *O*-acetylserine-dependent cystathionine β-synthase (*mccA*), then cystathionine γ-lyase (*mccB*). Asterisks indicate position of $^{13}C$ label in isotopic tracing experiments. **b**, Predicted presence of genes encoding cysteine biosynthesis enzymes within the cysteine synthase or reverse transsulfuration pathways in isolate genomes and MAGs of the indicated FGT *Lactobacillus* species (*n* = number of genomes; detailed results in Supplementary Table 12). Assessment of gene presence was based on annotations from eggNOG 5.0 performed using eggNOG-mapper v.2[82,83]. Results were supported by BLAST[84] searches of the genomes for sequences of interest[81]. **c**, Serine isotopic enrichment in *L. crispatus* and *L. iners* grown for 24 h in MRSQ broth containing no isotopic tracer (supplemented with unlabelled L-Cys, 4 mM) or supplemented with 1-$^{13}C$-L-serine ($^{13}C$-L-Ser) + homocysteine (both 4 mM). Plot depicts fractional abundance of unlabelled (m + 0, Ser) and labelled (m + 1, $^{13}C$-Ser) isotopologues in cellular hydrolysates. **d**, Cystine isotopic enrichment in lysates of *L. crispatus* or *L. iners* cultured as shown in **c** or in broth supplemented with 1-$^{13}C$-L-cysteine ($^{13}C$-L-Cys, 4 mM). An oxidation step during sample processing converts Cys to cystine for quantification. Plot depicts fractional abundance of cystine isotopologues: unlabelled (m + 0, cystine), partially labelled (m + 1, $^{13}C_1$-L-cystine) and fully labelled (m + 2, $^{13}C_2$-L-cystine) isotopologues. Data in **c** and **d** depict median values of 3 replicates per group and were unreplicated (full data in Supplementary Tables 13 and 14).

Tables 16 and 17). Other taxa, including various BV-associated bacteria, showed no correlation or significant negative correlation with Cys. Results were similar for the Cys-containing peptides reduced glutathione (GSH) and cysteinylglycine (Extended Data Fig. 4, and Supplementary Tables 16 and 17). Thus, vaginal Cys concentration is higher in women without BV and correlates with abundance of both *L. iners* and *L. crispatus*, consistent with the hypothesis that Cys availability is important for *Lactobacillus* colonization success in vivo.

***L. iners* has limited capacity to use complex L-Cys sources.** The species-specific nature of *L. iners*'s Cys-dependence in vitro appeared discordant with the finding that all FGT *Lactobacillus* species lacked canonical Cys biosynthesis pathways and that in vivo abundance of both *L. iners* and *L. crispatus* correlated with vaginal Cys concentrations. We therefore hypothesized that *L. iners*'s unique in vitro growth phenotype was due to a restricted capacity to utilize exogenous Cys sources in media. Alternatively, Cys might support *L. iners* growth by acting as a chemical reducing agent rather than a direct

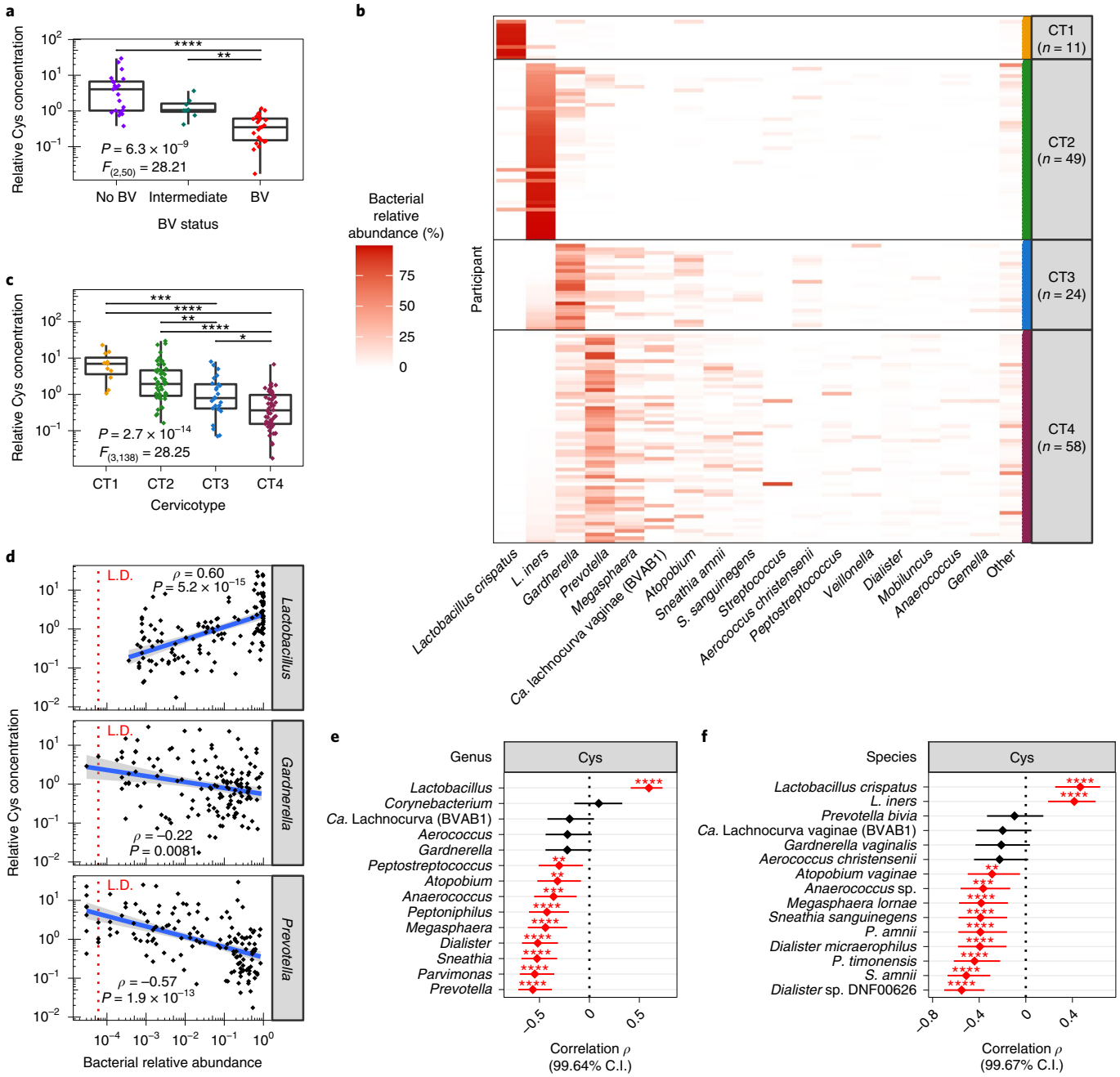



**Fig. 3 | Vaginal Cys concentrations are higher in women without BV and correlate with *Lactobacillus*-dominated microbiota. a**, Relative Cys concentration by BV status in CVL fluid from 53 South African women (21 without BV, 24 with BV and 8 intermediate by Nugent method[38]). Differences in log-transformed concentrations determined by one-way ANOVA with post-hoc Tukey's test; all significant pairwise differences at significance level 0.05 are shown (no BV-BV, $P<1\times10^{-7}$; intermediate-BV, $P=0.0060$). **b**, FGT bacterial microbiota composition among 142 HIV-uninfected South African women (including the 53 from **a**), determined by bacterial 16S rRNA gene sequencing (taxonomy assignments in Supplementary Table 15). Bacterial communities were classified into CTs using previously defined criteria[1]. **c**, Relative Cys concentrations per CT in CVL fluid from women in **b**. Significance determined by one-way ANOVA with post-hoc Tukey's test; all significant pairwise differences are shown. CT1-CT3, $P=0.000120$; CT1-CT4, $P<1\times10^{-7}$; CT2-CT3, $P=0.00898$; CT2-CT4, $P<1\times10^{-7}$; CT3-CT4, $P=0.0279$. In **a** and **c**, box centre lines, edges and whiskers signify median, interquartile range (IQR), and minima and maxima, respectively. **d**, Two-tailed Spearman rank correlation between Cys concentrations and bacterial relative abundances of the genera *Lactobacillus*, *Gardnerella* and *Prevotella*, showing correlation coefficients ($\rho$) with unadjusted $P$ values. Linear regression lines (solid blue) with 95% confidence intervals calculated from log-transformed abundances and concentrations are shown to assist visualization (*Lactobacillus*, $y=0.37+0.32x$; *Gardnerella*, $y=-0.25-0.15x$; *Prevotella*, $y=-0.47-0.27x$). Red dotted line represents L.D. **e**,**f**, Two-tailed Spearman correlation coefficients ($\rho$) with adjusted confidence intervals between Cys concentrations and relative abundances of each genus (**e**) or species (**f**) detected at >50% prevalence in cohort ($n=142$). $P$ values and confidence intervals in **e** and **f** were adjusted for multiple comparisons using the Bonferroni method at significance level $0.05/n$ (full statistical results in Supplementary Tables 16 and 17). Significance of adjusted $P$ values depicted as $^{*}P\leq0.05$, $^{**}P\leq0.01$, $^{***}P\leq0.001$, $^{****}P\leq0.0001$.

nutritional supplement. We found that concentrations of free molecular Cys and cystine (its oxidized counterpart) in MRSQ broth were ≤4.6 µM and 0.7 µM, respectively, which are below levels required by Cys-auxotrophic *Escherichia coli* strains[39]. L-cystine supplementation produced similar *L. iners* growth as L-Cys supplementation (Fig. 4a), including in the presence of the oxidizing agent hydrogen peroxide ($H_2O_2$; Extended Data Fig. 5a), demonstrating that L-cystine supports growth by acting as a direct nutritional supplement rather than by chemically reducing the media. Thus, *L. iners* does not require a reduced environment to grow when it has access to a bioavailable source of L-Cys (for example, L-cystine), but L-Cys bioavailability in unsupplemented MRSQ broth is inadequate for *L. iners*.

The ability of non-*iners* lactobacilli to grow in unsupplemented MRS broth implied the presence of Cys sources that *L. iners* failed to utilize (MRS contains yeast and beef extracts, which contribute peptides in addition to monomeric amino acids[30,40]). Adding chemical reducing agents to MRSQ broth substantially raised Cys and GSH concentrations (Extended Data Fig. 5b), demonstrating that most Cys and Cys-containing molecules in unsupplemented media exist as mixed disulfide compounds. Addition of chemical reducing agents also supported *L. iners* growth (Extended Data Fig. 5a,c), but in contrast to L-cystine, the oxidized counterparts of these reducing agents did not. Thus, MRS broth contains complex sources of L-Cys, including mixed disulfides, that *L. iners* has restricted capacity to utilize.

***L. iners* lacks transporters found in other FGT lactobacilli.** We hypothesized that *L. iners*'s restricted ability to use exogenous Cys sources reflected a limited repertoire of transport mechanisms for Cys and Cys-containing molecules in comparison with other lactobacilli. Uptake of Cys and Cys-containing molecules in bacteria occurs through multiple known or putative mechanisms. These include the well-characterized cystine transport systems TcyABC, TcyJKLMN, and TcyP[41], a putative Cys ABC transporter associated with Cys-binding protein CjaA[42,43], and the heterodimeric ABC transporter CydDC encoded by the redox-regulating locus *cydABCD*[44]—a system that exports both Cys and GSH to the periplasm in *E. coli*[44] and is proposed to perform glutathione uptake in lactobacilli[45]. Analysis of our *Lactobacillus* genome catalogues revealed that all *Lactobacillus* species, except for *L. iners*, encoded *cydABCD*, while *L. crispatus*, *L. vaginalis* and some *L. gasseri* each encoded one or more predicted *cjaA* orthologues, and *L. jensenii*, *L. gasseri* and *L. vaginalis* each encoded one or more cystine transport systems (Fig. 4b and Supplementary Table 12). Low-affinity Cys uptake also occurs via the branched-chain amino acid importer LivJ/LivKHMGF[43], but no genomes possessed *livKHMGF* orthologues except 2 *L. iners* MAGs, and no species possessed an intact GsiABCD glutathione transport system[46] (Extended Data Fig. 6a and Supplementary Table 12). Thus, each common FGT *Lactobacillus* species, except *L. iners*, encodes multiple mechanistically distinct known or putative systems to transport Cys or Cys-containing molecules, while *L. iners*'s Cys-related uptake occurs via currently unrecognized mechanisms.

***L. iners* growth is inhibited by cystine uptake inhibitors.** Experimental tools to perform genetic screens in *L. iners* do not currently exist[47], so we used phenotypic approaches to characterize its Cys-related transport mechanisms. Isotopic tracing (Fig. 4c) revealed that *L. iners* incorporated labelled L-cystine ($^{13}C_2$-L-cystine) at levels similar to those of *L. crispatus* but failed to take up labelled GSH ($^{13}$C-GSH; synthesis detailed in Supplementary Fig. 1), consistent with the predicted absence of GSH transport activity (Fig. 4b and Extended Data Fig. 6a). We therefore assessed whether *L. iners*'s limited repertoire of transport mechanisms would cause its growth in L-Cys-supplemented MRSQ broth to be uniquely impacted by the known cystine uptake inhibitors *S*-methyl-L-cysteine (SMC) and seleno-DL-cystine (SDLC)[41]. SMC caused species-specific, dose-dependent growth inhibition of a diverse collection of *L. iners* strains at concentrations of 32–64 mM, while SDLC caused selective inhibition at 0.25 mM (Fig. 4d,e and Extended Data Fig. 6b,c). Inhibition occurred for strains from both US and South African women, and from women both with and without BV (Supplementary Table 3). Inhibitor potency varies for different transporters, but these concentrations appear higher for SMC and slightly higher for SDLC than reported inhibitory concentrations for the TcyJKLMN and TcyP transporters in *Bacillus subtilis*[41]. To confirm that growth inhibition was not an artefact specific to MRS media, we assessed inhibition in NYCIII broth, a serum-containing nutrient-rich medium that supports growth of diverse fastidious FGT bacteria[48]. Both SMC and SDLC inhibited *L. iners* growth in NYCIII broth, but not growth of *L. crispatus* (Fig. 4f). They also selectively inhibited *L. iners* growth in MRSQ broth treated with chemical reducing agents, supporting the conclusion that reducing agents promote *L. iners* growth by enhancing Cys bioavailability (Extended Data Fig. 6d). Collectively, these results from isotopic tracing and growth inhibition assays confirm our genomic prediction that *L. iners* possesses a uniquely restricted repertoire of Cys-related transport mechanisms.

**SMC inhibits *L. iners* in competition with *L. crispatus*.** To assess the functional significance of *L. iners* growth inhibition by cystine uptake inhibitors, we assessed whether an inhibitor could shift the balance between *L. iners* and *L. crispatus* in culture competition assays. We tested pairwise strain combinations of *L. crispatus* and *L. iners* in L-Cys-supplemented MRSQ broth at varying SMC concentrations, then measured ratios of *L. iners* to *L. crispatus* in the mixed cultures by 16S rRNA gene sequencing. Experimental controls were included to confirm expected growth patterns and absence of serious contamination during sample processing (Extended Data Fig. 7). SMC significantly suppressed *L. iners* relative to *L. crispatus* in a dose-dependent fashion (representative experiments in Fig. 5a and Supplementary Table 18), confirming that it offers selective advantage to *L. crispatus* in direct competition.

**SMC plus MTZ enriches *L. crispatus* in mock BV communities.** To assess whether cystine uptake inhibitors have the potential to augment BV therapy by preferentially promoting *L. crispatus* expansion relative to *L. iners*, we investigated SMC's effects in combination with MTZ on mock BV-like bacterial communities cultured in a rich nutritional milieu in vitro. Multiple South African *L. crispatus* isolates unexpectedly failed to grow in NYCIII broth, so we developed a previously unreported nutrient-rich serum-containing formulation named 'S-broth' that supported South African *L. crispatus* strains, as well as diverse BV-associated bacteria (Extended Data Fig. 8a). Testing the effects of SMC and MTZ on pure cultures of *L. crispatus*, *L. iners*, *G. vaginalis*, *Prevotella bivia* and *Atopobium* (*Fannyhessea*) *vaginae* grown in S-broth revealed that neither MTZ nor SMC blocked *L. crispatus* growth, that SMC inhibited *L. iners*, and that MTZ suppressed both *G. vaginalis* and *P. bivia* (Fig. 5b). The *A. vaginae* strain was relatively MTZ-resistant, but susceptibility was enhanced by adding SMC.

We next cultured defined communities of these species in S-broth and assessed whether combining SMC with MTZ enhanced *L. crispatus* dominance. Each community contained a different pairwise combination of *L. crispatus* and *L. iners* strains (input ratios as in Supplementary Table 19) and we quantified mock community composition by 16S rRNA gene sequencing (controls shown in Extended Data Fig. 8b,c). Cultivation without inhibitors produced diverse mixtures of all 5 species (representative examples in Fig. 5c). As hypothesized, SMC alone diminished *L. iners* relative abundance, MTZ alone suppressed BV-associated species, and MTZ combined with SMC preferentially favoured *L. crispatus*. To quantify *L. crispatus*'s competitive advantage, we calculated the ratio

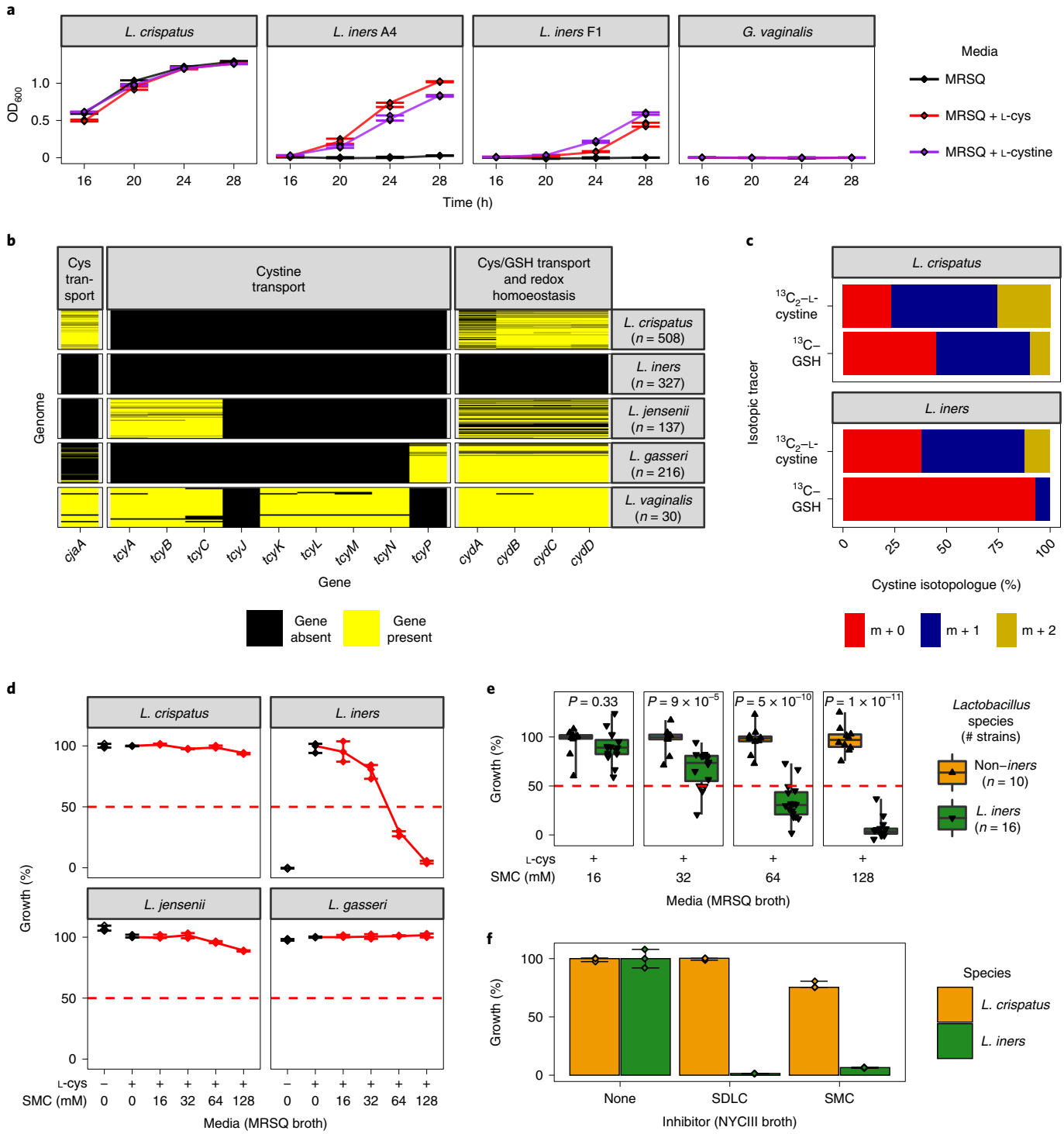



**Fig. 4 | *L. iners* lacks Cys-related transport mechanisms present in other lactobacilli and is selectively inhibited by cystine uptake inhibitors. a**, Growth of *L. crispatus*, *G. vaginalis* and representative *L. iners* strains in MRSQ broth ± L-Cys (4 mM) or L-cystine (2 mM). **b**, Predicted presence of putative Cys transport-related gene *cjaA*, cystine transport loci *tcyABC*, *tcyJKLMN* and *tcyP*, and the Cys/GSH transport/redox homoeostasis locus *cydABCD* in FGT *Lactobacillus* species genomes and MAGs ($n$ = number of genomes; detailed results in Supplementary Table 12). TcyBC takes up glutathione in some species. **c**, Cystine isotopic enrichment in *L. crispatus* or *L. iners* grown in MRSQ broth containing labelled L-cystine ($^{13}C_2$-L-cystine, 2 mM) or GSH ($^{13}$C-GSH, 4 mM; synthesized as described in Supplementary Fig. 1). Plot depicts isotopologue median fractional abundance and was unreplicated. (Full data in Supplementary Tables 13 and 14). **d**, Growth of representative *Lactobacillus* strains in L-Cys-supplemented MRSQ broth containing the cystine uptake inhibitor SMC. Percentages were calculated relative to the median $OD_{600}$ measurement in the L-Cys-containing, no-inhibitor control. **e**, Median growth inhibition of non-*iners Lactobacillus* strains (*L. crispatus*, $n = 7$; *L. jensenii*, $n = 2$; *L. gasseri*, $n = 1$) or *L. iners* ($n = 16$) by SMC in L-Cys-supplemented MRSQ broth. Significance of differences between *L. iners* and non-*iners* strains were determined by two-sided $t$-test. Box centre lines, edges and whiskers signify the median, IQR, and minima and maxima, respectively. **f**, Growth inhibition of *L. crispatus* and *L. iners* by SMC (128 mM) or SDLC (2 mM) in NYCIII broth. Plots in **a**, **d** and **f** depict median ± range for 3 replicates per condition and are representative of ≥2 independent experiments per condition.

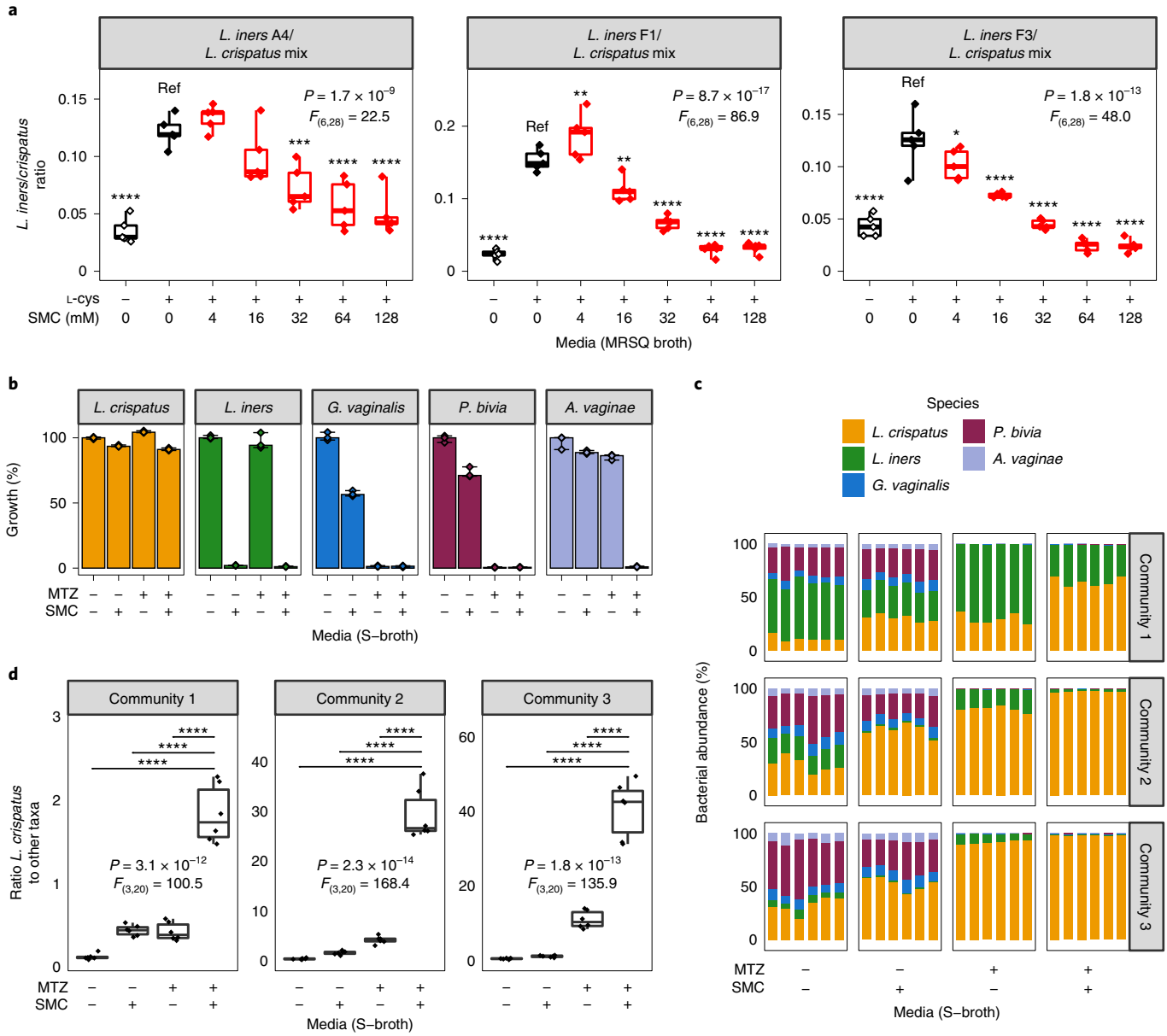



**Fig. 5 | SMC inhibits *L. iners* in competition with *L. crispatus* and combining SMC with metronidazole enhances *L. crispatus* dominance of mock BV-like communities. a**, Ratios of *L. iners* to *L. crispatus* in representative mixed culture competition assays after 28 h incubation in L-Cys-supplemented MRSQ broth with varying concentrations of SMC. Ratios were determined by sequencing the bacterial 16S rRNA gene in DNA isolated from the mixed cultures; plots depict results for 5 replicates per condition. Between-group differences were determined by one-way ANOVA. Pairwise comparisons to the reference condition ('Ref': no-inhibitor, L-Cys-supplemented control) were calculated using Dunnett's test. All significant comparisons are shown (full statistical results in Supplementary Table 18). Input ratios of c.f.u.s of *L. iners* to *L. crispatus* were 8.01:1, 1.08:1 and 5.96:1 for the mixes containing *L. iners* strains A4, F1 and F3, respectively. **b**, Growth inhibition of *L. crispatus*, *L. iners* and 3 BV-associated species at 48 h incubation in S-broth with or without SMC and/or MTZ. Plots depict median ± range for 3 replicates per condition and are representative of 2 independent experiments per condition. **c**, Relative abundance of bacterial species in 3 representative mock BV-like communities (each with a different *L. iners* strain) grown in S-broth with or without SMC and/or MTZ at 28 h incubation. Composition was determined by bacterial 16S rRNA gene sequencing. Input ratios of strains for each community are shown in Supplementary Table 19. **d**, Ratios of *L. crispatus* to the sum of all other taxa in the mock communities shown in **c**. **c** and **d** depict 6 replicates per condition. Between-group differences in **d** were determined by one-way ANOVA with post-hoc Tukey's test; selected pairwise differences are shown (full statistical results in Supplementary Table 20). In **a** and **e**, significance is depicted as $^{*}P \leq 0.05$, $^{**}P \leq 0.01$, $^{***}P \leq 0.001$, $^{****}P \leq 0.0001$. In **a** and **d,** box centre lines, edges and whiskers signify the median, IQR, and minima and maxima, respectively.

of *L. crispatus* relative abundance to the summed abundances of all other taxa (Fig. 5d and Supplementary Table 20). Combining MTZ with SMC increased this ratio significantly more than MTZ alone, suggesting that cystine uptake inhibitors could interact with MTZ to promote *L. crispatus* dominance during BV treatment.

## Discussion

*L. iners* has long been recognized as possessing diminished metabolic potential compared with other lactobacilli – indeed, the name '*iners*' refers to relative inertness in biochemical assays[31]. However, despite being the most abundant FGT species worldwide

and having adverse health associations compared with *L. crispatus*[1,3,7,11,12,16,18,49–54], *L. iners*'s biology remains poorly characterized because it fails to grow in *Lactobacillus* MRS media[31], resulting in a paucity of isolates and genomes[11,18]. Here we resolve this growth defect by identifying a species-specific requirement for Cys supplementation. Employing an isolate collection encompassing both reference and previously unreported strains and >1,200 *Lactobacillus* genomes from diverse geographical and clinical sources, we extend and experimentally confirm predictions from smaller-scale genomic studies suggesting that all common FGT lactobacilli lack canonical Cys biosynthesis pathways[55]. We further show that vaginal Cys concentrations strongly correlate with *Lactobacillus* relative abundance in vivo in a South African cohort. However, *L. iners* lacks transport mechanisms for Cys and Cys-containing molecules present in other lactobacilli, rendering its growth selectively susceptible to cystine uptake inhibitors. Combining a cystine uptake inhibitor with MTZ (a standard antibiotic used for BV treatment) enhances *L. crispatus* dominance in mock BV-like bacterial communities by suppressing *L. iners* growth. These advances establish a foundation for future study of *L. iners*'s biology, diversity and effects on human health while identifying a previously undescribed target for therapeutic vaginal microbiota modulation.

The in vivo correlation between Cys concentrations and *Lactobacillus* abundance suggests that Cys availability may influence FGT microbiota composition. A US-based case-control study comparing women with and without BV found that vaginal Cys concentrations positively correlated with non-*iners* lactobacilli and negatively correlated with various BV-associated species[56]. That study appeared to show a weaker correlation between *L. iners* and Cys than the correlation we identify here[56], which may reflect differences in study design and/or species abundances between the two populations. Since FGT lactobacilli appear to lack de novo Cys biosynthetic capacity, we hypothesize that the high Cys concentrations in *Lactobacillus*-dominated states are probably primarily host-derived. Possible mechanisms explaining the correlation could thus include preferential *Lactobacillus* colonization of hosts with high levels of mucosal Cys secretion, induction of host Cys secretion by lactobacilli, Cys degradation by host cells in setting of BV-associated inflammation[57], and/or degradation of Cys by BV-associated bacteria[56] as a means to outcompete lactobacilli. Further investigation of FGT microbiome–metabolome relationships will be required to fully assess these possibilities.

A striking aspect of our findings is the paucity of transport mechanisms for Cys and Cys-containing molecules in *L. iners* compared with other FGT lactobacilli. In contrast to cystine transporters[41], bacterial Cys transporters are not well characterized due to technical challenges inherent to redox chemistry[43]. Our genome analysis predicted that non-*iners* FGT *Lactobacillus* species each encode multiple known or putative transporters for Cys, cystine and/or Cys-containing molecules, whereas we identified no predicted transporters for these molecules in *L. iners*. However, both L-Cys and L-cystine supported *L. iners* growth, and isotopic tracing demonstrated uptake when *L. iners* was grown in the presence of labelled L-Cys or L-cystine, pointing to one or more currently unrecognized uptake mechanisms. The finding that cystine uptake inhibitors selectively inhibit *L. iners* growth provides further evidence that it lacks the multiple mechanistically distinct Cys-related transporters present in other species.

In addition to providing mechanistic insights, our results have important implications for BV treatment. Standard BV therapy with MTZ often results in disease recurrence[19,20,25]. This may be partially due to MTZ's tendency to promote *L. iners*-dominated FGT bacterial communities[17,21–24], which transition back to BV-like states more frequently than *L. crispatus*-dominated communities[13–17]. We recently showed in a South African population that *L. iners*-dominated communities are a 'gateway' for transitions from *Lactobacillus* dominance to more diverse BV-like communities, and our modelling of microbiota dynamics suggested that interventions shifting community transition probabilities towards *L. crispatus* over *L. iners* would most effectively increase the prevalence of *L.* crispatus dominance within the population[16]. Therapies to promote *L. crispatus* dominance during BV treatment are under active investigation. In a trial of a live *L. crispatus* biotherapeutic administered after MTZ treatment, BV recurred less frequently among biotherapeutic recipients than placebo recipients[25]. However recurrence in the treatment arm remained high (30% within 3 months)[25]. A small Israeli cases series reported achieving *L. crispatus* dominance and long-term BV remission in four of five women treated for refractory BV with MTZ plus vaginal microbiome transplants from healthy donors[58], but most patients required multiple rounds of therapy. Interestingly, that study did not report detecting *L. iners* in any donors or recipients – a notable difference from typical FGT microbiota composition in North and South America, Europe, Asia, Africa and Australia[1–5,12,13,17,21–24,48–54,56,59,60]. Thus, live biotherapeutic and microbiome transplant approaches for BV show promise but are unlikely to be a panacea. Our study shows that combining MTZ with a previously undescribed *L. iners* growth inhibitor was superior to MTZ alone in promoting *L. crispatus* community dominance of defined BV-like bacterial communities. The specific inhibitors identified here have potential limitations for use in humans, including low potency (SMC) and possible host toxicity (SDLC), and their direct effects (if any) on the FGT mucosa remain to be characterized. However, our results demonstrate proof of concept for a previously unrecognized therapeutic approach and provide a rationale to identify additional inhibitors targeting *L. iners* through this mechanism. Further development of this approach as an adjunct to antibiotics has potential to improve therapies for microbiota-linked reproductive health conditions worldwide, including for women in low- and middle-income countries where the burden of adverse reproductive health outcomes is greatest.

## Methods

**Overview of cohorts and sources of metagenomic data.** This study includes previously unreported FGT microbiota sequencing data, FGT metabolite concentration data, bacterial isolates, isolate genomes and/or MAGs assembled from whole-genome shotgun metagenomic data (WGS) from the Females Rising through Education, Support, and Health (FRESH) study—a South African cohort (details below). It also includes previously unreported bacterial isolates, isolate genomes and/or MAGs assembled from WGS data from participants in 6 US-based cohorts, as well as previously unreported MAGs assembled de novo from published WGS data from 5 additional cohorts from the US, Italy and China (details below).

**FRESH cohort characteristics.** The FRESH cohort is an ongoing prospective observational study based in Umlazi, South Africa, that enrols 18–23-year-old, HIV-uninfected, non-pregnant, otherwise healthy women willing to have frequent HIV testing and 3-monthly collection of blood and mucosal specimens during study enrolment. Exclusion criteria included haemoglobin level <10 g dl$^{-1}$, any chronic medical condition or other conflict likely to prevent adherence to the study protocol, and/or enrolment in any other study that involves frequent blood sampling or that might otherwise interfere with the FRESH study protocol. Study characteristics and inclusion and exclusion criteria have also been described in detail elsewhere[1,4,37]. The study protocol was approved by the Massachusetts General Hospital (MGH) Institutional Review Board (IRB, 2012P001812/MGH) and the Biomedical Research Ethics Committee of the University of KwaZulu-Natal (UKZN; Ethics Reference Number BF131/11). Participants were recruited at local sites frequented by young people including shopping malls, cafes and nightclubs. All participants provided written informed consent. The study protocol includes non-monetary benefits encompassing job-skills, health and empowerment training, a light meal during study site visits and cumulative monetary compensation over a 36-week period of ZAR3,700 (~US$280), which helps defray transportation expenses to twice-weekly study site visits. Participant self-selection bias may potentially occur on the basis of participants' perception of their own HIV risk, but it is unknown whether or how such a bias would affect FGT microbiota composition or metabolite concentrations.

FGT sample collection within the FRESH study included multiple cervical swab samples (Puritan 6" sterile standard foam swab with polystyrene handle), which were collected by swabbing the ectocervix in two full revolutions under direct visualization during speculum exam. The swabs were then used to make

a slide preparation for Gram stain analysis, cryopreserved in thioglycolate broth with 20% glycerol for bacterial isolation, or frozen without cryopreservatives for bacterial isolation or nucleic acid extraction and microbiota profiling. CVL samples were collected using a flexible plastic bulb pipette to dispense 5 ml of sterile normal saline into the vaginal vault, wash the cervix four times, then re-aspirate the fluid into a 15 ml conical tube. CVL and swab samples were stored on ice for 1–4 h during transport to the processing laboratory at the Nelson R. Mandela School of Medicine at UKZN, after which swabs were stored at −80 °C and CVL samples were centrifuged at 700 × *g* for 10 min at 4 °C. Supernatants were then transferred to cryovials and stored at −80 °C.

**FRESH cohort microbiota and metabolite profiling data.** Paired bacterial microbiota composition and cervicovaginal metabolite concentrations (Fig. 3, and Extended Data Figs. 3 and 4) were determined for 143 HIV-uninfected FRESH study participants consecutively sampled in May through October 2017 (methodological details below). The sample size for microbiota profiling was initially selected on the basis of a separate, unpublished investigation of microbiome-immune parameters; paired CVL samples from these participants were later used for FGT metabolite concentration measurements due to availability of paired microbiota sequencing data. No a priori sample size calculation was performed in relation to analysis of FGT microbiota–cysteine correlations because there was (to our knowledge) no available investigation of this association in a sub-Saharan African cohort and most studies in non-African populations have used case-control designs that are not adapted to evaluate general population-level correlations. We therefore lacked information upon which to make informed estimates about metabolite distributions for sample size calculations. However, the cohort was of similar or larger size as other reported studies of the FGT metabolome[56,61,62]. One of the 143 participants was excluded from analysis of vaginal metabolite concentrations on the basis of the pre-specified criterion of low bacterial 16S rRNA gene sequence read count (<10,000 reads). Assessment of BV status among this group of FRESH study participants using the Nugent scoring method[38] was performed sporadically before September 2017, when a protocol change instituted universal BV testing for all participants. We restricted analysis of BV–metabolite associations (Fig. 3a and Extended Data Fig. 4a,b) to the subset of 53 out of the 143 participants assessed under the universal testing protocol to minimize risk of bias. Nugent scoring was performed at Global Laboratories (now Neuberg Global Laboratories, Durban, South Africa), an accredited commercial laboratory diagnostics company, by trained laboratory technologists blinded to all participant information other than study ID.

**Sources of bacterial isolates and MAGs.** This study reports previously undescribed bacterial isolates used in experiments (experimentally studied strains are listed in Supplementary Table 3, isolation details are described below) and analyses a genome catalogue including previously unreported *Lactobacillus* bacterial isolate genomes and MAGs (Figs. 1b, 2b and 4b, and Extended Data Figs. 2 and 6a; see also Supplementary Tables 4–11). The genome catalogue was abstracted from a more extensive unpublished global characterization of FGT microbiome phylogenetic structure and function analysing a combination of published bacterial isolate genome and WGS data, as well as previously unreported US and South African genomic and metagenomic data (M.R.H. et al., manuscript in preparation; additional methodological details are provided below). The previously undescribed datasets contributing to the analysis by M.R.H. et al. include isolates and MAGs from a total of 138 FRESH study participants (6 of whom contributed both isolates and MAGs, 11 contributed isolates only and 121 contributed MAGs only). This number included 35 participants from among the 143 participants listed above, as well as 103 additional FRESH study participants for whom microbiota composition based on bacterial 16S rRNA gene-based profiling has previously been described[1,4].

Additional previously unreported US-based primary isolates tested experimentally (Supplementary Table 3), as well as bacterial isolate genomes and WGS sequence data that served as the source of MAGs (Supplementary Tables 5, 6, 8, 9 and 11) were obtained via the Vaginal Microbiome Research Consortium (VMRC; https://vmrc4health.org/). The VMRC data and isolates were derived from two different IRB-approved protocols. The first of these IRB-approved protocols (Stanford University IRB protocol 21956; principal investigator D.A.R.) encompasses two previously described[63] cohorts that enroled 135 pregnant women aged at least 18 years presenting for obstetrical care at either Lucille Packard Children's Hospital at Stanford University or at the University of Alabama, Birmingham, who provided written informed consent. Exclusion criteria included antibiotic or probiotic use within 12 weeks before enrolment, significant immunosuppression, or (retrospectively) delivery of a baby who had congenital defects. Vaginal swab specimens were collected during pregnancy and (in the Stanford cohort) postpartum. Cumulative monetary compensation totalled as much as US$350 per participant. Isolates and MAGs obtained from participants enroled via this protocol were derived from 18 total study participants, 11 of whom contributed both isolates and MAGs, 5 contributed isolates only and 2 contributed MAGs only. The second VMRC-associated cohort (University of Maryland Baltimore IRB protocol HP-00041351; principal investigator J. Ravel), which has also been previously described[23], enroled 135 non-pregnant, HIV-uninfected women at the University of Alabama, Birmingham, who were recruited via newspaper advertisements, fliers, general gynaecologist clinics and the Jefferson County Sexually Transmitted Diseases clinic. Participants were aged 18–45, had regular menses and provided written informed consent. Exclusion criteria included antibiotic or antimycotic use within 30 d before enrolment, pregnancy or current lactation, active infection with *Trichomonas vaginalis*, *Neisseria gonorrhoeae*, *Chlamydia trachomatis*, or syphilis, symptomatic BV or vulvovaginal candidiasis for which treatment was requested at the time of enrolment, immunocompromised status (including HIV), significant affective or psychotic disorders or intellectual limitations precluding informed consent, history of previous or pending hysterectomy or pelvic radiotherapy, non-fluency in English, current use of NuvaRing contraception, recent receipt of a bacterial vaccine, or concurrent participation in other trials involving blinded administration of antibiotics or topical microbicides. Participants self-collected daily vaginal swabs longitudinally over 10 weeks. Cumulative monetary compensation totalled US$440 per participant. Isolates and MAGs obtained from participants enroled via this protocol were derived from 43 total study participants, 11 of whom contributed both isolates and MAGs, 12 contributed isolates only and 20 contributed MAGs only.

Additional previously unreported US-derived genomes of 111 non-*iners* *Lactobacillus* isolates (Supplementary Tables 5, 6, 8, 9 and 11) were generated from samples collected as part of the Vaginal Health Project (VHP) study and the isolates were provided as a kind gift from VHP principal investigator Dr Jeanne M. Marrazzo[64]. The VHP was an IRB-approved longitudinal study of BV dynamics based in Seattle, Washington, that enroled women aged 16–30 years who reported having sex with ≥1 female partner in the previous year. Recruitment occurred through media, advertisements and community referrals. Participants provided written informed consent and attended at least quarterly study visits over a year during which mid-vaginal swabs were collected, with more frequent study visits if BV was diagnosed via a combination of Amsel criteria[65] and Nugent scoring[38,64]. Cumulative monetary compensation totalled US$100 per participant. Bacteria were subsequently isolated from cryopreserved swabs via cultivation on *Lactobacillus* MRS agar (thus *L. iners* was not isolated). For isolate genome sequencing, a subset of 6 participants was retrospectively selected from the overall VHP cohort on the basis of presence of both BV and cultivable lactobacilli during the baseline study visits, and genome sequencing was performed on 15 to 32 isolates per participant, isolation being done from samples collected at between 4 and 6 study visits per participant.

Other previously unreported US-derived genomes of 4 non-*iners* *Lactobacillus* isolates (Supplementary Tables 5, 6, 8, 9 and 11) were derived from 4 participants in the V2 cohort[66]—an IRB-approved cohort at MGH (MGH IRB Protocol 2014P001066) that enroled women presenting for gynaecological care at MGH who provided written informed consent for sample collection. The V2 cohort excluded those with HIV or those who were otherwise immunocompromised. Participants received no monetary compensation and were 24–38 years of age. Samples were collected via swabs.

MAGs were assembled from previously unreported US-based WGS data from a single participant (Supplementary Tables 5, 6, 8, 9 and 11) in an IRB-approved female reproductive tract biopsy protocol based at the Ragon Institute of MGH, MIT and Harvard in Cambridge, Massachusetts (MGH IRB Protocol 2012P001812). The protocol enrols non-pregnant, premenopausal women aged 18–50 years with normal Pap smears, normal results on routine screening laboratory tests and willingness to provide informed consent. Exclusion criteria included history of severe gynaecologic disease, symptomatic vaginal infection, treatment for vaginal infection within the prior 30 d, being <8 weeks postpartum, sexual intercourse, tampon use or application of vaginal medicines within the previous 2 d, use of immunomodulatory therapy, previous receipt of an HIV vaccine, history of significant medical illness, or history of requiring antibiotics prophylaxis for invasive procedures. Samples collected included cervical and mid-vaginal swabs, with optional consent for cervical and vaginal biopsy. Participants were compensated US$150 if no biopsy was performed and US$250 if biopsy was performed. The participant for whom MAGs were assembled from WGS data was 26 years old and HIV-uninfected.

Additional previously unreported MAGs (Supplementary Tables 5, 6, 8 and 9–11) were assembled from previously reported FGT WGS data from three additional US cohorts (95 total participants; National Center for Biotechnology Information (NCBI) BioProject accession numbers PRJNA48479 and PRJNA288562)[67–70], one Italian cohort (10 participants; NCBI BioProject accession number PRJNA352475)[71] and one Chinese cohort (4 participants; NCBI BioProject accession number PRJEB24147)[72].

**Bacterial culture conditions and additives for bacterial growth media.** Bacterial cultivation was performed at 35–37 °C under anaerobic conditions using an AS-580 anaerobic chamber (Anaerobe Systems) in an atmosphere of 5% carbon dioxide, 5% hydrogen and 90% nitrogen (Airgas). All media and other culture reagents were pre-reduced (deoxygenated) by placement overnight in the anaerobic chamber before use. Nutritional additives or inhibitors (reagents listed in Supplementary Table 1; additives prepared as in Supplementary Table 2) were added to autoclave-sterilized broth media after cooling to room temperature (broth media), then broth was re-sterilized by passage through a 0.22 µm vacuum filter before being transferred to the anaerobic chamber for pre-reduction (deoxygenation)

before use in experiments. For testing of nutrient pools (Extended Data Fig. 1), a single concentrated stock comprising all pool constituents was added to media as per Supplementary Table 2, rather than stocks of each additive being added separately. In contrast to other additives, stocks of $H_2O_2$ and MTZ were added to pre-reduced broth media as freshly prepared, filter-sterilized stock solutions immediately before inoculation of bacterial cultures due to potential volatility or chemical instability. For preparation of solid media, filter-sterilized additives were added to autoclave-sterilized agar media after cooling to 50 °C, mixed using a magnetic stir/hot plate, then used to pour agar plates.

**Bacterial broth culture media.** Growth characteristics and nutritional requirements distinguishing *L. iners* from other *Lactobacillus* species were investigated using *Lactobacillus* MRS broth. We tested two different commercial MRS formulations (BD Difco and Hardy Criterion). The BD Difco-formulated media is referred to as 'MRS' while the Hardy Criterion-formulated media is referred to as 'HMRS' throughout this paper. The broth was prepared by autoclaving according to manufacturer instructions and allowed to cool to room temperature. We confirmed that *L. crispatus*, *L. jensenii* and *L. gasseri* grew rapidly and robustly in both formulations, whereas *L. iners* failed to grow even after >10 d in HMRS and grew only after a prolonged delay of ~48–72 h in BD Difco-formulated MRS. Supplementation of MRS broth (BD Difco) with IsoVitaleX (2% v/v), L-Cys or L-Cys-containing mixtures (augmented by L-Gln) enabled robust growth of *L. iners* by 20–36 h, depending on the experimental strain (Fig. 1a,c and Extended Data Fig. 1a–e), while supplementation of HMRS broth had similar pro-growth effects after a more prolonged lag phase (Extended Data Fig. 1f). Unless otherwise indicated, figures depict growth in MRS (BD Difco) broth; key growth results were reproduced where shown using HMRS broth. MRS broth supplemented with L-Gln (1.1 mM) is abbreviated 'MRSQ' (or 'HMRSQ' for L-Gln-supplemented HMRS broth) in figures and text.

NYCIII broth (American Type Culture Collection (ATCC) medium 1685) was prepared using a slightly modified version of the standard ATCC protocol. Pre-media consisted of 4 g $l^{-1}$ HEPES, 15 g $l^{-1}$ proteose peptone no. 3 and 5 g $l^{-1}$ sodium chloride in 875 ml distilled water, was pH-adjusted to 7.3 and autoclaved. Before use, complete NYCIII broth was prepared from autoclaved cooled pre-media by adding dextrose (3 g per 45 ml) at 7.5% v/v, yeast extract solution at 2.5% v/v and heat-inactivated horse serum at 10% v/v, then sterilized by passage through a 0.22 µm vacuum filter. Where indicated, NYCIII broth was supplemented with IsoVitaleX (2% v/v), vitamin $K_1$-Hemin solution (5% v/v) and/or Tween-80 (1 g $l^{-1}$); volumes of distilled water in the pre-media were decreased accordingly to ensure all other components were present at standard concentrations in the final solution.

S-broth pre-media consisting of 37 g $l^{-1}$ brain heart infusion broth, 10 g $l^{-1}$ yeast extract (powder) and 1 g $l^{-1}$ dextrose was brought to a boil in 880 ml distilled water, then autoclaved (121 °C for 15 min) and cooled, followed by addition of 5% v/v foetal bovine serum, 5% v/v vitamin $K_1$-Hemin solution and 2% v/v IsoVitaleX enrichment to a final volume of 1 l. The complete broth was then sterilized by passage through a 0.22 µm vacuum filter. Unless otherwise indicated, S-broth was supplemented with 1 g $l^{-1}$ Tween-80.

Except where otherwise indicated, lactobacilli were propagated in MRSQ broth with L-Cys (4 mM), *Prevotella* species were propagated in Wilkins-Chalgren anaerobe broth (Thermo Scientific), and other species including *G. vaginalis* were propagated in NYCIII broth.

**Bacterial solid culture media.** Non-*iners Lactobacillus* isolates were maintained and quantitatively titred on *Lactobacillus* MRS agar plates (purchased as prepared media from Hardy Diagnostics). Most other species, including *L. iners*, were maintained and titred on Columbia blood agar (CBA, purchased as prepared media from Hardy Diagnostics). *Prevotella* species were maintained on CBA agar or CDC anaerobe laked sheep blood agar with Kanamycin and Vancomycin (LKV, purchased as prepared media from BD BBL). For bacterial isolations from primary clinical samples, we used commercial CBA and LKV agar plates and prepared in-house MRS agar plates by autoclaving *Lactobacillus* MRS Broth (BD DIFCO) in Noble agar (BD DIFCO, 2% w/v), cooling to 50 °C with agitation, then supplementing with IsoVitaleX enrichment (BD BBL, 2% v/v) with or without vitamin $K_1$-Hemin solution (BD BBL, 5% v/v), or with freshly filter-sterilized stocks of L-Cys and L-Gln (at final concentrations of 4 mM and 1.1 mM, respectively). All solid media was stored at 4 °C in the dark after purchase or preparation until being introduced to the anaerobic chamber for deoxygenation and use.

**Bacterial isolate sources and isolation protocols.** Experiments were performed using a combination of previously reported bacterial strains ($n = 9$ strains), previously unreported South African primary isolates from cervicovaginal samples from the FRESH cohort ($n = 20$ strains) and previously unreported US strains from cervicovaginal samples ($n = 4$ strains; See Supplementary Table 3 and additional details below). Isolation of South African strains from FRESH cohort samples was performed at the Ragon Institute. Cervicovaginal swabs were thawed in the anaerobic chamber. Swabs that had been frozen dry (without cryopreservatives) were immersed in 500 ul of room temperature pre-reduced Dulbecco's phosphate-buffered saline (PBS). Swabs that had been frozen in cryopreservative solution were rapidly thawed directly without additional dilution. Thawed swabs were agitated vigorously within the solution for 30 s using sterile forceps to dislodge bacteria, then removed from the cryovial using a sterile micropipette tip to strain excess fluid from the swab. The solution was then serially diluted in 10-fold dilutions in sterile pre-reduced PBS and dilutions were plated on *Lactobacillus* MRS agar plates (with or without nutritional additives), CBA, and LKV agar. After 2–5 d incubation (depending on growth rate), colonies were selected on the basis of distinct morphology, and individual representative colonies were subcultured onto reduced agar plates, then further clonally subcultured into broth media. Aliquots from the purified broth cultures were frozen at −80 °C in 20% glycerol, with additional aliquots reserved for nucleic acid extraction and sequence-based taxonomic identification as described below. *L. iners* colonies were preferentially identified for isolation from primary swab cultures on supplemented MRS or CBA plates on the basis of characteristic morphology after 3–4 d incubation (circular, white-translucent colonies between 0.5 and 2 mm in diameter with entire vs slightly irregular edges, smooth surfaces, and convex/umbonate contours that were non-haemolytic on CBA after ~4 d). In experienced hands (S.M.B. and N.A.M.), morphology-based identification of *L. iners* colonies for isolation from CBA plates resulted in equally high or higher rates of *L. iners* recovery from clinical samples compared with isolation using plates with supplemented MRS media (which had lower c.f.u. yield than CBA), enabling us to perform targeted isolations of *L. iners* from >80% of attempted clinical samples, including BV-associated anaerobe-dominated samples known to have <5% *L. iners* relative abundance on the basis of 16S rRNA gene sequencing. Isolated bacteria were identified on the basis of Illumina-based sequencing of genomic DNA and/or Sanger sequencing of the full-length bacterial 16S rRNA gene as described below. BV status of the samples from which isolates were derived was determined on the basis of Nugent scoring when available (Supplementary Table 3).

The following 4 experimental strains were obtained through BEI Resources, National Institute of Allergy and Infectious Diseases (NIAID), NIH as part of the Human Microbiome Project: *L. iners*, strain LEAF 2053A-b, HM-705; *L. iners*, strain LEAF 3008A-a, HM-708; *L. iners*, strain UPII 143-D, HM-126; *L. crispatus*, strain JV-V01, HM-103. Other strains (4 in total) were obtained from the ATCC or the Culture Collection University of Gothenburg culture repositories as indicated (Supplementary Table 3). Genomes corresponding to these strains were obtained from RefSeq. Information about the BV status of the samples from which these isolates and other isolate genomes were derived was obtained (when available) from metadata accompanying the isolate genome entries in RefSeq, the Genomes OnLine Database (GOLD; https://gold.jgi.doe.gov/) or from associated information from the reference strain repositories from which they were obtained. One previously reported US-based experimental strain of *L. crispatus*[1] was genome-sequenced for this project using the methods described below. Additional previously unreported US-based strains used experimentally in this study were provided by Jacques Ravel and the VMRC (4 strains; strain details are reported in Supplementary Table 3). Unless otherwise indicated, all growth, growth inhibition, isotopic tracing and mock community experiments employed *L. crispatus* strain 233, *L. iners* strain F1 and *G. vaginalis* strain ATCC 14018. Reference bacterial strains used experimentally in this study are available from the indicated culture repositories (see Supplementary Table 3) and previously unreported strains used in these experiments are available upon reasonable request to the corresponding author.

**Nucleic acid extraction.** Total nucleic acids from cervicovaginal swab samples were extracted using the phenol–chloroform method, which includes a bead beating process to disrupt bacteria as previously described[4,60]. Swabs were thawed on ice, transferred into a solution consisting of phenol:chloroform:isoamyl alcohol (PCI, 25:24:1, pH 7.9, Ambion) and 20% sodium dodecyl sulfate (Fisher) in Tris-EDTA buffer with sterile 0.1 mm glass beads (BioSpec), vigorously rubbed against the walls of the tube to dislodge microbial material, then incubated on ice for 5–10 min. Swabs were then removed, squeezing out excess fluid by pressing the swab against the side of the tube using a sterile pipette tip as it was withdrawn. A bead beater was used to homogenize samples for 2 min at 4 °C, then centrifuged at 6,800 × *g* for 3 min at 4 °C. The aqueous phase was transferred to a clean tube with equal volume of PCI solution, mixed by vortexing, centrifuged again at 16,000 × *g* for 5 min at 4 °C, and the aqueous phase was transferred to a clean tube, precipitated using 0.8 volume of −20 °C isopropanol (molecular biology grade, Sigma) with 0.08 volume (relative to initial sample) 3 M sodium acetate at pH 5.5 (Life Technologies), inverted to mix and incubated overnight at −20 °C. Samples were then centrifuged for 30 min at 21,100 × *g* at 4 °C, washed in 0.5 ml 100% ethanol (Decon) and centrifuged for 15 min at 21,100 × *g* at 4 °C. The ethanol supernatant was discarded, the pellet was allowed to dry and the sample was resuspended in 20 µl molecular-grade Tris-EDTA buffer. Genomic DNA (gDNA) from bacterial isolates or mock communities cultured in vitro was extracted using a plate-based adaptation of the above protocol that included a bead beating process and combined phenol–chloroform isolation with Qiagen QIAamp 96 DNA QIAcube HT kit (Qiagen) procedures.

**Bacterial full-length 16S rRNA gene PCR and Sanger sequencing.** Bacterial isolate identity was confirmed based on the bacterial 16S rRNA gene sequence. The near-full-length bacterial 16S rRNA gene was PCR-amplified

from gDNA of individual bacterial isolates using the broad-range primers Bact-8F (5'-AGAGTTTGATCCTGGCTCAG-3') and Bact-1510R (5'-CGGTTACCTTGTTACGACTT-3'; Integrated DNA Technologies). PCR amplicons were confirmed via agarose gel electrophoresis, then purified and directly Sanger-sequenced on an ABI3730XL DNA analyser at the MGH Center for Computational and Integrative Biology DNA Core using the Bact-8F and Bact-1510R PCR primers as sequencing primers in separate forward and reverse sequencing reactions. Isolate identities were determined by BLAST searches against the NCBI nucleotide collection (nt) database and Bacterial 16S Ribosomal RNA RefSeq Targeted Loci Project database (NCBI accession PRJNA33175). Isolate identities were further confirmed in most cases by bacterial whole-genome sequencing as detailed below.

**Shotgun library preparation.** Shotgun sequencing libraries for cultured bacterial isolate genomes or culture-independent swab samples were prepared following the modified protocol of Baym et al.[73] using the Nextera DNA Library preparation kit (Illumina) and KAPA HiFi Library amplification kit (Kapa Biosystems). Briefly, the DNA concentration of each sample was standardized to 0.6 ng μl$^{-1}$ after quantification with SYBR Green I. Simultaneous fragmentation and sequencing adaptor incorporation was performed by mixing 0.6 ng DNA with 1.25 μl TD buffer and 0.25 μl TDE1 (Tagment DNA Enzyme, Nextera) and incubating for 10 min at 55 °C. Tagmented DNA fragments were PCR-amplified using KAPA high-fidelity library amplification reagents and primers incorporating Illumina adaptor sequences and sample barcodes. Products were pooled, purified with magnetic beads and paired-end sequenced on an Illumina NextSeq with a 300-cycle kit.

**Culture-independent bacterial 16S rRNA gene amplification and Illumina MiSeq sequencing.** Bacterial microbiota composition in study participants and results of bacterial growth competition experiments were determined via Illumina-based amplicon sequencing of the V4 region of the bacterial 16S rRNA gene. The V4 region of the bacterial 16S rRNA gene was PCR-amplified following standard protocols[4,60,74]. Samples were amplified using 0.5 units of Q5 high-fidelity DNA polymerase (NEB) in 25 μl reaction with 1X Q5 reaction buffer, 0.2 mM deoxyribonucleotide triphosphate mix (Sigma), 200 pM 515F primer (5'- AATGATACGGCGACCACCGAGACGTACGTACGGTGTGCCAGCMGCCGCGGTAA-3', where the underlined sequence represents the region of complementarity to the bacterial 16S rRNA gene; IDT) and 200 pM barcoded 806R primer (5'-CAAGCAGAAGACGGCATACGAGATXXXXXXXXXXXXAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT-3', where the underlined sequence represents the region of complementarity to the bacterial 16S rRNA gene and the X characters represent the barcode position; IDT) in PCR-clean water (Invitrogen ultra pure DNase/RNase-free distilled water). A water-template negative control reaction was performed in parallel for each barcode master mix. Additional blank extraction and amplification controls performed using unique barcoded primers were also included in sequencing libraries. DNA from clinical swab samples for FGT microbiota profiling was amplified in triplicate reactions that were then combined before library pooling to minimize stochastic amplification biases. DNA from in vitro mock community experiments (for which each experimental condition had been cultured in multiple replicates) was amplified in a single reaction per replicate culture. Amplification was performed at 98 °C for 30 s, followed by 30 cycles of 98 °C for 10 s, 60 °C for 30 s, and 72 °C for 20 s, with a final 2 min extension at 72 °C. PCR products were checked via agarose gel electrophoresis in parallel with the matching water-template control reactions to confirm successful target amplification and absence of background amplification. Gel band strength was used to semi-quantitatively estimate relative amplicon concentrations for library pooling. To prepare the sequencing libraries, 3–20 μl of individual PCR products (adjusted on the basis of estimated relative amplicon concentration) were combined into 100 μl subpools and purified using an UltraClean 96 PCR cleanup kit (Qiagen). Blank extractions, water-template and (for in vitro experiments) blank media controls were included in the sequencing libraries, although they did not produce visible PCR bands. Concentrations of the subpools were quantified using a Nanodrop (Thermo Scientific), then pooled at equal molar concentrations to assemble the final library. The pooled library was diluted and supplemented with 10% PhiX according to standard Illumina protocols, then single-end sequenced on an Illumina MiSeq using a v2 300-cycle sequencing kit with addition of custom Earth Microbiome Project sequencing primers (read 1 sequencing primer: 5'-ACGTACGTACGGTGTGCCAGCMGCCGCGGTAA-3'; read 2 sequencing primer: 5'-ACGTACGTACCCGGACTACHVGGGTWTCTAAT-3'; index sequencing primer: 5'- ATTAGAWACCCBDGTAGTCCGGCTGACTGACT-3'; IDT)[74].

**Genome and metagenome sequence processing and assembly.** We constructed *Lactobacillus* genome catalogues from 127 reference isolate genomes, as well as from genome sequences of previously unreported bacterial strains and MAGs assembled from culture-independent WGS sequences of genital tract samples derived from the cohorts detailed above (M.R.H. et al., manuscript in preparation; see cohort descriptions above and details in Supplementary Tables 4–11). Briefly, we retrieved all reference genomes annotated as *L. crispatus*, *L. iners*, *L. jensenii*, *L. gasseri* or *L. vaginalis* that were deposited in the NCBI RefSeq database[35] as of February 2020 (Supplementary Table 7). Genomes of previously unreported FGT strains from the South African FRESH cohort (isolated as described above) and of non-*iners Lactobacillus* strains isolated from the US-based VHP and V2 cohorts were sequenced on an Illumina NextSeq at the Ragon Institute as described above. Additional US FGT *Lactobacillus* isolate genomes were provided by the VMRC from the cohorts detailed above. FGT WGS metagenomic samples from cohorts in South Africa, the US, Italy and China (see above) were used to generate MAGS. Sequence reads from isolate genomes and WGS samples were trimmed and filtered to high-quality reads. Human reads from WGS samples were removed by mapping to GRCh38 (GenBank accession GCA_000001405.15). We assembled genomes and MAGs from high-quality reads, then binned contigs and removed contamination. Bin completeness, contamination and strain heterogeneity metrics were determined, then genomes and MAGs were assigned quality scores on the basis of criteria from the Genome Standards Consortium[75]. Species genome bins (SGBs) were determined on the basis of 95% pairwise absolute nucleotide identity (ANI).

***Lactobacillus* genome taxonomy assignment.** We assigned taxonomy to SGBs with FGT *Lactobacillus* genomes on the basis of presence of reference genomes with defined NCBI taxonomy, confirming that each contained reference genomes only from a single species. Of note, for both *L. jensenii* and *L. gasseri*, publicly available reference genomes classified in RefSeq as belonging to these species actually segregate into two separate SGBs per species, indicating that *L. jensenii* and *L. gasseri* (as traditionally defined) each comprise two distinct genomic species (M.R.H. et al., manuscript in preparation)[76]. However, we analysed *L. jensenii* and *L. gasseri* as single species units in this study to correspond with prevailing paradigms in literature on 16S rRNA gene-based FGT microbiota profiling[77].

**Cervicovaginal *Lactobacillus* pan-genome construction and analysis.** We constructed a cervicovaginal *Lactobacillus* pan-genome using genes from catalogues of *L. iners*, *L. crispatus*, *L. jensenii*, *L. gasseri* and *L. vaginalis* genomes and MAGs. To maximize comprehensiveness of the pan-genome, we included genomes and MAGs classified as high-quality (>90% completeness and <5% contamination) or medium-quality (≥50% completeness and <10% contamination) assemblies on the basis of Genome Standards Consortium criteria[75]. Some of the medium-quality assemblies are classified as 'partial' by NCBI standards on the basis of size criteria. Genes were identified within individual genomes by Prokka v.1.14.5[78], which predicts genes using Prodigal v.2.6.3[79]. The resulting genes were clustered into a comprehensive pan-genome at 95% nucleotide identity using Roary v.3.13.0[80] to generate a multi-fasta file containing gene sequences and a per-genome gene presence–absence table[81]. The pan-genome was annotated with eggNOG 5.0 using eggNOG-mapper v.2[82,83]. We used custom R scripts to parse the eggNOG output for genes predicted to encode enzymatic or transporter activities of interest on the basis of gene names, Kyoto Encyclopedia of Genes and Genomes (KEGG) numbers, KEGG reaction numbers, Clusters of Orthologous Groups (COGs), Enzyme Commission numbers, Transporter Classification (TC) numbers and Gene Ontology (GO) terms, followed by manual curation of the initial search results and BLAST-based sequence review for genes with unclear annotations[81]. We then determined the number of genomes from each species that were predicted to encode each gene function of interest. Since the Genome Standards Consortium criteria for high- and medium-quality genomes and MAGs allow for low-level sequence contamination within assemblies[75], we restricted gene presence–absence analysis to gene sequences of interest that were detected in at least two genomes or MAGs from a species in order to exclude singleton contaminant sequences from our analysis. Importantly, including partial genomes and MAGs with completeness as low as 50% maximizes genome catalogue diversity, thus increasing pan-genome size and sensitivity for detecting genes of interest within each species, but also results in a fraction of genomes and MAGs appearing to lack universally present genes due to incompleteness of assemblies. The absence of intact Cys synthase and reverse transsulfuration pathways in these species determined by EggNOG-based analysis (Fig. 2b) was further supported by results of BLAST[84] searches against the genome collections using gene sequences of interest from related species (see custom code posted on Zenodo[81]).

**Maximum likelihood phylogenetic distances and phylogenetic reconstruction.** Phylogenetic reconstruction of *L. iners* genomes and MAGs (Fig. 1b) was performed using assemblies with >60% completeness and <5% contamination to maximize robustness of the analysis. fetchMG v.1.0[85] was used to extract DNA sequences for each of 40 single-copy universal bacterial marker genes from each genome (called using Prodigal[79]). Genome assemblies containing fewer than 10 of the 40 universal marker genes were omitted from the subsequent alignment. A phylogenetic reconstruction was then produced using ETE3 v.3.1.1 (parameters: 'ete3 build -w clustalo_default-trimal-gappyout-none-none -m cog_85-alg_concat_default-fasttree_default')[86], and the tree was visualized using iTOL v.4[87]. An additional phylogenetic reconstruction encompassing other major FGT *Lactobacillus* species (Extended Data Fig. 2a) was performed using *L. iners* genomes and MAGs, as well as isolate genomes from *L. crispatus*, *L. jensenii*, *L. gasseri* and *L. vaginalis*, filtered on the basis of the same quality and completeness criteria.

**Preparation of bacterial inocula for growth experiments.** For initial experiments identifying cysteine and cystine as key nutrient requirements for *L. iners* growth (Figs. 1a,c and 4a, and Extended Data Fig. 1), bacteria from frozen stocks were plated on solid media, incubated for 3 d, then suspended in sterile pre-reduced

PBS, adjusted to an $OD_{600}$ of 0.3 ± 0.05, and then inoculated into the indicated broth media formulations for measurement of growth kinetics. In subsequent growth experiments for *Lactobacillus* species, experimental bacterial inocula were prepared from liquid starter cultures in MRSQ broth with L-Cys (4 mM) that were incubated for 18–20 h. The starter cultures were then pelleted by centrifugation for 10 min at 3,716 × *g*, spent media was decanted, and bacteria were washed 2 times by re-resuspending in sterile pre-reduced PBS followed by centrifugation to avoid carryover of nutrients from the original starter culture media. Washed bacteria were resuspended in PBS, adjusted to an $OD_{600}$ of 0.3 ± 0.05, then inoculated into experimental media at 3.5% (v/v), equating to bacterial titre ranging from ~$1 \times 10^5$ to $1 \times 10^6$ colony-forming units (c.f.u.s) per ml, depending on the experimental species and strain.

**Growth kinetics quantification.** We found that *Lactobacillus* growth kinetics in broth media were adversely affected by either continuous or intermittent agitation, therefore we grew broth cultures without agitation. Monoculture growth kinetics were assessed using separate cultures prepared in parallel for each experimental timepoint to avoid repeatedly agitating a single culture by performing serial measurements. For timepoint series lasting ≤48 h, broth cultures were grown in a volume of 250 µl in technical triplicate in clear 96-well flat-bottom plates (Falcon), with a blank for each media condition per plate. Measurements were taken at ~16, 20, 24, 28, ~42 and 48 h by removing plates from the anaerobic chamber, pipetting to resuspend bacteria, and measuring $OD_{600}$ on a Tecan Infinite M1000 PRO plate reader. Due to slower growth kinetics in the HMRS broth formulation, *L. iners* growth measurements in HMRS were taken daily for up to 10 d, precluding use of 96-well plates due to evaporation at later timepoints. Bacteria were therefore incubated in low-evaporation 1.2 ml 96-well cluster tubes (Corning) in triplicate, with separate parallel culture plates used for each experimental timepoint, then 250 µl from each culture was transferred to a clear 96-well flat-bottom Falcon plate for optical density measurement. Data from monoculture growth experiments are depicted as median ± range for the 3 replicates. For each bacterial strain and media condition, figures depict representative results from 1 of ≥2 independent experiments with distinct batches of freshly prepared media and freshly prepared bacterial input inocula, unless otherwise indicated. Growth data collection and analysis were not performed blind to the conditions of the experiments.

**Inhibitor experiments and analysis.** To test monoculture growth inhibition, bacteria were cultured in media containing inhibitors at varying concentrations as indicated, including a reference (no-inhibitor) control. Since growth kinetics differed between species and strains, inhibition was determined for each strain at the first timepoint, fulfilling the European Commission on Antimicrobial Susceptibility Testing (EUCAST) criterion of 'definite turbidity' in the reference control (which we defined experimentally as $OD_{600} > 0.2$), unless otherwise specified. At the selected timepoint for each strain, the median $OD_{600}$ value for the reference control was set to a reference value of 100%, which was used to calculate percentage growth in the other conditions. Growth experiments employing MTZ used a concentration of 50 µg ml$^{-1}$, approximating the concentration in 0.75% intravaginal MTZ gel (a first-line treatment for BV)[25]. Data from inhibition experiments are depicted as median ± range for 3 technical replicates. For each bacterial strain and media condition, figures depict representative results from 1 of ≥2 independent experiments with distinct batches of freshly prepared media and freshly prepared bacterial input inocula, unless otherwise indicated. Inhibition data collection and analysis were not performed blind to the conditions of the experiments.

**Competition and mock community culture experiments.** For pairwise competition between *L. iners* and *L. crispatus* strains in MRS broth ± SMC and for mock BV-like community experiments with *L. iners* (multiple strains), *L. crispatus* (multiple strains), *G. vaginalis* (ATCC 14018), *P. bivia* (ATCC 29303) and *A. vaginae* (0795_578_1_1_BHK4; see Supplementary Table 3) in S-broth ± SMC and/or MTZ, bacteria were initially prepared in individual suspensions in broth as described above for bacterial monoculture growth and inhibition experiments. Aliquots of the monobacterial suspensions were mixed in defined ratios, then mixtures were divided into replicate cultures for incubation (5 replicates per condition for MRSQ experiments; 6 replicates per condition for S-broth experiments) by adding 150 µl of mixture per well into V-bottom 96-well plates (Falcon). C.f.u. titres were determined for each input monobacterial suspension as described above and used to calculate starting ratios within the mixed cultures. At 28 h, cultures were harvested by centrifuging at 4,700 × *g* for 25 min at 4 °C, spent media was decanted, and pellets were frozen for later DNA extraction and analysis. Relative growth within mixed cultures was assessed by bacterial 16S rRNA gene sequencing as described below. Aliquots of monobacterial suspensions in the corresponding media type were cultured separately to confirm expected growth patterns (for example, Fig. 5b and Extended Data Fig. 7a).

**Processing of Illumina MiSeq 16S rRNA gene sequences.** Demultiplexing of Illumina MiSeq bacterial 16S rRNA gene sequence data was performed using QIIME 1 version 1.9.1[88]. Mapping files in standard QIIME 1 format were created and validated using validate_mapping_file.py, sequences were demultiplexed using split_libraries_fastq.py with parameter store_demultiplexed_fastq and no quality filtering or trimming, and demultiplexed sequences were organized into individual fastq files using split_sequence_file_on_sample_ids.py[60]. We then used dada2 version 1.6.0[89] in R to filter and trim reads at positions 10 (left) and 230 (right) using the filterAndTrim function with parameters truncQ = 11, MaxEE = 2 and MaxN = 0. Sequences were then inferred and initial taxonomy assigned using the dada2 assignTaxonomy function, employing the Ribosomal Database Project training database rdp_train_set_16.fa.gz (https://www.mothur.org/wiki/RDP_reference_files). Taxonomic assignments were refined and extended via extensive manual review (see Supplementary Table 15 for amplicon sequence variant (ASV) taxonomy). The denoised dada2 results with final taxonomic assignment were analysed in R using phyloseq version 1.30.0[90] and custom R scripts. Initial sequencing-based analysis of FGT microbiota composition of FRESH cohort participants, taxonomy assignment and cervicotype assignments were performed blinded to participants' corresponding Nugent scores and metabolite concentrations in CVL fluid.

**16S rRNA gene-based microbiome analysis.** For 16S rRNA gene-based microbiome profiling of clinical specimens, microbial communities were classified into 4 cervicotypes (CTs) as previously defined in a non-overlapping subset of participants from the FRESH cohort: CT1 includes communities with >50% relative abundance of non-*iners Lactobacillus* species (which consists almost entirely of *L. crispatus* in this population[4]); CT2 consists of communities in which *L. iners* is the most dominant taxon; CT3 consists of communities in which the genus *Gardnerella* is the most dominant taxon; and CT4 consists of communities dominated by other species, typically featuring high abundance of one or more *Prevotella* species[1]. Although *L. jensenii* and *L. gasseri* were detected in microbial communities of some FRESH study participants by both sequencing and cultivation, they were not dominant in any individuals, a common finding among studies of sub-Saharan African populations[1,4,6,50] that contrasts with observations in most North American and European cohorts, where a consistent minority of women tend to have *L. gasseri*-dominated or *L. jensenii*-dominated communities[12,49,59,77]. For further sequence processing and analysis, ASVs that could not be defined at least to the level of taxonomic class were pruned from the dataset and one sample was excluded due to a sequence read count <10,000. The remaining 142 samples were rarefied without replacement to a uniform depth of 16,603 reads (the minimum read count among remaining samples), yielding a relative abundance limit of detection (L.D.) of $6.02 \times 10^{-5}$. ASVs were collapsed at the species or genus level as indicated for further visualization and statistical analysis. In plots displaying taxon relative abundance using a logarithmic scale, a pseudocount equal to 0.5 × L.D. was added to assist visualization of any taxon with read count 0. Correlation analysis between bacterial relative abundances and cervicovaginal metabolite concentrations was restricted to taxa detected (≥1 sequence read after rarefaction) in ≥50% of samples (Extended Data Fig. 3c,d).

**Sequence analysis for in vitro competition experiments and mock communities.** For pairwise competitions between *L. crispatus* and *L. iners*, 16S rRNA gene sequences were generated, processed and annotated as described above. Ratios of *L. iners* reads to *L. crispatus* reads in each sample were calculated. Significance of between-group differences for each mixture was determined by one-way analysis of variance (ANOVA). Significance of pairwise differences between the positive control condition (MRSQ + L-Cys without inhibitor) and each experimental condition was calculated using Dunnett's test and all significant pairwise comparisons were plotted (see Supplementary Table 18 for numeric *P* values). Results of representative competition experiments are shown in Fig. 5a.

For mock BV-like community experiments, 16S rRNA gene sequences were generated, processed and annotated as described above. Relative abundances of each experimental strain were determined for individual replicates and displayed in Fig. 5c. To assess *L. crispatus* enrichment, read counts of all experimental strains except *L. crispatus* were summed for each replicate sample, then ratios of *L. crispatus* to the summed taxa were calculated. Significance of between-group differences for each mixture was determined by one-way ANOVA, and significance of pairwise comparisons was calculated using Tukey's test, with selected pairwise comparisons plotted in Fig. 5d and corresponding numeric *P* values reported in Supplementary Table 20. Competition and mock community assay data collection and analysis were not performed blind to the conditions of the experiments.

**Measurement and analysis of metabolites in cervicovaginal lavage fluid.** Concentrations of Cys, reduced glutathione (GSH, γ-L-glutamyl-L-cysteinyl-glycine) and cysteinylglycine (Cys-Gly) were measured in CVL supernatants from the 143 FRESH cohort participants whose microbiome composition was concurrently profiled by 16S rRNA gene sequencing in this study. CVL supernatant samples underwent a single freeze–thaw cycle during preparation of aliquots for metabolite analysis. Metabolites of interest were quantified using ultraperformance liquid chromatography/tandem mass spectrometry (UPLC–MS/MS) by Metabolon, as part of an untargeted metabolomics analysis. To avoid bias in assay performance, measurements were performed after randomized reordering of samples by laboratory staff at Metabolon, who were blinded to all participant information except sample ID. Metabolon's method generates relative instead of absolute metabolite concentrations. For analysis, concentrations were volume-normalized and median values were adjusted to 1, then percent

missingness (reflecting samples with analyte concentrations below the L.D.) was calculated for each analyte. For analytes with missingness >0%, analyte L.D. was inferred as equalling the lowest measured relative concentration in the cohort, then missing values were imputed at half the L.D. The Shapiro–Wilk test of normality was then performed on relative concentration values after log-transformation. In subsequent analysis of concentration differences between cervicotypes, differences were analysed by one-way ANOVA with post-hoc Tukey's test for Cys (which had 0% missingness and was normally distributed, Fig. 3a,c) and by Kruskal–Wallis test with post-hoc Dunn's test for GSH and Cys-Gly (which had 21.7% and 7.0% missingness, respectively, and were therefore not normally distributed after imputation, Extended Data Fig. 4a–d). We calculated Spearman rank-order correlations between individual metabolites and bacterial taxa at both the genus and species levels, adjusting *P* values for multiple comparisons where indicated, using the Bonferroni method via the R stats package function p.adjust(). The R package DescTools was used to calculate confidence intervals for the Spearman correlation coefficients ($\rho$), including Bonferroni-corrected confidence levels at $(1 - 0.05/n)$, where $n$ represents the number of taxa (Fig. 3e,f, Extended Data Fig. 4e,f, and Supplementary Tables 16 and 17).

**Quantification of oxidized and reduced cysteine and glutathione in MRS broth.** Oxidized and reduced glutathione and cysteine were quantified in MRS broth by UPLC–MS/MS (Waters Acquity/TQ-S), according to a modification of the protocol by Sutton et al.[91]. Briefly, 100 µl anaerobic broth was mixed with 90 µl buffer (100 mM ammonium bicarbonate, pH 7.4), followed by 10 µl fresh *N*-ethylmaleimide solution (25 mg ml$^{-1}$ in ethanol), and allowed to react at room temperature for 10 min. Samples were then diluted 100-fold in water before analysis by UPLC–MS/MS, with a 1 µl injection volume. Separation was carried out with an Acquity BEH/C18 UPLC column (Waters 186002350), with the following gradient: 0 min, 2% B; 6 min, 98% B; 7 min, 98% B; 7.1 min, 2% B; 9 min, 2% B. Mobile phase A was water with 0.1% formic acid, and B was acetonitrile with 0.1% formic acid. Compounds were quantified by comparison of peak area to authentic standards processed via the same method.

**Synthesis and analysis of isotopically labelled cystine, oxidized glutathione (GSSG) and GSH.** $^{13}$C-labelled cystine (1,1'-$^{13}C_2$-L-cystine) was prepared from 1-$^{13}$C-L-cysteine (Cambridge Isotope Labs, CLM-3852-0.5) via a modification of the procedure of Hill et al.[92]. Briefly, 100 mg labelled cysteine was mixed with 200 µl 10% NaOH and 800 µl 3% hydrogen peroxide. The resulting precipitate was filtered and washed extensively with ice-cold water, then finally redissolved in 1 M aqueous HCl. Complete oxidation was confirmed by UPLC–MS/MS analysis of the resulting solution, as described above.

Labelled glutathione-(1-$^{13}$C-cysteine) was prepared enzymatically using *E. coli* GshA and GshB overexpressed and purified as N-terminal 6xHis fusions. *gshA* and *gshB* were amplified from genomic DNA of *E. coli* BL21(DE3) using Q5 polymerase (New England Biolabs) and the following primers (purchased from Millipore Sigma): HIS-gshA-F, gcctggtgccgcgcggcagcATCCCGGACGTATCACAG; HIS-gshA-R, cagcttcctttcgggctttgTCAGGCGTGTTTTTCCAGCC; HIS-gshB-F, gcctggtgccgcgcggcagcATCAAGCTCGGCATCGTGAT; and HIS-gshB-R, cagcttcctttcgggctttgTTACTGCTGCTGTAAACGTGC, according to the manufacturer's instructions. The expression vector backbone pET28 was amplified from purified stock using the primers: pET28-F, GCTGCCGCGCGGCACCAG; and pET28-R, CAAAGCCCGAAAGGAAGCTG. After DpnI digest and purification, expression plasmids pETgshA and pETgshB were assembled using HiFi Assembly MasterMix (NEB) and transformed into chemically competent *E. coli* DH5a. Positive colonies were verified via Sanger sequencing, and correctly assembled plasmids were transformed into *E. coli* BL21(DE3). Cells for protein expression were inoculated from overnight culture with 100-fold dilution into 4 l flasks containing 1 l LB broth, and grown at 37 °C and 200 r.p.m. until reaching $OD_{600}$ ~ 0.5, after which they were induced with 0.5 mM isopropyl ß-D-1-thiogalactopyranoside and grown for a further 3 h at 37 °C. Pellets were harvested by centrifugation at 6,000 × *g* for 10 min, then resuspended in ice-cold 98% buffer A (50 mM HEPES, 300 mM KCl, 10% glycerol, pH 7.5) and 2% buffer B (buffer A with 500 mM imidazole). Cells were lysed by 3 passages through an Emulsiflex C5 (Avestin) at 15,000 psi, and lysates were clarified by centrifugation at 20,000 × *g* and 4 °C for 30 min. Target proteins were bound to 1 ml Ni-NTA His-Bind resin (Millipore Sigma) in gravity columns, then washed with 5 ml wash buffer (90% buffer A, 10% buffer B) and eluted with 3 ml elution buffer (100% buffer B). Fractions containing purified proteins were identified via SDS–PAGE, then combined and dialysed overnight against buffer A with 3.5k MWCO Slide-A-Lyzer cassettes (ThermoFisher), then concentrated by centrifugation through 3k MWCO Microsep spin filters (Pall).

Glutathione (γ-L-glutamyl-L-cysteinyl-glycine) synthesis was carried out in 100 mM sodium phosphate buffer (pH 7.2), with 80 mM 1-$^{13}$C-L-cysteine, 120 mM glycine, 120 mM glutamic acid, 40 mM $MgCl_2$ and 100 mM adenosine triphosphate (ATP), using 50 mg 1-$^{13}$C-L-cysteine and 1 mg each of GshA and GshB, incubating overnight at 37 °C. The resulting glutathione-(1-$^{13}$C-cysteine) was treated with 3% hydrogen peroxide to produce GSSG for purification via preparative HPLC on a Thermo Scientific Dionex UltiMate 3000 HPLC system with a Thermo Scientific Hypersil GOLD aQ C18 preparative column (20 × 20 mm, 5 µm) using a gradient from 100% A (water + 0.1% formic acid) to 70% B (acetonitrile + 0.1% formic acid). Fractions containing labelled GSSG ($^{13}C_2$-GSSG) were pooled and lyophilised. Analysis of the purified product was conducted by $^1$H NMR (400 MHz) at the Magnetic Resonance Laboratory of the Harvard University Department of Chemistry and Chemical Biology on a Jeol J-400, as well as by LC–MS/MS as described above, and by high-resolution LC–MS using an Agilent 6530 Q-TOF (confirming expected product characteristics as detailed in the Human Metabolome Database: https://doi.org/10.13018/BMSE000906). This analysis confirmed the authenticity of the desired product and complete conversion of cysteine, and also revealed contamination with adenosine diphosphate (ADP), which was not removed by preparative HPLC (Supplementary Fig. 1). This material was used without further purification. Reduced labelled glutathione ($^{13}$C-GSH, 4 mM) for use in the experiment measuring GSH uptake was produced from the oxidized form ($^{13}C_2$-GSSG) by treating a stock solution of 81.6 mM labelled GSSG with Tris(2-carboxyethyl)phosphine at a 0.9:1 molar ratio to reduce the disulfide bonds before addition to MRSQ broth.

**Quantification and isotopic analysis of amino acids.** For isotopic analysis of proteinogenic amino acids, *L. crispatus* (strain 233) and *L. iners* (strain F1) grown in MRSQ broth supplemented with isotopically labelled or unlabelled substrates were harvested via centrifugation. The pellets were washed 3 times with water and resuspended in 0.6 ml 6 N aqueous HCl and heated at 100 °C overnight to facilitate cell lysis, protein hydrolysis and Cys oxidation. The hydrolysate was air-dried, then resuspended in 100 µl of 90% acetonitrile and 10% water. For determination of amino acid concentration in media, samples were prepared by diluting media 10–100-fold in 90:10 acetonitrile:water, followed by centrifugation to remove precipitated material. All amino acid samples were analysed via UPLC–MS/MS using instrumentation described above. Separation was carried out with an Acquity BEH/Amide UPLC column (Waters 186004800) in HILIC mode, with the following gradient: 0 min, 90% B; 0.3 min, 73% B; 1 min, 73% B; 1.5 min, 30% B; 2 min, 30% B; 2.2 min, 90% B; 4 min, 90% B. Mobile phase A was water with 0.1% formic acid, and B was acetonitrile with 0.1% formic acid. Quantification of amino acids was confirmed by comparison of peak area to authentic standards (Sigma AAS18-5ML) processed via the same method. Isotopic distribution data were corrected for the natural abundance of $^{13}$C using IsoCor v.2.1.3[93] (Supplementary Tables 13 and 14). Isotopic assay data collection and analysis were not performed blind to the conditions of the experiments.

**Statistics, software and visualization.** Data analysis, statistics and visualization were performed in R v.3.6.3, except where otherwise indicated, using packages including seqinr v.4.2.5, tidyverse, v.1.3.1, knitr v.1.33, ggpubr v.0.4.0, DescTools v.0.99.41, gtools v.3.8.2, gridExtra v.2.3, cowplot v.1.1.1, scales v.1.1.1, grid v.3.6.3, broom v.0.7.6, e1071 v.1.7.6, and table1 v.1.4. All *P* values are two-sided with statistical significance defined at $\alpha = 0.05$, unless otherwise indicated.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Compressed directories containing data files sufficient to reproduce (1) analysis of pan-genome composition and gene content, (2) analysis for human FGT microbiota–metabolite analysis and (3) analysis for competition cultures of *L. iners* and *L. crispatus* and mixed community cultures are posted at Zenodo.org under https://zenodo.org/record/5900469[81]. The dataset containing the raw Illumina MiSeq read data for genital tract bacterial 16S rRNA gene profiling analysed in this study (Fig. 3 and Extended Data Figs. 3 and 4) is available in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA729907. The dataset containing the raw Illumina MiSeq read data for the bacterial 16S rRNA gene sequences from competition cultures of *L. iners* and *L. crispatus* (Fig. 5a) and from mixed community cultures (Fig. 5c,d) is available in the NCBI SRA under BioProject PRJNA777644. The taxonomic assignments used for amplicon sequence variants (ASVs) from bacterial 16S rRNA gene sequencing are supplied in Supplementary Table 15. The *Lactobacillus* genomic catalogues included a total of 1,091 previously unreported isolate genomes, partial genomes and MAGs from multiple human cohorts, as detailed above. The assemblies are available in the NCBI SRA under BioProjects PRJNA799384, PRJNA799634, PRJNA799626, PRJNA799445, PRJNA799630, PRJNA799633, PRJNA799642, PRJNA799746, PRJNA799744, PRJNA799737, PRJNA799762 and PRJNA797778; additional details on the individual studies associated with these BioProjects are contained in Supplementary Tables 5, 6 and 10, and individual NCBI BioSample accession numbers for each of the 1,091 assemblies are listed in Supplementary Table 8. In addition, the genome catalogues included 127 previously reported isolate genomes that were retrieved from RefSeq; the individual accession numbers for these genomes are listed in Supplementary Table 7. The raw and corrected cystine and serine isotopologue measurements associated with Figs. 2c,d and 4c are available in Supplementary Tables 13 and 14. Some metadata related to previously reported isolate genomes were obtained from corresponding entries in RefSeq (https://www.ncbi.nlm.nih.gov/refseq/) or the Genomes OnLine Database (GOLD; https://gold.jgi.doe.gov/). Source data are supplied for plots and phylogenetic trees, including for Figs. 1–5 and Extended Data Figs. 1–8.

## Code availability

Compressed directories containing R analysis code sufficient to reproduce (1) analysis of pan-genome composition and gene content, (2) analysis for human FGT microbiota–metabolite analysis and (3) analysis for competition cultures of *L. iners* and *L. crispatus* and mixed community cultures are available at Zenodo.org under https://zenodo.org/record/5900469[81]. Each compressed directory contains a README file describing dependencies and other details, an R Project file, an R Markdown file containing the analysis code with additional information, and associated subdirectories used in the analysis.

Received: 16 June 2021; Accepted: 27 January 2022;
Published online: 3 March 2022

## References

1. Gosmann, C. et al. *Lactobacillus*-deficient cervicovaginal bacterial communities are associated with increased HIV acquisition in young South African women. *Immunity* **46**, 29–37 (2017).
2. McClelland, R. S. et al. Evaluation of the association between the concentrations of key vaginal bacteria and the increased risk of HIV acquisition in African women from five cohorts: a nested case-control study. *Lancet Infect. Dis.* **18**, 554–564 (2018).
3. Fettweis, J. M. et al. The vaginal microbiome and preterm birth. *Nat. Med.* **25**, 1012–1021 (2019).
4. Anahtar, M. N. et al. Cervicovaginal bacteria are a major modulator of host inflammatory responses in the female genital tract. *Immunity* **42**, 965–976 (2015).
5. Jespers, V. et al. A longitudinal analysis of the vaginal microbiota and vaginal immune mediators in women from sub-Saharan Africa. *Sci. Rep.* **7**, 11974 (2017).
6. Lennard, K. et al. Microbial composition predicts genital tract inflammation and persistent bacterial vaginosis in South African adolescent females. *Infect. Immun.* **86**, e00410–e00417 (2018).
7. Norenhag, J. et al. The vaginal microbiota, human papillomavirus and cervical dysplasia: a systematic review and network meta-analysis. *BJOG* **127**, 171–180 (2020).
8. Anahtar, M. N., Gootenberg, D. B., Mitchell, C. M. & Kwon, D. S. Cervicovaginal microbiota and reproductive health: the virtue of simplicity. *Cell Host Microbe* **23**, 159–168 (2018).
9. McKinnon, L. R. et al. The evolving facets of bacterial vaginosis: implications for HIV transmission. *AIDS Res. Hum. Retroviruses* **35**, 219–228 (2019).
10. Kenyon, C., Colebunders, R. & Crucitti, T. The global epidemiology of bacterial vaginosis: a systematic review. *Am. J. Obstet. Gynecol.* **209**, 505–523 (2013).
11. Vaneechoutte, M. *Lactobacillus iners*, the unusual suspect. *Res. Microbiol.* **168**, 826–836 (2017).
12. Kindinger, L. M. et al. The interaction between vaginal microbiota, cervical length, and vaginal progesterone treatment for preterm birth risk. *Microbiome* **5**, 6 (2017).
13. Brooks, J. P. et al. Changes in vaginal community state types reflect major shifts in the microbiome. *Microb. Ecol. Health Dis.* **28**, 1303265 (2017).
14. DiGiulio, D. B. et al. Temporal and spatial variation of the human microbiota during pregnancy. *Proc. Natl Acad. Sci. USA* **112**, 11060–11065 (2015).
15. Lambert, J. A., John, S., Sobel, J. D. & Akins, R. A. Longitudinal analysis of vaginal microbiome dynamics in women with recurrent bacterial vaginosis: recognition of the conversion process. *PLoS ONE* **8**, e82599 (2013).
16. Munoz, A. et al. Modeling the temporal dynamics of cervicovaginal microbiota identifies targets that may promote reproductive health. *Microbiome* **9**, 163 (2021).
17. Srinivasan, S. et al. Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. *PLoS ONE* **5**, e10197 (2010).
18. Petrova, M. I., Reid, G., Vaneechoutte, M. & Lebeer, S. *Lactobacillus iners*: friend or foe? *Trends Microbiol.* **25**, 182–191 (2017).
19. Schwebke, J. R. et al. Treatment of male sexual partners of women with bacterial vaginosis: a randomized, double-blind, placebo-controlled trial. *Clin. Infect. Dis.* **73**, e672–e679 (2021).
20. Bradshaw, C. S. et al. High recurrence rates of bacterial vaginosis over the course of 12 months after oral metronidazole therapy and factors associated with recurrence. *J. Infect. Dis.* **193**, 1478–1486 (2006).
21. Joag, V. et al. Impact of standard bacterial vaginosis treatment on the genital microbiota, immune milieu, and ex vivo human immunodeficiency virus susceptibility. *Clin. Infect. Dis.* **68**, 1675–1683 (2019).
22. Mitchell, C. et al. Behavioral predictors of colonization with *Lactobacillus crispatus* or *Lactobacillus jensenii* after treatment for bacterial vaginosis: a cohort study. *Infect. Dis. Obstet. Gynecol.* **2012**, 706540 (2012).
23. Ravel, J. et al. Daily temporal dynamics of vaginal microbiota before, during and after episodes of bacterial vaginosis. *Microbiome* **1**, 29 (2013).
24. Verwijs, M. C., Agaba, S. K., Darby, A. C. & van de Wijgert, J. H. H. M. Impact of oral metronidazole treatment on the vaginal microbiota and correlates of treatment failure. *Am. J. Obstet. Gynecol.* **222**, 157.e1–157.e13 (2020).
25. Cohen, C. R. et al. Randomized trial of lactin-V to prevent recurrence of bacterial vaginosis. *N. Engl. J. Med.* **382**, 1906–1915 (2020).
26. Rampersaud, R. et al. Inerolysin, a cholesterol-dependent cytolysin produced by *Lactobacillus iners*. *J. Bacteriol.* **193**, 1034–1041 (2011).
27. Antonio, M. A. D., Rabe, L. K. & Hillier, S. L. Colonization of the rectum by *Lactobacillus* species and decreased risk of bacterial vaginosis. *J. Infect. Dis.* **192**, 394–398 (2005).
28. France, M. T., Mendes-Soares, H. & Forney, L. J. Genomic comparisons of *Lactobacillus crispatus* and *Lactobacillus iners* reveal potential ecological drivers of community composition in the vagina. *Appl. Environ. Microbiol.* **82**, 7063–7073 (2016).
29. Witkin, S. S. et al. Influence of vaginal bacteria and D- and L-lactic acid isomers on vaginal extracellular matrix metalloproteinase inducer: implications for protection against upper genital tract infections. *mBio* **4**, e00460–13 (2013).
30. De Man, J. C., Rogosa, M. & Sharpe, M. E. A medium for the cultivation of lactobacilli. *J. Appl. Bacteriol.* **23**, 130–135 (1960).
31. Falsen, E., Pascual, C., Sjödén, B., Ohlén, M. & Collins, M. D. Phenotypic and phylogenetic characterization of a novel *Lactobacillus* species from human sources: description of *Lactobacillus iners* sp. nov. *Int. J. Syst. Bacteriol.* **49**, 217–221 (1999).
32. Damelin, L. H. et al. Identification of predominant culturable vaginal *Lactobacillus* species and associated bacteriophages from women with and without vaginal discharge syndrome in South Africa. *J. Med. Microbiol.* **60**, 180–183 (2011).
33. Manhanzva, M. T. et al. Inflammatory and antimicrobial properties differ between vaginal *Lactobacillus* isolates from South African women with non-optimal versus optimal microbiota. *Sci. Rep.* **10**, 6196 (2020).
34. Matsumoto, A. et al. Characterization of the vaginal microbiota of Japanese women. *Anaerobe* **54**, 172–177 (2018).
35. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
36. Guédon, E. & Martin-Verstraete, I. in *Amino Acid Biosynthesis ~ Pathways, Regulation and Metabolic Engineering* (ed. Wendisch, V. F.) 195–218 (Springer, 2006); https://doi.org/10.1007/7171_2006_060
37. Dong, K. L. et al. Detection and treatment of Fiebig stage I HIV-1 infection in young at-risk women in South Africa: a prospective cohort study. *Lancet HIV* **5**, e35–e44 (2018).
38. Nugent, R. P., Krohn, M. A. & Hillier, S. L. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. *J. Clin. Microbiol.* **29**, 297–301 (1991).
39. McIver, C. J. & Tapsall, J. W. Cysteine requirements of naturally occurring cysteine auxotrophs of *Escherichia coli*. *Pathology* **19**, 361–363 (1987).
40. Proust, L. et al. Insights into the complexity of yeast extract peptides and their utilization by *Streptococcus thermophilus*. *Front. Microbiol.* **10**, 906 (2019).
41. Burguière, P., Auger, S., Hullo, M.-F., Danchin, A. & Martin-Verstraete, I. Three different systems participate in L-cystine uptake in *Bacillus subtilis*. *J. Bacteriol.* **186**, 4875–4884 (2004).
42. Müller, A. et al. An ATP-binding cassette-type cysteine transporter in *Campylobacter jejuni* inferred from the structure of an extracytoplasmic solute receptor protein. *Mol. Microbiol.* **57**, 143–155 (2005).
43. Zhou, Y. & Imlay, J. A. *Escherichia coli* k-12 lacks a high-affinity assimilatory cysteine importer. *mBio* **11**, e01073–20 (2020).
44. Poole, R. K., Cozens, A. G. & Shepherd, M. The CydDC family of transporters. *Res. Microbiol.* **170**, 407–416 (2019).
45. Pophaly, S. D., Singh, R., Pophaly, S. D., Kaushik, J. K. & Tomar, S. K. Current status and emerging role of glutathione in food grade lactic acid bacteria. *Microb. Cell Fact.* **11**, 114 (2012).
46. Suzuki, H., Koyanagi, T., Izuka, S., Onishi, A. & Kumagai, H. The yliA, -B, -C, and -D genes of *Escherichia coli* K-12 encode a novel glutathione importer with an ATP-binding cassette. *J. Bacteriol.* **187**, 5861–5867 (2005).
47. Rampersaud, R. *Identification and Characterization of Inerolysin, the Cholesterol Dependent Cytolysin produced by Lactobacillus iners* (Columbia University, 2014).
48. Klatt, N. R. et al. Vaginal bacteria modify HIV tenofovir microbicide efficacy in African women. *Science* **356**, 938–945 (2017).
49. Borgdorff, H. et al. The association between ethnicity and vaginal microbiota composition in Amsterdam, the Netherlands. *PLoS ONE* **12**, e0181135 (2017).
50. Borgdorff, H. et al. *Lactobacillus*-dominated cervicovaginal microbiota associated with reduced HIV/STI prevalence and genital HIV viral load in African women. *ISME J.* **8**, 1781–1793 (2014).
51. Chen, Y. et al. Association between the vaginal microbiome and high-risk human papillomavirus infection in pregnant Chinese women. *BMC Infect. Dis.* **19**, 677 (2019).

52. Marconi, C. et al. Characterization of the vaginal microbiome in women of reproductive age from 5 regions in Brazil. *Sex. Transm. Dis.* **47**, 562–569 (2020).
53. Plummer, E. L. et al. Sexual practices have a significant impact on the vaginal microbiota of women who have sex with women. *Sci. Rep.* **9**, 19749 (2019).
54. Vargas-Robles, D. et al. Changes in the vaginal microbiota across a gradient of urbanization. *Sci. Rep.* **10**, 12487 (2020).
55. Mendes-Soares, H., Suzuki, H., Hickey, R. J. & Forney, L. J. Comparative functional genomics of *Lactobacillus* spp. reveals possible mechanisms for specialization of vaginal lactobacilli to their environment. *J. Bacteriol.* **196**, 1458–1470 (2014).
56. Srinivasan, S. et al. Metabolic signatures of bacterial vaginosis. *mBio* **6**, e00204–e00215 (2015).
57. Castellano, F., Correale, J. & Molinier-Frenkel, V. Editorial: immunosuppressive amino acid catabolizing enzymes in heallth and disease. *Front. Immunol.* **12**, 689864 (2021).
58. Lev-Sagie, A. et al. Vaginal microbiome transplantation in women with intractable bacterial vaginosis. *Nat. Med.* **25**, 1500–1504 (2019).
59. Fettweis, J. M. et al. Differences in vaginal microbiome in African American women versus women of European ancestry. *Microbiology* **160**, 2272–2282 (2014).
60. Hoang, T. et al. The cervicovaginal mucus barrier to HIV-1 is diminished in bacterial vaginosis. *PLoS Pathog.* **16**, e1008236 (2020).
61. Vitali, B. et al. Vaginal microbiome and metabolome highlight specific signatures of bacterial vaginosis. *Eur. J. Clin. Microbiol. Infect. Dis.* **34**, 2367–2376 (2015).
62. McMillan, A. et al. A multi-platform metabolomics approach identifies highly specific biomarkers of bacterial diversity in the vagina of pregnant and non-pregnant women. *Sci. Rep.* **5**, 14174 (2015).
63. Callahan, B. J. et al. Replication and refinement of a vaginal microbial signature of preterm birth in two racially distinct cohorts of US women. *Proc. Natl Acad. Sci. USA* **114**, 9966–9971 (2017).
64. Marrazzo, J. M., Thomas, K. K., Fiedler, T. L., Ringwood, K. & Fredricks, D. N. Relationship of specific vaginal bacteria and bacterial vaginosis treatment failure in women who have sex with women: a cohort study. *Ann. Intern. Med.* **149**, 20 (2008).
65. Amsel, R. et al. Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations. *Am. J. Med.* **74**, 14–22 (1983).
66. Mitchell, C. M. et al. Vaginal microbiota and mucosal immune markers in women with vulvovaginal discomfort. *Sex. Transm. Dis.* **47**, 269–274 (2020).
67. Methé, B. A. et al. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
68. Consortium, H. M. P. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
69. Goltsman, D. S. A. et al. Metagenomic analysis with strain-level resolution reveals fine-scale variation in the human pregnancy microbiome. *Genome Res.* **28**, 1467–1480 (2018).
70. Hudson, P. L. et al. Comparison of the vaginal microbiota in postmenopausal black and white women. *J. Infect. Dis*. https://doi.org/10.1093/infdis/jiaa780 (2020).
71. Ferretti, P. et al. Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. *Cell Host Microbe* **24**, 133–145.e5 (2018).
72. Li, F. et al. The metagenome of the female upper reproductive tract. *Gigascience* **7**, giy107 (2018).
73. Baym, M. et al. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS ONE* **10**, e0128036 (2015).
74. Caporaso, J. G. et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl Acad. Sci. USA* **108**, 4516–4522 (2011).
75. Bowers, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
76. Rocha, J. et al. *Lactobacillus mulieris* sp. nov., a new species of *Lactobacillus delbrueckii* group. *Int. J. Syst. Evol. Microbiol.* **70**, 1522–1527 (2020).
77. France, M. T. et al. VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition. *Microbiome* **8**, 166 (2020).
78. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
79. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
80. Page, A. J. et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
81. Bloom, S. M. & Kwon, D. S. *Cysteine Dependence of Lactobacillus iners is a Potential Therapeutic Target for Vaginal Microbiota Modulation* (Zenodo, 2021); https://doi.org/10.5281/ZENODO.5900469
82. Huerta-Cepas, J. et al. EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
83. Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
84. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
85. Milanese, A. et al. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat. Commun.* **10**, 1014 (2019).
86. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
87. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
88. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
89. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
90. McMurdie, P. J. & Holmes, S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
91. Sutton, T. R. et al. A robust and versatile mass spectrometry platform for comprehensive assessment of the thiol redox metabolome. *Redox Biol.* **16**, 359–380 (2018).
92. Hill, J. W., Coy, R. B. & Lewandowski, P. E. Oxidation of cysteine to cystine using hydrogen peroxide. *J. Chem. Educ.* **67**, 172 (1990).
93. Millard, P. et al. IsoCor: isotope correction for high-resolution MS labeling experiments. *Bioinformatics* **35**, 4484–4487 (2019).


## Acknowledgements

We thank study participants for donating clinical samples used in this study; study staff at the FRESH cohort; laboratory staff at the HIV Pathogenesis Programme at UKZN for sample processing; J. A. Elsherbini (Ragon Institute) for bioinformatic support; L. Froehle (Ragon Institute) for helpful discussions of analysis; and D. Jenkins (Harvard Department of Chemistry and Chemical Biology) and M. Farcasanu, K. Jackson, L. Froehle and J. Bramante (Ragon Institute) for assay and sample assistance. J. Ravel and M. France (University of Maryland) and the Vaginal Microbiome Research Consortium (VMRC) kindly provided 4 experimental isolates (details in Supplementary Table 3, referred to as 'VMRC') as well as WGS data from 34 study participants and 312 previously unreported isolate genomes (details summarized in Supplementary Tables 5, 6, 8, 9 and 11; referred to as 'VMRC'). J. Marrazzo (University of Alabama, Birmingham) kindly provided 111 strains of non-*iners* lactobacilli that were sequenced for genomic analysis (details summarized in Supplementary Tables 5, 6, 8, 9 and 11; referred to as 'Vaginal Health Project'). This work was supported in part by National Institutes of Health grant NIH 1R01AI111918-01 to D.S.K. and by NIH grant T32 AI007387 to S.M.B.; additional NIH support was provided by grants to S.M.B. and M.S.G. from the Harvard University Center for AIDS Research (CFAR), an NIH funded programme (P30 AI060354) supported by the following NIH co-funding and participating institutes and centres: NIAID, NCI, NICHD, NIDCR, NHLBI, NIDA, NIMH, NIA, NIDDK, NINR, NIMHD, FIC and OAR. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The work was also supported in part by: Bill and Melinda Gates Foundation grants OPP1189208 to D.S.K. and OPP1158186 to E.P.B. and D.A.R.; a Burroughs Wellcome Career Award for Medical Scientists to D.S.K.; Vincent Memorial Research Funds and a Domolky Innovation Award (Massachusetts General Hospital) to C.M.M.; the South African Research Chairs Initiative through the National Research Foundation and the Victor Daitz Foundation to T.N.; funds from the Thomas C. and Joan M. Merigan Endowment at Stanford to D.A.R.; funding from the Harvard Program for Research in Science and Engineering (PRISE) and the Harvard Microbial Sciences Initiative to A.B.A.; and the Ragon Institute Summer Program Fellowship to X.W.

## Author contributions

S.M.B. and D.S.K. conceived the overall study and guided it throughout, with input from B.M.W., E.P.B. and C.M.M.; S.M.B., N.A.M. and J.K.R. performed primary bacterial isolations; S.M.B., N.A.M., J.F.F., B.M.W., A.J.M., X.W., N.C. and C.M.M. contributed to media design and production and/or bacterial growth and inhibition experiments; B.M.W. and E.P.B. synthesized labelled glutathione; B.M.W., S.M.B., N.A.M. and E.P.B. designed, performed and/or analysed measurements of media composition and isotopic tracing experiments; S.M.B., N.A.M. and J.X. performed nucleic acid extractions and sequencing; S.M.B. performed bacterial 16S rRNA gene sequencing analysis; M.R.H. and D.A.R. performed bacterial isolate genomic and metagenomic sequence analysis and assembly, genome catalogue development, and phylogenetic reconstructions; S.M.B., M.R.H., F.A.H. and B.M.W. conceived and/or performed genomic pathway analysis; S.M.B. and A.B.A. performed analysis of in vivo metabolite data; K.L.D., M.D., T.G., F.X.C., T.N., N.I., S.M.B., N.X., M.S.G. and D.S.K. contributed to clinical cohort design, cohort performance and/or sample acquisition and processing efforts; S.M.B., B.M.W., M.R.H., N.A.M. and D.S.K. wrote the paper, and all authors reviewed, offered input to the writing and approved the manuscript.

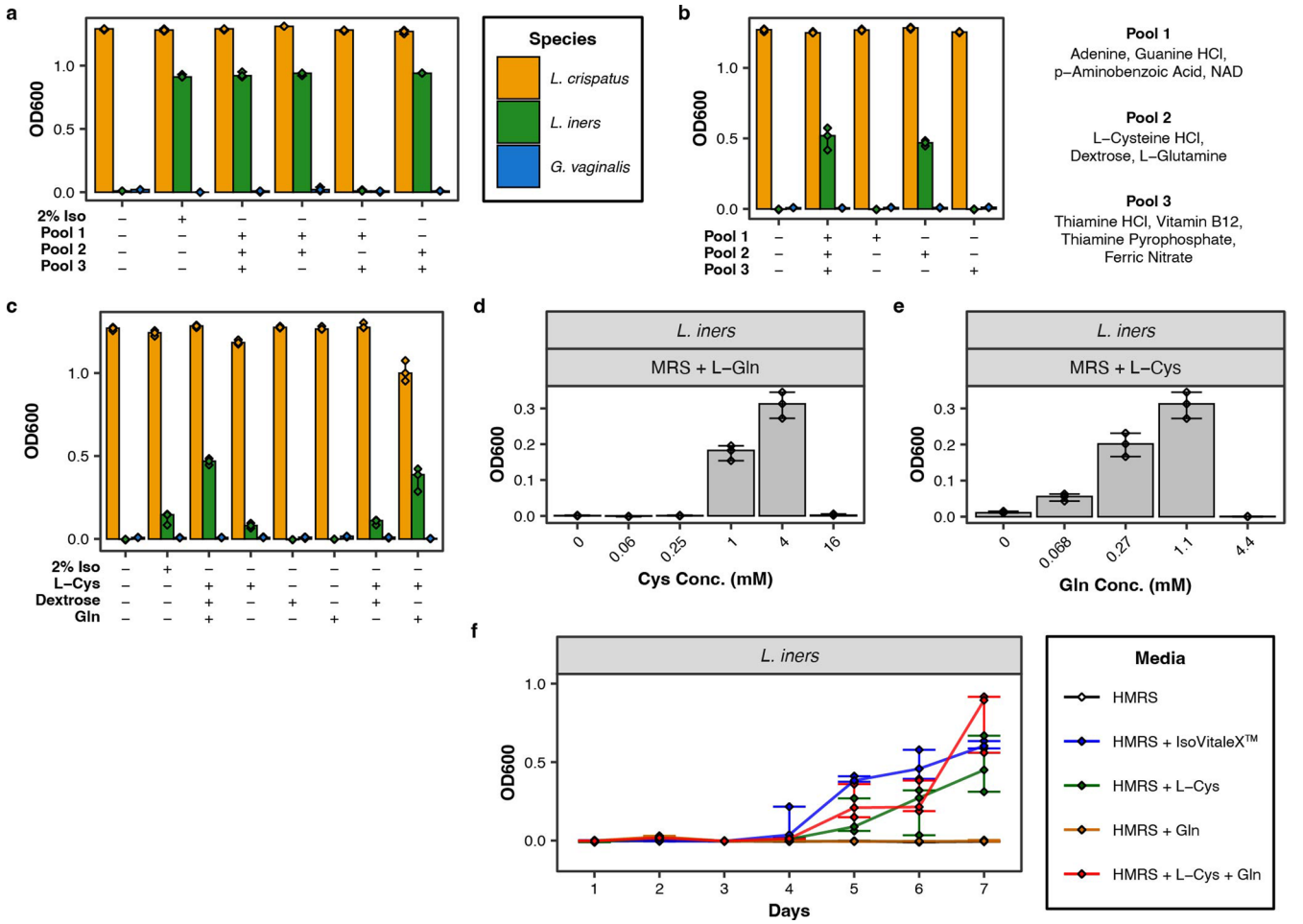
## Competing interests
All authors declare no competing interests.

## Additional information
**Extended data** is available for this paper at https://doi.org/10.1038/s41564-022-01070-7.

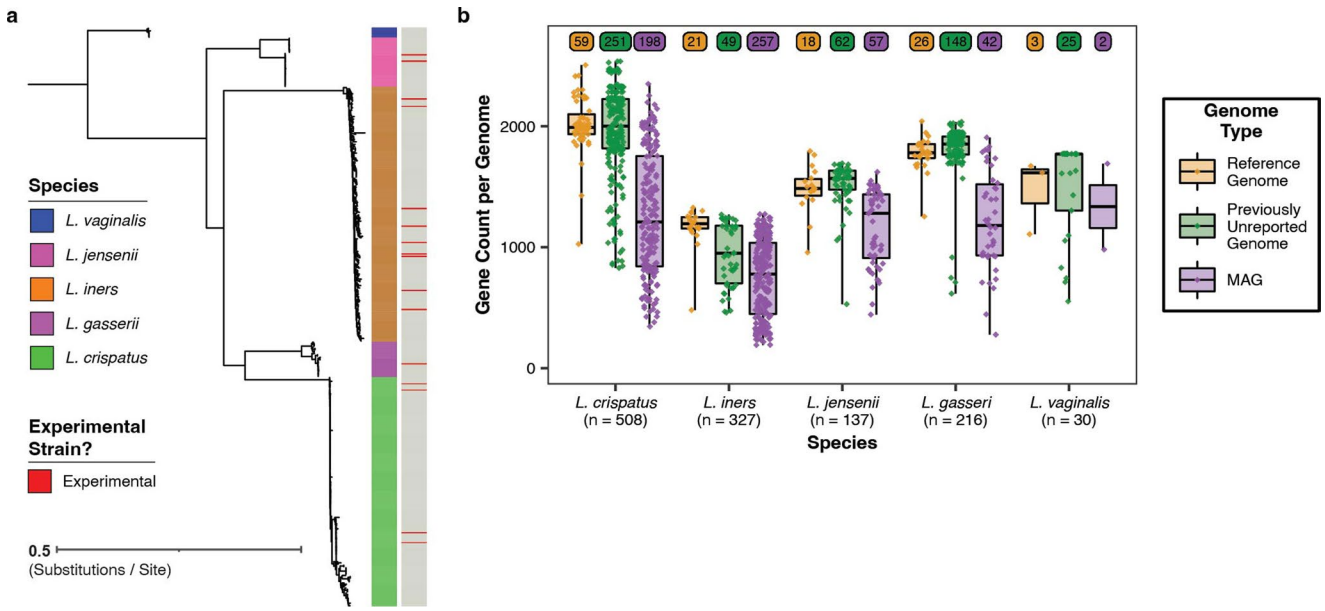
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41564-022-01070-7.

**Correspondence and requests for materials** should be addressed to Douglas S. Kwon.

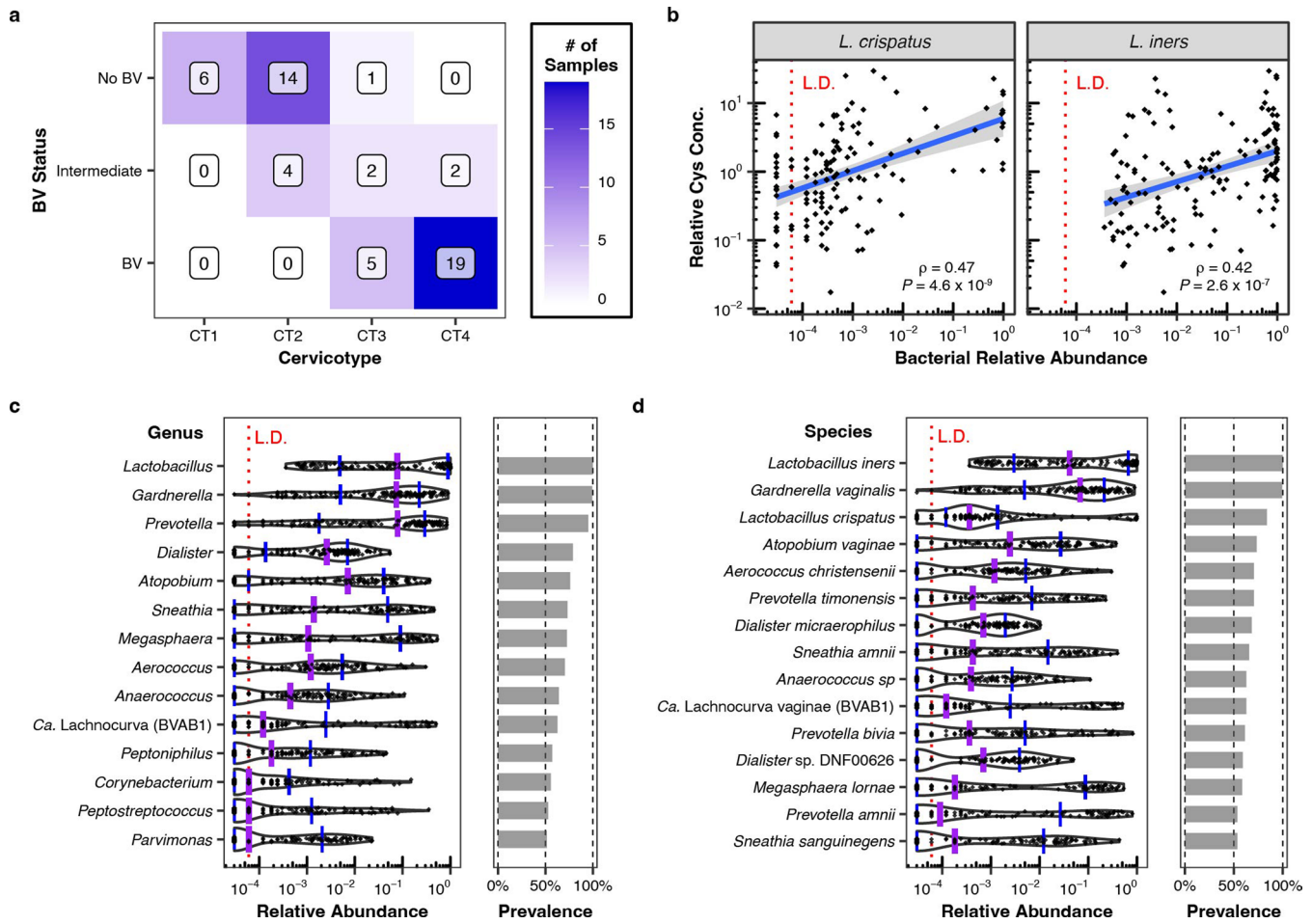
**Peer review information** *Nature Microbiology* thanks Lindi Masson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



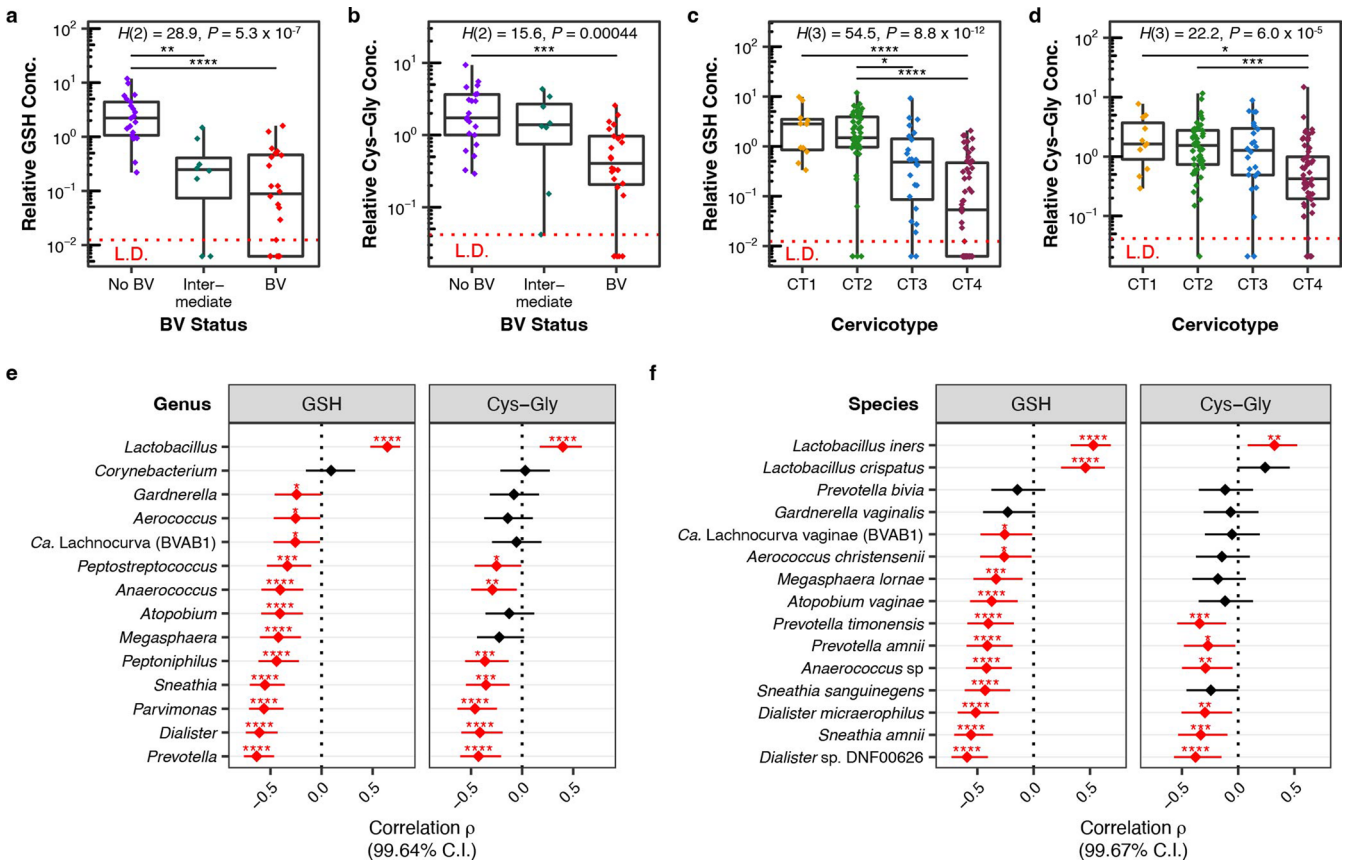

**Extended Data Fig. 1 | L-cysteine supplementation supports *L. iners* growth in *Lactobacillus* MRS broth, augmented by L-glutamine. a,b** Growth of *L. crispatus*, *L. iners*, and *G. vaginalis* at 24 hours incubation in MRS broth (BD DIFCO) ± supplementation with 2% IsoVitaleX ("Iso") or with the indicated sub-pools of IsoVitaleX components. **c**, Growth in MRS broth supplemented with 2% IsoVitaleX or various combinations of the nutrients in "Pool 2". **d**, Growth of *L. iners* at 24 hours in MRS broth + L-Gln (1.1 mM) supplemented with varying concentrations of L-Cys or **e**, in MRS broth + L-Cys (4 mM) supplemented with varying concentrations of L-Gln. Experiments in (**a-e**) all used BD DIFCO-formulated MRS broth base. (**f**) Growth of *L. iners* in Hardy Criterion-formulated MRS ("HMRS") broth supplemented with IsoVitaleX 2% v/v, L-Cys (4 mM), and/or L-Gln (1.1 mM) produced similar results, although with a substantially longer lag phase. Each plot depicts median (± range) for 3 replicates per condition and each plot is representative 1 of ≥2 independent experiments per strain and media condition except (**a**), which was un-replicated.



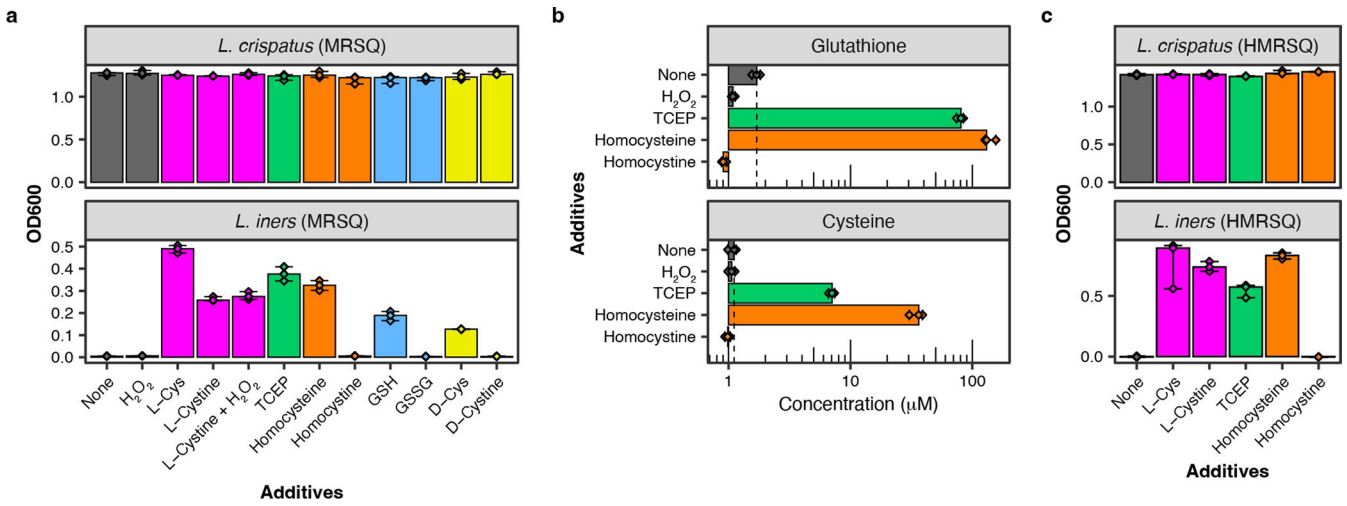

**Extended Data Fig. 2 | Phylogeny and per-genome gene content of genomes and MAGs within FGT *Lactobacillus* genome catalogs. a**, Phylogenetic tree of *L. iners* isolate genomes and MAGs as in Fig. 1b, plus isolate genomes of *L. crispatus* (n = 182), *L. jensenii* (n = 39), *L. gasseri* (n = 28), and *L. vaginalis* (n = 8). Isolates further experimentally studied in this work are indicated. The tree depicts only genome assemblies exceeding certain quality thresholds to ensure robustness of the phylogenetic reconstruction (see Methods); additional strains and genomes were included in other analyses. **b**, Number of genes per genome for each genome retrieved from RefSeq ("Reference Genome"), previously unreported isolate genome ("Novel Genome"), or MAG within the *Lactobacillus* genome catalogs analyzed in Figs. 2b and 4b and Extended Data Fig. 6a. All MAGs present in the analysis represent previously unreported assemblies (Hayward et al, manuscript in preparation, see additional information in Methods and Supplementary Tables 4-11). To maximize comprehensiveness of the pan-genomes, the catalogs included high- and medium-quality genomes and MAGs, defined as assemblies with minimum estimated completeness >50% (some of which are classified as partial assemblies by NCBI size criteria) and maximum estimated contamination <10%. Gene count analysis excludes gene sequences observed in only 1 genome or MAG per species to eliminate singleton contaminating sequences within individual genome assemblies. Box center lines, edges, and whiskers signify the median, interquartile range (IQR), minima and maxima respectively.



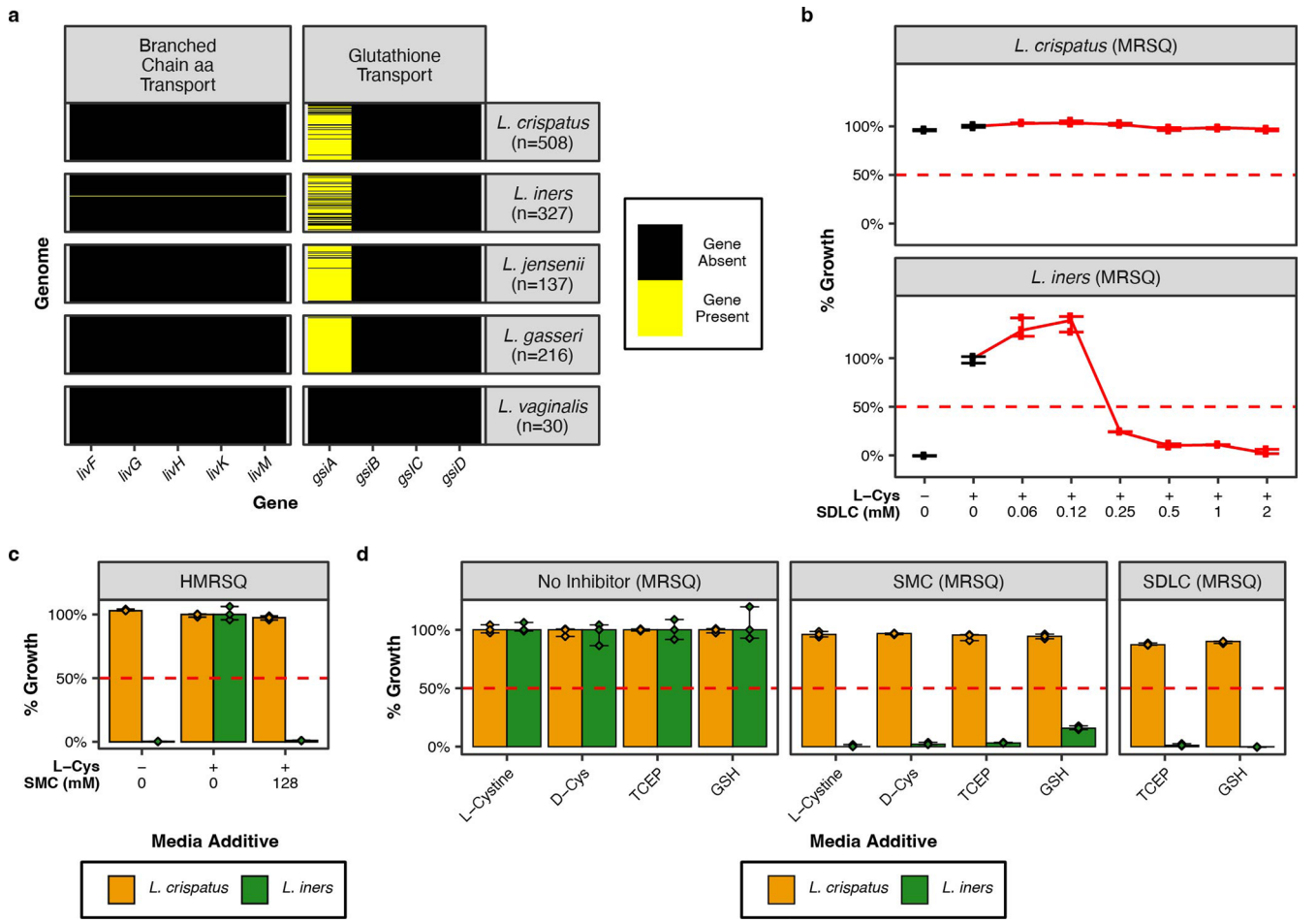

**Extended Data Fig. 3 | *In vivo* association of BV status and vaginal Cys concentrations with microbiota composition. a**, Relationship between Nugent score-based BV status and cervicotype among the 53 women depicted in Fig. 3a. BV status and cervicotype were significantly associated ($P = 1.902 \times 10^{-11}$; two-sided Fisher's Exact Test). **b**, Two-tailed Spearman correlation between relative Cys concentrations in cervicovaginal lavage (CVL) fluid and relative abundances of the species *L. iners*, and *L. crispatus* among the 142 women depicted in Fig. 3b-f, showing correlation coefficients (ρ) with unadjusted p-values. Linear regression lines (solid blue) with 95% confidence intervals calculated based on log-transformed abundances and concentrations are shown to assist visualization (*L. crispatus*: $y = 0.77 + 0.25x$; *L. iners*: $y = 0.30 + 0.22x$). The red dotted line represents the bacterial limit of detection (L.D.). **c**,**d** Per-sample relative abundances and cohort-level prevalence (fraction of samples from the cohort in which each taxon was detected) for each genus (**c**) or species (**d**) with ≥50% prevalence (panels correspond to main Figs. 3e and f, respectively). Purple and blue lines respectively represent median and interquartile range of relative abundances for each taxon.



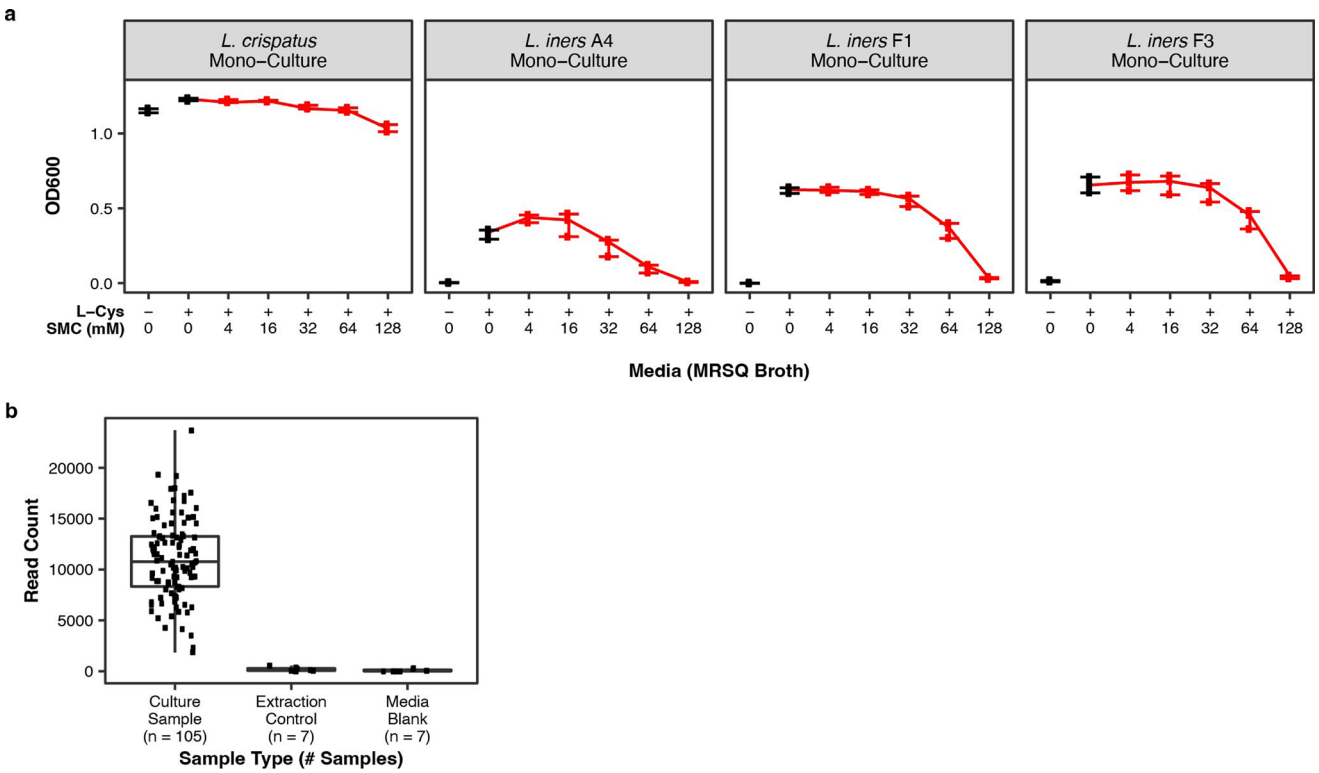

**Extended Data Fig. 4 | Vaginal concentrations of the Cys-containing peptides reduced glutathione (GSH) and cysteinylglycine (Cys-Gly) in cervicovaginal fluid are higher in women without BV and correlate with *Lactobacillus* dominance of the microbiota. a,b** Relative concentrations of (**a**) GSH and (**b**) Cys-Gly by BV status in CVL fluid from the 53 women in (Fig. 3a). **c,d** Relative concentrations of (**c**) GSH and (**d**) Cys-Gly by cervicotype in CVL fluid from the 142 women depicted in (Fig. 3b,c). In (**a-d**) the red dotted line represents the metabolite limit of detection (L.D.). For samples in which an analyte was below the L.D., concentrations were imputed at 0.5 x L.D. Log-transformed concentrations were not normally distributed due to the imputed values, so between-group differences were determined via Kruskal-Wallis test with post-hoc Dunn's test, adjusting for multiple comparisons using the Bonferroni method. All significant pairwise differences are displayed. **a** No BV-BV: $P = 5.7 \times 10^{-7}$; No BV-Intermediate: P = 0.0037. **b** No BV-BV: P = 0,0032. **c** CT1-CT4: $P = 7.7 \times 10^{-5}$; CT2-CT3: P = 0.0292; CT2-CT4: $P = 2.9 \times 10^{-11}$. **d** CT1-CT4: P = 0.0191; CT2-CT4: P = 0.00014. Box center lines, edges, and whiskers signify the median, IQR, minima and maxima respectively. **e,f** Forest plots depicting Spearman correlation coefficients (ρ) between concentrations of GSH and Cys-Gly and relative abundances of each bacterial genus (**e**) or species (**f**) detected at >50% prevalence in the cohort (n = 142). P-values and confidence intervals in **e,f** were adjusted for multiple comparisons using the Bonferroni method at significance level 0.05/n (full statistical results in Supplementary Tables 14 & 15). Significance is depicted for adjusted p-values as * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$.



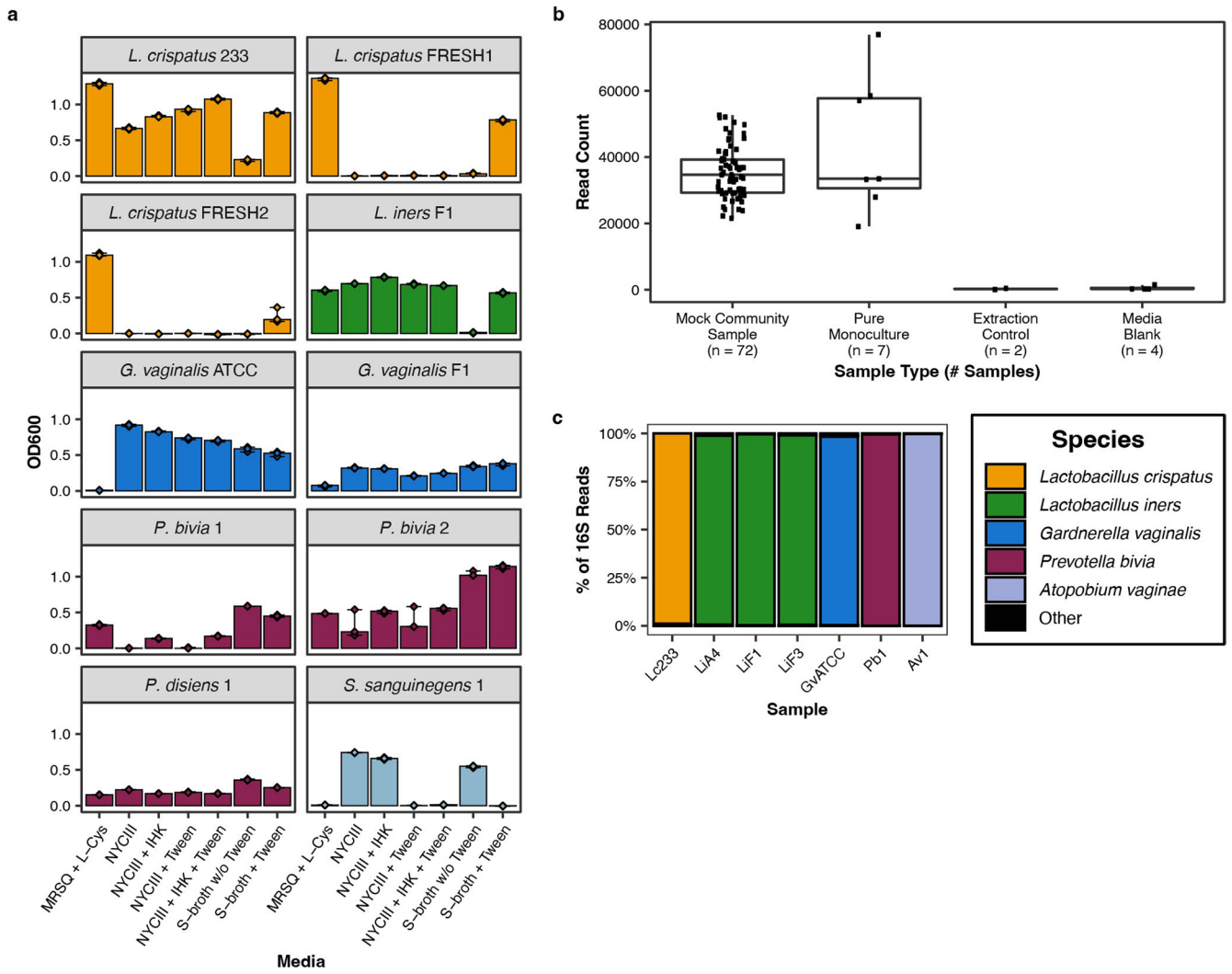

**Extended Data Fig. 5 | Cys and Cys-containing molecules in MRS exist primarily as mixed disulfides and addition of chemical reducing agents permits *L. iners* growth. a**, Growth of *L. iners* and *L. crispatus* at 28 hours in MRSQ broth supplemented as indicated with the reduced thiols L-Cys, D-Cys (the non-physiological enantiomer of L-Cys), GSH, or homocysteine (each 4 mM), with their oxidized counterparts L-cystine, D-cystine, oxidized glutathione (GSSG), or homocystine (each 2 mM), or with the non-sulfur-containing reducing agent Tris(2-carboxyethyl)phosphine (TCEP; 4 mM), $H_2O_2$ (0.4 mM), or L-cystine + $H_2O_2$. **b**, Concentrations of reduced Cys (baseline median concentration 1.11 μM) and glutathione (GSH; baseline median concentration 1.70 μM) in MRSQ broth supplemented with the oxidizing agent $H_2O_2$ (0.4 mM), the reducing agents TCEP or homocysteine (each 4 mM), or homocysteine's oxidized counterpart homocystine (2 mM). **c**, Growth at 7 days of *L. crispatus* and *L. iners* in HMRS broth with 1.1 mM L-Gln ("HMRSQ") supplemented as indicated with L-Cys, L-cystine, TCEP, homocysteine, or homocystine at the above concentrations. All plots depict median ± range for 3 replicates per condition and each plot is representative of 1 of ≥2 independent experiments per strain and media condition. Bar coloring highlights the pairing of media conditions with each thiol-containing reducing agent and its oxidized counterpart.




**Extended Data Fig. 6 | FGT *Lactobacillus* genomes lack predicted alternate Cys and GSH transporters and *L. iners* is selectively inhibited by cystine uptake inhibitors. a**, Predicted presence of the branched-chain amino acid transport locus *livFGHKM* (which has low-affinity Cys transport activity in *E. coli*) and the glutathione transport locus *gsiABCD* in isolate genomes and MAGs of common FGT *Lactobacillus* species (n = number of genomes. Detailed statistics are in Supplementary Table 12). **b**, Selective growth inhibition of *L. iners* in MRSQ broth with or without L-Cys (4 mM) and varying concentrations of the cystine uptake inhibitor seleno-DL-cystine (SDLC). **c**, Growth of *L. crispatus* and *L. iners* in HMRSQ broth with or without L-Cys (4 mM) ± SMC. **d**, Growth inhibition of *L. crispatus* and *L. iners* by SMC (128 mM) or SDLC (2 mM) in MRSQ supplemented with L-cystine (2 mM) or D-Cys, TCEP, or GSH (4 mM each). For each growth additive, percentage growth in in presence of inhibitor was calculated relative to median growth in broth containing that additive without inhibitor. Plots in **b-d** depict median ± range for 3 replicates per condition and each plot is representative 1 of ≥2 independent experiments per strain and media condition.




**Extended Data Fig. 7 | Sample preparation and controls for *L. iners / L. crispatus* growth competition assays. a**, Mono-culture growth of *L. crispatus* (strain FRESH1) and 3 representative *L. iners* strains at 28 hrs incubation in MRSQ broth with or without L-Cys (4 mM) and varying concentrations of the cystine uptake inhibitor *S*-methyl-L-cysteine (SMC), exhibiting expected growth patterns for the respective species. Plots depict median ± range for 3 replicates per condition and each plot is representative 1 of ≥2 independent experiments per strain and media condition. The competition assays between *L. iners* and *L. crispatus* depicted in Fig. 5a were prepared by mixing the input inocula from each of these *L. iners* monocultures pairwise with the input inoculum for the *L. crispatus* monoculture at the colony-forming unit (c.f.u.) ratios listed in the legend of Fig. 5a. **b**, Bacterial 16 S rRNA gene read counts in cultured samples (n = 105 individual cultures), with negligible background read counts in blank growth media controls (n = 7 controls, one each per media type) and extraction controls (n = 7 controls) associated with the competition experiments in main Fig. 5a. (Control sample reactions were included in the sequencing library despite absence of visible PCR bands on agarose gel electrophoresis.) Box center lines, edges, and whiskers signify the median, IQR, minima and maxima respectively.




**Extended Data Fig. 8 | Development of "S-broth" and controls for mock BV-like community growth experiments. a**, Monoculture growth at 48 hours of experimental US (strain "233") and South African (strains FRESH1 and FRESH2) *L. crispatus* strains, as well as *L. iners*, *G. vaginalis*, *Prevotella bivia*, *Prevotella disiens*, and *Sneathia sanguinegens* strains in various broth media including MRSQ broth + L-Cys (4 mM), NYCIII broth with or without 2% IsoVitaleX plus 5% Vitamin $K_1$-Hemin solution ("IHK") and/or Tween-80 (1 g/L), and in "S-broth" (see Methods) with or without Tween-80 (1 g/L). S-broth + Tween was used in subsequent experiments. (Detailed strain information is in Supplementary Table 3). Plots depict median ± range for 3 replicates per condition and each plot is representative 1 of ≥2 independent experiments per strain and media condition. **b**, Bacterial 16 S rRNA gene read counts in mock mixed communities (n = 72 individual cultures) and pure strain monocultures (n = 7), with negligible background read counts in blank growth media controls (n = 4 controls, one each per media type) and extraction controls (n = 2 controls) associated with the experiments in Fig. 5c,d. (Control sample reactions were included in the sequencing library despite absence of visible PCR bands on agarose gel electrophoresis.) Box center lines, edges, and whiskers signify the median, IQR, minima and maxima respectively. **c**, Confirmation of identity of input strains in mock BV-like communities by 16 S rRNA gene sequencing of bacterial monoculture controls, prepared from the same input inocula used for the bacterial mock communities shown in Fig. 5c,d.

nature research

Corresponding author(s): Kwon, Douglas S.

Last updated by author(s): Nov 23, 2021

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ A description of all covariates tested |
| ☐ | ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | No software was used for primary data collection in this study. |
| Data analysis | * For pan-genome generation, genes were identified within individual genomes by Prokka v1.14.5 (Seemann, 2014), which predicts genes using Prodigal v2.6.3 (Hyatt et al., 2010). The resulting genes were clustered at 95% nucleotide identity into a comprehensive pan-genome using Roary v3.13.0 (Page et al., 2015). The pan-genome was annotated with eggNOG 5.0 using eggNOG-mapper v2 (Huerta-Cepas et al., 2017, 2019). Analysis of presence of genes of interest in the eggNOG-annotated pan-genome and within individual genomes was performed using custom R code; the custom R analysis code is available at zenodo.org under DOI https://doi.org/10.5281/zenodo.5715682.<br>* Demultiplexing of Illumina MiSeq bacterial 16S rRNA gene sequence data was performed using QIIME 1 version 1.9.1(Caporaso et al., 2010). Mapping files in standard QIIME 1 format were created and validated using validate_mapping_file.py, sequences were demultiplexed using split_libraries_fastq.py with parameter store_demultiplexed_fastq and no quality filtering or trimming, and demultiplexed sequences were organized into individual fastq files using split_sequence_file_on_sample_ids.py(Hoang et al., 2020). We then used dada2 version 1.6.0(Callahan et al., 2016) in R used to filter and trim reads at positions 10 (left) and 230 (right) using the filterAndTrim function with parameters truncQ = 11, MaxEE = 2, and MaxN = 0. Sequences were then inferred and initial taxonomy assigned using the dada2 assignTaxonomy function employing the RDP training database rdp_train_set_16.fa.gz (obtained from https://www.mothur.org/wiki/RDP_reference_files). Taxonomic assignments were refined and extended via extensive manual review; see Supplementary Table 13 for amplicon sequence variant (ASV) taxonomy. The denoised dada2 results with final taxonomic assignment were analyzed in R using phyloseq version 1.30.0(McMurdie & Holmes, 2013) and custom R scripts available at zenodo.org under DOI https://zenodo.org/record/5900469.<br>* Genome phylogenetic reconstructions were performed using fetchMG v1.0 (Milanese et al., 2019) and Ete3 v3.1.1 (Huerta-Cepas, Serra, & Bork, 2016), and the tree was visualized using iTOL v4 (Letunic & Bork, 2019).<br>* Isotope correction for mass spectrometry labeling experiments was performed using IsoCor v2.1.3.<br>* Compressed directories containing R analysis code sufficient to reproduce (1) analysis of pan-genome composition and gene content, (2) analysis for human FGT microbiota-metabolite analysis, and (3) analysis for competition cultures of L. iners and L. crispatus and mixed community cultures is available at zenodo.org under DOI https://zenodo.org/record/5900469. Each compressed directory contains a README |

file describing dependencies and other details, an R Project file, an R Markdown file containing the analysis code with additional information, and associated sub-directories used in the analysis.
* Data analysis and visualization was performed in R v3.6.3 using packages including phyloseq version 1.30.0, seqinr v4.2.5, tidyverse, v1.3.1, knitr v1.33, ggpubr v0.4.0, DescTools v0.99.41, gtools v3.8.2, gridExtra v2.3, cowplot v1.1.1, scales v1.1.1, grid v3.6.3, broom v0.7.6, e1071 v1.7.6, and table1 v1.4.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Compressed directories containing data files sufficient to reproduce (1) analysis of pan-genome composition and gene content, (2) analysis for human FGT microbiota-metabolite analysis, and (3) analysis for competition cultures of L. iners and L. crispatus and mixed community cultures are posted at zenodo.org under DOI https://zenodo.org/record/5900469. The dataset containing the raw Illumina MiSeq read data for genital tract bacterial 16S rRNA gene profiling analyzed in this study (Figure 3 and Extended Data Figures 3 & 4) is available in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA729907. The dataset containing the raw Illumina MiSeq read data for the bacterial 16S rRNA gene sequences from competition cultures of L. iners and L. crispatus (Fig. 5a) and from mixed community cultures (Figs 5c&d) is available in the NCBI SRA under BioProject PRJNA777644. The taxonomic assignments used for amplicon sequence variants (ASVs) from bacterial 16S rRNA gene sequencing are supplied in Supplementary Table 15. The Lactobacillus genomic catalogs included a total of 1,091 previously unreported isolate genomes, partial genomes, and MAGs from multiple human cohorts, as detailed above. The assemblies are available in the NCBI SRA under BioProjects PRJNA799384, PRJNA799634, PRJNA799626, PRJNA799445, PRJNA799630, PRJNA799633, PRJNA799642, PRJNA799746, PRJNA799744, PRJNA799737, PRJNA799762, and PRJNA797778; additional details on the individual studies associated with these BioProjects are contained in Supplementary Tables 5, 6, and 10, and individual NCBI BioSample accession numbers for each of the 1,091 assemblies are listed in Supplementary Table 8. In addition, the genome catalogs included 127 previously reported isolate genomes that were retrieved from RefSeq; the individual accession numbers for these genomes are listed in Supplementary Table 7. The raw and corrected cystine and serine isotopologue measurements associated with Fig 2c,d and 4c are available in Supplementary Tables 13 and 14. Some metadata related to previously reported isolate genomes was obtained from corresponding entries in RefSeq (https://www.ncbi.nlm.nih.gov/refseq/) or the Genomes OnLine Database (GOLD; https://gold.jgi.doe.gov/). Source data are supplied for plots and phylogenetic trees, including for Figures 1-5 and Extended Data Figures 1-8.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Paired bacterial microbiota composition and cervicovaginal metabolite concentrations (Figure 3 and Extended Data Figures 4 and 5) were determined for 143 HIV-uninfected FRESH study participants consecutively sampled in May through October of 2017 (methodological details below). The sample size was initially selected for microbiota profiling based on a separate, unpublished investigation of microbiome-immune parameters; paired CVL samples from these participants were later used for FGT metabolite concentration measurements due to availability of paired microbiota sequencing data. No a priori sample size calculation was performed in relation to analysis of FGT microbiota-cysteine correlations because it was (to our knowledge) the first investigation of this association in a sub-Saharan African cohort and most studies in non-African populations have used case-control designs that are not adapted to evaluate general population-level correlations – we therefore lacked information upon which to make informed estimates about metabolite distributions for power calculations. However the cohort was of similar or larger size as other reported studies of the FGT metabolome(McMillan et al., 2015; Srinivasan et al., 2015; Vitali et al., 2015).<br><br>Assessment of BV status in this cohort by the Nugent scoring method(Nugent et al., 1991) was performed sporadically prior to September 2017, when a protocol change instituted universal BV testing for all participants. We restricted analysis of BV-metabolite associations (Figure 3a and Extended Data Figures 5a,b) to the subset of 53 out of 143 participants assessed under the universal testing protocol to minimize risk of bias. |
| Data exclusions | One of the initial 143 participants in the analysis of vaginal microbiota composition was excluded from analysis based on the pre-specified criterion of low 16S sequence read count (<10,000 reads in the sample) in order to maximize sensitivity of analyses of microbiome/metabolite correlations. |
| Replication | The distribution of microbiota composition observed in FRESH cohort participants closely mirrors previously published reports regarding microbiota composition among different sub-cohorts of participants in this ongoing study (e.g. Gosmann, 2017). We are aware of no prior data on these FGT metabolite-taxon correlations in this population or in other sub-Saharan African populations for comparison or replication.<br><br>Results shown for in vitro experiments all depict 1 of >=2 independent experiments in which results were successfully replicated unless the text otherwise explicitly states that the experiment was not replicated (isotopic tracing experiments in Figures 2c,d and 4c; the growth experiment shown in Extended Data Figure 1a, and the confirmatory analyses of glutathione synthesis in Extended Data Figures 8b-d. |

| | |
|---|---|
| Randomization | FRESH study participants were not randomized because they were part of an observational cohort. |
| Blinding | Nugent scoring for samples from FRESH cohort participants was performed at Global Laboratories (now Neuberg Global Laboratories, Durban, South Africa), an accredited commercial laboratory diagnostics company, by trained laboratory technologists blinded to all participant information other than study ID.<br><br>In order to avoid bias in assay performance, metabolite measurements from clinical samples were performed after randomized reordering of samples by laboratory staff at Metabolon, Inc., who were blinded to all participant information except sample ID.<br><br>Initial sequencing-based analysis of FGT microbiota composition of FRESH cohort participants, taxonomy assignment, and cervicotype assignments were performed blinded to participants' corresponding Nugent scores and metabolite concentrations in CVL fluid.<br><br>Researchers performing measurement and data analysis of bacterial growth patterns, growth inhibition, chemical concentrations, and sequencing-based analysis of in vitro bacterial growth competition assays were not blinded to experimental group due to impracticality of setting up and performing the assays in blinded fashion. However results of all of these experiments were measured via quantitative metrics (e.g. optical density or sequence read counts), reducing risk of bias in the results. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☐ | ☒ Human research participants |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Human research participants

Policy information about studies involving human research participants

| | |
|---|---|
| Population characteristics | The Females Rising through Education, Support, and Health (FRESH) study is an ongoing prospective observational study based in Umlazi, South Africa, that enrolls 18-23 year-old, HIV-uninfected, non-pregnant, otherwise healthy women willing to have frequent HIV testing and 3-monthly collection of blood and mucosal specimens during study enrollment. Exclusion criteria included hemoglobin level <10 g/dL, any chronic medical condition or other conflict likely to prevent adherence to the study protocol, and/or enrollment in any other study that involves frequent blood sampling or that might otherwise interfere with the FRESH study protocol. Study characteristics and inclusion and exclusion criteria have also been described in detail elsewhere(Anahtar et al., 2015; Dong et al., 2018; Gosmann et al., 2017). Samples from 143 consecutively sampled FRESH study participants were used for analysis of FGT microbiota-metabolite correlations; one of these participants was excluded due to low sequence read count.<br><br>Previously unreported Lactobacillus isolates and MAGs were derived from samples obtained from participants in the FRESH cohort (138 total participants) and from participants in additional cohorts (see below).<br><br>Two previously described(Callahan et al., 2017) cohorts linked to the Vaginal Microbiome Research Consortium (VMRC; https://vmrc4health.org/) were approved under the same IRB protocol at Stanford University (principal investigator D.A.R.). This protocol cumulatively enrolled 135 pregnant women aged at least 18 years presenting for obstetrical care at either Lucille Packard Children's Hospital at Stanford University or at the University of Alabama, Birmingham, who provided written informed consent. Exclusion criteria included antibiotic or probiotic use within twelve weeks prior to enrollment, significant immunosuppression, or (retrospectively) delivery of a baby who had congenital defects. Samples from 18 total study participants collected across multiple timepoints were analyzed here.<br><br>The second VMRC-associated cohort, which has also been previously described(Ravel et al., 2013), enrolled 135 non-pregnant, HIV-uninfected women at University of Alabama, Birmingham. The study enrolled participants aged 18-45 with regular menses who provided written informed consent. Exclusion criteria included antibiotic or antimycotic use within 30 days prior to enrollment, pregnancy or current lactation, active infection with Trichomonas vaginalis, Neisseria gonorrheae, Chlamydia trachomatis, or syphilis, symptomatic BV or vulvovaginal candidiasis for which treatment was requested at time of enrollment, immunocompromised status (including HIV), significant affective or psychotic disorders or intellectual limitations precluding informed consent, history of prior or pending hysterectomy or pelvic radiotherapy, non-fluency in English, current use of NuvaRing contraception, recent receipt of a bacterial vaccine, or concurrent participation in other trials involving blinded administration of antibiotics or topical microbicides. Samples from 43 total study participants collected across multiple timepoints were analyzed here. |

The Vaginal Health Project (VHP) cohort was a previously described (Marrazzo et al., 2008), IRB-approved longitudinal study of BV dynamics based in Seattle, Washington, USA, that enrolled women aged 16-30 years who reported having sex with >=1 female partner in the prior year. Participants provided written informed consent and attended at least quarterly study visits over a year at which mid-vaginal swabs were collected, with more frequent study visits if BV was diagnosed via a combination of Amsel criteria(Amsel et al., 1983) and Nugent scoring(Marrazzo et al., 2008; Nugent et al., 1991). Samples from 6 VHP cohort participants collected over multiple visits were analyzed here.

The V2 cohort(C. M. Mitchell et al., 2020) was an IRB-approved cohort at Massachusetts General Hospital (MGH) in Boston, MA, that enrolled women presenting for gynecological care at MGH who provided written informed consent, excluding those with HIV or who were otherwise immunocompromised. Samples from 4 V2 cohort participants were analyzed here; ages ranged from 24-38 years.

The Ragon Institute Female Reproductive Tract biopsy protocol enrolls non-pregnant, pre-menopausal women aged 18 to 50 years with normal Pap smears, normal results on routine screening laboratory tests, and willingness to provide informed consent. Exclusion criteria include history of severe gynecologic disease, symptomatic vaginal infection, treatment for vaginal infection within the prior 30 days, being <8 weeks postpartum, sexual intercourse, tampon use or application of vaginal medicines within the prior 2 days, use of immunomodulatory therapy, prior receipt of an HIV vaccine, history of significant medical illness, or history of requiring antibiotics prophylaxis for invasive procedures. The single participant from this protocol whose samples were analyzed here was 26 years old and HIV uninfected.

Additional MAGs were assembled from previously reported FGT WGS data from three different US cohorts (95 total participants)(Consortium, 2012; Goltsman et al., 2018; Hudson et al., 2020; Methé et al., 2012), one Italian cohort (10 participants)(Ferretti et al., 2018), and one Chinese cohort (4 participants)(Li et al., 2018).

Recruitment

FRESH cohort participants were recruited at local sites in Umlazi, South Africa, frequented by young people including shopping malls, cafes, and nightclubs. All participants provided written informed consent. The study protocol includes non-monetary benefits encompassing job-skills, health, and empowerment training, a light meal during study site visits, and cumulative monetary compensation over a 36-week period of 3,700 rand (~$280 US dollars), which helps defray transportation expenses to twice-weekly study site visits. Participant self-selection bias may potentially occur based on participants' perception of their own HIV risk, but it is unknown whether or how such a bias would affect FGT microbiota composition or metabolite concentrations.

Samples from FRESH participants and participants in other studies (detailed below) were used to generated of MAGs from metagenomic sequencing and/or to isolate bacterial strains for genome sequencing; the resulting genomic information was then used to explore species gene content. It is unknown how biases in enrollment or other differences in cohort characteristics might relate to bacterial gene content and thus bias results of bacterial genomic sampling and analysis. Notably, although culture-independent microbiota surveys have demonstrated that the vast majority of FGT microbiota communities in human populations worldwide contain one or more Lactobacillus species at varying level of abundance, a wide range of sample-, species-, and study-level biological and technical factors affect whether Lactobacillus isolates or MAGs can be obtained from any given sample. Thus, recovery of genomes or MAGs is not expected from all samples or participants in a cohort and differences in yield for different samples may be due to to technical issues rather than biological or epidemiological bias. The fact that the species pan-genomes we produced by sampling from geographically, clinically, and racially/ethnically diverse cohorts enrolled under varying protocols yielded consistent bacterial species-level results in our analysis suggests that our observations are broadly generalizable and robust to individual selection biases that may have occurred within individual cohorts.

For the Stanford-based, VMRC-associated pregnancy cohorts, participants were recruited from among women presenting for obstetrical care at either Lucille Packard Children's Hospital at Stanford University or at the University of Alabama, Birmingham. Cumulative monetary compensation totaled as much as $350 per participant.

The non-pregnancy VMRC-associated cohort enrolled women at University of Alabama, Birmingham, who were recruited via newspaper advertisements, fliers, general gynecology clinics, and the Jefferson County STD clinic. Cumulative monetary compensation totaled $440 per participant.

Recruitment for the VHP cohort occurred through media, advertisements, and community referrals in Seattle, Washington, USA. Cumulative monetary compensation totaled $100 per participant.

Participants in the V2 cohort were recruited from women presenting for gynecological care at MGH, Boston, Massachusetts, USA. Participants received no monetary compensation.

The Ragon Female Reproductive Tract biopsy protocol recruited participants via the MGH and Brigham & Women's Hospital Infectious Disease clinics in Boston, Massachusetts, USA, as well as recruiting participants who were referred by their medical providers or word of mouth for study enrollment, as well as by contacting participants in other studies at the Ragon Institute who gave permission to be contacted for possible new study recruitment. Participants were compensated $150 if no biopsy was performed and $250 if biopsy was performed.

Ethics oversight

The FRESH study protocol was approved by the MGH Institutional Review Board (2012P001812/MGH) and the Biomedical Research Ethics Committee of the University of KwaZulu-Natal (UKZN; Ethics Reference Number BF131/11). The Stanford-VMRC pregnancy protocol was approved by the Stanford University IRB (protocol #21956). The non-pregnancy VMRC-associated cohort was approved by the University of Maryland Baltimore IRB (protocol #HP-00041351). The V2 cohort protocol was approved by the MGH IRB (Protocol #2014P001066). The Ragon Female Reproductive Tract biopsy protocol was approved by the MGH IRB (Protocol #2012P001812).

Note that full information on the approval of the study protocol must also be provided in the manuscript.