Molecular definition of vaginal microbiota in East African commercial sex workers

Supplementary Information

- Figures S1-S4
- Tables S1-S7

Fig. S1: Tiers of abundance

Range of relative abundance across orders of magnitude for OTU detected in deep sequencing read datasets expressed as “tiers of abundance”: Tier 1=>1%, Tier 2=0.1-0.99%, Tier 3=0.01-0.099%, Tier 4=<0.01%. Only a small proportion of OTU are detected at >1% of total reads (Tier 1), with most detected at <0.1% (Tiers 3/4) (A). Inverse relationship observed between OTU frequency and read frequency, with few OTU but most reads in Tier 1 and vice versa (B). Proportion of individuals in study group colonized by each OTU (C). Median and range for each OTU distribution across all individuals in study group, showing low median values (<1%) for all OTU and range across orders of magnitude (D).

Fig. S2: Richness, diversity and phylotype abundance by HIV serostatus

Chao1 richness (A) and Simpson's diversity index (B) by HIV serostatus as calculated using mothur. Relative abundance of phylum and order-level phylotypes by HIV serostatus (C). Boxplots were created and p values calculated based on non-parametric significance tests (wilcox.test) using R statistical software (v.2.11.0).

Fig. S3: Richness, diversity and phylotype abundance by mBV status

Chao1 richness (A) and Simpson's diversity index (B) by mBV status as calculated using mothur. Relative abundance of phylum and order-level phylotypes by mBV status (C). Boxplots were created and p values calculated based on non-parametric significance tests (wilcox.test) using R statistical software (v.2.11.0).

Fig. S4: Heat-map and hierarchical clustering of study sub-groups and OTU

Individual samples were collapsed into 10 groups (HESN/BV+, HESN/mBVI, HESN/BV–, HIV-N/BV+, HIV-N/mBVI, HIV-N/BV–, HIV+/BV+, HIV+/mBVI, HIV+/mBV–, HIV+/Outlier) and heat-map generated using Genespring software based on OTU observed in at least 25% of the study group. Hierarchical clustering of groups and OTU was generated using Genespring software.

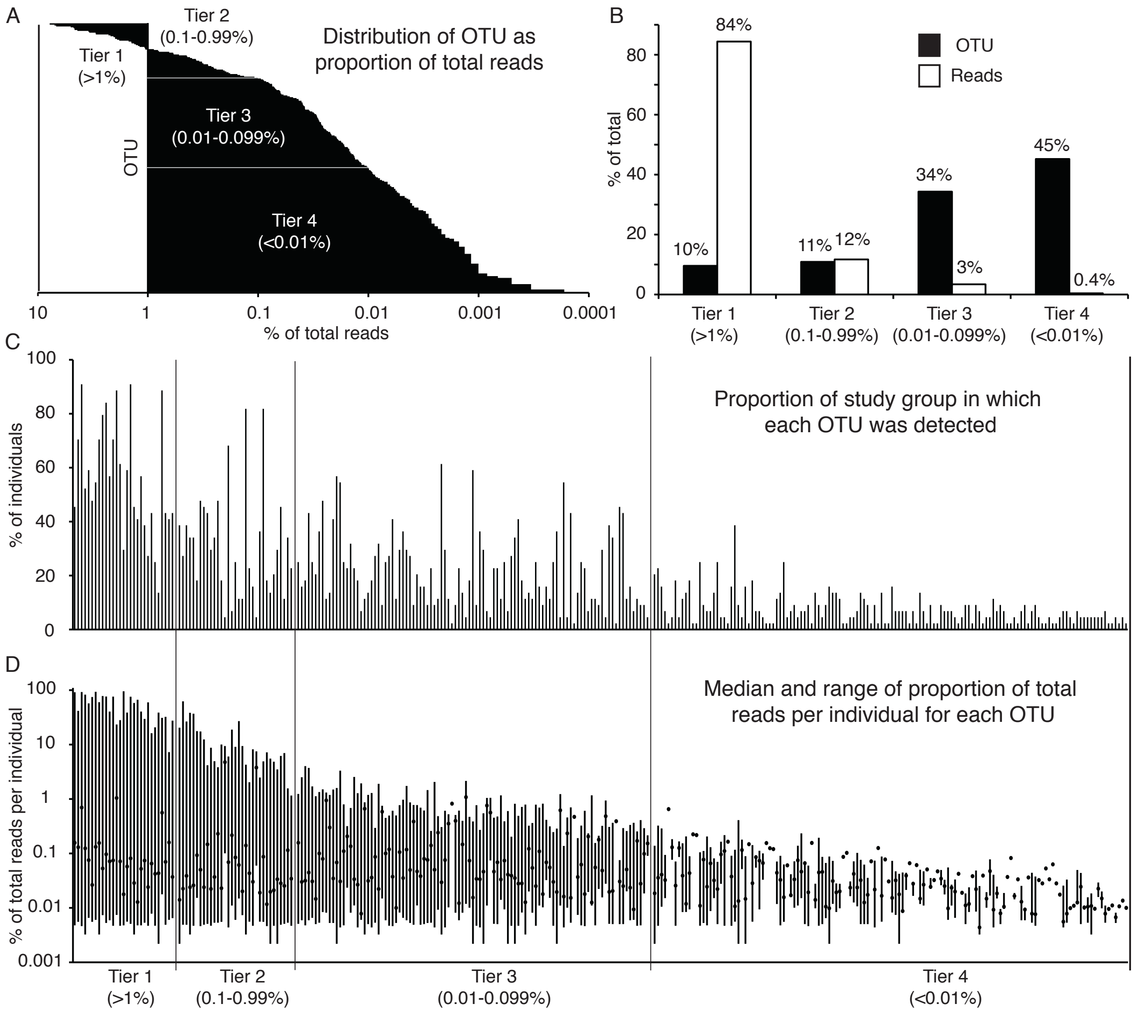

Fig. S1: Tiers of abundance

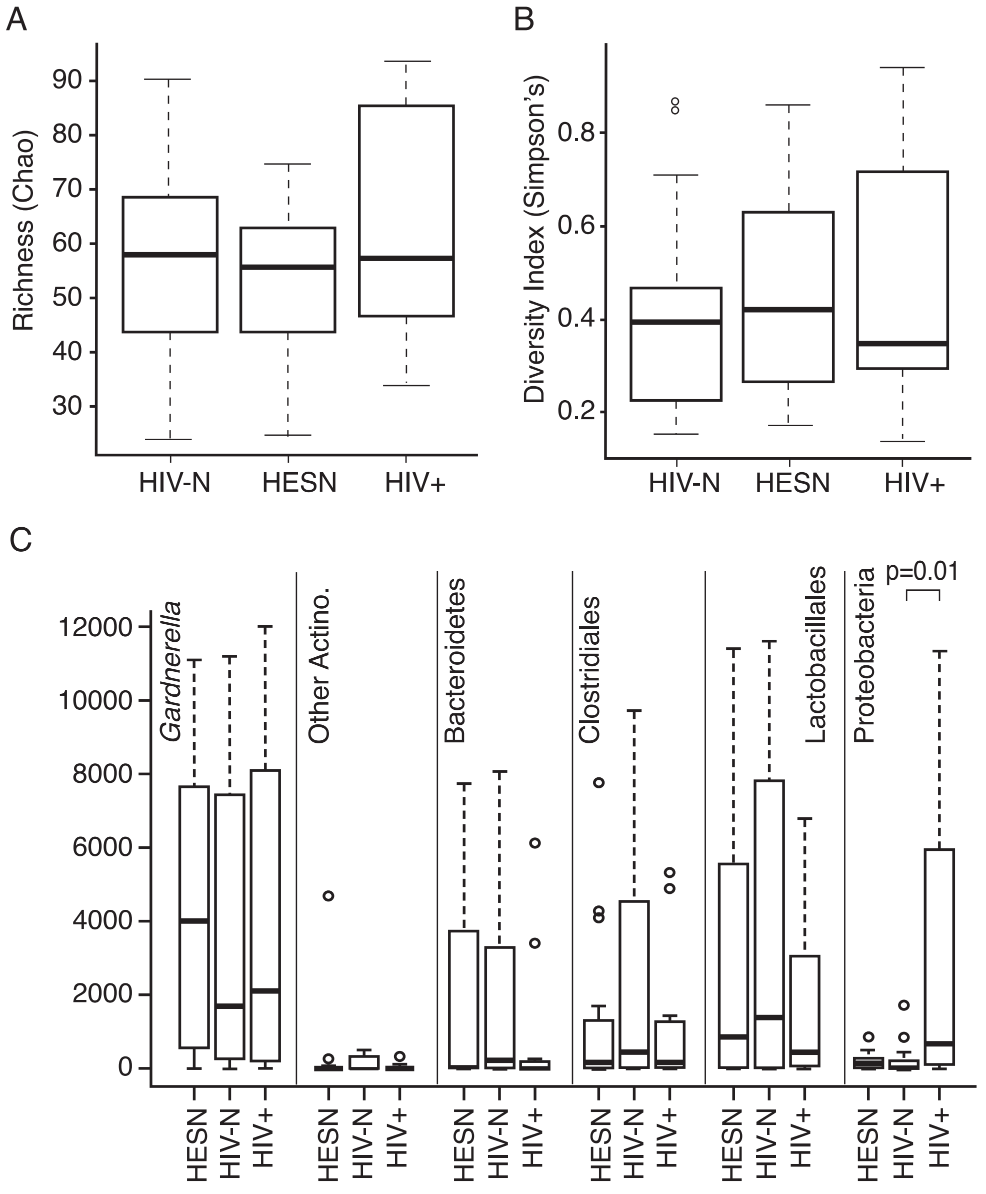

Fig. S2: Richness, diversity and phylotype abundance by HIV serostatus

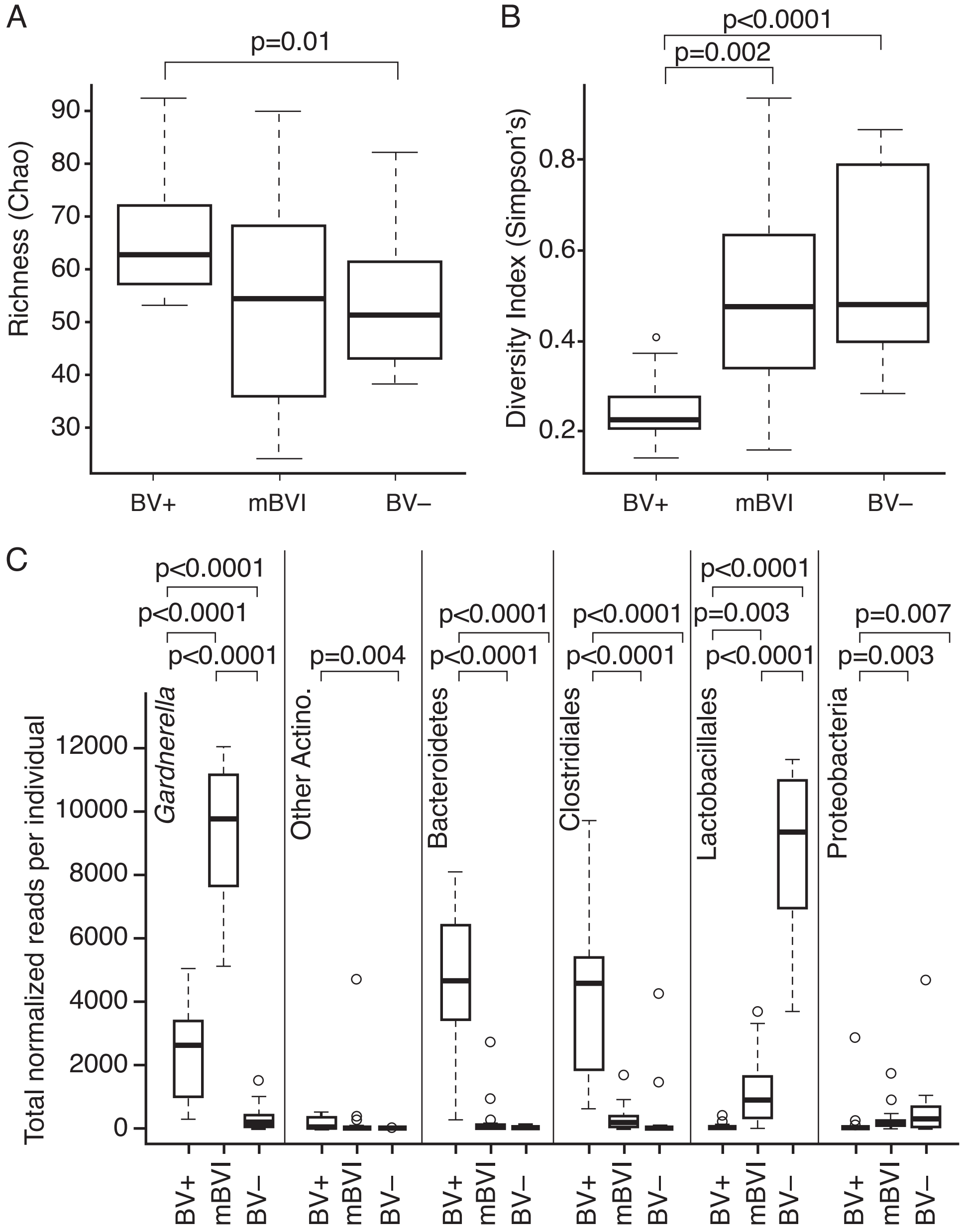

Fig. S3: Richness, diversity and phylotype abundance by molecular BV status

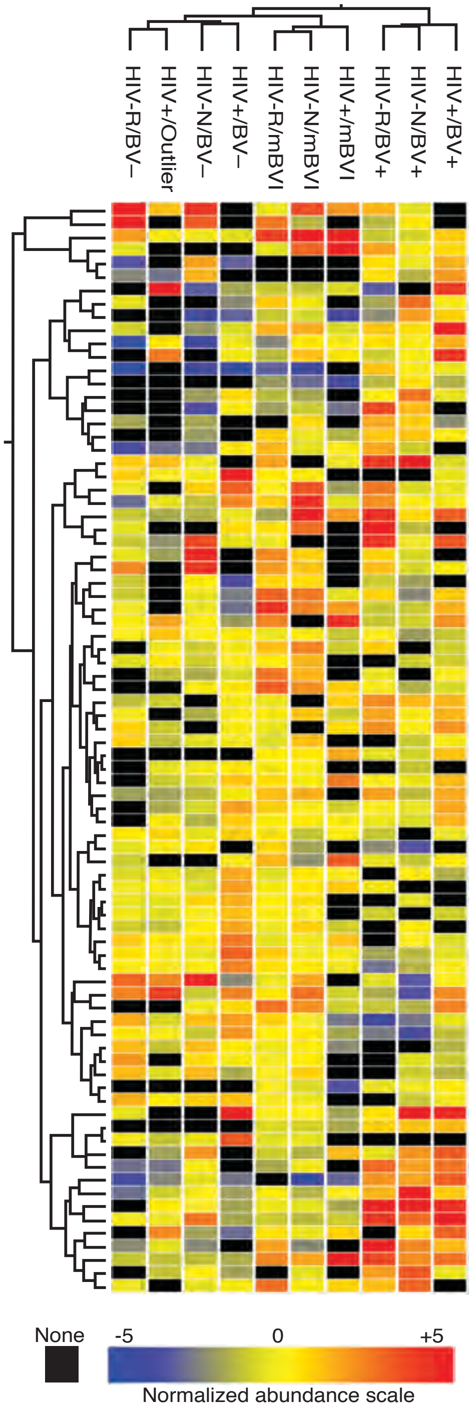

Fig. S4: Heat-map and hierarchical clustering of study sub-groups and OTU

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
	1	NC032	*Gardnerella vaginalis* ATCC 14018	90.9	0	1	0.013986	0	0
	2	NC172	*Gardnerella vaginalis* ATCC 14018	88.4	0	3	0.041958	0	0
	3	NC034	*Gardnerella vaginalis* ATCC 14018	89.6	0	2	0.027972	0	0
	4	ND631	*Gardnerella vaginalis* ATCC 14018	86.5	0	0	0	6	0.0009962
	5	NC028	*Gardnerella vaginalis* ATCC 14018	89.2	0	13	0.1818182	207	0.0343681
	6	ND717	*Gardnerella vaginalis* ATCC 14018	88.5	0	0	0	55	0.0091316
	7	NC042	*Gardnerella vaginalis* ATCC 14018	88.6	0	2	0.027972	0	0
	8	ND578	*Gardnerella vaginalis* ATCC 14018	86.5	0	0	0	7	0.0011622
	9	NC031	*Gardnerella vaginalis* ATCC 14018	89.5	0	1	0.013986	0	0
	10	NC177	*Gardnerella vaginalis* ATCC 14018	88.9	0	2	0.027972	0	0
	11	NC179	*Gardnerella vaginalis* ATCC 14018	91.6	0	2	0.027972	0	0
	12	NC180	*Gardnerella vaginalis* ATCC 14018	90.2	0	2	0.027972	0	0
	13	ND071	*Gardnerella vaginalis* ATCC 14018	87.5	0	0	0	528	0.0876637
	14	ND624	*Gardnerella vaginalis* ATCC 14018	87	0	0	0	6	0.0009962
1						**28**	**0.3916084**	**809**	**0.134318**
	15	ND528	*Gardnerella vaginalis* ATCC 14018	86.5	0	0	0	17	0.0028225
	16	ND035	*Gardnerella vaginalis* ATCC 14018	87.5	0	0	0	18	0.0029885
	17	ND454	*Gardnerella vaginalis* ATCC 14018	86.6	0	0	0	12	0.0019924
	18	NC186	*Gardnerella vaginalis* ATCC 14018	90.2	0	5	0.0699301	28175	4.6778858
	19	NC138	*Gardnerella vaginalis* ATCC 14018	89.2	0	47	0.6573427	0	0
	20	ND753	*Gardnerella vaginalis* ATCC 14018	86.5	0	0	0	13	0.0021584
	21	ND844	*Gardnerella vaginalis* ATCC 14018	86.5	0	0	0	31	0.0051469
	22	ND554	*Gardnerella vaginalis* ATCC 14018	87.5	0	0	0	99	0.0164369
	23	NC143	*Gardnerella vaginalis* ATCC 14018	91.1	0	2	0.027972	0	0
	24	ND694	*Gardnerella vaginalis* ATCC 14018	86	0	0	0	160	0.0265647

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
	25	NC033	*Gardnerella vaginalis* ATCC 14018	89.4	0	170	2.3776224	0	0
	26	N153	*Gardnerella vaginalis* ATCC 14018	89.7	2	0	0	0	0
2					**2**	**224**	**3.1328671**	**28525**	**4.7359962**
	27	NC176	*Gardnerella vaginalis* ATCC 14018	88	0	1	0.013986	35	0.005811
	28	NC171	*Gardnerella vaginalis* ATCC 14018	89.5	0	1	0.013986	0	0
	29	ND142	*Gardnerella vaginalis* ATCC 14018	86.5	0	0	0	9	0.0014943
	30	N144	*Gardnerella vaginalis* ATCC 14018	90	2	5	0.0699301	11862	1.9694439
	31	ND022	*Gardnerella vaginalis* ATCC 14018	88.5	0	0	0	59	0.0097958
	32	ND462	*Gardnerella vaginalis* ATCC 14018	91.5	0	0	0	7	0.0011622
	33	NC093	*Gardnerella vaginalis* ATCC 14018	92.4	0	1	0.013986	0	0
	34	NC182	*Gardnerella vaginalis* ATCC 14018	92.4	0	1	0.013986	0	0
	35	ND655	*Gardnerella vaginalis* ATCC 14018	89	0	0	0	25	0.0041507
	36	ND725	*Gardnerella vaginalis* ATCC 14018	92	0	0	0	87	0.0144446
	37	ND242	*Gardnerella vaginalis* ATCC 14018	88.5	0	0	0	40	0.0066412
	38	ND094	*Gardnerella vaginalis* ATCC 14018	88	0	0	0	167	0.027727
	39	ND390	*Gardnerella vaginalis* ATCC 14018	87.5	0	0	0	9	0.0014943
	40	ND239	*Gardnerella vaginalis* ATCC 14018	87.5	0	0	0	178	0.0295533
	41	ND393	*Gardnerella vaginalis* ATCC 14018	87	0	0	0	30087	4.9953346
3					**2**	**9**	**0.1258741**	**42565**	**7.0670527**
	42	ND535	*Gardnerella vaginalis* ATCC 14018	97	0	0	0	10	0.0016603
	43	N156	*Gardnerella vaginalis* ATCC 14018	88.9	2	0	0	2480	0.4117536
	44	NC036	*Gardnerella vaginalis* ATCC 14018	88.6	0	1	0.013986	0	0
	45	NC174	*Gardnerella vaginalis* ATCC 14018	88.2	0	1	0.013986	0	0
	46	ND301	No near neighbour	<85	0	0	0	89	0.0147766
	47	N072	*Gardnerella vaginalis* ATCC 14018	88.9	5	0	0	20818	3.4564056

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
	48	ND457	*Gardnerella vaginalis* ATCC 14018	87	0	0	0	6	0.0009962
	49	NC054	*Gardnerella vaginalis* ATCC 14018	89.7	0	1	0.013986	0	0
	50	ND733	*Gardnerella vaginalis* ATCC 14018	92	0	0	0	18	0.0029885
	51	ND748	*Gardnerella vaginalis* ATCC 14018	90	0	0	0	24	0.0039847
4					**7**	**3**	**0.041958**	**23445**	**3.8925655**
	52	NC049	*Gardnerella vaginalis* ATCC 14018	87.8	0	1	0.013986	0	0
	53	N134	*Gardnerella vaginalis* ATCC 14018	88.9	3	105	1.4685315	0	0
	54	NC048	*Gardnerella vaginalis* ATCC 14018	90.2	0	1	0.013986	0	0
	55	ND002	*Gardnerella vaginalis* ATCC 14018	86	0	0	0	15	0.0024904
	56	ND061	*Gardnerella vaginalis* ATCC 14018	86.5	0	0	0	12454	2.0677335
	57	NC050	*Gardnerella vaginalis* ATCC 14018	94.2	0	6	0.0839161	0	0
	58	NC105	*Gardnerella vaginalis* ATCC 14018	91.1	0	2	0.027972	0	0
	59	ND493	*Gardnerella vaginalis* ATCC 14018	86.5	0	0	0	7	0.0011622
	60	N137	*Gardnerella vaginalis* ATCC 14018	89.5	1	0	0	7433	1.2340985
	61	NC073	*Gardnerella vaginalis* ATCC 14018	88.7	0	1	0.013986	0	0
	62	ND532	*Gardnerella vaginalis* ATCC 14018	87.5	0	0	0	13	0.0021584
	63	NC055	*Gardnerella vaginalis* ATCC 14018	88.8	0	1	0.013986	0	0
	64	NC103	*Gardnerella vaginalis* ATCC 14018	89.9	0	1	0.013986	0	0
	65	NC063	*Gardnerella vaginalis* ATCC 14018	88.8	0	1	0.013986	0	0
	66	NC069	*Gardnerella vaginalis* ATCC 14018	89.3	0	2	0.027972	0	0
	67	NC064	*Gardnerella vaginalis* ATCC 14018	90.8	0	1	0.013986	0	0
	68	NC052	*Gardnerella vaginalis* ATCC 14018	90.2	0	1	0.013986	0	0
	69	NC030	*Gardnerella vaginalis* ATCC 14018	89.1	0	5	0.0699301	2415	0.4009616
	70	NC088	*Gardnerella vaginalis* ATCC 14018	92	0	4	0.0559441	0	0
5					**4**	**132**	**1.8461538**	**22337**	**3.7086047**

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
	71	ND203	*Gardnerella vaginalis* ATCC 14018	90	0	0	0	52	0.0086335
	72	ND628	*Gardnerella vaginalis* ATCC 14018	89.5	0	0	0	7	0.0011622
	73	NC112	*Gardnerella vaginalis* ATCC 14018	92.2	0	1	0.013986	0	0
	74	ND235	*Gardnerella vaginalis* ATCC 14018	93	0	0	0	28	0.0046488
	75	ND463	*Gardnerella vaginalis* ATCC 14018	94.5	0	0	0	6	0.0009962
	76	ND622	*Gardnerella vaginalis* ATCC 14018	93	0	0	0	17	0.0028225
	77	ND886	*Gardnerella vaginalis* ATCC 14018	92	0	0	0	24	0.0039847
	78	ND258	*Gardnerella vaginalis* ATCC 14018	93.5	0	0	0	172	0.0285571
	79	ND662	*Gardnerella vaginalis* ATCC 14018	96.5	0	0	0	288	0.0478165
	80	ND458	*Gardnerella vaginalis* ATCC 14018	95.5	0	0	0	7	0.0011622
	81	ND507	*Gardnerella vaginalis* ATCC 14018	97	0	0	0	1073	0.1781498
	82	NC151	*Gardnerella vaginalis* ATCC 14018	94	0	13	0.1818182	0	0
	83	NC113	*Gardnerella vaginalis* ATCC 14018	93.1	0	1	0.013986	0	0
	84	ND766	*Gardnerella vaginalis* ATCC 14018	97.5	0	0	0	236	0.039183
	85	ND067	*Gardnerella vaginalis* ATCC 14018	97.5	0	0	0	1244	0.2065409
	86	ND369	*Gardnerella vaginalis* ATCC 14018	98.5	0	0	0	8	0.0013282
6						**15**	**0.2097902**	**3162**	**0.5249858**
	87	ND234	*Gardnerella vaginalis* ATCC 14018	97.5	0	0	0	6	0.0009962
	88	ND404	*Gardnerella vaginalis* ATCC 14018	98	0	0	0	9174	1.5231562
	89	ND664	*Gardnerella vaginalis* ATCC 14018	98	0	0	0	16	0.0026565
	90	N164	*Gardnerella vaginalis* ATCC 14018	99.1	1	0	0	0	0
	91	ND231	*Gardnerella vaginalis* ATCC 14018	100	0	0	0	13011	2.160212
	92	ND233	*Gardnerella vaginalis* ATCC 14018	98	0	0	0	158	0.0262327
	93	ND316	*Gardnerella vaginalis* ATCC 14018	97	0	0	0	20713	3.4389725
	94	ND302	*Gardnerella vaginalis* ATCC 14018	97.5	0	0	0	422	0.0700645

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
	95	N165	*Gardnerella vaginalis* ATCC 14018	97.3	1	0	0	0	0
	96	ND642	*Gardnerella vaginalis* ATCC 14018	97	0	0	0	37	0.0061431
	97	NC062	*Gardnerella vaginalis* ATCC 14018	94.9	0	1	0.013986	0	0
	98	NC097	*Gardnerella vaginalis* ATCC 14018	97.6	0	8	0.1118881	0	0
	99	ND260	*Gardnerella vaginalis* ATCC 14018	96.5	0	0	0	7961	1.3217622
	100	ND243	*Gardnerella vaginalis* ATCC 14018	96	0	0	0	16	0.0026565
	101	ND573	*Gardnerella vaginalis* ATCC 14018	96	0	0	0	14	0.0023244
7					**2**	**24**	**0.3356643**	**56178**	**9.3272146**
	102	ND102	No near neighbour	<85	0	0	0	5005	0.8309785
	103	NC057	*Gardnerella vaginalis* ATCC 14018	89.3	0	1	0.013986	0	0
	104	NC070	*Gardnerella vaginalis* ATCC 14018	87.6	0	1	0.013986	0	0
	105	NC089	*Gardnerella vaginalis* ATCC 14018	86.2	0	1	0.013986	0	0
	106	ND533	*Gardnerella vaginalis* ATCC 14018	87.5	0	0	0	16	0.0026565
	107	ND724	*Gardnerella vaginalis* ATCC 14018	93.5	0	0	0	40	0.0066412
	108	NC076	*Gardnerella vaginalis* ATCC 14018	92.9	0	2	0.027972	0	0
	109	NC085	*Gardnerella vaginalis* ATCC 14018	90.9	0	1	0.013986	0	0
	110	NC102	*Gardnerella vaginalis* ATCC 14018	91.1	0	1	0.013986	0	0
	111	NC074	*Gardnerella vaginalis* ATCC 14018	91.8	0	1	0.013986	0	0
	112	NC051	*Gardnerella vaginalis* ATCC 14018	91.3	0	2	0.027972	0	0
	113	NC101	*Gardnerella vaginalis* ATCC 14018	90	0	1	0.013986	0	0
	114	NC065	*Gardnerella vaginalis* ATCC 14018	88.4	0	14	0.1958042	1522	0.2526972
	115	ND692	*Gardnerella vaginalis* ATCC 14018	86	0	0	0	9	0.0014943
	116	NC084	*Gardnerella vaginalis* ATCC 14018	88.7	0	1	0.013986	0	0
	117	NC107	*Gardnerella vaginalis* ATCC 14018	93.3	0	1	0.013986	0	0
8						**27**	**0.3776224**	**6592**	**1.0944676**

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
	118	ND865	*Gardnerella vaginalis* ATCC 14018	86	0	0	0	80	0.0132824
	119	NC130	*Gardnerella vaginalis* ATCC 14018	90.4	0	140	1.958042	1922	0.319109
	120	NC131	*Gardnerella vaginalis* ATCC 14018	88.2	0	15	0.2097902	0	0
	121	NC141	*Gardnerella vaginalis* ATCC 14018	88.8	0	4	0.0559441	0	0
	122	NC140	*Gardnerella vaginalis* ATCC 14018	90	0	2	0.027972	0	0
	123	NC183	*Gardnerella vaginalis* ATCC 14018	88.7	0	1	0.013986	0	0
	124	NC092	*Gardnerella vaginalis* ATCC 14018	88.4	0	1	0.013986	0	0
	125	NC142	*Gardnerella vaginalis* ATCC 14018	89.9	0	60	0.8391608	0	0
	126	NC185	*Gardnerella vaginalis* ATCC 14018	88.9	0	1	0.013986	0	0
	127	NC145	*Gardnerella vaginalis* ATCC 14018	90.9	0	1	0.013986	0	0
	128	ND358	*Gardnerella vaginalis* ATCC 14018	95	0	0	0	51	0.0084675
	129	ND530	*Gardnerella vaginalis* ATCC 14018	94	0	0	0	8	0.0013282
	130	NC058	*Gardnerella vaginalis* ATCC 14018	95.7	0	1	0.013986	0	0
	131	ND492	*Gardnerella vaginalis* ATCC 14018	90.5	0	0	0	47	0.0078034
9						**226**	**3.1608392**	**2108**	**0.3499905**
	132	NC083	*Gardnerella vaginalis* ATCC 14018	88.6	0	1	0.013986	0	0
	133	N160	*Gardnerella vaginalis* ATCC 14018	90.9	1	454	6.3496503	646	0.1072552
	134	NC071	*Gardnerella vaginalis* ATCC 14018	88.6	0	1	0.013986	0	0
	135	NC068	*Gardnerella vaginalis* ATCC 14018	89.1	0	1	0.013986	0	0
	136	NC072	*Gardnerella vaginalis* ATCC 14018	88.4	0	2	0.027972	0	0
	137	ND527	*Gardnerella vaginalis* ATCC 14018	88.6	0	0	0	368	0.0610989
	138	NC137	*Gardnerella vaginalis* ATCC 14018	90.6	0	7	0.0979021	0	0
	139	NC144	*Gardnerella vaginalis* ATCC 14018	90.2	0	2	0.027972	0	0
	140	ND436	*Gardnerella vaginalis* ATCC 14018	87.5	0	0	0	12606	2.09297
	141	NC095	*Gardnerella vaginalis* ATCC 14018	91.3	0	1	0.013986	0	0

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
	142	NC135	*Gardnerella vaginalis* ATCC 14018	89.5	0	8	0.1118881	0	0
10					**1**	**477**	**6.6713287**	**13620**	**2.2613241**
	143	ND416	*Bifidobacterium pseudocatenulatum* D2CA	99.5	0	0	0	23	0.0038187
	144	ND840	*Bifidobacterium dentium* ATCC 27679	100	0	0	0	92	0.0152747
	145	ND726	*Bifidobacterium adolescentis* L2-32	95	0	0	0	8	0.0013282
	146	N033	*Bifidobacterium bifidum* JCM 1255	99.3	1	0	0	0	0
	147	N010	*Bifidobacterium longum* ATCC 15697	99.8	1	0	0	0	0
	148	N081	*Bifidobacterium longum* CCUG 52486	100	7	0	0	0	0
	149	N103	*Bifidobacterium longum* CCUG 52486	99.1	1	0	0	0	0
11					**10**			**123**	**0.0204216**
	150	N159	No near neighbour	<85	1	0	0	0	0
12					**1**				
	151	ND020	*Mobiluncus mulieris* ATCC 35239	100	0	0	0	4187	0.6951662
	152	ND096	*Mobiluncus mulieris* ATCC 35239	99.5	0	0	0	148	0.0245724
13								**4335**	**0.7197386**
	153	ND306	*Mobiluncus curtisii holmesii* ATCC 35242	100	0	0	0	538	0.089324
	154	ND611	*Mobiluncus* sp. 192_1_67	97.5	0	0	0	4	0.0006641
14								**542**	**0.0899881**
	155	ND771	No near neighbour	<85	0	0	0	84	0.0139465
15								**84**	**0.0139465**
	156	ND160	*Propionibacterium acnes* KPA171202	100	0	0	0	31	0.0051469
16								**31**	**0.0051469**
	157	ND768	No near neighbour	<85	0	0	0	1	0.000166
17								**1**	**0.000166**
	158	ND739	No near neighbour	<85	0	0	0	45	0.0074713

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
18								**45**	**0.0074713**
	159	ND484	*Corynebacterium tuberculostearicum* SK141	99.5	0	0	0	129	0.0214178
	160	ND537	*Corynebacterium tuberculostearicum* SK141	96	0	0	0	16	0.0026565
	161	ND460	*Corynebacterium tuberculostearicum* SK141	95.6	0	0	0	3	0.0004981
19								**148**	**0.0245724**
	162	ND847	No near neighbour	<85	0	0	0	25	0.0041507
20								**25**	**0.0041507**
	163	ND682	No near neighbour	<85	0	0	0	6	0.0009962
21								**6**	**0.0009962**
	164	ND542	*Corynebacterium jeikeium* ATCC 43734	96.5	0	0	0	7	0.0011622
	165	ND548	*Corynebacterium jeikeium* ATCC 43734	93	0	0	0	8	0.0013282
22								**15**	**0.0024904**
	166	ND700	Unknown sp. 192_1_65	99.5	0	0	0	19	0.0031546
23								**19**	**0.0031546**
	167	ND451	No near neighbour	<85	0	0	0	6	0.0009962
	168	NC122	No near neighbour	<85	0	1	0.013986	0	0
	169	ND148	No near neighbour	<85	0	0	0	23	0.0038187
24						**1**	**0.013986**	**29**	**0.0048149**
	170	ND031	*Slackia exigua* ATCC 700122	99.5	0	0	0	189	0.0313796
25								**189**	**0.0313796**
	171	ND145	*Eggerthella lenta* DSM 2243	89	0	0	0	33	0.005479
26								**33**	**0.005479**
	172	ND836	*Atopobium vaginae* PB189-T1-4	100	0	0	0	15	0.0024904
	173	NC119	*Atopobium vaginae* PB189-T1-4	98.6	0	1	0.013986	0	0
27						**1**	**0.013986**	**15**	**0.0024904**

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	Branch[2]	Label[3]	Nearest Neighbour[4]	%ID	Isolates	Clones	% of total clones	Reads	% of total reads
	174	NC045	*Atopobium vaginae* DSM15829	98	0	5	0.0699301	0	0
	175	ND128	*Atopobium vaginae* DSM15829	100	0	0	0	184	0.0305495
	176	NC108	*Atopobium vaginae* DSM15829	96.4	0	1	0.013986	0	0
	177	ND106	*Atopobium vaginae* DSM15829	97	0	0	0	2625	0.4358279
	178	ND010	*Atopobium vaginae* DSM15829	96.5	0	0	0	3942	0.6544889
	179	NC047	*Atopobium vaginae* DSM15829	97.5	0	117	1.6363636	0	0
28						**123**	**1.7202797**	**6751**	**1.1208663**
	180	ND538	*Streptococcus* sp. (v5264)	99.5	0	0	0	105	0.0174331
	181	N142	*Streptococcus pseudopneumoniae* ATCC BAA-960	93.7	2	0	0	0	0
	182	ND280	*Streptococcus* sp. (v5264)	99	0	0	0	7	0.0011622
	183	N163	*Streptococcus mitis* ATCC 49456	94.6	1	0	0	0	0
	184	N080	*Streptococcus mitis* ATCC 49456	94.4	1	0	0	0	0
	185	N127	*Streptococcus pneumoniae* SP14-BS69	94.6	1	0	0	0	0
	186	ND446	*Streptococcus infantis* ATCC 700779	89	0	0	0	26	0.0043168
29					**5**			**138**	**0.0229121**
	187	N085	*Streptococcus anginosus* ATCC 33397	95.3	2	0	0	0	0
	188	N121	*Streptococcus anginosus* ATCC 33397	96.2	1	0	0	0	0
	189	N124	*Streptococcus constellatus* ATCC 27823	96.9	4	0	0	0	0
30					**7**				
	190	NC132	*Streptococcus salivarius* ATCC 7073	95.8	0	1	0.013986	0	0
	191	NC184	*Streptococcus salivarius* ATCC 7073	95.8	0	17	0.2377622	0	0
31						**18**	**0.2517483**		
	192	N011	*Streptococcus gallolyticus* CCUG35224	99.3	5	0	0	19	0.0031546
	193	N001	*Streptococcus infantarius* CCUG 43822	100	3	0	0	0	0
	194	N009	*Streptococcus lutetiensis* ATCC BAA-103	99.8	1	0	0	0	0

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
	195	N128	*Streptococcus agalactiae* ATCC 12386	100	4	0	0	0	0
32					**13**			**19**	**0.0031546**
	196	N040	*Enterococcus raffinosus* ATCC 49427	100	2	0	0	0	0
	197	N162	*Enterococcus faecalis* ATCC 49332	99.6	1	0	0	0	0
33					**3**				
	198	N167	*Bacillus cereus* ATCC 10876	98	1	0	0	0	0
	199	N099	*Gemella haemolysans* ATCC 10379	92.2	1	0	0	0	0
	200	N100	*Gemella haemolysans* z1643	98	1	0	0	0	0
34					**3**				
	201	N118	*Staphylococcus lugdunensis* CRSN850412	99.8	1	0	0	0	0
	202	N122	*Staphylococcus lugdunensis* CRSN850412	98.9	1	0	0	0	0
	203	N082	*Staphylococcus lugdunensis* CRSN850412	95.1	1	0	0	0	0
	204	N111	*Staphylococcus muscae* CCS767	98.9	1	0	0	0	0
	205	N113	*Staphylococcus muscae* CCS767	99.8	1	0	0	0	0
35					**5**				
	206	N119	*Staphylococcus piscifermentans* ATCC 51136	99.5	1	0	0	0	0
	207	N106	*Staphylococcus simulans* ATCC 27848	99.1	1	0	0	0	0
	208	N104	*Staphylococcus hominis* ATCC 27844	99.8	1	0	0	0	0
36					**3**				
	209	N102	*Staphylococcus haemolyticus* ATCC 29970	100	1	0	0	0	0
	210	N125	*Staphylococcus haemolyticus* ATCC 29970	98.9	1	0	0	0	0
37					**2**				
	211	ND032	No near neighbour	<85	0	0	0	871	0.1446118
38								**871**	**0.1446118**
	212	N065	No near neighbour	<85	1	0	0	0	0

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
39					**1**				
	213	N018	*Lactobacillus* sp. N6	98.9	2	7	0.0979021	375	0.0622611
	214	N023	*Lactobacillus* sp. N6	99.8	5	13	0.1818182	0	0
	215	ND307	*Lactobacillus* sp. N6	100	0	0	0	13057	2.1678494
	216	N006	*Lactobacillus* sp. N6	99.8	1	0	0	0	0
	217	N139	*Lactobacillus* sp. N6	97.8	1	0	0	0	0
	218	N068	*Lactobacillus* sp. N6	96.6	4	0	0	0	0
40					**13**	**20**	**0.2797203**	**13432**	**2.2301105**
	219	N019	*Lactobacillus reuteri* JCM 1112	92.8	1	0	0	0	0
	220	N047	*Lactobacillus reuteri* JCM 1112	100	1	0	0	0	0
	221	N020	*Lactobacillus coleohominis* 101-4-CH	97.5	5	0	0	187	0.0310475
	222	N073	*Lactobacillus coleohominis* 101-4-CH	99.8	1	0	0	0	0
41					**8**			**187**	**0.0310475**
	223	N027	No near neighbour	<85	1	0	0	0	0
	224	N054	No near neighbour	<85	1	0	0	802	0.1331558
42					**2**			**802**	**0.1331558**
	225	N017	*Lactobacillus plantarum* I18A	99.8	2	0	0	0	0
	226	N078	*Pediococcus pentosaceus* ATCC 25745	99.3	1	0	0	0	0
	227	N039	*Lactobacillus ruminis* ATCC 25644	99.3	4	0	0	0	0
	228	N035	No near neighbour	<85	3	0	0	0	0
43					**10**				
	229	N003	*Lactobacillus salivarius* DSM 20555	100	12	13	0.1818182	1703	0.2827485
	230	N042	*Lactobacillus salivarius* JCM 1042	99.8	2	0	0	0	0
44					**14**	**13**	**0.1818182**	**1703**	**0.2827485**
	231	ND111	*Lactobacillus iners* DSM 13335	98	0	0	0	151	0.0250705

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
	232	ND246	*Lactobacillus iners* DSM 13335	98.5	0	0	0	17730	2.943706
	233	NC009	*Lactobacillus iners* DSM 13335	98.2	0	16	0.2237762	732	0.1215337
	234	ND279	*Lactobacillus iners* DSM 13335	99	0	0	0	85	0.0141125
	235	NC011	*Lactobacillus iners* DSM 13335	96.4	0	3	0.041958	0	0
	236	ND471	*Lactobacillus iners* DSM 13335	99.5	0	0	0	12	0.0019924
	237	NC013	*Lactobacillus iners* DSM 13335	98.9	0	2	0.027972	446	0.0740492
	238	NC014	*Lactobacillus iners* DSM 13335	98.9	0	1	0.013986	40617	6.743627
	239	ND347	*Lactobacillus iners* DSM 13335	98.5	0	0	0	106	0.0175991
	240	NC041	*Lactobacillus iners* A3M7	99.3	0	439	6.1398601	25	0.0041507
	241	ND019	*Lactobacillus iners* A3M7	99	0	0	0	460	0.0763736
	242	NC005	*Lactobacillus iners* A3M7	98.2	0	2715	37.972028	0	0
	243	ND006	*Lactobacillus iners* DSM 13335	99.5	0	0	0	10	0.0016603
	244	NC010	*Lactobacillus iners* DSM 13335	97.5	0	2	0.027972	0	0
45						**3178**	**44.447552**	**60374**	**10.023875**
	245	ND675	*Lactobacillus crispatus* CECT4840	99.5	0	0	0	81	0.0134484
	246	ND734	*Lactobacillus crispatus* CECT4840	99.5	0	0	0	13	0.0021584
	247	ND562	*Lactobacillus crispatus* CECT4840	99.5	0	0	0	12	0.0019924
	248	N028	*Lactobacillus crispatus* CECT4840	100	5	1089	15.230769	36472	6.055434
	249	ND863	*Lactobacillus crispatus* CECT4840	99	0	0	0	3	0.0004981
	250	ND021	*Lactobacillus crispatus* CECT4840	99.5	0	0	0	5	0.0008301
	251	ND319	*Lactobacillus crispatus* CECT4840	99.5	0	0	0	13	0.0021584
	252	ND623	*Lactobacillus crispatus* CECT4840	99.5	0	0	0	9	0.0014943
	253	NC002	*Lactobacillus crispatus* CECT4840	98.2	0	1	0.013986	0	0
	254	ND443	*Lactobacillus crispatus* CECT4840	99.5	0	0	0	6	0.0009962
	255	NC001	*Lactobacillus crispatus* CECT4840	97.6	0	10	0.1398601	0	0

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
46						**1100**	**15.384615**	**36614**	**6.0790102**
	256	N012	*Lactobacillus gasseri* MV-22	100	1	0	0	0	0
	257	N038	*Lactobacillus gasseri* MV-22	99.7	4	0	0	0	0
	258	N022	*Lactobacillus gasseri* ATCC 9857	98.9	1	0	0	0	0
	259	N048	*Lactobacillus gasseri* ATCC 9857	98.7	5	0	0	0	0
	260	N002	*Lactobacillus johnsonii* NCC 533	99.3	1	0	0	0	0
47					**12**				
	261	N046	*Lactobacillus delbrueckii* PB2003/044-T3-4	99.8	2	0	0	2	0.0003321
48					**2**			**2**	**0.00033**
	262	ND439	*Lactobacillus jensenii* ATCC 25258	100	0	0	0	11029	1.8311412
	263	N154	*Lactobacillus jensenii* ATCC 25258	100	3	0	0	0	0
	264	ND667	*Lactobacillus jensenii* ATCC 25258	96	0	0	0	1695	0.2814203
	265	ND401	*Lactobacillus jensenii* ATCC 25258	95	0	0	0	8568	1.4225422
	266	N004	*Lactobacillus jensenii* ATCC 25258	95.7	4	0	0	0	0
49					**7**			**21292**	**3.5351037**
	267	ND633	Clone unknown hpoC_006_a04	100	0	0	0	6	0.0009962
	268	ND860	*Faecalibacterium prausnitzii* A2-165	98	0	0	0	9	0.0014943
	269	NC124	*Clostridium* genomosp. BVAB3 UPII9-5	86.4	0	3	0.041958	0	0
50						**3**	**0.041958**	**15**	**0.0024904**
	270	ND550	No near neighbour	<85	0	0	0	41	0.0068072
	271	NC015	No near neighbour	<85	0	1	0.013986	0	0
	272	ND784	*Bacteroides coagulans* JCM 12528	98	0	0	0	77	0.0127843
51						**1**	**0.013986**	**118**	**0.0195915**
	273	ND116	No near neighbour	<85	0	0	0	12	0.0019924
	274	NC053	No near neighbour	<85	0	295	4.1258741	27121	4.5028906

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
	275	ND756	No near neighbour	<85	0	0	0	6	0.0009962
	276	NC029	No near neighbour	<85	0	19	0.2657343	0	0
	277	NC173	No near neighbour	<85	0	4	0.0559441	0	0
	278	ND059	No near neighbour	<85	0	0	0	6312	1.0479793
	279	ND332	No near neighbour	<85	0	0	0	133	0.0220819
52						**318**	**4.4475524**	**33584**	**5.5759403**
	280	ND318	*Megasphaera* sp. genomosp. type_1 str. 28L	99.5	0	0	0	6754	1.1213644
	281	ND722	*Megasphaera* sp. genomosp. type_1 str. 28L	99	0	0	0	5	0.0008301
	282	ND635	*Megasphaera* sp. genomosp. type_1 str. 28L	99	0	0	0	8	0.0013282
	283	ND630	*Megasphaera* sp. genomosp. type_1 str. 28L	99	0	0	0	21	0.0034866
	284	NC044	*Megasphaera* sp. genomosp. type_1 str. 28L	99.6	0	200	2.7972028	42452	7.0482914
	285	ND479	*Megasphaera* sp. genomosp. type_1 str. 28L	99.5	0	0	0	54	0.0089656
	286	ND657	*Megasphaera* sp. genomosp. type_1 str. 28L	99.5	0	0	0	7	0.0011622
	287	ND665	*Megasphaera* sp. genomosp. type_1 str. 28L	99.5	0	0	0	7	0.0011622
	288	NC077	*Megasphaera* sp. genomosp. type_1 str. 28L	98.9	0	1	0.013986	0	0
	289	ND353	*Megasphaera* sp. genomosp. type_1 str. 28L	96.5	0	0	0	1428	0.2370904
	290	NC086	No near neighbour	<85	0	1	0.013986	0	0
53						**202**	**2.8251748**	**50736**	**8.4236811**
	291	ND625	No near neighbour	<85	0	0	0	21	0.0034866
	292	ND839	Clostridia sp. 192_1_59	98.5	0	0	0	352	0.0584424
54								**373**	**0.0619291**
	293	NC146	No near neighbour	<85	0	4	0.0559441	0	0
	294	NC150	No near neighbour	<85	0	3	0.041958	0	0
	295	NC147	No near neighbour	<85	0	1	0.013986	0	0
55						**8**	**0.1118881**		

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
	296	ND375	No near neighbour	<85	0	0	0	161	0.0267308
	297	NC040	No near neighbour	<85	0	100	1.3986014	13882	2.3048238
	298	ND198	No near neighbour	<85	0	0	0	7	0.0011622
	299	NC158	No near neighbour	<85	0	2	0.027972	0	0
	300	ND752	*Dialister invisus* DSM 15470	96.5	0	0	0	61	0.0101278
56						**102**	**1.4265734**	**14111**	**2.3428446**
	301	NC168	No near neighbour	<85	0	1	0.013986	0	0
	302	NC021	No near neighbour	<85	0	1	0.013986	0	0
	303	ND065	No near neighbour	<85	0	0	0	117	0.0194255
	304	NC118	No near neighbour	<85	0	3	0.041958	0	0
	305	ND291	No near neighbour	<85	0	0	0	2655	0.4408088
	306	NC154	No near neighbour	<85	0	1	0.013986	0	0
	307	ND431	No near neighbour	<85	0	0	0	224	0.0371906
	308	NC153	No near neighbour	<85	0	5	0.0699301	0	0
	309	NC038	No near neighbour	<85	0	1	0.013986	0	0
	310	NC039	No near neighbour	<85	0	81	1.1328671	2349	0.3900037
57						**93**	**1.3006993**	**5345**	**0.8874286**
	311	N169	No near neighbour	<85	1	0	0	0	0
	312	ND638	No near neighbour	<85	0	0	0	2	0.0003321
	313	NC096	No near neighbour	<85	0	1	0.013986	0	0
58					**1**	**1**	**0.013986**	**2**	**0.0003321**
	314	ND038	*Sphingobium japonicum* UT26S	89.5	0	0	0	1613	0.2678059
	315	ND658	*Sphingomonas wittichii* RW1	89.5	0	0	0	408	0.0677401
	316	ND856	*Sphingomonas wittichii* RW1	92.5	0	0	0	4	0.0006641
	317	ND857	*Sphingomonas wittichii* RW1	92.5	0	0	0	417	0.0692344

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
59								**2442**	**0.4054444**
	318	ND057	*Methylobacterium radiotolerans* JCM 2831	98.5	0	0	0	79	0.0131163
60								**79**	**0.0131163**
	319	ND614	*Rhodopseudomonas palustris* HaA2	92.5	0	0	0	80	0.0132824
	320	ND690	*Bradyrhizobium japonicum* USDA110	89	0	0	0	13	0.0021584
61								**93**	**0.0154408**
	321	ND652	*Mesorhizobium amorphae* DSM 21831	90.5	0	0	0	95	0.0157728
	322	ND056	No near neighbour	<85	0	0	0	14	0.0023244
62								**109**	**0.0180972**
	323	ND251	*Bradyrhizobium japonicum* USDA110	90	0	0	0	79	0.0131163
	324	ND788	*Bradyrhizobium japonicum* USDA110	92.5	0	0	0	18	0.0029885
63								**97**	**0.0161049**
	325	ND559	*Mesorhizobium opportunistum* WSM 2075	87.5	0	0	0	91	0.0151087
	326	ND643	*Rhizobium etli* CFN 42	89	0	0	0	16	0.0026565
	327	ND447	No near neighbour	<85	0	0	0	1	0.000166
64								**108**	**0.0179312**
	328	ND253	*Variovorax paradoxus* z1389	98	0	0	0	7216	1.1980701
	329	ND713	No near neighbour	<85	0	0	0	139	0.0230781
	330	ND427	*Comamonas terrigena* ATCC 8461	88	0	0	0	1219	0.2023902
	331	ND765	*Comamonas terrigena* ATCC 8461	91	0	0	0	447	0.0742153
65								**9021**	**1.4977536**
	332	ND481	*Moraxella* sp. z390	99	0	0	0	55	0.0091316
66								**55**	**0.0091316**
	333	ND674	*Morganella morganii* A7-39	99.5	0	0	0	47	0.0078034
67								**55**	**0.0091316**

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
	334	ND653	*Klebsiella pneumoniae* ATCC 13884	96	0	0	0	11	0.0018263
	335	N110	*Klebsiella pneumoniae* ATCC 13884	99.6	1	0	0	2078	0.3450096
68					**1**			**2089**	**0.346836**
	336	ND499	*Escherichia coli* MS 21-1	99.5	0	0	0	24	0.0039847
	337	ND491	*Escherichia coli* MS 21-1	99.5	0	0	0	52	0.0086335
	338	ND289	*Escherichia coli* MS 21-1	99.5	0	0	0	429	0.0712267
	339	ND040	*Escherichia coli* MS 21-1	99.5	0	0	0	21	0.0034866
	340	N098	*Escherichia coli* MS 198-1	100	2	0	0	46470	7.7153986
	341	ND072	*Escherichia coli* MS 21-1	99.5	0	0	0	470	0.0780339
	342	ND281	*Escherichia coli* ATCC 700336	99	0	0	0	27	0.0044828
	343	ND805	*Escherichia coli* MS 21-1	99.5	0	0	0	6	0.0009962
	344	ND064	*Escherichia coli* K12	100	0	0	0	3488	0.5791115
	345	ND698	*Escherichia coli* K12	99.5	0	0	0	13	0.0021584
69					**2**			**51000**	**8.467513**
	346	ND088	*Porphyromonas uenonis* JCM 13868	98.5	0	0	0	3	0.0004981
	347	ND425	*Porphyromonas uenonis* JCM 13868	97.5	0	0	0	143	0.0237422
	348	ND151	*Porphyromonas uenonis* JCM 13868	99.5	0	0	0	74	0.0122862
	349	ND503	*Porphyromonas uenonis* JCM 13868	99	0	0	0	2	0.0003321
	350	ND257	*Porphyromonas uenonis* JCM 13868	99	0	0	0	225	0.0373567
	351	NC035	*Porphyromonas uenonis* JCM 13868	99	0	3	0.041958	6473	1.07471
	352	ND315	*Porphyromonas uenonis* JCM 13868	97.5	0	0	0	1871	0.3106415
	353	NC117	*Porphyromonas uenonis* JCM 13868	98.2	0	34	0.4755245	0	0
	354	ND104	*Porphyromonas uenonis* JCM 13868	98	0	0	0	2176	0.3612806
	355	NC123	*Porphyromonas uenonis* JCM 13868	97.5	0	1	0.013986	0	0
	356	NC126	*Porphyromonas uenonis* JCM 13868	98.4	0	3	0.041958	0	0

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
	357	ND469	*Porphyromonas uenonis* JCM 13868	96	0	0	0	6	0.0009962
	358	ND468	*Porphyromonas asaccharolytica* JCM 6326	97.5	0	0	0	14	0.0023244
70						**41**	**0.5734266**	**10987**	**1.8241679**
	359	ND504	No near neighbour	<85	0	0	0	70	0.0116221
71								**70**	**0.0116221**
	360	NC148	No near neighbour	<85	0	1	0.013986	0	0
72						**1**	**0.013986**		
	361	NC159	No near neighbour	<85	0	3	0.041958	0	0
73						**3**	**0.041958**		
	362	ND574	*Bacteroides forsythus* ATCC 43037	99	0	0	0	69	0.011456
74								**69**	**0.011456**
	363	ND651	*Porphyromonas gingivalis* ATCC 33277	100	0	0	0	28	0.0046488
75								**28**	**0.0046488**
	364	ND632	No near neighbour	<85	0	0	0	1	0.000166
76								**1**	**0.000166**
	365	ND834	No near neighbour	<85	0	0	0	48	0.0079694
77								**48**	**0.0079694**
	366	ND069	*Prevotella bivia* JCVIHMP010	100	0	0	0	101	0.016769
	367	N114	*Prevotella bivia* JCM 6332	99.8	2	0	0	0	0
	368	NC121	*Prevotella bivia* JCVIHMP010	99.3	0	31	0.4335664	0	0
	369	NC161	*Prevotella bivia* JCVIHMP010	92.3	0	1	0.013986	0	0
78					**2**	**32**	**0.4475524**	**101**	**0.016769**
	370	ND880	*Prevotella amnii* JCM 14753	98.5	0	0	0	1109	0.1841269
	371	NC037	*Prevotella amnii* JCM 14753	100	0	10	0.1398601	3190	0.5296346
	372	NC175	*Prevotella melaninogenica* D18	92.1	0	2	0.027972	0	0

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
79						**12**	**0.1678322**	**4299**	**0.7137615**
	373	ND485	*Prevotella corporis* JCM 8529	100	0	0	0	121	0.0200896
	374	N083	*Prevotella disiens* ATCC 29426	99.1	1	0	0	38	0.0063091
	375	ND453	*Prevotella bergensis* JCM 13869	98.5	0	0	0	2	0.0003321
80					**1**			**161**	**0.0267308**
	376	ND181	No near neighbour	<85	0	0	0	33	0.005479
	377	NC115	No near neighbour	<85	0	79	1.1048951	0	0
	378	ND150	No near neighbour	<85	0	0	0	4804	0.7976065
	379	NC162	No near neighbour	<85	0	1	0.013986	0	0
81						**80**	**1.1188811**	**4837**	**0.8030855**
	380	NC020	*Prevotella melaninogenica* D18	98.6	0	76	1.0629371	8564	1.4218781
	381	NC043	*Prevotella melaninogenica* D18	96	0	1	0.013986	0	0
	382	ND384	*Prevotella zoogleoformans* ATCC 33285	99	0	0	0	1500	0.2490445
	383	N090	*Bacteroides fragilis* ATCC 25285	100	2	0	0	0	0
82					**2**	**77**	**1.0769231**	**10064**	**1.6709226**
	384	ND182	*Prevotella buccalis* ATCC 35310	96	0	0	0	564	0.0936407
	385	NC023	*Prevotella buccalis* ATCC 35310	97	0	11	0.1538462	0	0
	386	NC129	*Prevotella buccalis* ATCC 35310	96	0	5	0.0699301	0	0
	387	ND718	*Prevotella buccalis* ATCC 35310	96	0	0	0	374	0.0620951
	388	ND271	*Prevotella buccalis* ATCC 35310	96.5	0	0	0	62	0.0102938
	389	ND256	*Prevotella buccalis* ATCC 35310	96.5	0	0	0	482	0.0800263
	390	ND450	*Prevotella buccalis* ATCC 35310	97.5	0	0	0	35	0.005811
83						**16**	**0.2237762**	**1517**	**0.251867**
	391	NC163	*Prevotella timonensis* CRIS 5C-B1	94.1	0	1	0.013986	0	0
	392	NC170	*Prevotella timonensis* CRIS 5C-B1	92.8	0	2	0.027972	0	0

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
	393	NC125	*Prevotella buccalis* ATCC 35310	90.7	0	179	2.5034965	0	0
	394	NC149	*Prevotella buccalis* ATCC 35310	90.7	0	1	0.013986	0	0
	395	ND420	*Prevotella buccalis* ATCC 35310	88	0	0	0	5338	0.8862664
	396	NC166	*Prevotella buccalis* ATCC 35310	89.4	0	89	1.2447552	0	0
84						**272**	**3.8041958**	**5338**	**0.8862664**
	397	NC169	No near neighbour	<85	0	2	0.027972	0	0
	398	ND728	*Prevotella timonensis* CRIS 5C-B1	99	0	0	0	1	0.000166
	399	ND287	*Prevotella timonensis* CRIS 5C-B1	100	0	0	0	203	0.033704
	400	ND426	*Prevotella timonensis* CRIS 5C-B1	99.5	0	0	0	28237	4.6881797
	401	NC160	*Prevotella buccalis* ATCC 35310	90.7	0	4	0.0559441	0	0
	402	NC165	*Prevotella buccalis* ATCC 35310	88.5	0	1	0.013986	0	0
	403	NC167	*Prevotella timonensis* CRIS 5C-B1	95.7	0	83	1.1608392	0	0
	404	ND100	*Prevotella timonensis* CRIS 5C-B1	99	0	0	0	161	0.0267308
	405	ND282	*Prevotella timonensis* CRIS 5C-B1	96.5	0	0	0	9	0.0014943
	406	ND684	*Prevotella timonensis* JCM 15640	96	0	0	0	15	0.0024904
	407	ND050	*Prevotella timonensis* JCM 15640	96.5	0	0	0	605	0.1004479
	408	ND762	*Prevotella timonensis* JCM 15640	97.5	0	0	0	34	0.005645
	409	ND681	*Prevotella timonensis* JCM 15640	99	0	0	0	56	0.0092977
	410	ND103	*Prevotella timonensis* JCM 15640	97	0	0	0	3523	0.5849225
	411	ND313	*Prevotella timonensis* JCM 15640	98	0	0	0	1932	0.3207693
	412	ND741	*Prevotella timonensis* JCM 15640	97.5	0	0	0	8	0.0013282
	413	NC078	*Prevotella timonensis* CRIS 5C-B1	98.4	0	194	2.7132867	0	0
	414	ND737	*Prevotella timonensis* JCM 15640	97.5	0	0	0	16	0.0026565
	415	ND122	*Prevotella timonensis* JCM 15640	98	0	0	0	8	0.0013282
	416	ND051	*Prevotella timonensis* JCM 15640	98	0	0	0	10	0.0016603

Table S1: Nearest cpnDB neighbours and relative abundance of deep sequencing OTU (ND), library clones (NC) and isolates (N)

Cluster[1]	*Branch*[2]	*Label*[3]	*Nearest Neighbour*[4]	*%ID*	*Isolates*	*Clones*	*% of total clones*	*Reads*	*% of total reads*
	417	ND305	*Prevotella timonensis* JCM 15640	98.5	0	0	0	17912	2.9739234
	418	ND403	*Prevotella timonensis* JCM 15640	98	0	0	0	82	0.0136144
	419	ND331	*Prevotella timonensis* JCM 15640	97.5	0	0	0	25	0.0041507
	420	ND472	*Prevotella timonensis* JCM 15640	98	0	0	0	5	0.0008301
85						**282**	**3.9440559**	**52842**	**8.7733396**

[1] Clusters of branches summarized in circles at branch ends in Fig. 1; [2] All branches of Fig. 1 phylogeny shown clockwise from bottom left;
[3] N=Nairobi isolate, NC=Nairobi clone, ND=Nairobi deep sequencing OTU; [4] Nearest neighbour in cpnDB.

Table S2: Most frequently detected operational taxonomic units by abundance of reads detected

			Proportion of reads		*Proportion of group*	
Rank	*OTU*[1]	*Nearest neighbour*[2]	*# reads*	*% total*	*# colonized*	*% total*
1	ND348	*Escherichia coli* 100%	46470	7.70	20	45
2	ND175	*Megasphaera sp.* 99.6%	42076	6.97	31	70
3	ND255	*Lactobacillus iners* 98.9%	40617	6.73	40	91
4	ND393	*Gardnerella vaginalis* 87%	29972	4.97	23	52
5	ND426	*Prevotella timonensis* 99.5%	28237	4.68	26	59
6	ND320	*Lactobacillus crispatus* 100%	27717	4.59	21	48
7	ND254	*Gardnerella vaginalis* 90.2%	27641	4.58	24	55
8	ND317	No near neighbour (NC053)	26090	4.32	31	70
9	ND316	*Gardnerella vaginalis* 97%	20713	3.43	35	80
10	ND252	*Gardnerella vaginalis* 88.9%	20350	3.37	37	84

[1] OTU: Operational taxonomic unit, ND: Nairobi deep sequencing OTU; [2] Nearest neighbour in cpnDB

Table S3: Most frequently detected operational taxonomic units by proportion of group colonized

Rank	*OTU*[1]	*Nearest neighbour*[2]	*Proportion of group*		*Proportion of reads*	
			# colonized	*% total*	*# reads*	*% total*
1	ND263	*Gardnerella vaginalis* 90%	40	91	11862	1.97
1	ND255	*Lactobacillus iners* 98.9%	40	91	40617	6.73
2	ND237	No near neighbour (NC040)	39	89	13608	2.25
2	ND253	*Variovorax paradoxus* 98%	39	89	7110	1.18
3	ND252	*Gardnerella vaginalis* 88.9%	37	84	20350	3.37
4	ND038	*Sphingobium japonicum* 89.5%	36	82	1588	0.26
4	ND427	*Comamonas terrigena* 88%	36	82	1202	0.20
5	ND316	*Gardnerella vaginalis* 97%	35	80	20713	3.43
6	ND317	No near neighbour (NC053)	31	70	26090	4.32
6	ND175	*Megasphaera sp.* 99.6%	31	70	42076	6.97
6	ND246	*Lactobacillus iners* 98.5%	31	70	17730	2.94
7	ND313	*Prevotella timonensis* 98%	30	68	1932	0.32
8	ND435	No near neighbour (NC040)	27	61	176	0.03
8	ND307	*Lactobacillus sp. N6* 100%	27	61	13057	2.16
9	ND231	*Gardnerella vaginalis* 100%	26	59	11869	1.97
9	ND426	*Prevotella timonensis* 99.5%	26	59	28237	4.68
9	ND375	No near neighbour (NC040)	26	59	161	0.03
10	ND404	*Gardnerella vaginalis* 98%	25	57	8938	1.48
10	ND305	*Prevotella timonensis* 98.5%	25	57	17815	2.95
10	ND857	*Sphingomonas wittichii* 92.5%	25	57	417	0.07

[1] OTU: Operational taxonomic unit, ND: Nairobi deep sequencing OTU; [2] Nearest neighbour in cpnDB

Table S4: Distribution of operational taxonomic units by HIV serostatus and significance testing

			% colonized			*Median Tier*[4]			*Significance 1 (p)*[3]			*Significance 2 (p)*[4]		
Row[1]	*OTU*[2]	*Nearest neighbour*[3]	*HESN*	*HIV-N*	*HIV+*	*HESN*	*HIV-N*	*HIV+*	*HESN/N*	*N/+*	*HESN/+*	*HESN/N*	*N/+*	*HESN/+*
1	ND348	*Escherichia coli* 100%	25	25	100	5	5	2		0.00004	0.00008			
2	ND765	*Comamonas terrigena* 91%	50	44	8	4.5	5	5		0.03	0.02			
3	ND713	No near neighbour <85%	43	25	0	5	5	5			0.01			
4	ND427	*Comamonas terrigena* 88%	93	69	83	3	3	3						
5	ND251	*Bradyrhizobium japonicum* 90%	37	25	42	5	5	5						
6	ND670	*Variovorax paradoxus* 98%	43	38	33	5	5	5						
7	ND326	*Variovorax paradoxus* 98%	62	25	42	3.5	5	5						
8	ND253	*Variovorax paradoxus* 98%	87	94	83	2	2	2						
9	ND308	No near neighbour <85%	56	31	50	3	5	4.5						
10	ND057	*Methylobacterium radiotolerans* 98.5%	37	31	50	5	5	4.5						
11	ND559	*Mesorhizobium opportunistum* 87.5%	43	31	58	5	5	4						
12	ND614	*Rhodopseudomonas palustris* 92.5%	25	25	42	5	5	5						
13	ND389	*Sphingomonas wittichii* 92.5%	6	44	25	5	5	5	0.02					
14	ND658	*Sphingomonas wittichii* 89.5%	43	50	75	5	4.5	3.5						
15	ND857	*Sphingomonas wittichii* 92.5%	56	44	75	3	5	3						
16	ND038	*Sphingobium japonicum* 89.5%	87	75	83	2.5	3	2.5						
17	ND401	*Lactobacillus jensenii* 95%	25	44	8	5	5	5		0.03				
18	ND439	*Lactobacillus jensenii* 100%	25	69	25	5	4	5	0.04	0.04				
19	ND296	*Lactobacillus crispatus* 100%	25	44	0	5	5	5		0.01				
20	ND320	*Lactobacillus crispatus* 100%	50	75	8	4.5	3	5		0.001	0.03			
21	ND085	*Lactobacillus iners* 99%	37	31	42	5	5	5					0.03	
22	ND294	*Lactobacillus iners* 99%	37	44	17	5	5	5						
23	ND295	*Lactobacillus iners* 98.2%	25	44	17	5	5	5						
24	ND019	*Lactobacillus iners* 99%	37	38	33	5	5	5						
25	ND877	*Lactobacillus iners* 98.9%	43	44	58	5	5	4						
26	ND284	*Lactobacillus iners* 98.2%	37	44	50	5	5	4.5						
27	ND246	*Lactobacillus iners* 98.5%	75	50	92	3	4.5	2	0.05	0.006				
28	ND255	*Lactobacillus iners* 98.9%	87	94	92	2.5	2	2						
29	ND303	*Lactobacillus sp. N6* 98.9%	37	44	8	5	5	5						
30	ND307	*Lactobacillus sp. N6* 100%	68	69	42	3	3.5	5						
31	ND032	No near neighbours	25	63	50	5	4	4.5						
32	ND089	No near neighbour (NC039)	37	25	8	5	5	5						
33	ND278	No near neighbour (NC039)	37	31	17	5	5	5						
34	ND398	No near neighbour (NC039)	56	44	42	4	5	5						
35	ND291	No near neighbour (NC039)	43	56	33	5	4	5						
36	ND360	*Megasphaera sp.* 99.5%	25	25	33	5	5	5						

Table S4: Distribution of operational taxonomic units by HIV serostatus and significance testing

			% colonized			*Median Tier*[4]			*Significance 1 (p)*[3]			*Significance 2 (p)*[4]		
Row[1]	*OTU*[2]	*Nearest neighbour*[3]	*HESN*	*HIV-N*	*HIV+*	*HESN*	*HIV-N*	*HIV+*	*HESN/N*	*N/+*	*HESN/+*	*HESN/N*	*N/+*	*HESN/+*
37	ND044	No near neighbour (NC040)	50	63	50	4.5	4	4.5						
38	ND375	No near neighbour (NC040)	56	69	50	4	4	4.5						
39	ND435	No near neighbour (NC040)	56	69	58	3.5	3	4						
40	ND237	No near neighbour (NC040)	93	94	75	2	1.5	2						
41	ND084	*Megasphaera sp.* 99.5%	25	31	25	5	5	5						
42	ND318	*Megasphaera sp.* 99.5%	25	44	58	5	5	3.5						
43	ND175	*Megasphaera sp.* 99.6%	68	63	83	3	3	3						
44	ND286	No near neighbour (NC053)	18	44	17	5	5	5						
45	ND344	No near neighbour (NC053)	25	38	8	5	5	5						
46	ND288	No near neighbour (NC053)	18	38	17	5	5	5					0.05	
47	ND059	No near neighbour (NC053)	50	31	50	4.5	5	4						
48	ND317	No near neighbour (NC053)	62	81	67	4	2	3.5				0.04		
49	ND257	*Porphyromonas uenonis* 99%	18	50	8	5	4.5	5		0.02				
50	ND315	*Porphyromonas uenonis* 97.5%	6	63	0	5	3.5	5	0.001	0.001				
51	ND297	*Porphyromonas uenonis* 99%	18	69	42	5	3	5	0.01	0.02			0.03	0.02
52	ND069	*Prevotella bivia* 100%	18	13	50	5	5	4.5		0.02	0.04			0.03
53	ND023	*Prevotella amnii* 100%	43	50	50	5	4	4.5						
54	ND429	*Prevotella melaninogenica* 98.6%	62	31	17	3.5	5	5			0.03			
55	ND313	*Prevotella timonensis* 98%	75	63	67	3	3.5	3.5						
56	ND103	*Prevotella timonensis* 97%	56	6	42	4	5	5	0.005	0.04				
57	ND305	*Prevotella timonensis* 98.5%	81	50	33	3	4.5	5			0.03			
58	ND426	*Prevotella timonensis* 99.5%	31	75	75	5	3	3	0.03					
59	ND150	No near neighbour (NC115)	31	50	33	5	4.5	5						
60	ND256	*Prevotella buccalis* 96.5%	25	19	33	5	5	5						
61	ND182	*Prevotella buccalis* 96%	12	44	17	5	5	5						
62	ND420	*Prevotella buccalis* 88%	31	50	33	5	4.5	5						
63	ND406	*Atopobium vaginae* 96.5%	18	31	25	5	5	5						
64	ND106	*Atopobium vaginae* 97%	31	50	50	5	4.5	4.5						
65	ND010	*Atopobium vaginae* 96.5%	56	25	17	4	5	5						
66	ND484	*Coryne. tuberculostearicum* 99.5%	43	19	67	5	5	3.5		0.006				
67	ND739	No near neighbour <85%	25	25	25	5	5	5					0.05	
68	ND298	*Gardnerella vaginalis* 100%	50	13	25	4.5	5	5	0.04					
69	ND231	*Gardnerella vaginalis* 100%	81	25	75	3	5	3	0.01	0.04				
70	ND233	*Gardnerella vaginalis* 98%	31	44	33	5	5	5						
71	ND302	*Gardnerella vaginalis* 97.5%	37	38	50	5	5	4.5						
72	ND067	*Gardnerella vaginalis* 97.5%	37	25	50	5	5	4.5						

Table S4: Distribution of operational taxonomic units by HIV serostatus and significance testing

			% colonized			Median Tier[4]			Significance 1 (p)[3]			Significance 2 (p)[4]		
Row[1]	OTU[2]	Nearest neighbour[3]	HESN	HIV-N	HIV+	HESN	HIV-N	HIV+	HESN/N	N/+	HESN/+	HESN/N	N/+	HESN/+
73	ND316	*Gardnerella vaginalis* 97%	81	100	50	3	3	4		0.02	0.03			
74	ND285	*Gardnerella vaginalis* 98%	43	44	33	5	5	5						
75	ND404	*Gardnerella vaginalis* 98%	56	63	50	3	3.5	4.5						
76	ND260	*Gardnerella vaginalis* 96.5%	18	56	58	5	3.5	3			0.05			
77	ND102	No near neighbour (NC070)	6	69	0	5	3.5	5	0.0005	0.0006				
78	ND301	No near neighbour (N72)	31	25	17	5	5	5						
79	ND304	*Gardnerella vaginalis* 88.9%	31	31	25	5	5	5						
80	ND394	*Gardnerella vaginalis* 88.9%	37	31	17	5	5	5						
81	ND252	*Gardnerella vaginalis* 88.9%	87	88	75	2.5	3	3						
82	ND042	*Gardnerella vaginalis* 89.1%	43	25	33	5	5	5						
83	ND061	*Gardnerella vaginalis* 86.5%	25	13	58	5	5	4		0.02				
84	ND054	*Gardnerella vaginalis* 90.9%	25	44	33	5	5	5						
85	ND436	*Gardnerella vaginalis* 87.5%	50	44	42	4.5	5	5						
86	ND412	*Gardnerella vaginalis* 88.9%	43	19	25	5	5	5						
87	ND328	*Gardnerella vaginalis* 90%	37	25	33	5	5	5						
88	ND263	*Gardnerella vaginalis* 90%	93	94	83	3	2.5	3						
89	ND393	*Gardnerella vaginalis* 87%	37	81	33	5	3	5	0.04	0.02				
90	ND254	*Gardnerella vaginalis* 90.2%	75	38	50	4	5	4.5	0.03					

[1] Row (from left to right) in Fig.3; [2] OTU: Operational taxonomic unit, ND: Nairobi deep sequencing OTU; [3] Nearest neighbour in cpnDB; [4] Median tier of abundance observed in each group (1=>1%, 2=0.1-0.99%, 3=0.01-0.099%, 4=0.001-0.0099%, 5=undetected); [5] Non-parametric test conducted in R, shaded cells indicate p value <0.05 after Bonferroni correction (p*90 tests), [6] Non-parametric test conducted in Genespring, no p value <0.05 after correction for multiple comparisons, shaded cell indicates significance by both methods.

Table S5: Distribution of operational taxonomic units by molecular BV status and significance testing

			% colonized			Median Tier[4]			Significance 1 (p)[5]			Significance 2 (p)[6]		
Row[1]	OTU[2]	Nearest neighbour[3]	BV+	mBVI	BV–	BV+	mBVI	BV–	BV+/mBVI	BV–/mBVI	BV+/BV–	BV+/mBVI	BV–/mBVI	BV+/BV–
1	ND348	*Escherichia coli* 100%	29	53	42	5	4	5						
2	ND765	*Comamonas terrigena* 91%	7	53	58	5	3	4	0.03		0.03			
3	ND713	No near neighbour <85%	7	40	33	5	5	5	0.03					
4	ND427	*Comamonas terrigena* 88%	71	93	92	3	3	2	0.005		0.004	0.009		0.02
5	ND251	*Bradyrhizobium japonicum* 90%	7	47	58	5	5	4			0.007			
6	ND670	*Variovorax paradoxus* 98%	7	53	67	5	4	4			0.007			
7	ND326	*Variovorax paradoxus* 98%	14	60	67	5	3	3	0.03		0.001			
8	ND253	*Variovorax paradoxus* 98%	79	93	100	3	2	1	0.003		0.0004	0.001		0.003
9	ND308	No near neighbour <85%	36	40	67	5	5	3		0.04	0.03			
10	ND057	*Methylobacterium radiotolerans* 98.5%	0	60	58	5	4	3.5	0.02		0.004			
11	ND559	*Mesorhizobium opportunistum* 87.5%	14	60	58	5	4	4	0.006		0.003			
12	ND614	*Rhodopseudomonas palustris* 92.5%	7	40	50	5	5	4			0.03			
13	ND389	*Sphingomonas wittichii* 92.5%	7	40	33	5	5	5						
14	ND658	*Sphingomonas wittichii* 89.5%	29	80	58	5	3	3	0.03		0.03			
15	ND857	*Sphingomonas wittichii* 92.5%	21	80	75	5	3	3	0.02		0.003			
16	ND038	*Sphingobium japonicum* 89.5%	71	93	92	3.5	2	2	0.005		0.002	0.01		0.01
17	ND401	*Lactobacillus jensenii* 95%	29	33	25	5	5	5						
18	ND439	*Lactobacillus jensenii* 100%	50	33	42	4.5	5	5						0.009
19	ND296	*Lactobacillus crispatus* 100%	29	20	33	5	5	5						
20	ND320	*Lactobacillus crispatus* 100%	64	27	58	3.5	5	4	0.05					
21	ND085	*Lactobacillus iners* 99%	7	47	67	5	5	3.5		0.04	0.002			
22	ND294	*Lactobacillus iners* 99%	7	47	58	5	5	3.5			0.02			
23	ND295	*Lactobacillus iners* 98.2%	0	47	50	5	5	4.5	0.01		0.002			
24	ND019	*Lactobacillus iners* 99%	7	47	67	5	5	3			0.007			
25	ND877	*Lactobacillus iners* 98.9%	14	60	83	5	3	2.5	0.007	0.04	0.0005			
26	ND284	*Lactobacillus iners* 98.2%	7	60	75	5	4	3	0.03		0.0003			
27	ND246	*Lactobacillus iners* 98.5%	57	73	75	4	3	3						
28	ND255	*Lactobacillus iners* 98.9%	93	87	100	3	1	0.5			0.01			0.02
29	ND303	*Lactobacillus sp. N6* 98.9%	14	27	67	5	5	2.5		0.01	0.002			
30	ND307	*Lactobacillus sp. N6* 100%	50	67	83	4.5	3	1.5		0.006	0.001			0.007
31	ND032	No near neighbour	57	40	25	4	5	5						

Table S5: Distribution of operational taxonomic units by molecular BV status and significance testing

			% colonized			Median Tier[4]			Significance 1 (p)[5]			Significance 2 (p)[6]		
Row[1]	OTU[2]	Nearest neighbour[3]	BV+	mBVI	BV–	BV+	mBVI	BV–	BV+/mBVI	BV–/mBVI	BV+/BV–	BV+/mBVI	BV–/mBVI	BV+/BV–
32	ND089	No near neighbour (NC039)	71	7	0	3	5	5	0.0004		0.0003			
33	ND278	No near neighbour (NC039)	86	7	0	3	5	5	0.0001		0.0003			
34	ND398	No near neighbour (NC039)	86	27	33	2	5	5	0.00008		0.00004			0.005
35	ND291	No near neighbour (NC039)	93	27	17	3	5	5	0.00003		0.00006	0.009		
36	ND360	*Megasphaera sp.* 99.5%	50	7	33	4	5	5	0.001	0.03				
37	ND044	No near neighbour (NC040)	86	53	17	4	4	5		0.02	0.001			
38	ND375	No near neighbour (NC040)	79	67	25	4	3	5		0.008	0.006			
39	ND435	No near neighbour (NC040)	100	60	17	3	3	5		0.007	0.0006			
40	ND237	No near neighbour (NC040)	100	80	83	1	1	3		0.02	0.0002		0.04	0.001
41	ND084	*Megasphaera sp.* 99.5%	57	13	17	4	5	5	0.0004		0.02			
42	ND318	*Megasphaera sp.* 99.5%	50	27	33	3.5	5	5	0.006		0.04	0.005		
43	ND175	*Megasphaera sp.* 99.6%	71	67	67	1.5	3	3	0.004			0.02		
44	ND286	No near neighbour (NC053)	57	20	0	3.5	5	5	0.007		0.0008	0.03		
45	ND344	No near neighbour (NC053)	57	13	8	3	5	5	0.004		0.005			
46	ND288	No near neighbour (NC053)	57	13	0	3.5	5	5	0.003		0.0008			
47	ND059	No near neighbour (NC053)	64	33	25	3	5	5	0.01		0.05			
48	ND317	No near neighbour (NC053)	79	60	75	2	3	3.5	0.03		0.03	0.02		0.006
49	ND257	*Porphyromonas uenonis* 99%	50	13	17	4	5	5	0.02					
50	ND315	*Porphyromonas uenonis* 97.5%	43	13	25	5	5	5						0.02
51	ND297	*Porphyromonas uenonis* 99%	71	27	33	3	5	5	0.005		0.02			
52	ND069	*Prevotella bivia* 100%	29	27	8	5	5	5						
53	ND023	*Prevotella amnii* 100%	71	53	17	2	4	5	0.0003		0.0002	0.04		
54	ND429	*Prevotella melaninogenica* 98.6%	50	27	50	4.5	5	4.5						
55	ND313	*Prevotella timonensis* 98%	79	60	58	3	3	3			0.04			
56	ND103	*Prevotella timonensis* 97%	50	27	25	4.5	5	5						
57	ND305	*Prevotella timonensis* 98.5%	64	40	67	4	5	3.5						
58	ND426	*Prevotella timonensis* 99.5%	57	73	42	4	3	5				0.02		
59	ND150	No near neighbour (NC115)	64	27	25	3	5	5	0.004		0.003			0.02
60	ND256	*Prevotella buccalis* 96.5%	36	20	17	5	5	5						
61	ND182	*Prevotella buccalis* 96%	36	20	17	5	5	5				0.04		
62	ND420	*Prevotella buccalis* 88%	57	20	42	4	5	5	0.01					

Table S5: Distribution of operational taxonomic units by molecular BV status and significance testing

			% colonized			*Median Tier*[4]			*Significance 1 (p)*[5]			*Significance 2 (p)*[6]		
Row[1]	*OTU*[2]	*Nearest neighbour*[3]	*BV+*	*mBVI*	*BV–*	*BV+*	*mBVI*	*BV–*	*BV+/mBVI*	*BV–/mBVI*	*BV+/BV–*	*BV+/mBVI*	*BV–/mBVI*	*BV+/BV–*
63	ND406	*Atopobium vaginae* 96.5%	43	27	0	5	5	5			0.002			
64	ND106	*Atopobium vaginae* 97%	64	33	17	2.5	5	5			0.002			
65	ND010	*Atopobium vaginae* 96.5%	43	27	33	5	5	5						
66	ND484	*Coryne. tuberculostearicum* 99.5%	14	47	67	5	5	3			0.002			
67	ND739	No near neighbour <85%	43	20	0	5	5	5		0.03	0.01			
68	ND298	*Gardnerella vaginalis* 100%	57	20	8	4	5	5			0.007			
69	ND231	*Gardnerella vaginalis* 100%	64	60	50	3	4	4						0.003
70	ND233	*Gardnerella vaginalis* 98%	50	33	25	4.5	5	5				0.008		
71	ND302	*Gardnerella vaginalis* 97.5%	43	47	33	5	5	5					0.02	
72	ND067	*Gardnerella vaginalis* 97.5%	64	20	8	4	5	5	0.02		0.004			
73	ND316	*Gardnerella vaginalis* 97%	93	73	83	3	2	3				0.05	0.03	
74	ND285	*Gardnerella vaginalis* 98%	36	53	42	5	4	5						
75	ND404	*Gardnerella vaginalis* 98%	71	33	67	3	5	3						
76	ND260	*Gardnerella vaginalis* 96.5%	43	53	33	5	3	5					0.01	0.02
77	ND102	No near neighbour (NC070)	29	27	33	5	5	5						
78	ND301	No near neighbour (N72)	43	27	0	5	5	5			0.005			
79	ND304	*Gardnerella vaginalis* 88.9%	50	27	8	4.5	5	5			0.006			
80	ND394	*Gardnerella vaginalis* 88.9%	43	40	0	5	5	5		0.01	0.005			
81	ND252	*Gardnerella vaginalis* 88.9%	93	87	75	2	3	3			0.008		0.02	
82	ND042	*Gardnerella vaginalis* 89.1%	36	20	50	5	5	4.5		0.03				
83	ND061	*Gardnerella vaginalis* 86.5%	29	27	25	5	5	5						
84	ND054	*Gardnerella vaginalis* 90.9%	36	27	42	5	5	5						
85	ND436	*Gardnerella vaginalis* 87.5%	71	27	33	4	5	5				0.001		
86	ND412	*Gardnerella vaginalis* 88.9%	43	27	17	5	5	5						
87	ND328	*Gardnerella vaginalis* 90%	29	47	17	5	5	5						
88	ND263	*Gardnerella vaginalis* 90%	100	93	75	2.5	2	3.5		0.01	0.005		0.03	0.04
89	ND393	*Gardnerella vaginalis* 87%	57	47	50	4	5	4						
90	ND254	*Gardnerella vaginalis* 90.2%	86	40	33	3.5	5	5			0.05			

[1] Row (from left to right) in Fig. 3; [2] OTU: Operational taxonomic unit, ND: Nairobi deep sequencing OTU; [3] Nearest neighbour in cpnDB; [4] Median tier of abundance observed in each group (1=>1%, 2=0.1-0.99%, 3=0.01-0.099%, 4=0.001-0.0099%, 5=undetected); [5] Non-parametric test conducted in R, shaded cells indicate p value <0.05 after Bonferroni correction (p*90 tests), [6] Non-parametric test conducted in Genespring, no p value <0.05 after correction for multiple comparisons, shaded cell indicates significance by both methods.

Table S6: K-means clustering of operational taxonomic units

Cluster	*OTU*[1]	*Nearest neighbour*[2]	*Group with highest abundance*
0	ND042	*Gardnerella vaginalis* 89.1%	BV+
0	ND260	*Gardnerella vaginalis* 96.5%	BV+
0	ND231	*Gardnerella vaginalis* 100%	BV+
0	ND420	*Prevotella buccalis* 88%	BV+
0	ND150	No near neighbour (NC115)	BV+
0	ND426	*Prevotella timonensis* 99.5%	BV+
0	ND103	*Prevotella timonensis* 97%	BV+
0	ND023	*Prevotella amnii* 100%	BV+
0	ND297	*Porphyromonas uenonis* 99%	BV+
0	ND317	No near neighbour (NC053)	BV+
0	ND059	No near neighbour (NC053)	BV+
0	ND175	*Megasphaera sp.* 99.6%	BV+
1	ND394	*Gardnerella vaginalis* 88.9%	BV+
1	ND106	*Atopobium vaginae* 97%	BV+
1	ND182	*Prevotella buccalis* 96%	BV+
1	ND315	*Porphyromonas uenonis* 97.5%	BV+
1	ND288	No near neighbour (NC053)	BV+
1	ND344	No near neighbour (NC053)	BV+
1	ND286	No near neighbour (NC053)	BV+
1	ND318	*Megasphaera sp.* 99.5%	BV+
1	ND237	No near neighbour (NC040)	BV+
1	ND360	*Megasphaera sp.* 99.5%	BV+
1	ND291	No near neighbour (NC039)	BV+
1	ND398	No near neighbour (NC039)	BV+
1	ND089	No near neighbour (NC039)	BV+
1	ND348	*Escherichia coli* 100%	BV+
2	ND254	*Gardnerella vaginalis* 90.2%	mBVI
2	ND393	*Gardnerella vaginalis* 87%	mBVI
2	ND263	*Gardnerella vaginalis* 90%	mBVI
2	ND328	*Gardnerella vaginalis* 90%	mBVI
2	ND054	*Gardnerella vaginalis* 90.9%	mBVI
2	ND304	*Gardnerella vaginalis* 88.9%	mBVI
2	ND285	*Gardnerella vaginalis* 98%	mBVI
2	ND316	*Gardnerella vaginalis* 97%	mBVI
2	ND302	*Gardnerella vaginalis* 97.5%	mBVI
2	ND233	*Gardnerella vaginalis* 98%	mBVI

Table S6: K-means clustering of operational taxonomic units

Cluster	*OTU*[1]	*Nearest neighbour*[2]	*Group with highest abundance*
2	ND406	*Atopobium vaginae* 96.5%	mBVI
2	ND256	*Prevotella buccalis* 96.5%	mBVI
2	ND305	*Prevotella timonensis* 98.5%	mBVI
2	ND313	*Prevotella timonensis* 98%	mBVI
2	ND069	*Prevotella bivia* 100%	mBVI
2	ND257	*Porphyromonas uenonis* 99%	mBVI
2	ND084	*Megasphaera sp.* 99.5%	mBVI
2	ND375	No near neighbour (NC040)	mBVI
2	ND032	No near neighbour (NC039)	mBVI
2	ND765	*Comamonas terrigena* 91%	mBVI
3	ND436	*Gardnerella vaginalis* 87.5%	mBVI
3	ND061	*Gardnerella vaginalis* 86.5%	mBVI
3	ND252	*Gardnerella vaginalis* 88.9%	mBVI
3	ND102	No near neighbour (NC070)	mBVI
3	ND298	*Gardnerella vaginalis* 100%	mBVI
3	ND010	*Atopobium vaginae* 96.5%	mBVI
3	ND429	*Prevotella melaninogenica* 98.6%	mBVI
4	ND301	No near neighbour (N72)	BV–
4	ND404	*Gardnerella vaginalis* 98%	BV–
4	ND484	*Corynebacterium tuberculostearicum* 99.5%	BV–
4	ND435	No near neighbour (NC040)	BV–
4	ND044	No near neighbour (NC040)	BV–
4	ND303	*Lactobacillus sp. N6* 98.9%	BV–
4	ND255	*Lactobacillus iners* 98.9%	BV–
4	ND246	*Lactobacillus iners* 98.5%	BV–
4	ND284	*Lactobacillus iners* 98.2%	BV–
4	ND877	*Lactobacillus iners* 98.9%	BV–
4	ND019	*Lactobacillus iners* 99%	BV–
4	ND295	*Lactobacillus iners* 98.2%	BV–
4	ND294	*Lactobacillus iners* 99%	BV–
4	ND085	*Lactobacillus iners* 99%	BV–
4	ND038	*Sphingobium japonicum* 89.5%	BV–
4	ND857	*Sphingomonas wittichii* 92.5%	BV–
4	ND658	*Sphingomonas wittichii* 89.5%	BV–
4	ND389	*Sphingomonas wittichii* 92.5%	BV–
4	ND614	*Rhodopseudomonas palustris* 92.5%	BV–

Table S6: K-means clustering of operational taxonomic units

Cluster	*OTU*[1]	*Nearest neighbour*[2]	*Group with highest abundance*
4	ND559	*Mesorhizobium opportunistum* 87.5%	BV–
4	ND057	*Methylobacterium radiotolerans* 98.5%	BV–
4	ND308	No near neighbour <85%	BV–
4	ND253	*Variovorax paradoxus* 98%	BV–
4	ND326	*Variovorax paradoxus* 98%	BV–
4	ND670	*Variovorax paradoxus* 98%	BV–
4	ND251	*Bradyrhizobium japonicum* 90%	BV–
4	ND427	*Comamonas terrigena* 88%	BV–
4	ND713	No near neighbour <85%	BV–
5	ND412	*Gardnerella vaginalis* 88.9%	BV–
5	ND067	*Gardnerella vaginalis* 97.5%	BV–
5	ND739	No near neighbour <85%	BV–
5	ND278	No near neighbour (NC039)	BV–
5	ND307	*Lactobacillus sp. N6* 100%	BV–
5	ND320	*Lactobacillus crispatus* 100%	BV–
5	ND296	*Lactobacillus crispatus* 100%	BV–
5	ND439	*Lactobacillus jensenii* 100%	BV–
5	ND401	*Lactobacillus jensenii* 95%	BV–

[1] OTU: Operational taxonomic unit, ND: Nairobi deep sequencing OTU; [2] Nearest neighbour in cpnDB

Table S7: Tagged deep sequencing primers for multiplex GS-FLX

Primer	emPCR primer[1]	Tag	*cpn*60 UT primer
P279gatg:	GCCTCCCTCGCGCCATCAG	AGATGC	GAIIIIGCIGGIGAYGGIACIACIAC
P279gctg:	GCCTCCCTCGCGCCATCAG	AGCTGC	GAIIIIGCIGGIGAYGGIACIACIAC
P279gtga:	GCCTCCCTCGCGCCATCAG	AGTGAC	GAIIIIGCIGGIGAYGGIACIACIAC
P279ctct:	GCCTCCCTCGCGCCATCAG	ACTCTC	GAIIIIGCIGGIGAYGGIACIACIAC
P279ctga:	GCCTCCCTCGCGCCATCAG	ACTGAC	GAIIIIGCIGGIGAYGGIACIACIAC
P279caga:	GCCTCCCTCGCGCCATCAG	ACAGAC	GAIIIIGCIGGIGAYGGIACIACIAC
P279tctg:	GCCTCCCTCGCGCCATCAG	ATCTGC	GAIIIIGCIGGIGAYGGIACIACIAC
P279tgag:	GCCTCCCTCGCGCCATCAG	ATGAGC	GAIIIIGCIGGIGAYGGIACIACIAC
P279tgca:	GCCTCCCTCGCGCCATCAG	ATGCAC	GAIIIIGCIGGIGAYGGIACIACIAC
P279tagt:	GCCTCCCTCGCGCCATCAG	ATAGTC	GAIIIIGCIGGIGAYGGIACIACIAC
P279caca:	GCCTCCCTCGCGCCATCAG	ACACAC	GAIIIIGCIGGIGAYGGIACIACIAC
P279gaga:	GCCTCCCTCGCGCCATCAG	AGAGAC	GAIIIIGCIGGIGAYGGIACIACIAC
P279gtgt:	GCCTCCCTCGCGCCATCAG	AGTGTC	GAIIIIGCIGGIGAYGGIACIACIAC
P279gaca:	GCCTCCCTCGCGCCATCAG	AGACAC	GAIIIIGCIGGIGAYGGIACIACIAC
P279gtct:	GCCTCCCTCGCGCCATCAG	AGTCTC	GAIIIIGCIGGIGAYGGIACIACIAC
P279tcta:	GCCTCCCTCGCGCCATCAG	ATCTAC	GAIIIIGCIGGIGAYGGIACIACIAC
P280:	GCCTTGCCAGCCCGCTCAG	------	YKIYKITCICCRAAICCIGGIGCYTT
P1612gatg:	GCCTCCCTCGCGCCATCAG	AGATGC	GAIIIIGCIGGYGACGGYACSACSAC
P1612gctg:	GCCTCCCTCGCGCCATCAG	AGCTGC	GAIIIIGCIGGYGACGGYACSACSAC
P1612gtga:	GCCTCCCTCGCGCCATCAG	AGTGAC	GAIIIIGCIGGYGACGGYACSACSAC
P1612ctct:	GCCTCCCTCGCGCCATCAG	ACTCTC	GAIIIIGCIGGYGACGGYACSACSAC
P1612ctga:	GCCTCCCTCGCGCCATCAG	ACTGAC	GAIIIIGCIGGYGACGGYACSACSAC
P1612caga:	GCCTCCCTCGCGCCATCAG	ACAGAC	GAIIIIGCIGGYGACGGYACSACSAC
P1612tctg:	GCCTCCCTCGCGCCATCAG	ATCTGC	GAIIIIGCIGGYGACGGYACSACSAC
P1612tgag:	GCCTCCCTCGCGCCATCAG	ATGAGC	GAIIIIGCIGGYGACGGYACSACSAC
P1612tgca:	GCCTCCCTCGCGCCATCAG	ATGCAC	GAIIIIGCIGGYGACGGYACSACSAC
P1612tagt:	GCCTCCCTCGCGCCATCAG	ATAGTC	GAIIIIGCIGGYGACGGYACSACSAC
P1612caca:	GCCTCCCTCGCGCCATCAG	ACACAC	GAIIIIGCIGGYGACGGYACSACSAC
P1612gaga:	GCCTCCCTCGCGCCATCAG	AGAGAC	GAIIIIGCIGGYGACGGYACSACSAC
P1612gtgt:	GCCTCCCTCGCGCCATCAG	AGTGTC	GAIIIIGCIGGYGACGGYACSACSAC
P1612gaca:	GCCTCCCTCGCGCCATCAG	AGACAC	GAIIIIGCIGGYGACGGYACSACSAC
P1612gtct:	GCCTCCCTCGCGCCATCAG	AGTCTC	GAIIIIGCIGGYGACGGYACSACSAC
P1612tcta:	GCCTCCCTCGCGCCATCAG	ATCTAC	GAIIIIGCIGGYGACGGYACSACSAC
P1613:	GCCTTGCCAGCCCGCTCAG	------	CGRCGRTCRCCGAAGCCSGGIGCCTT

---

[1] emPCR primer: Sequence on left is primer for emulsion PCR, sequence in center is tag unique to each sample, and sequence on right is *cpn*60 UT-specific.