nature communications

Article

https://doi.org/10.1038/s41467-024-52102-1

Insights into the assembly of the neovaginal microbiota in Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome patients

Received: 16 May 2024

Accepted: 23 August 2024

Published online: 06 September 2024

Check for updates

Na Chen[1,17], Lilan Hao[2,3,4,17], Zhe Zhang[2,17], Chenglu Qin[5,17], Zhuye Jie[2,3,17], Hongxin Pan[6], Jiali Duan[7], Xincheng Huang[2,8], Yunhong Zhang[9], Hongqin Gao[10], Ruike Lu[10], Tianshu Sun[1], Hua Yang[1], Jinqiu Shi[5], Maolian Liang[5], Jianbin Guo[1], Qianqian Gao[1], Xiaoyue Zhao[1], Zhiyuan Dou[1], Liang Xiao[2,4,11], Shaoqiao Zhang[12], Xin Jin[2], Xun Xu[2], Huanming Yang[2,13], Jian Wang[2,13], Huijue Jia[2,14,15], Tao Zhang[12,16] ✉, Karsten Kristiansen[2,3,4] ✉, Chen Chen[2,3] ✉ & Lan Zhu[1] ✉

Neovaginas are surgically constructed to correct uterovaginal agenesis in women with Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome or as part of gender-affirming surgery for transfeminine individuals. Understanding the assembly of the neovaginal microbiota is crucial for guiding its management. To address this, we conducted a longitudinal study on MRKH patients following laparoscopic peritoneal vaginoplasty. Our findings reveal that the early microbial assemblage exhibited stochastic characteristics, accompanied with a notable bloom of *Enterococcus faecalis* and genital *Mycoplasmas*. While both the pre-surgery dimple microbiota and the fecal microbiota constituted the primary species pool, the neovaginal microbiota developed into a microbiota that resembled that of a normal vagina at 6–12 months post-surgery, albeit with a bacterial vaginosis (BV)-like structure. By 2–4 years post-surgery, the neovaginal microbiota had further evolved into a structure closely resembling with the homeostatic pre-surgery dimple microbiota. This concords with the development of the squamous epithelium in the neovagina and highlights the pivotal roles of progressive selective forces imposed by the evolving neovaginal environment and the colonization tropism of vaginal species. Notably, we observed that strains of *Lactobacillus crispatus* colonizing the neovagina primarily originated from the dimple. Since *L. crispatus* is generally associated with vaginal health, this finding suggests potential avenues for future research to promote its colonization.

Neovaginas are surgically constructed vaginas for correcting uterovaginal agenesis in women with Mayer–Rokitansky–Küster–Hauser (MRKH) syndrome[1] or as part of gender-affirming surgery for transfeminine individuals[2]. In recent years, few cross-sectional studies have examined the neovaginal microbiota using primarily cultivation and microscopy[3,4]. In most cases, focus has been on the neovagina of transgender individuals, where a polymicrobial environment generally was identified[3,4]. A recent study utilizing metaproteomic analyses revealed that the polymicrobial communities harbored in penile skin-lined neovagina were quite distinct from the microbiota in the vagina

A full list of affiliations appears at the end of the paper. ✉e-mail: tao.zhang@genomics.cn; kk@bio.ku.dk; chenchen1@genomics.cn; zhu_julie@vip.sina.com

of adult cisgender women (in the following referred to as a normal vagina) and showed similarities in composition to an uncircumcised penis[5]. MRKH syndrome women are characterized by hypoplasia of the uterus, the cervix, and the upper part of the vagina, but with a normal female karyotype (46, XX), and generally, normal steroid hormone production[1]. A study on the microbiota in the neovagina of MRKH patient reported on a dysbiotic microbiota years after surgery, with a boost of *Lactobacillus* following squamous epithelialization[6].

The host epithelium constitutes a critical interface between the microbial communities and the host[7]. In the normal vagina, the stratified squamous epithelium coated with cervicovaginal mucus provides a habitat and nutrient source for colonizing microbes[8,9]. During reproductive age, increased glycogen accumulation in the epithelium promotes a dominance of *Lactobacillus* species[10,11]. By contrast, reduced glycogen accumulation following reduced estrogen production and tissue atrophy during menopause leads to a decreased colonization by *Lactobacillus* species accompanied with the presence of and/or increased relative abundances of anaerobes commonly associated with bacterial vaginosis (BV)[10–14]. A few studies evaluating neovaginas created with different techniques suggested that the neovaginas in MRKH women lined with amniotic membrane, peritoneum or tissue-engineered biomaterial during creation would complete metaplasia and be coated in time with a glycogen-containing stratified squamous metaplastic epithelium[15–18]. However, a comprehensive characterization of the neovaginal microbiota and detailed information on how the microbial community develops and assembles in this process are hitherto lacking. The initial assembly of the human gut microbiota has been extensively studied in infants[19–23], while much less is known in relation to the establishment of the vaginal microbiota. The surgically constructed neovagina in MRKH patients provides a unique opportunity for following the establishment of a vaginal microbiota. Here, we present a detailed spatiotemporal analysis of the development of the neovaginal microbiota in women with MRKH syndrome following laparoscopic peritoneal vaginoplasty based on multi-site longitudinal sampling and deep shotgun metagenomic sequencing. Our findings reveal intricate connections between the neovaginal microbiota and the pre-surgery dimple microbiota, emphasizing the pivotal roles of progressive selective forces imposed by the evolving neovaginal environment and colonization tropism of vaginal species.

Results

Study design and morphological characterization of the neovagina

In this study, 39 Chinese women with MRKH syndrome (at the time of surgery aged 17–35 years (24.85 ± 4.06 years, this format across the manuscript consistently represents mean ± standard deviation)) were recruited and followed with multi-site longitudinal samples collected before (PRE: 0–2 days pre-surgery), and four times up to 2 ~ 4 years after laparoscopic peritoneal vaginoplasty (P14D: 14.11 ± 2.04 days; P90D: 91.56 ± 7.76 days; P6/12 M: 8.79 ± 2.50 months; and P2/4Y: 42.81 ± 8.18 months) (Fig. 1a, Table 1 and Methods). In total, 456 samples were collected and analyzed with deep shotgun metagenomic sequencing, including 142 vaginal samples, 31 peritoneal fluid samples, 111 fecal samples, 108 tongue coating samples, 33 abdominal skin samples, and 31 control saline samples (Supplementary Table 1). Metadata were collected via peri-surgery clinical examination and questionnaires for neovaginal management, function, and health evaluation (Table 1 and Supplementary Data 1). All patients had normal female karyotype (46, XX). Pre-surgery estradiol measurements were all within the normal range except for patient LH042, whose hormone levels were then re-examined at P14D and turned out to be normal (Supplementary Data 1).

Before surgery, the external genital appeared normal and the vagina presented as a dimple with variable length between 1 and 3 cm, and similar to a normal vagina, coated with a glycogen-containing stratified squamous epithelium (Fig. 1b–e, Supplementary Video). During surgery, a canal between the bladder and the rectum was created and lined with dissected peritoneum. In this way, the neovaginal cavity was created. The neovagina was ~9 cm in length when first constructed and 9.4 ± 1.11 cm at P2/4Y (Table 1 and Methods). The original dimple, acting as the neovaginal opening post-surgery, formed the initiating site for spontaneous upward squamous epithelization in the neovagina. Squamous epithelium coating became discernible as early as P14D, and extended to the apex by P90D (Fig. 1b–e, Supplementary Video). It has been suggested that the neovagina matures about 12 months post-surgery[15], forming an epithelium resembling that of the dimple and the normal vagina. In our study, we observed further progression towards a normal epithelium 2–4 years post-surgery (Fig. 1b–e). Colposcopy revealed a pink and smooth mucosa (Fig. 1b), lacking the folds or 'rugae' that characterize the normal vagina[16]. Furthermore, the apex of the neovagina was not connected to a cervix and a uterus, and thus, the neovagina would not be coated with cervical mucus as in the normal vagina.

Along the progression of epithelization in the neovagina, there emerged a certain degree of similarity and relevance to the conditions found in the normal vagina of premenopausal and postmenopausal women. We have thus additionally collected data on the microbiota of the normal vagina from healthy Chinese women. In order to avoid the fluctuations during the peri-menopausal phase to confound our grouping, we excluded individuals between 45 and 55 years of age and finally included 237 premenopausal (H_R, $n = 237$, aged 39.99 ± 2.54 years) and 235 postmenopausal (H_M, $n = 235$, aged 58.09 ± 3.86 years) women, respectively, to serve as references for comparison (Fig. 1a).

Microbiota assembly in the neovagina of MRKH patients

We first characterized the pre-surgery dimple and post-surgery neovaginal microbiota of MRKH patients in comparison to those of pre- and postmenopausal women (Fig. 2 and Supplementary Fig. 1a). The most abundant species in the dimple, including *Lactobacillus iners*, *Prevotella timonensis*, *Atopobium vaginae*, *Prevotella bivia*, *Lactobacillus crispatus*, *Prevotella disiens*, and *Gardnerella vaginalis* (Fig. 2a), comprised over 75% mean relative abundance and were also among the top 10 species in the healthy references (Supplementary Data 2). Overall, the pre-surgery dimple microbiota exhibited some resemblance to both H_R and H_M women, with a high abundance of *L. iners* as in H_R women and high representation of some BV-associated species including *G. vaginalis* and *Prevotella* spp. as in H_M women (Fig. 2). Considering species accounting for >1‰ mean relative abundance, bacteria harbored in the pre-surgery dimple and the post-surgery neovagina of MRKH patients were largely common 'native' vaginal species (Supplementary Fig. 1b). Nevertheless, the overall compositional structure at different stages differed significantly from the healthy references (Supplementary Fig. 1c), and consistently exhibited more species and higher Shannon index except for the early assemblage at P14D (Supplementary Fig. 1d, e).

We characterized the assembly progression more detailed in relation to the pre-surgery dimple microbiota (Fig. 2 and Supplementary Fig. 1a). At P14D, the local environment in the neovagina was under the influence of dramatic changes and peri-surgery medications, including laxatives, paraffin oil, and prophylactic antibiotics (cefoxitin in combination with metronidazole or clindamycin) (Table 1). At this time-point, we observed a precipitous decline in species number (Two-sided Wilcoxon rank-sum test, $P < 0.0001$) (Supplementary Fig. 1e) and in the relative abundances of *L. iners, P. timonensis*, and *A. vaginae* (Two-sided Wilcoxon rank-sum test, P adjust <0.1) (Fig. 2a, b and Supplementary Data 3), but in some patients blooming of several species intrinsically insensitive or resistant to the applied cephalosporins, including *Enterococcus faecalis, Ureaplasma parvum*, *Ureaplasma urealyticum*, and *Mycobacterium hominis*[24,25] (Fig. 2g). Despite

https://doi.org/10.1038/s41467-024-52102-1

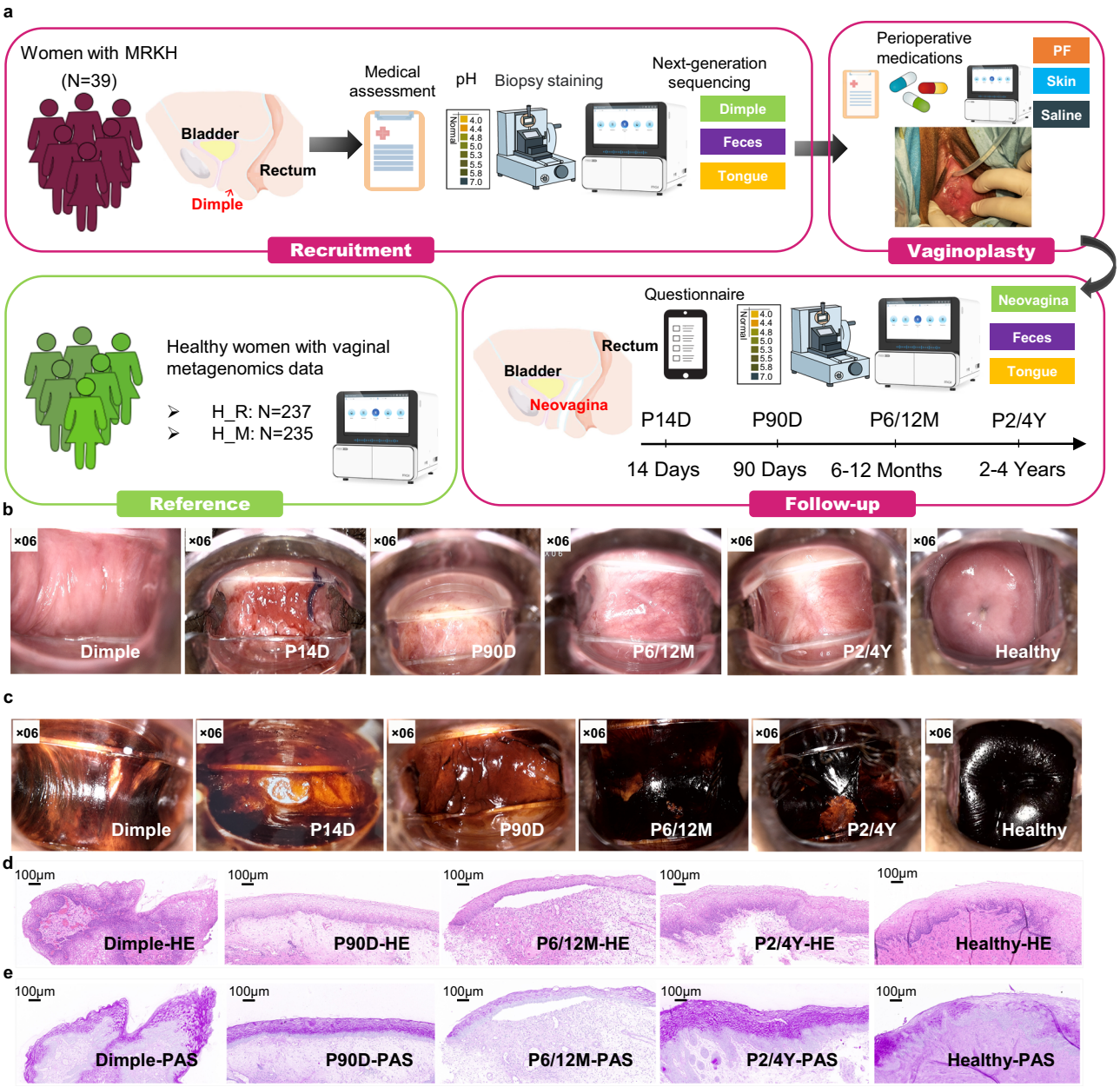


Fig. 1 | Schematic overview of the study design and neovaginal epithelium development. a Multi-site longitudinal sampling in 39 MRKH syndrome patients subjected to laparoscopic peritoneal vaginoplasty. Samples from the vagina (if not otherwise indicated, MRKH patient's vagina refers to the pre-surgery dimple and the post-surgery neovagina collectively for simplicity), stool, and tongue coating were collected 0–2 days before the surgery with no prior perturbation, and 14 days, 3 months, 6–12 months, and 2–4 years after the surgery, with time points denoted as PRE, P14D, P90D, P6/12 M, and P2/4Y, respectively. Peritoneal fluid (PF), abdominal skin samples, and control saline were collected during the laparoscopic Davydov procedure. Normal vaginal microbiota data of 472 healthy adult women were included as references for the neovaginal microbiota. **b** Colposcopic images of the pre-surgery dimple, neovaginal apex (P14D, P90D, P6/12 M, and P2/4Y) and healthy cervix. **c** Iodine staining of the pre-surgery dimple, neovaginal apex (P90D, P6/12 M, and P2/4Y) and healthy cervix. Cells of vaginal origin contain glycogen, which is stained brown or dark brown by iodine. Biopsies taken from the vaginas of three MRKH women (dimple, the upper third neovagina at P90D, P6/12 M, and P2/4Y), and the normal vagina of three control women (the upper third). After formalin fixation, paraffin embedding and sectioning (section thickness: ~3 μm), the upper, middle and lower sections of each biopsy were subjected to Hematoxylin and eosin (HE) staining (**d**) and PAS staining (**e**), and similar results were obtained. HE-stained vaginal biopsies (**d**) showing fibrous stroma lined with a stratified squamous epithelium in all samples. PAS-stained vaginal biopsies (**e**) showing glycogen-accumulated in a stratified squamous epithelium in all samples.

of the drastic change at P14D, intra-individual dissimilarities between adjacent time points were still significantly lower than inter-individual dissimilarities (Two-sided Wilcoxon rank-sum test, $P < 0.05$) (Fig. 2h), indicating an inherent continuity of the microbiota during assembly. At P90D, when the squamous epithelium reached the apex of the neovagina, the microbiota developed into a poly-microbial state with BV-associated species becoming abundant (Fig. 2c). Of note, the genus *Prevotella*, primarily represented by *P. timonensis* and *P. bivia* (Fig. 2c), exhibited a significant increase to account for ~29% mean relative abundance at P90D (Two-sided Wilcoxon rank-sum test, *P* adjust <0.1) (Supplementary Data 3). The mean relative abundance of *G. vaginalis* and *A. vaginae*, which are critical for the manifestation of BV[26,27], also increased, with *G. vaginalis* even becoming the most abundant species among the 11 patients we managed to follow at P6/12 M during the

Table 1 | Summary of the most relevant metadata in the 39 women with MRKH syndrome

Metadata category	Metadata	Value in the MRKH cohort
Age	Age at vaginoplasty	24.846 ± 4.064[a]
MRKH type	MRKH type	Type I: 26; Type II: 13
pH	PRE pH	pH 4.8: 1; pH 5.8: 1; pH 7: 1
	P14D pH	pH 4.8: 1; pH 5–5.8: 10; pH 6: 1; pH 7: 24
	P90D pH	pH 5–5.8: 8; pH 7: 9
	P2/4Y pH	pH 4–4.8: 18; pH 5: 4; pH 7: 5
Sampling time	Pre-surgery days of PRE	0.205 ± 0.528[a]
	Post-surgery days of P14D	14.105 ± 2.038[a]
	Post-surgery days of P90D	91.556 ± 7.757[a]
	Post-surgery days of P6/12 M	263.636 ± 72.207[a]
	Post-surgery days of P2/4Y	1275.731 ± 246.233[a]
Peri-surgery medications	Pre-surgery bowel preparation	Oral laxatives: 39
	Types of inpatient peri-surgery antibiotic usage	Cefoxitin Sodium and metronidazole: 38; Clindamycin: 1
	Days of peri-surgery antibiotic usage	4.46 ± 0.854[a]
	Lubricating material on dilator during the initial week	Paraffin oil: 39
	Post-surgery painkiller usage	No: 37; Yes: 2
	Post-surgery painkiller usage frequency	No: 37; Once: 2
Questionaire at P2/4Y	Vaginal length at P2/4Y (cm)	9.359 ± 1.106[a]
	Expand frequency (mould or intercourse) within a month	No: 6 ;1 time a month: 5; 2-3 times a month: 3; 1 time a week: 7; 2 times a week: 5; 1 time every two days: 1
	Mould types within a month	No use: 12; Plastic: 10; Rubber: 5
	Condom use with a month	No: 12; Yes: 15
	Lubricating oil use with one month	No: 8; Yes: 19
	Sexual intercourse after surgery	No: 5; Yes: 22
	Vaginal douche with one month	No: 24; Yes: 3
	Medication usage within a month	No: 25; Yes: 2
	Sexual activity within a month	No: 12; Yes: 15
Morphological test at P2/4Y	BV assessment by Nugent score	BV: 7; Intermediate BV: 9; Normal: 10

The superscript 'a' in the table represents mean ± SD.

COVID pandemic (Supplementary Fig. 1a). Together, the intermediate stages at P90D and P6/12M were characterized by low levels of *Lactobacillus* spp. (<16% relative abundance on average) and high levels of BV-associated anaerobes with only few patients presenting a microbial composition dominated by *Lactobacillus* spp. (P90D 3/27; P6/12M 0/11) (Fig. 2g).

Two to four years post-surgery (P2/4Y), the neovaginal microbiota, intriguingly, did not carry over the BV-like state, but assumed a state most close to the pre-surgery condition characterizing the dimple, where the relative abundance of just one (*P. buccalis*) out of the top 20 species differed significantly (Two-sided Wilcoxon rank-sum test, *P* adjust <0.1) (Fig. 2a, d, i, Supplementary Fig. 1c, and Supplementary Data 3). At this time-point, *L. iners* had colonized the neovagina of all the recalled patients. The relative abundance of *L. iners, L. crispatus*, and the genus *Lactobacillus* altogether, significantly increased compared to the early-intermediate stages (Two-sided Wilcoxon rank-sum test, *P* adjust <0.1) (Supplementary Data 3), rendering the *Lactobacillus* level at P2/4Y comparable to the dimple level (40.8% at PRE and 38.3% at P2/4Y), but lower than that in the normal vagina of H_R women (75.6%) and higher than in H_M women (17.1%) (Supplementary Data 2). The ratio of *Lactobacillus*-dominated patients at P2/4Y was close to PRE, but with more attributed to *L. crispatus* and *L. gasseri* (Fig. 2g), which are considered to have a better protective effect on vaginal health than *L. iners*[8,28,29]. Concurrently, the neovaginal pH also showed an acidifying trend, where the proportion of women having a vaginal pH lower than 5 increased over time (P14D 1/36; P90D 10/26; P2/4Y 15/20) (Supplementary Data 1). Additionally, compared to P90D, there was also a significant increase in the relative abundance of *P. timonensis* (Two-sided Wilcoxon rank-sum test, *P* adjust <0.1) (Fig. 2c, d, Supplementary Data 3).

To gain insight into the neovaginal microbiota assembly from an ecological perspective, we performed community typing based on Dirichlet Multinomial Mixtures providing evidence for three distinct clusters (Fig. 3a and Supplementary Fig. 2a). *L. iners, U. parvum*, and *P. bivia* were major contributors to cluster 1, *P. timonensis, Finegoldia magna*, and *Peptoniphilus harei* to cluster 2, and *L. iners* to cluster 3 (Fig. 3a). Interestingly, cluster 1 dominated from pre-surgery to P90D. Then at P2/4Y, the microbiota showed an obviously biased transition towards cluster 2 (19 out of 27, 70.4%) (Fig. 3b), with 7 out of 13 (53.9%), 8 out of 9 (88.9%), and 4 out of 5 (80%) individuals initially classified as clusters 1, 2, and 3, respectively, adopting this cluster at P2/4Y (Fig. 3c and Supplementary Fig. 2b). Thus, individuals belonging to cluster 2 at PRE exhibited significantly less deviations from their initial state at P2/4Y compared to individuals belonging to cluster 1 or 3. (Fig. 3d). Despite that none of the *Lactobacillus* spp. appeared among the top three contributors to cluster 2, nearly half (13/27) of the individuals harbored communities dominated by *Lactobacillus* spp. at P2/4Y (Fig. 2g). Further, analyses of co-occurrence networks demonstrated a developmental process mirroring that observed in the compositional structure and DMM community typing (Fig. 3c, e). Specifically, the network developed from a sparse structure with few interactions detected at P14D, through the intermediate stage at P90D with increasing significant interactions, to a structure at P2/4Y with the highest number of significant interactions (Fig. 3e and Supplementary Fig. 3). Along this process, while an increasing number of interactions characterizing the pre-surgery microbiota were re-captured, some new

https://doi.org/10.1038/s41467-024-52102-1

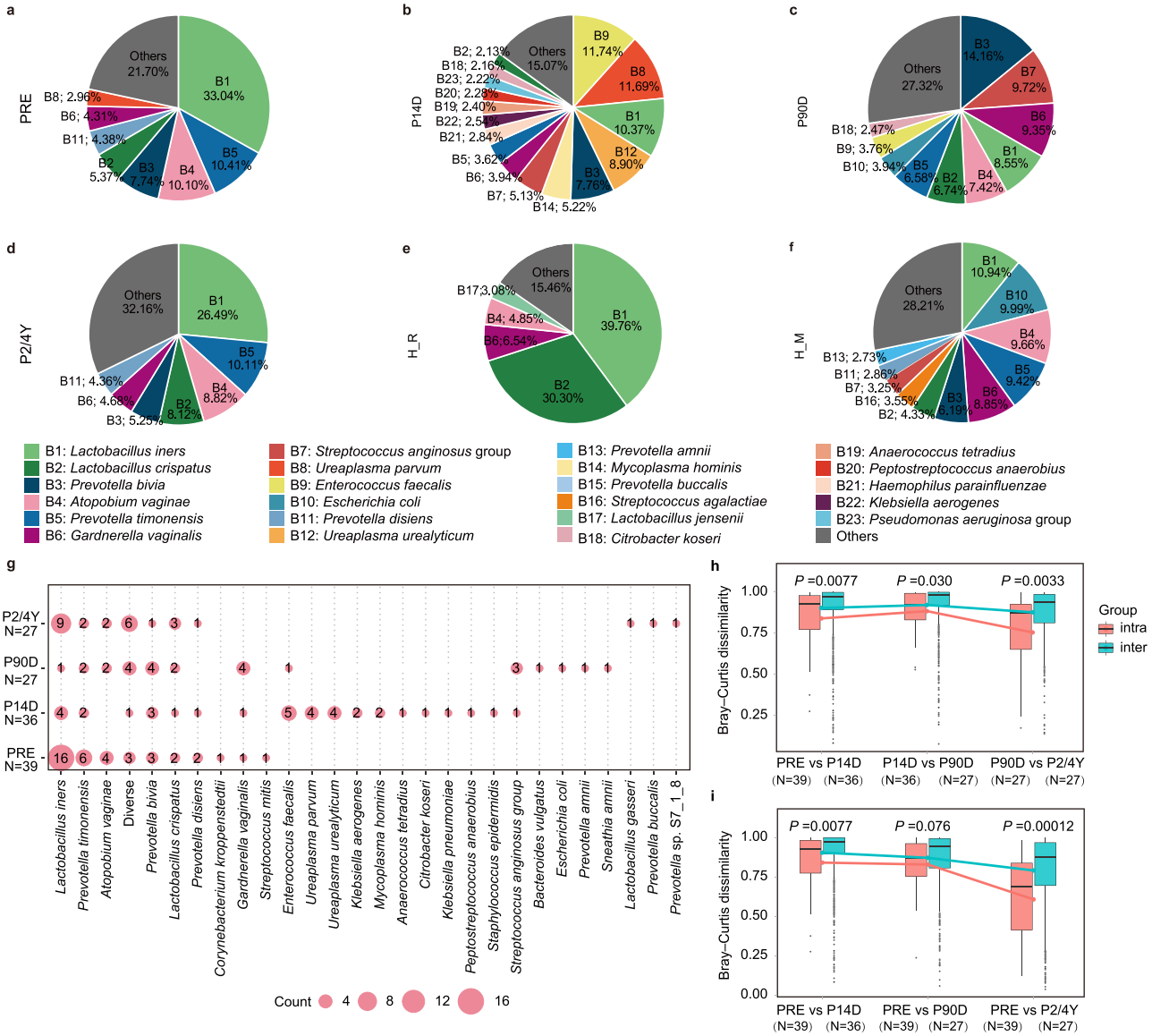


Fig. 2 | Compositional developments along the assembly of the microbiota in the neovagina in MRKH patients. Overall composition of the vaginal microbiota in MRKH women at four stages, including PRE (**a**), P14D (**b**), P90D (**c**), and P2/4Y (**d**), premenopausal women (H_R) (**e**) and postmenopausal women (H_M) (**f**). **g** The distribution of the vagitypes in MRKH patients at different time-points. Vagitype was assigned according to the species with the highest relative abundance in each vaginal sample. If the maximal relative abundance was less than 30%, the type was defined as 'Diverse'[49]. Circles are scaled according to the number of patients for the given vagitype at the given time-point. Intra- and inter-individual Bray–Curtis dissimilarities in the vaginal microbiota of MRKH patients between adjacent time points (**h**) and between PRE and different post-surgery time-points (**i**). p, Wilcoxon rank-sum test. Box plot elements include: Centerline for median, box limits for upper and lower quartiles; whiskers indicate 1.5× interquartile range; and points for outliers.

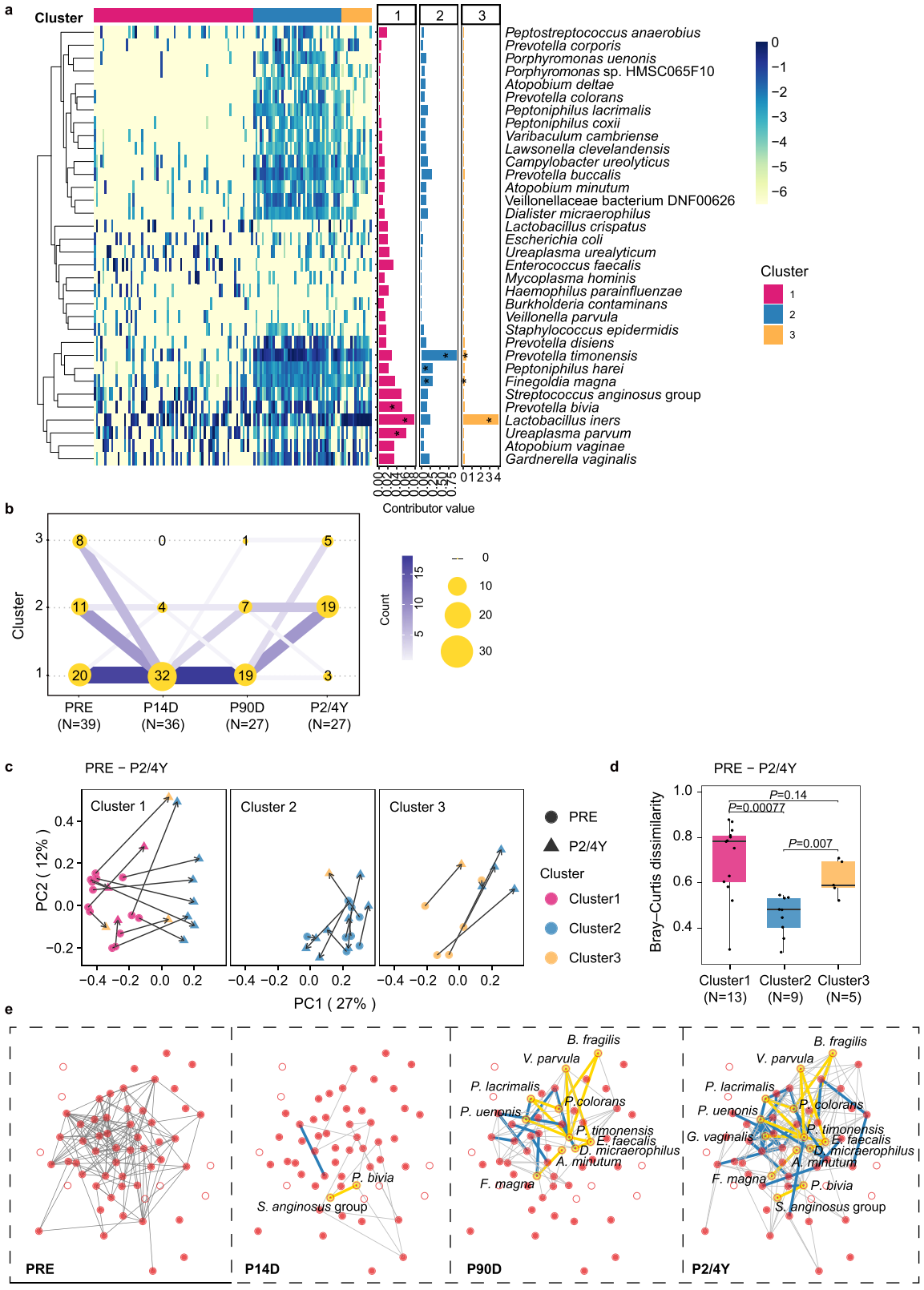
interactions formed at P14D and P90D lasted until 2–4 years post-surgery (Fig. 3e). Notably, while cluster 2, most prominently characterized by *P. timonensis* in the DMM-based community typing, became the dominant type in the cohort only at P2/4Y, a nascent network structure with *P. timonensis* at its core was already taking shape at P90D. Together, these results demonstrated that while the neovaginal microbiota at P2/4Y to a large extent resembled the homeostatic pre-surgery dimple microbiota, it also displayed unique characteristics. This may be pointing to a role for resilience and evolution in the plasticity of the microbial ecosystems with regard to the conditions of the evolving inhabiting environment.

The origin of species colonizing the neovagina

To infer the origin of species colonizing the neovagina, we carried out source tracking analyses first with a coarse-grained approach using FEAST[30] to estimate potentially contributing sources, then with a fine-grained approach using StrainPhlAn[31] for putative strain transmission identification.

FEAST estimated that the largest percentage of the neovaginal microbiota was attributed to the pre-surgery dimple microbiota, steadily increasing from 40.76% at P14D to 79.30% at P2/4Y (Supplementary Fig. 4a), followed by the fecal microbiota accounting for an average of 22.09%. The peritoneal fluid, in direct contact with the peritoneum for covering the dissection during neovagina construction reached a maximum of 2.17% at P90D and decreased to a marginal level at P2/4Y. In the early-intermediate stages, the tongue coating microbiota was also predicted to account for minute percentages. Interestingly, while the percentages potentially attributed to peritoneal fluid, the oral cavity, and unknown sources decreased, the estimates for the skin microbiota and environmental controls became noticeable at P2/

https://doi.org/10.1038/s41467-024-52102-1



4Y. Overall, the FEAST estimation is highly consistent with the mean relative abundances accounted for in each sample type by the 10 most abundant species in the neovagina (Supplementary Fig. 4b).

Strain-level analyses showed that for many species, cross-body site strains exhibited significantly lower phylogenetic distances within than between individuals (Supplementary Fig. 5). This was particularly pronounced for dimple versus neovagina, and stool versus neovagina, suggesting potential transmission routes from the dimple and the stool. We counted the number of patients with strain sharing for the neovagina and classified the species into four categories (Fig. 4 and Supplementary Figs. 6,7). 1) Species with only intravaginal (dimple-neovagina) strain sharing (4/75, 5.3%) including the widely regarded keystone species for a healthy vaginal microbiota[13,14], *L. crispatus*, suggesting a primary origin from the dimple in this observational

Fig. 3 | Ecosystem developments along the assembly of the microbiota in the neovagina of MRKH patients. a Heatmap distribution of the relative abundances of the 15 most abundant species within each of the three DMM clusters. The contribution of each species to each cluster is distinctly represented in three separate bar graph panels on the right-hand side. An asterisk (*) in each bar graph panel signifies the top three contributing species within each respective cluster. **b** DMM type transition of the vaginal microbiota in MRKH patients from PRE to P2/4Y with circles and lines scaled according to the number of patients. **c** PCoA showing the shifts in the vaginal microbiota from PRE to P2/4Y. Patients are divided into three panels based on the DMM type at PRE. Red, blue, and green indicate DMM type 1, 2, and 3, respectively. **d** Intra-individual Bray–Curtis dissimilarities between the vaginal microbiota at PRE and P2/4Y among MRKH patients within three DMM types. Two-sided Wilcoxon rank-sum test were used. Box plot elements include: Centerline for median, box limits for upper and lower quartiles; whiskers indicate 1.5× interquartile range; and points for outliers. **e** Co-occurrence network between bacterial species at each time-point (PRE, P14D, P90D, and P2/4Y) with nodes representing the species and edges representing significant correlations ($P < 0.05$). Blue edges indicate interactions characterizing the pre-surgery state and re-captured along the assembly in the neovagina. Yellow edges indicate new interactions formed in the neovagina at P14D and P90D and lasted until 2–4 years post-surgery. Nodes outlined in yellow, along with the corresponding microbial species, have been accentuated to delineate the newly identified interactions post-surgery, as indicated by the yellow edges.

setting. 2) Species exhibiting both intravaginal and extravaginal (stool, skin and tongue coating) sharing (21/75, 28%). The species of this category accounted for most strain sharing events, which were primarily observed for the dimple and the fecal microbiota, and occasionally for the skin microbiota as in the case of *L. iners* and *Finegoldia magna*. Of these species, *L. iners*, *U. parvum, P. bivia, S. antinosus group*, and *P. disiens* had the highest sharing counts and sharing rates (50–100%) for both the dimple and the stool. This suggests that these species are well adapted to both the gut and the vaginal environments. 3) Species with only extravaginal sharing (35/75, 46.7%). In this category, all strain sharing events, except for one, were with the stool, and most exhibited low counts, but high rates, especially for the species typically found in the gut microbiota[32]. 4) Species without detected strain sharing with all surveyed sources (15/75, 20%). Comparing strains at P14D with pre-surgery strains in the dimple and the fecal microbiota, we observed sharing events more often with the dimple, and solely with the dimple in the case of *L. iners, U. parvum*, *G. vaginalis, Sneathia amnii*, and *P. amnii* (Fig. 5a). Sharing with the pre-surgery fecal microbiota was however less and often comprised *Prevotella* species (Fig. 5a). Additionally, bacteria- and host-dependent long-term strain maintenance was also observed in many cases (Fig. 5b). Of particular note, the strains of *L. crispatus* in two of the four patients with strain sharing detected, persisted up to P2/4Y (Supplementary Fig. 6). Among the six patients with long-term strain preservation for more than one species, LH046, LH024, and LH021 never reported vaginal intercourse (Supplementary Data 1). Together, these results suggested that while the fecal microbiota made a significant contribution to colonize the neovagina, either as a direct source or a relay, the dimple microbiota served as the major reservoir. This points to colonization tropism and niche specificity for vaginal species.

Factors shaping the neovaginal microbiota assembly in MRKH patients

The vaginal microbiota in MRKH patients exhibited substantial heterogeneity (Supplementary Fig. 1f). To investigate factors influencing the development of the neovaginal microbiota, we first analyzed the impact of the variable metadata. We found that MRKH type, even though it did not affect the overall structural variation (PERMANOVA test, P adjust >0.05), significantly correlated with the relative abundance of several species, including *Atopobium minutum*, *Bifidobacterium dentium*, *P. timonensis*, and *Streptococcus agalactiae* (Fig. 6a).

Despite the temporal increase in the similarity with the pre-surgery dimple microbiota on the cohort-level (Fig. 2i, Supplementary Fig. 1c), within-individual similarities were not consistently observed across the patients (Fig. 3c, Supplementary Fig. 8). Given that the dimple and the fecal microbiota both significantly contributed to the species pool for colonizing the neovagina, we tested the impact of the respective pre-surgery status of individual species (presence/absence or relative abundance) in the dimple or the stool on their relative abundance in the neovagina. Notably, all significant correlations showed a positive relationship and most were associated with the prior status in the dimple (10 solely with dimple, 3 solely with stool and 3 with both) (geeglm[33], P adjust <0.1) (Fig. 6b). Thus for *L. crispatus*, prior colonization in the dimple not only largely determined acquisition of this species in the neovagina, especially in the 1st post-surgery year (Supplementary Fig. 9a), but also significantly affected its post-surgery relative abundance (Fig. 6b). For other species, such as *U. parvum, P. bivia*, and *A. vaginae*, although strain sharing were detected with both the dimple and the stool, their relative abundances in the neovagina were only significantly correlated with their prior status in the dimple (Figs. 4 and 6b, and Supplementary Fig. 9b, c). Interestingly, such an association pattern was even observed for some species with only stool-neovagina strain sharing events, such as *Faecalibacterium prausnitzii* (Figs. 4 and 6b). Of the six species exhibiting significant association with the pre-surgery levels in the stool, including *L. iners, P. corporis, P. copri, Peptoniphilus duerdenii, Anaerococcus vaginalis*, and *Anaerococcus prevotii*, *L. iners* was most intriguing (Fig. 6b). It was highly prevalent and most abundant at both PRE and P2/4Y, and also showed higher rate of putative transmission from the dimple. Nonetheless, its relative abundance in the neovagina was not correlated with its prior status in the dimple (Fig. 4 and Supplementary Fig. 9d).

Other species, such as *P. timonensis, P. disiens, G. vaginlis, C. ureolyticus, F. magna*, and *Peptoniphilus harei*, although exhibiting frequent sharing with both the dimple and the stool, showed no significant association with their pre-surgery status in either the dimple or the stool (Figs. 4 and 6b, Supplementary Fig. 9e, f and Supplementary Data 4). However, multivariate analysis using random forest model to predict temporal changes in relative abundance based on the bacterial composition in the preceding time point showed that these species, followed by *L. iners*, had the highest prediction accuracy (Fig. 6c), underscoring the complexity in the interplay between the dimple microbiota and the neovaginal microbiota. More importantly, the predictability results resonate with the biased transition towards cluster 2 at P2/4Y, with the most predictive species, including *P. timonensis, F. magna*, and *P. harei*, aligning with the top 3 contributors to cluster 2 in the DMM-based community typing (Fig. 3a–d).

Discussion

The surgically created neovagina in MRKH patients through laparoscopic peritoneal vaginoplasty is characterized by unique features, making it particularly interesting for the study of microbial community assembly. When initially created, the neovagina represents a nearly blank ecological niche, facilitating the influx of microbes from adjacent anatomical regions with the microbiota of the dimple and the anus area constituting the primary species pools. Concurrently, the local environment undergoes rapid transformation from an initial peritoneal lining to a squamous epithelium coating becoming discernible as early as P14D, reaching the apex by P90D, and still showing further progression until P2/4Y. The glycogen within the proliferating squamous cells serves as a vital nutrient source for the microbiota[34], creating a dynamic nutritional milieu posing a progressive selection force on the colonization of the arriving microbes. Influenced by the

https://doi.org/10.1038/s41467-024-52102-1

Fig. 4 | Strain-level source tracking of the bacteria colonizing the neovagina. The surveyed potential sources of neovagina in strain-level analysis included pre-surgery dimple, stool, skin, and tongue coating. Species were classified into four categories according to their potential transmission routes and indicated by different colors: 1) only intravaginal (pre-surgery dimple-neovagina) strain transmission (in green); 2) both intravaginal and extravaginal (stool/skin/tongue coating-neovagina) transmission (in orange); 3) only extravaginal transmission (in violet); 4) no strain transmission from all surveyed sources (in pink). Solid cycles are scaled according to the number of patients with strain transmission events detected, and colored based on strain transmission rate, which is defined as the number of patients with transmission (same strain) detected divided by the number of patients with transmission (same strain) and non-transmission (different strain) detected between the potential source and the neovagina and shown in fractions next to the circle.

https://doi.org/10.1038/s41467-024-52102-1

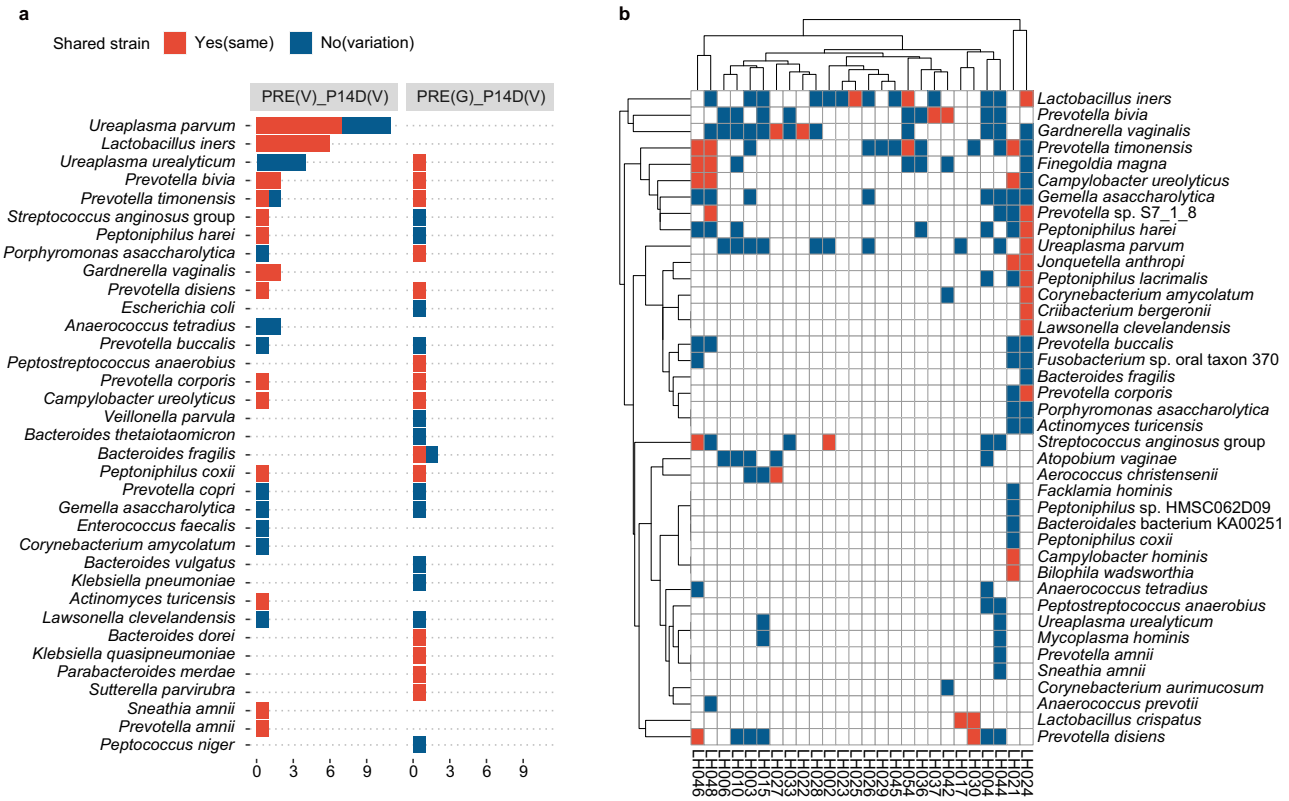


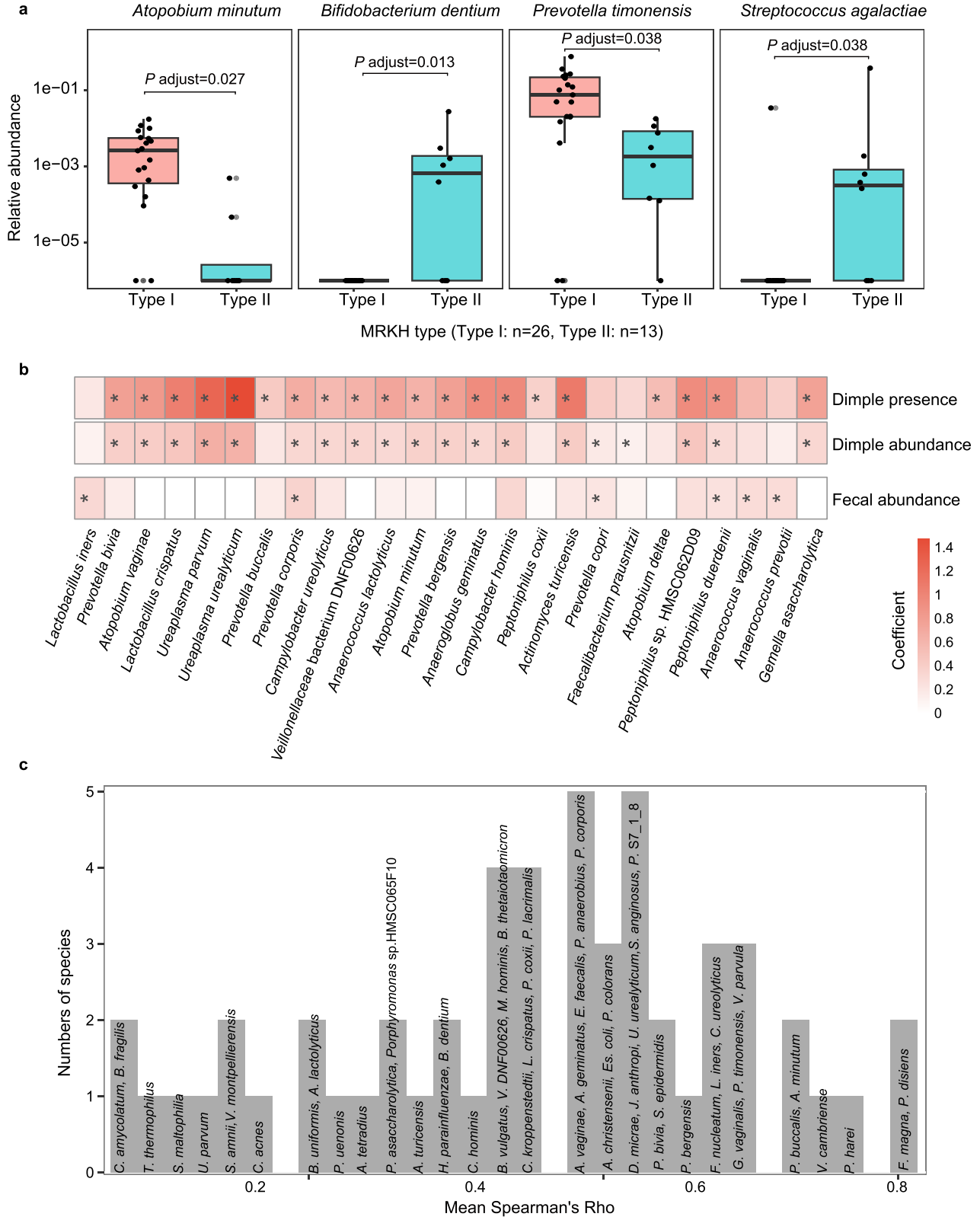
Fig. 5 | Temporal analysis of strain transmission in the neovagina. a Putative strain transmission for the newly constructed neovagina detected at P14D from the pre-surgery dimple (PRE(V)_P14D(V)), and pre-surgery stool (PRE(G)_P14D(V)), respectively. **b** Putative long-term maintenance of strains presented at PRE or acquired within one year post-surgery until 2–4 years post-surgery. In both (**a**) and (**b**), transmission (same strain) is indicated in red, non-transmission (different strain) is indicated in blue.

transient nature of the local environment and individualized responses to peri-surgery medications, the early assemblage at P14D exhibits stochastic characteristics with deviating microbial structures, where the bacterial species exhibited rather few interactions, accompanied with blooming of species intrinsically insensitive or resistant to the applied cephalosporins. Among these, *E. faecalis* possibly arose from random colonization of strains present in the hospital environment. At P90D and P6/12 M, when the nutritional conditions had significantly improved, the neovaginal microbiota began to resemble that of a normal vaginal microbiota, though exhibiting a BV-like structure similar to that found in H_M women. While it is difficult to escape from BV-like states in a normal vagina[35,36], we observed that as the squamous epithelium further developed, the neovaginal microbiota at P2/4Y evolved into a structure with a composition resembling that of the homeostatic pre-surgery dimple microbiota. These findings suggest that the neovaginal microbiota assembly follows a pattern governed by the initial species pool in the vaginal dimple and by the local epithelium's contribution of nutritional sources, which together with host physiological factors promote selective colonization. Similarly, in the assembly of the infant gut microbiota, infants born by caesarean section harbor, compared to vaginally delivered infants, a perturbed gut microbiota at early days, but acquire a microbiota composition showing similarities to vaginally delivered infant after about one year[21]. In the case of a penile skin-lined neovagina in transgender women, it is thus not surprising that it was found to be characterized by a microbiota similar to that of an uncircumcised penile (coronal sulcus) microbiota[5,37,38] since such a neovagina has been shown to retain the morphological and histochemical characteristics of the grafted skin[16]. Nevertheless, we cannot rule out the possibility of priority effects[39] in our study, considering that the neovagina at P14D had already captured many important vaginal species also present in the dimple.

Ecological network analyses have demonstrated general patterns in the gut microbiota of healthy individuals being absent in disrupted states such as recurrent *Clostridium difficile* infection[40]. Similarly, the neovaginal microbiota at P2/4Y showed high consistency among the patients based on DMM community typing. *Prevotella* was shown to be most heritable among vaginal bacteria and acts as a hub for vaginal microbiota influenced by host genetics and further associated with obesity[41]. Interestingly, though cluster 2, characterized prominently by *P. timonensis*, prevailed only until P2/4Y, a nascent network structure with this species at its core was already established at P90D. This may be pointing to a role for *P. timonensis* as a keystone taxon in the assembly of the neovaginal microbiota in MRKH patients setting a possible foundation for future ecological engineering efforts. In the current study, colonization with *L. crispatus* in the neovagina primarily occurred in women harboring this species in the pre-surgery dimple, and the relative abundance strongly associated with its prior status in the pre-surgery dimple. In addition, we also observed an increase in the proportion of *L. crispatus*-dominated vagitype at P2/4Y compared to PRE. This may reflect a more growth supportive environment within the deeper neovaginal cavity. Considering potential priority effects[39] and the recent findings that the long-term persistence of individual strains in the vagina does not necessarily correspond to the consistency in species composition[42], topical introduction of *L. crispatus* strains in the dimple or the neovagina early post-surgery may be able to promote the colonization of this species and confer benefits on the outcome of neovaginal microbiota assembly following vaginoplasty.

Several aspects merit further discussion. Firstly, while external measures undertaken were largely consistent for standard peri-surgery management protocols, individual responses can still vary. Our data are limited for assessing the impact of many of these factors. Secondly, the Covid pandemic severely hindered sample collection at P6/12 M.

https://doi.org/10.1038/s41467-024-52102-1



Still, we managed to obtain samples and data from 70% (27/39) of the patients in the last recall at P2/4Y. Of these patients, there were some bias with regard to the pre-surgery dimple microbiota in that non-*Lactobacillus*-dominated patients had a higher recall ratio than *Lactobacillus*-dominated patients (76.2% [16/21] vs. 61.1% [11/18]). Since the abundance of *Lactobacillus* at P2/4Y significantly correlated with the dimple status, we speculate that the true ground average *Lactobacillus* level would be even higher. Thirdly, our study is not generalizable to vaginal microbiota assembly in women with a normal vagina or transgender women due to differences in physiological conditions in peri-surgery drug regimens, and in inflammatory responses resulting from surgery. Fourthly, MRKH patients generally have normal

Fig. 6 | Factors shaping the neovaginal microbiota in MRKH patients. a Pair-wise association for individual species significantly correlated with MRKH type. Two-sided Wilcoxon rank-sum test was performed with *P* values adjusted with the Benjamini–Hochberg Procedure (FDR). Box plot elements include: centerline for median, box limits for upper and lower quartiles; whiskers indicate 1.5× interquartile range; and points for outliers. **b** Pre-surgery status of individual species (presence/absence or relative abundance) in the dimple or the stool significantly correlated with their relative abundance in the neovagina identified by geeglm. An asterisk (*) indicates *P* adjust <0.1 (Supplementary Data 4). **c** Distribution of the prediction accuracy for predicting the changes in the relative abundance of individual species in the neovaginal microbiota, illustrated by the correlation coefficients between predicted and detected values from random forest models with 10-fold cross validation in x-axis. Due to space constraints, 'Veillonellaceae bacterium DNF00626' is abbreviated to 'V. DNF00626', '*S. anginosus* group' to '*S. anginosus*', '*Prevotella* sp. S7_1_8' to 'P. S7_1_8', and '*D. micraerophilus*' to '*D. micrae*'.

development of ovaries and steroid hormone production. However, their hormonal cycles were difficult to assess because MRKH patients do not have menses, which may in part explain why we did not find correlations between the neovaginal microbiota and peri-surgery hormone levels. Lastly, there might be some bias in our reference data given that age, geographic location, ethnicity, socioeconomic status, and disease conditions, can all impact *Lactobacillus* levels and overall structure of the vaginal microbiota.

Methods

Participants

The study was approved by the ethics committee at Peking Union Medical College Hospital in Beijing (PUMCH) (ZS-1657), Luohu Hospital in Shenzhen (LHQRMYY-KYLL-018), and BGI-Research (BGI-IRB 21061). 39 patients diagnosed with the MRKH syndrome (aged 17–35 years, 24.98 ± 4.64 years at the time of surgery) and underwent the laparoscopic Davydov procedure between June 2018 and Dec 2020 were enrolled at the two hospital centers in China. The surgical procedure was performed by experienced gynecologists in Luohu Hospital to avoid bias of study site. All participants met the inclusion criteria: with normal female karyotype (46, XX), and without any other types of surgical vaginoplasty before enrollment. Of them, 26 (66.7%) were with type I and 13 (33.3%) were with type II MRKH syndrome[43]. Ten (25.6%) patients had concomitant skeletal anomalies (Idiopathic Scoliosis (IS) in 8, Klipple-Feil syndrome in 2), four (10.3%) had concomitant renal anomalies (Renal Agenesis (RA) in 2, Ectopic Kidney in 1, Renal Dysplasia in 1), two had Ventricular Septal Defect (VSD) (1 with IS and RA), one had conductive hearing loss (with IS, RA and VSD). The study was explained in detail to all participants and written informed consent was obtained from all.

Surgical procedures and post-surgery management

The steps of the laparoscopic Davydov procedure have been described in detail previously. In the laparoscopic Davydov procedure[44], a 10-cm-long needle was inserted at the center of the vestibule, followed by an injection of adrenaline diluted in saline into the rectovesical space. An additional 100–300 mL isotonic saline solution was injected into the space when the needle was near the bottom of the pelvic peritoneum. Under laparoscopy, the peritoneum floor could be seen bulging gradually and eventually the pelvic peritoneum separating from the wall. Then a tunnel between the rectum and the bladder was created. With the use of a peritoneum pusher the bulging pelvic peritoneum was then pushed downward until the latter could be seen from the vestibular incision. The peritoneum was then incised transversely, and the incision margin of peritoneum sutured to the vaginal opening. A purse-string stitch of the bladder peritoneum and rectal peritoneum was performed to close the top of neovagina. After surgery, a soft mold (9 cm in length and 3 cm in width) made of 1–2 vaseline gauze (each gauze being 90 cm long and 6 cm wide) was placed in the neovagina. 5–7 days after surgery, the soft mold was removed and regular vaginal dilation (5–10 min a day for the first 3 months post-surgery) with a silicone dilator was performed. Three months post-surgery, the patient was recommended to start trying vaginal intercourse if she had a sexual partner and gradually reduce the frequency of dilation to 2–3 times a week, 5–10 min per time to prevent contraction of the neovagina until the patient had regular sexual activities.

Peri-surgery medications

Oral laxatives (Polyethylene Glycol-Electrolyte Powder) were given for bowel preparation the day before surgery. 38 patients were intravenously infused with cefoxitin sodium half an hour pre-surgery and cefoxitin sodium and metronidazole for 3–7 days post-surgery. One patient with cephalosporin allergy was treated with clindamycin half an hour pre-surgery and for 4 days post-surgery. Nonsteroidal anti-inflammatory drugs (NSAIDs) were used to relieve mild-to-severe pain if necessary. During vaginal expansion, paraffin oil was applied to dilators to alleviate pain if needed.

Sample collection

Before surgery (PRE) and 14 days after surgery (P14D), the vaginal (pre-surgery dimple and post-surgery neovagina), stool and tongue samples were collected from the 39 recruited patients by the hospital medical staff. During surgery, the peritoneal fluid, abdominal skin and saline were collected by the medical staff. For the next three time points (90 days, 6–12 months, 2–4 years), Vaginal samples were collected by the medical staff when patients returned to the hospital for a clinical examination three-month post-surgery visit (P90D), and vaginal samples were self-collected at 6–12 months post-surgery (P6/12M) and 2–4 years post-surgery (P2/4Y); Stool and tongue samples were self-collected in the morning. All samples were immediately placed into a stabilizing reagent enabling preservation of the microbiota at room temperature[45], and then stored at −80 °C before transfer to the laboratory on dry ice. All participants completed an online questionnaire covering the full range of demographic characteristics at each sample collection.

After obtaining ethical approval and written informed consent, three women (aged 22, 25, and 27) with normal vaginal development were enrolled as controls. These women tested positive for human papillomavirus and required colposcopy at the time of enrollment. Biopsies were taken from the upper third of their vaginas. Concurrently, biopsies were taken from the vaginal dimple during surgery and the upper third of the neovaginas in three MRKH patients at 3 months post-surgery (P90D), 9 months post-surgery (P6/12 M), and 3 years post-surgery (P2/4Y), respectively. The samples were processed by formalin fixation, paraffin embedding, and sectioning at approximately 3 μm thickness. Subsequently, the upper, middle and lower sections of each biopsy were stained with hematoxylin-eosin and PAS stains.

The vaginal microbiota from 472 healthy adult Chinese women

To compare the vaginal microbiota of MRKH patients with the normal vaginal microbiota, we randomly selected 472 adult Chinese women (aged 36–65 years, 49 ± 9.6 years) from the Peacock cohort[46] who displayed no abnormalities in routine gynecological examinations. The exclusion criteria were: age between 45 and 55 years, positive for any of the following gynecological conditions including bacterial vaginosis, vulvovaginal candidiasis, *Trichomonas vaginalis*, and positive HPV screening. Written informed consent was obtained from all participants. The cervicovaginal samples were collected by the medical staff during the examinations, and were immediately placed into a stabilizing reagent enabling preservation of the microbiota at room temperature[45]. They were then stored at −80 °C before transfer to the laboratory on dry ice.

DNA extraction, sequencing, and data preprocessing

DNA extraction of samples from MRKH patients and healthy women was performed as previously described[47]. Briefly, take 1 mL of the sample solution together with the sampling swab and centrifuge it at 18,000 × *g* for 5 minutes at 4 °C. Discard the supernatant. To the pellet, add 300 μL of glass beads and 600 μL of ATL buffer (Cat. No. / ID: 939011, Qiagen) mixture containing 20 μL of lysozyme. Grind the mixture at a frequency of 30 r/s for 5 minutes, then incubate it at 70 °C for 20 min. Next, add 600 μL of Phenol-Chloroform-Isoamyl Alcohol mixture (25:24:1), grind at a frequency of 15 r/s for 30 s, and let the mixture stand for 2 min. Centrifuge at 14,000 × *g* for 5 minutes at 4 °C, and collect the supernatant. Finally, proceed with magnetic bead purification of the supernatant. Library construction and metagenomic shotgun sequencing were performed using the BGI-DIPSEQ platform with 100 bp or 150 bp paired-end reads[48]. 138,685,839,710 raw sequencing reads were obtained for MRKH patients and 157,580,154,680 reads for healthy women. Quality control was first performed with strict standards for filtering and trimming the sequencing reads and the reads were retained with average Phred quality score ≥20 and length ≥30 using fastp v0.19.4[49]. Reads originating from human were removed using Bowtie2 v2.3.5[50] (human genome GRCh38) with default parameters except for '--end-to-end --very-sensitive'. Finally, 25,073,576,842 metagenomic reads were obtained from the MRKH patients and 1,738,415,044 reads were obtained from the healthy women. The average amount of the high-quality metagenomic clean data is listed in Supplementary Table 1.

Microbial community

Taxonomic assignment of the high-quality shotgun metagenomic reads after quality control and removal of host reads for all samples, including both of MRKH patients and healthy women, was performed using MetaPhlAn3 with default parameters (--read_min_len 70) and database v30[51]. Rare species present in less than three samples were discarded. Bacterial microbes were retained and renormalized, which rendered the data of 2 neovaginal samples from P14D unavailable for subsequent analysis.

Microbiota alpha diversity, Shannon–Wiener index, and the number of observed species were calculated based on the relative abundance profile of species using the function 'diversity' (R vegan v2.6-4)[52]. The Bray–Curtis dissimilarity was calculated with the relative abundance profile of species using the function 'vegdist' (R vegan v2.6–4). Permutational Multivariate Analysis of Variance (PERMANOVA) was used to assess the statistically significant differences between entire microbial community at P2/4Y and metadata including MRKH type and variable metadata at P2/4Y, which had enough variables for statistical analysis based on Bray–Curtis dissimilarity using the function 'adonis' (R vegan v2.6-4). To assess the shared species in vagina among MRKH patients and healthy woman, a Venn diagram was created based on the species whose mean relative abundance was detected as >0.001 in at least one group using the function 'venn.diagram' (R VennDiagram v1.7.3).

Statistical tests

Comparisons among multiple groups in terms of diversity indices (the number of observed species, Shannon–Wiener index, and Bray–Curtis dissimilarity), and relative abundance of *Lactobacillus spp*. were performed by using the Kruskal–Wallis test with *P* values adjusted with the Benjamini–Hochberg Procedure (False Discovery Rate, FDR). Comparisons between two groups in terms of diversity indices (the number of observed species, Shannon–Wiener index, and Bray–Curtis dissimilarity), species/genus relative abundance and normalized phylogenetic distance were performed by using a Wilcoxon rank-sum test (Mann–Whitney U test). Multiple testing corrections were calculated using the FDR, and an *P* adjust value < 0.1 was considered significant.

Vaginal community type

Vagitype: in this study, vagitypes were assigned according to the species with the highest relative abundance in each vaginal sample. When the predominant species did not exceed a relative abundance of 30%, the sample was classified as 'Diverse'. This method of classification draws on the precedent set by a prior study[53].

DMM type: The relative abundances of species in all MRKH vaginal samples (except 6–12 months) were used to determine the vaginal community type for MRKH patients according to the Dirichlet multinomial mixture model (DMM)[54].

SparCC network

To assess the variability of networks among different time points (pre-surgery, 14 days, 90 days and 2–4 years post-surgery), we established co-occurrence network by using the SparCC method (Sparse Correlations for Compositional data)[55] for each time points obtaining empirical *p*-values from 1000 times bootstrapping. Significant correlations with prevalence >10% and FDR < 0.05 were visualized using the R package igraph. Species included in this analysis passed two rounds of filtering. First, species were filtered out in the data of all vaginal samples from MRKH patients based on: 1) prevalence <10%; 2) mean relative abundance <1‰; and 3) in <3 samples with a relative abundance lower than 1‰. Next, in the data of each time-point, a species would be filtered out if its relative abundance had less than five non-zero values.

Source tracking

Microbial source tracking software FEAST (Version 1.0)[30] was used to predict the potential contributions to the neovaginal microbiota from bacteria from different body sites, i.e., the pre-surgery dimple, peritoneal fluid, stool, tongue, and skin of MRKH patients. The microbiota of saline was also included. Samples of each type from their respective initial sampling point were examined for being potential sources. The species that occurred in more than 10% of the samples for the specific sample type were selected, except for a 5% occurrence limit for peritoneal fluid. Samples were filtered by more than 500 reads and less than 500,000,000 reads, 3 samples were omitted due to less than 500 reads.

Strain analysis

Strain-level analysis of single-nucleotide variant profiling was performed using StrainPhlAn[31]. Strains of a certain species were selected for construction of a database only when at least 80% with ≥20 markers detected (-print_clades_only with default parameter) were selected for construction of a database. The species-specific MetaPhlAn markers were selected, with the parameters '--marker_in_n_samples 10 --sample_with_n_markers 10 --secondary_sample_with_n_markers 10', and 220 strains were finally selected across all samples. (n = 105 strains in vaginal samples, n = 174 strains in stool samples, n = 58 strains in tongue samples, n = 49 strains in skin samples, n = 13 strains in saline samples). These markers were then processed to produce the multiple sequence alignment (MSA) and then to build the phylogenetic tree (Supplementary Data 5). The percentage of used markers amongst available markers was >80% for the majority of strains (n = 185/220 strains) (Supplementary Data 5). The average coverage across strains was 12.7×.

Species-specific phylogenetic distances were generated using 'strain_transmission.py' as input for the phylogenetic trees[51]. Comparison among different trees was performed by normalizing distances in each tree by its median value, and using a threshold of the normalized pairwise phylogenetic distance <0.1 for the definition of identity of a pair of strains[20]. The phylogenetic trees were visualized using the 'ggtree' (R ggtree v3.6.2)[56]. Identical strains were detected between the pre-surgery dimple and the neovagina at each timepoint after surgery (14 days, 90 days, 6–12 months, 2–4 years post-surgery); Identity

calculations were implemented to identify strain-sharing events within the same individual between the neovagina and gut, the neovagina and tongue, and the neovagina and skin at any given time point. The strain transmissibility rate was calculated as the number of strain-sharing events across the pre-surgery dimple-neovagina, gut-neovagina, tongue-neovagina, and skin-neovagina detected for the species, divided by the total potential number of strain-sharing events based on the presence of a strain-level profile by StrainPhlAn.

Generalized estimating equations

To assess the influence of the pre-surgery dimple and the stool microbiota on the neovagina microbiota, we fitted generalized estimating equations (GEE) using the 'gee.glm' function (R geepack v1.3.9[33]. Firstly, we selected the species with occurrence rates over 10% in both observation groups. Next, data normalization was conducted by using the 'scale' function based on a log10 transformation of the relative abundances. The categorical variable (indicating the presence or absence of a species in the dimple microbiota) was transformed into a binary variable with values of 0 or 1. Models for each species were as follows, 1) the relative abundance of samples taken at multiple time points in the neovagina (14 days, 90 days, and 2–4 years post-surgery) ~the presence or absence in the dimple microbiota; 2) the relative abundance of samples taken at multiple time points in the neovagina (14 days, 90 days, 2–4 years post-surgery) ~the relative abundance in the dimple microbiota; 3) the relative abundance of samples taken at multiple time points in the neovagina (14 days, 90 days, 2–4 years post-surgery) ~the relative abundance in the gut microbiota pre-surgery. Multiple testing corrections were calculated using the FDR, and an P adjust <0.1 was considered significant. The species detected as significant at least once among the above models were visualized using the 'pheatmap' function (R pheatmap v1.0.12).

Random forest prediction

To check whether future alterations in the microbial features of the neovagina could be predicted, a random forest model with 10-fold cross validation was trained to predict the relative changes in the abundances of the neovagina microbiota between two time points (14 days, 90 days, 2–4 years post-surgery) based on the previous time point abundance profile. The best sets of input features were determined by the 10-fold cross validation. Species with prevalence >10%, 0.1% mean abundance and at least 3 samples with abundance >0.1% in MRKH samples by which 54 species remained. The abundances of the bacterial species were log-transformed and scaled in this model. The accuracy of prediction for a given species was determined by Spearman's correlation coefficients of true value and predict value[57].

Reporting summary

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

Data availability

The metagenomic data are publicly available in the CNGB Nucleotide Sequence Archive (CNSA: https://db.cngb.org/cnsa) under accession number CNP0004711 (456 samples from the MRKH patients) and CNP0006125 (472 healthy adult Chinese women).

Code availability

Analysis code is available via github (https://github.com/haolilan/Neovaginal-microbiota-assembly).

References

1. Herlin, M. K., Petersen, M. B. & Brännström, M. Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome: a comprehensive update. *Orphanet J. Rare Dis.* **15**, 1–16 (2020).
2. Horbach, S. E. R. et al. Outcome of Vaginoplasty in Male-to-Female Transgenders: A Systematic Review of Surgical Techniques. *J. Sex. Med.* **12**, 1499–1512 (2015).
3. Mora, R. M., Mehta, P., Ziltzer, R. & Samplaski, M. K. Systematic Review: The Neovaginal Microbiome. *Urology* **167**, 3–12 (2022).
4. Krakowsky, Y. et al. The Effect of Gender-Affirming Medical Care on the Vaginal and Neovaginal Microbiomes of Transgender and Gender-Diverse People. *Front. Cell Infect. Microbiol.* **11**, 769950 (2021).
5. Birse, K. D. et al. The neovaginal microbiome of transgender women post-gender reassignment surgery. *Microbiome* **8**, 61 (2020).
6. Qin, C. et al. Analysis of the artificial vaginal microecology in patients after laparoscopic peritoneal vaginoplasty. *Sci. Rep.* **9**, 8482 (2019).
7. McLoughlin, K., Schluter, J., Rakoff-Nahoum, S., Smith, A. L. & Foster, K. R. Host Selection of Microbiota via Differential Adhesion. *Cell Host Microbe* **19**, 550–559 (2016).
8. France, M., Alizadeh, M., Brown, S., Ma, B. & Ravel, J. Towards a deeper understanding of the vaginal microbiota. *Nat. Microbiol* **7**, 367–378 (2022).
9. Anderson, D. J., Marathe, J. & Pudney, J. The structure of the human vaginal stratum corneum and its role in immune defense. *Am. J. Reprod. Immunol.* **71**, 618–623 (2014).
10. Łaniewski, P. & Herbst-Kralovetz, M. M. Connecting microbiome and menopause for healthy ageing. *Nat. Microbiol* **7**, 354–358 (2022).
11. Graham, M. E. et al. Gut and vaginal microbiomes on steroids: implications for women's health. *Trends Endocrinol. Metab.* **32**, 554–565 (2021).
12. Ravel, J. et al. Vaginal microbiome of reproductive-age women. *Proc. Natl Acad. Sci.* **108**, 4680–4687 (2010).
13. Younes, J. A. et al. Women and Their Microbes: The Unexpected Friendship. *Trends Microbiol.* **26**, 16–32 (2018).
14. 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).
15. Fedele, L. et al. Neovaginal mucosa after Vecchietti's laparoscopic operation for Rokitansky syndrome: structural and ultrastructural study. *Am. J. Obstet. Gynecol.* **195**, 56–61 (2006).
16. Fedele, L. et al. Creation of a neovagina by Davydov's laparoscopic modified technique in patients with Rokitansky syndrome. *Am. J. Obstet. Gynecol.* **202**, 33.e1–6 (2010).
17. Piazza, M. J. Study and evaluation of neovagina epithelium. *JBRA Assist Reprod.* **25**, 581–585 (2021).
18. Zhu, L., Zhou, H., Sun, Z., Lou, W. & Lang, J. Anatomic and sexual outcomes after vaginoplasty using tissue-engineered biomaterial graft in patients with Mayer-Rokitansky-Küster-Hauser syndrome: a new minimally invasive and effective surgery. *J. Sex. Med.* **10**, 1652–1658 (2013).
19. Sprockett, D., Fukami, T. & Relman, D. A. Role of priority effects in the early-life assembly of the gut microbiota. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 197–205 (2018).
20. 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).
21. Shao, Y. et al. Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature* **574**, 117–121 (2019).
22. Olm, M. R. et al. Robust variation in infant gut microbiome assembly across a spectrum of lifestyles. *Science* **376**, 1220–1223 (2022).
23. Vatanen, T. et al. Mobile genetic elements from the maternal microbiome shape infant gut microbial assembly and metabolism. *Cell* **185**, 4921–4936.e15 (2022).

24. Patel, M. A. & Nyirjesy, P. Role of Mycoplasma and ureaplasma species in female lower genital tract infections. *Curr. Infect. Dis. Rep.* **12**, 417–422 (2010).
25. Kristich, C. J., Little, J. L., Hall, C. L. & Hoff, J. S. Reciprocal regulation of cephalosporin resistance in Enterococcus faecalis. *mBio* **2**, e00199–11 (2011).
26. Swidsinski, A. et al. Adherent biofilms in bacterial vaginosis. *Obstet. Gynecol.* **106**, 1013–1023 (2005).
27. Hardy, L. et al. A fruitful alliance: the synergy between Atopobium vaginae and Gardnerella vaginalis in bacterial vaginosis-associated biofilm. *Sex. Transm. Infect.* **92**, 487–491 (2016).
28. Petrova, M. I., Reid, G., Vaneechoutte, M. & Lebeer, S. Lactobacillus iners: Friend or Foe? *Trends Microbiol* **25**, 182–191 (2017).
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. Shenhav, L. et al. FEAST: fast expectation-maximization for microbial source tracking. *Nat. Methods* **16**, 627–632 (2019).
31. Truong, D. T., Tett, A., Pasolli, E., Huttenhower, C. & Segata, N. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res.* **27**, 626–638 (2017).
32. Almeida, A. et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* **39**, 105–114 (2021).
33. Halekoh, U., Højsgaard, S. & Yan, J. The R package geepack for generalized estimating equations. *J. Stat. Softw.* **15**, 1–11 (2006).
34. Smith, S. B. & Ravel, J. The vaginal microbiota, host defence and reproductive physiology. *J. Physiol.* **595**, 451–463 (2017).
35. Gajer, P. et al. Temporal dynamics of the human vaginal microbiota. *Sci. Transl. Med* **4**, 132ra52 (2012).
36. Chen, C. et al. Cervicovaginal microbiome dynamics after taking oral probiotics. *J. Genet. Genomics* **48**, 716–726 (2021).
37. Liu, C. M. et al. Penile anaerobic dysbiosis as a risk factor for HIV infection. *mBio* **8**, e00996–17 (2017).
38. Price, L. B. et al. The effects of circumcision on the penis microbiome. *PLoS One* **5**, e8422 (2010).
39. Debray, R. et al. Priority effects in microbiome assembly. *Nat. Rev. Microbiol* **20**, 109–121 (2022).
40. Bashan, A. et al. Universality of human microbial dynamics. *Nature* **534**, 259–262 (2016).
41. Si, J., You, H. J., Yu, J., Sung, J. & Ko, G. P. Prevotella as a Hub for Vaginal Microbiota under the Influence of Host Genetics and Their Association with Obesity. *Cell Host Microbe* **21**, 97–105 (2017).
42. France, M., Ma, B. & Ravel, J. Persistence and In Vivo Evolution of Vaginal Bacterial Strains over a Multiyear Time Period. *mSystems* **7**, e0089322 (2022).
43. Fontana, L., Gentilin, B., Fedele, L., Gervasini, C. & Miozzo, M. Genetics of Mayer–Rokitansky–Küster–Hauser (MRKH) syndrome. *Clin. Genet.* **91**, 233–246 (2017).
44. Kang, J. et al. Sexual function and quality of life after the creation of a neovagina in women with Mayer-Rokitansky-Küster-Hauser syndrome: comparison of vaginal dilation and surgical procedures. *Fertil. Steril.* **113**, 1024–1031 (2020).
45. Han, M. et al. A novel affordable reagent for room temperature storage and transport of fecal samples for metagenomic analyses. *Microbiome* **6**, 1–7 (2018).
46. Chen, C. et al. Genomic landscape in the human vaginal microbiome links to host geographics and genetics. https://doi.org/10.21203/rs.3.rs-4129268/v1 (2024).
47. Chen, C. et al. The microbiota continuum along the female reproductive tract and its relation to uterine-related diseases. *Nat. Commun.* **8**, 875 (2017).
48. Fang, C. et al. Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing. *Gigascience* **7**, 1–8 (2018).
49. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
50. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2013).
51. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* **10**, e65088 (2021).
52. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Vegetation Sci.* **14**, 927–930 (2003).
53. Fettweis, J. M. et al. The vaginal microbiome and preterm birth. *Nat. Med.* **25**, 1012–1021 (2019).
54. Ding, T. & Schloss, P. D. Dynamics and associations of microbial community types across the human body. *Nature* **509**, 357–360 (2014).
55. Friedman, J. & Alm, E. J. Inferring Correlation Networks from Genomic Survey Data. *PLoS Comput. Biol.* **8**, e1002687 (2012).
56. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
57. Jie, Z. et al. The Baseline Gut Microbiota Directs Dieting-Induced Weight Loss Trajectories. *Gastroenterology* **160**, 2029–2042.e16 (2021).

Acknowledgements

We thank all the patients for agreeing to join this study. Special thanks to Prof. Guangnan Luo from Shenzhen Luohu Hospital for his contributions to the outpatient follow-up of patients with MRKH syndrome in this study. Unfortunately, he passed away on February 4, 2023 due to Covid-19, at the age of 84. We are very grateful to the colleagues at BGI-Research for sample collection, and discussions, and China National Genebank (CNGB) for DNA extraction, library construction, and sequencing. Special thanks go to Jiang Fangfang, who has been in charge of supervising the DNA extraction process for our study. This work was supported in part by the National Key Research and Development Program of China (2021YFC2701401, 2021YFC2701405), the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2020-PT320-003 and 2021-PT320-001), the National Natural Science Foundation of China (82171614, 82271656 and 82171621), Beijing Natural Science Foundation (7232125), CAMS Initiative Fund for Medical Sciences (2022-I2M-C&T-B-029, 2022-PUMCH-A-231, and 2021-I2M-C&T-A-008).

Author contributions

L.Z., C.C., and T.Z. conceived and directed the study. C.Q. performed all the surgical procedures. N.C., C.Q., J.D., G.L., H.P., J.S., M.L., J.G., Q.G., Z.D., J.S., and L.Z. performed the clinical diagnosis, sample collection and result analyses. Y.Z., H.G., and R.L. performed the sample collection of Peacock cohort. X.H. was responsible for organizing the P2/4Y samples and related phenotypes. L.H., Z.Z., Z.J., T.S., Hua.Y., X.Z., and C.C. performed the bioinformatics analyses and prepared the display items. Z.Z. conceived the framework of the article and wrote the manuscript with contributions from C.C., N.C., Z.J., and L.H. K.K. extensively restructured and revised the manuscript while fully complying with the regulations of the Human Genetic Resources Administration of China in the project. H.J., L.X., S.Z., X.J., X.X., Huanming.Y., and J.W. contributed by fulfilling key roles in both strategic management and administrative oversight. All authors contributed to the revision of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-52102-1.

Correspondence and requests for materials should be addressed to Tao Zhang, Karsten Kristiansen, Chen Chen or Lan Zhu.

Peer review information *Nature Communications* thanks Kayla Carter and the other, anonymous, reviewer for their contribution to the peer review of this work. A peer review file is available.

[1]National Clinical Research Center for Obstetric and Gynecologic Diseases, Department of Obstetrics and Gynecology, Peking Union Medical College Hospital, State Key Laboratory for Complex Severe and Rare Diseases, Peking Union Medical College, Chinese Academy of Medical Sciences, No. 1 Shuaifuyuan, 100005 Beijing, China. [2]BGI-Research, Shenzhen 518083, China. [3]Department of Biology, Ole Maaløes Vej 5, University of Copenhagen, DK-2200 Copenhagen, Denmark. [4]Qingdao-Europe Advanced Institute for Life Sciences, BGI-Research, Qingdao 266555, China. [5]Department of Obstetrics and Gynaecology, The 3rd Affiliated Hospital of Shenzhen University, Luohu hospital, Shenzhen 518000 Guangdong, China. [6]Department of Gynecology, Southern University of Science and Technology Hospital, 6019 liuxian street, Shenzhen 518000, China. [7]Chinese Academy of Medical Sciences & Peking Union Medical College, 4+4 Medical Doctor Program, No.9 Dongdansantiao, 100730 Beijing, China. [8]China National GeneBank, BGI-Research, Shenzhen 518210, China. [9]Social Affairs Bureau of Suzhou National New and Hi-tech Industrial Development Zone, Suzhou 215163, China. [10]Suzhou National New and Hi-tech Industrial Development Zone Center for Maternal and Child Health and Family Planning Service, Suzhou 215163, China. [11]Shenzhen Engineering Laboratory of Detection and Intervention of Human Intestinal Microbiome, BGI-Research, Shenzhen 518083, China. [12]BGI Research, Wuhan 430074, China. [13]James D. Watson Institute of Genome Sciences, Hangzhou 310013, China. [14]School of Life Sciences, Fudan University, Shanghai 200433, China. [15]Greater Bay Area Institute of Precision Medicine (Guangzhou), Fudan University, Guangzhou 511458, China. [16]Shenzhen Key Laboratory of Human Commensal Microorganisms and Health Research, Shenzhen, Guangdong 518083, China. [17]These authors contributed equally: Na Chen, Lilan Hao, Zhe Zhang, Chenglu Qin, Zhuye Jie. ✉e-mail: tao.zhang@genomics.cn; kk@bio.ku.dk; chenchen1@genomics.cn; zhu_julie@vip.sina.com