### **Supplementary Information to:**

Man-made microbial resistances in built environments

Alexander Mahnert<sup>1\*</sup>, Christine Moissl-Eichinger<sup>2, 3</sup>, Markus Zojer<sup>4</sup>, David Bogumil<sup>5</sup>, Itzhak Mizrahi<sup>5</sup>, Thomas Rattei<sup>4</sup>, José Luis Martinez<sup>6</sup> and Gabriele Berg<sup>1, 3</sup>

<sup>1</sup>Institute of Environmental Biotechnology, Graz University of Technology, Graz, Austria <sup>2</sup>Department of Internal Medicine, Medical University Graz, Austria <sup>3</sup>BioTechMed Graz, Graz, Austria

<sup>4</sup>Division of Computational Systems Biology, Department of Microbiology and Ecosystem Science, University of Vienna, Austria

<sup>5</sup>Department of Life Sciences, Faculty of Natural Sciences, Ben-Gurion University of the Negev, Israel

<sup>6</sup>Centro Nacional de Biotecnologia, CSIC, Madrid, Spain

\*Correspondence: Alexander Mahnert Institute of Environmental Biotechnology Graz University of Technology Graz, Austria alexander.mahnert@gmail.com

#### **Supplementary Methods**

Detailed description of sampled built environments:

Public buildings as well as public and private houses located in the rural area of a wildlife park (in close touch with nature in Eekholt, Germany) represented the model for a totally unrestricted built environment (UB, Supplementary Fig. S19). The intensive care unit (ICU, Department of Internal Medicine, University hospital, Graz, Austria, Supplementary Fig. S20) in full operation and in a suburban area, act as a model for confined built environments (CB) with special attributes of its occupants <sup>1</sup>. Finally, a highly confined indoor environment (CB) was represented by a cleanroom facility<sup>2</sup> with its adjacent gowning area used for spacecraft assembly by the European Space Agency (ESA) in the urban area of Turin, Italy (Supplementary Fig. S21). Samples throughout these different built environments were obtained from floors to allow a high grade of comparability and sufficient biomass for the applied methods. In addition, samples from ICU floor surfaces were compared with samples from other ICU locations as well, using a targeted approach of the 16S rRNA gene. We also tried to reduce potential technical effects (sampling procedures, DNA extractions, sequencing run effects) as well as distortive environmental effects (sampling season and room locations) by a consistent experimental design and numerous controls throughout our study <sup>3–5</sup>. Such precautions minimize technical differences in our study that could overlay actual biological effects <sup>4</sup>.

In particular the following environmental influences should be considered for each sampled category:

The natural unrestricted indoor environment of public and private buildings and houses in a rural area were exposed to the highest degree to the surrounding outdoor environment. A variety of discrete landscapes with respectively associated ecosystems surrounded these structures – including bogs, wetlands, coniferous and deciduous wooded forests with a winding creek in grasslands and pasture including > 700 animals in ~100 different species of insects, fish, reptiles, birds and mammals (www.wildpark-eekholt.de).

The public buildings were located in all these different ecosystems. During the business hours of the wildlife park (from 9 am to 6 pm) these buildings were wide apart for a microbial input via the outdoor air (pollen, seeds, water droplets and other biotic and abiotic particles and compounds)

and various animals (insects, birds, mammals, humans and their pets) with brought along particles associated to soil, foods (snacks), and personal items (e.g. buggies etc.). The rooms of these structures were window ventilated or simply wide apart. Floor surface materials comprised concrete, tiles, polymers and wood. In contrast to the other sampled indoor environments the floors were cleaned only mechanically with a broom and people interacted only marginally with the floor environment, since they primarily observed wildlife, studied boards about the presented ecosystem or simply seek for shelter due to unfavorable weather conditions.

The public houses were so-called 'A-frame houses' surrounded by lawns, conifers and deciduous trees adjoining to broad fields of farmland. The houses comprised a recreation room (wooden floor), a kitchen and restrooms (tiled floors). Especially the floor surfaces of the kitchen and the restrooms were treated regularly with natural soaps and were also affected (food preparation, conducting daily hygiene) in a different way compared to the previous category of public buildings. The public houses were used by school classes, who occupied the rooms for a week to receive environmental education. Houses were closed during absence of occupants and window ventilated in the presence of pupils or the cleaning staff.

While a public house is frequented by numerous temporary residents, the private house in contrast is usually inhabited by a family or a reduced number of residents. Likewise, daily behavior and interaction with surfaces is highly personalized. Hence, compared to public houses, this personalized maintenance represents a first step of confinement of an indoor environment. The sampled private house was an old farmyard encircled by a garden with fruit trees and conifers of the adjacent forest. Sampled rooms comprised a kitchen (floor with polymer tiles), a barn mainly used for dining with guests (stone floor) and a conference and locker room (tiled floor) in an adjacent building used as a bird foster station. The house was window ventilated and occupied by the co-founder of the wildlife park - an elderly women and her dog. The resident was supported by a household help and regularly visited by employees of the wildlife park. Floors were cleaned with all-purpose cleaners, the dog received its food in a bowl in the kitchen and the barn was inhabited by swallows during the summertime. Due to the high frequentation of occupants it was uncommon to change shoes before entering the private house (in contrast to many other private households). The same behavior was true for the public built environments described above.

On the contrary to these UB environments, the investigated ICU was located in a suburban area. Compared to public buildings, public and private houses, indoor environments of hospitals are severely different in many aspects: First of all beside staff the majority of people are suffering from severe infections and other health problems. Therefore these patients might deliver specific microbes like opportunistic pathogens to their surroundings or acquire problematic germs from it. In this ICU critically affected patients suffering from all subspecialties of internal medicine and neurologic defects were treated in 15 beds including an isolation unit for severe immunocompromised patients. All rooms of the ICU were mechanical ventilated by an air conditioning system. Access to the ICU was restricted and controlled by the medical staff. Special garment regulations were applicable in dependence on the grade of interaction with patients or medical devices. A comprehensive maintenance, cleaning and disinfection plan especially characterizes this indoor environment: Waste was removed three times a day and waste containers were disinfected with a surface disinfection cleaner. In the same manner doorknobs and tiles in working height were cleaned twice per day. All-purpose cleaners were used to treat free surfaces and furniture (shelfs, chairs, tables, conduits, window boards, radiators, elevators), a surface disinfectant was used for dirty laundry buckets, handrails, cleaning cars, surface cleaning devices and sanitary glass cleaners and surface disinfectants were used at the bath place for showers and sinks once a day. All other furniture, objects or room installations were cleaned weekly, monthly, per quarterly period, semiannual or annually. Beside restrooms and shower heads all floors were cleaned twice a day or even more often if visible contaminants (blood, feces or other secreted liquids) were observed. If known hospital germs (e.g. Clostridium difficile) were detected additional hygiene protocols became effective up to individual isolation of patients. A detailed list of all cleaning products used for disinfecting hands, surfaces, instruments or antisepsis of skin and mucosa are listed in Supplementary Table 4 and Supplementary Table 5 beside other agents to treat known resistant hospital germs and viruses.

An additional level of microbial control is realized in cleanrooms during spacecraft assembly, especially if a celestial body with relevance to extraterrestrial life detection is targeted. According to planetary protection regulations <sup>6</sup> these spacecraft are not only assembled in cleanrooms, but are also subject to regular microbial monitoring, cleaning and removal of surface associated microbes (bioburden control and countermeasures) <sup>7</sup>. Most microbes are intensively reduced not only by the low water and nutrient availability, but especially through intensive HEPA air filtration and the reduction of air-borne particles through air recirculation. Hence, cleanrooms are categorized into different cleanroom classes by the amount and size of particles. Moreover, type and quality of gaseous substances, temperature, humidity, electromagnetics, electrostatics etc. can be controlled. Low emitting surfaces and materials (e.g. floors, doors, door handles, tables, shelves, trolleys) are

mopped with sterile cloths to remove dust, before surfaces are cleaned several times a day with alcohols (e.g. 70% isopropanol), detergents based on hydrogen peroxide (e.g. Klercide-CR), high alkaline cleaning reagents (e.g. Jaminal Plus or Kleenol 30) and sometimes even with vapor-phase hydrogen peroxide or cold plasmas to reduce the bioburden of spacecraft. However, due to the complexity of working processes during spacecraft assembly, engineers still have to interact with spacecraft components and are therefore a main source of microbial dispersal. As a counter measurement strict gowning protocols with special cleanroom garment have to be followed in repetitive steps and sluices and air showers define transitions through different cleanroom classes and rooms. Finally strict protocols (e.g. slow body movements) regulate the interaction of engineers and staff with the cleanroom environment, spacecraft components and the assembled spacecraft itself to minimize the dispersal of human associated microbes into the cleanroom.

All sampled built environments could be characterized by distinct environmental features such as the geographic location, microclimate, architecture, room maintenance and usage (Table 1). Whereas samples from unrestricted environments were located at low elevations in northern Germany, samples from confined built environments originated from higher altitudes in south-east Austria and north-west Italy. Concerning respective microclimates, unrestricted environments of public buildings and public as well as private houses were specified by higher relative air humidity and lower temperatures, compared to controlled environments of the ICU, and the cleanroom facility with its gowning areas. The ICU highlighted most opposed climatic conditions: with the highest room temperature (24°C) and the lowest relative air humidity (32%). Further discriminating attributes were represented by architectural features of sampled built environments. Here, cleanrooms represented very large indoor room surfaces compared to small spaced infrastructure in the case of gowning areas, rooms in the ICU, public buildings and public and private houses. The smallest room volumes were represented in private houses. Likewise, more controlled built environments harbored only a few defined synthetic materials on floors, compared to a higher variety of floor materials including materials of natural sources like wood and stone in unrestricted built environments. Moreover, occupancy and intensity of cleaning were obviously different between public buildings, public and private houses, the ICU and the cleanroom facility with its gowning areas. Hence, a higher grade of environmental confinement resulted in more elaborative maintenance and cleaning procedures, but a lower grade of occupancy.

#### Sample processing:

Alpha Wipes® (TX1009, VWR International GmbH, Vienna, Austria) were extracted in 40 mL 0.9% sterile DNA-free sodium chloride solution by vortexing for 10 seconds and sonication at 40 kHz for 2 min in an ultrasonic bath (Transonic Digitals, Elma, United States). Extracted sampling liquids were concentrated by repeated centrifugation cycles at 3220 g and 4°C for 5 min using UV sterilized Amicon filter tubes (Amicon Ultra-15 Centrifugal Filter Units, Merck, Germany) to a final volume of 500 μl.

#### PMA treatment of samples:

An aliquot of selected biological samples and spot tests of used reagents were treated with propidium monoazide (PMA) according to manufacturer instructions (GenIUL, S.L., Terassa, Spain). Samples were treated with a final concentration of 50  $\mu$ M PMA for 30 min. in the dark. Afterwards treated and non-treated samples were exposed in parallel to the PhAST blue-Photo activation system for tubes (GenIUL, S.L., Terassa, Spain) for 15 min.

#### DNA extraction:

The DNA extraction procedure included mechanical, thermal and chemical lysis of the cells. All cell suspensions were homogenized with a FastPrep-24 Classic Instrument (MP Biomedicals, United States) in two cycles for 30 seconds at 6.5 m/s. Homogenized samples were then added to an equal volume of freshly prepared XS buffer (2x). A  $\sim$ 20 ml stock solution of the XS buffer (2x) contained: 4 ml of 1M Tris/HCl (pH 7.4); 4.56 ml of 7M ammonium acetate; 3.2 ml of 250 mM ethylene diamine tetraacetic acid; 4 ml of 10% (w/v) sodium dodecyl sulfate; 0.4 g of potassium ethyl xanthogenate; and 4.99 ml of PCR-grade water. The XS buffer was incubated at 65°C for 15 min. to dissolve the xanthogenate completely. The homogenized cell suspension in the XS buffer was incubated at 65°C for 2 hours and gently mixed by inverting the tube every 30 min. After the incubation period, the tubes were vortexed for 10 seconds and placed on ice for 10 min, followed by centrifugation at 100 g and 4°C, for 5 min. The supernatant was transferred into a PhaseLock Gel tube (Eppendorf, Hamburg, Germany), and an equal volume of phenol: chloroform:isoamyl alcohol (25:24:1) was added. Sample suspensions were mixed and centrifuged at 2000 g and 15°C for 5 min. The upper aqueous layer was transferred into a new tube. For DNA precipitation, the same volume of cold 100% isopropanol and 1/10 volume of 4M ammonium acetate was added. Suspensions were gently mixed and incubated at -20°C overnight. On the next day the precipitated DNA was centrifuged at 13600 g and 4°C for 30 min. Invisible DNA pellets were washed with 1 ml of ice-cold 70% ethanol and centrifuged again at 13600 g and 4°C for 30 min. DNA pellets were dried completely in a clean bench and finally dissolved in 50 µl of PCR-grade water.

### qPCR:

Quantitative molecular measures were based on the primer pair 515F - 927R (10µM each, Supplementary Table 6). For the qPCR run DNA templates were amplified in 40 cycles with denaturation at 95°C for 20 sec., annealing at 54°C for 15 sec. and elongation at 72°C for 30 sec. on a Rotor-Gene<sup>TM</sup> 6000 real – time rotary analyzer (Corbett Research, Sydney, Australia). A melt curve from 72°C to 95°C served as a quality control for the amplified products. One qPCR reaction mix was constituted as follows: 1.06 µl PCR grade water, 3.5 µl KAPA Plant PCR buffer (KAPA3G Plant PCR Kit, Peqlab, VWR International GmbH, Erlangen, Germany), 0.42 µl forward and reverse primers, 0.056 µl of KAPA3G Plant DNA-polymerase (2.5 u per µl), 0.78 µl of SYBR<sup>®</sup> Green (4x concentrate, Invitrogen<sup>TM</sup>, Eugene, OR, USA), and 0.8 µl of the extracted DNA template. qPCR runs with a mean reaction efficiency of 0.84 and mean standard curve R<sup>2</sup> values of 0.99 (16S rRNA gene product of *Bacillus subtilus* B2G) were evaluated in triplicate and counts in negative controls were subtracted from all other samples in their respective qPCR runs.

#### Plasmidome assembly methods:

The plasmidome of each sample was assembled using the recycler pipeline <sup>8</sup>. Paired end reads were assembled de novo into contigs and the assembly graph, connecting these contigs via shared kmers, was obtained using SPAdes <sup>9</sup>. In order to calculate the read coverage for each contig from the assembly, paired end reads were aligned to the contigs using the Burrows-Wheeler aligner (BWA) <sup>10</sup>. Contigs were subsequently assembled into cycles defined by connectivity in the assembly graph form the de novo assembly and uniform coverage distribution around the circular combination of contigs, using the recycler Software <sup>8</sup>. Plasmid encoded open reading frames (ORFs) were then predicted upon these cycles using Metagenemark <sup>11</sup>. Annaotation of ORFs was performed by blast+ <sup>12</sup> searches against the KEGG FTP release 2017-03-27 database <sup>13</sup>, the Uniref90 release 2016\_10 database <sup>14</sup> and the CARD 1.1 database <sup>15</sup>.

Synteny analysis:

The genomic context of resistance genes (20 kb, 10 kb up -and downstream), pan-genomes, and virulence was analyzed in MaGe <sup>16</sup>. Annotation of virulence factors was based on VFDB <sup>17</sup> and VirulenceFinder <sup>18</sup>. Pan-genomes were calculated with the SiLiX software <sup>19</sup>. IntegronFinder <sup>20</sup> was applied to detect integron clusters. Regions of genome plasticity (RGP) were based on PkGDB organisms and NCBI RefSeq organisms showing highest similarity with the query genome by using RGP Finder and the tools AlienHunter <sup>21</sup> and SIGI-HMM <sup>22</sup>.

#### 16S rRNA gene amplicons:

Four individual PCR reactions à 50  $\mu$ l (17.6  $\mu$ l PCR grade water, 25  $\mu$ l KAPA Plant PCR buffer, 0.4  $\mu$ l KAPA3G Plant DNA-polymerase (2.5 u per  $\mu$ l), 3  $\mu$ l forward and reverse barcoded primers 515f and 806r (Supplementary Table 6) and 1  $\mu$ l extracted DNA template) were pooled after successful amplification (40 cycles of 95°C for 30 sec. denaturation, 60°C for 15 sec. annealing and elongation at 72°C for 12 sec.) on a TECHNE TC-PLUS gradient thermocycler (Bibby Scientific Ltd, Stone, UK) and validated by gel electrophoresis. Pooled PCR products were cleaned with the Wizard SV Gel and PCR Clean-Up System kit (Promega, Madison, WI, USA) and measured on the NanoDrop instrument (Thermo Scientific, Wilmington, DE, USA). Equimolar concentrations of PCR amplicons were pooled and sent for sequencing at Eurofins Genomics GmbH, Ebersberg, Germany. The pool was additionally purified by gel extraction and quantified before the sequencing library was prepared by adaptor ligation, PCR amplification according to PCR product insert size and a final library purification and quality control. Sequencing of amplicon samples was performed on an Illumina MiSeq instrument with v3 chemistry and the 2 x 300 bp paired-end read module. Amplicon sequences of the ICU were generated and sequenced as described in <sup>1</sup>.

16S rRNA gene amplicon data analysis:

Sequences of 16S rRNA gene amplicons of all indoor environments were analyzed in QIIME 1.9. and QIIME 2 (versions 2017.10 to 2018.11)<sup>23,24</sup>, according to <sup>25,26</sup> and developer provided tutorials. Forward and reverse reads were joined with a minimum overlap of 100 bp and a maximum allowed difference of 3%. Barcodes were extracted for demultiplexing and quality filtering of reads with default parameters. Reads in controls were removed by blast (100% alignment cutoff). 454 reads

of ICU samples were denoised with mothur <sup>27</sup>, demultiplexed and quality filtered before they were concatenated with the filtered Illumina reads. All sequences were then trimmed to the same 16S rRNA gene regions and lengths (FASTX-toolkit by Assaf Gordon, accessed on Galaxy, galaxyproject.org). Reads were additional quality filtered to remove primer sequences before chimeric sequences were removed with USEARCH<sup>28</sup> giving both Silva 119 and the Greengenes 13\_8 release as a reference. OTUs were picked with USEARCH to the same reference and every sequence not present was clustered denovo at 97% similarity level. For phylogenetic based metrics and measures a phylogenetic tree was calculated. The OTU table was filtered for singletons, doubletons, assigned reads to chloroplasts or mitochondria and sorted for following read normalizations in alpha and beta diversity analysis and statistics. Since jackknife supported bootstrapped trees showed higher confidence for weighted than unweighted unifrac measures, weighted metrics were preferred throughout the analysis as also recommended by <sup>3</sup>. OTUs present in the reference were examined for their functional potential with PICRUSt <sup>29</sup>. Potential microbial phenotypes were predicted with BugBase<sup>30</sup>. OTU networks were based on the assigned taxonomy, calculated in QIIME and visualized in Cytoscape <sup>31</sup> as described earlier <sup>32</sup>. Sample metadata predictions were based on random Forest classification and regression <sup>33</sup>. Associations between microbial composition and environmental variables were assessed by bioenv analysis <sup>34</sup> and verified by multivariate linear models using MaAsLin<sup>35</sup>, balances in gneiss<sup>36</sup> and linear mixed effects <sup>37</sup>.

#### Statistics on sequencing reads:

3.0 - 6.7 x 10<sup>7</sup> sequences per sample could be obtained via shotgun Illumina HiSeq sequencing. After filtering, a range of 7 x 10<sup>6</sup> to 2.5 x 10<sup>7</sup> quality sequences (phred score  $\geq$  q36) could be retained for following assemblies and binning attempts (Supplementary Table 7). Assemblies on Ray Meta resulted in a satisfying amount of small ( $\geq$  100 nt) and large ( $\geq$  500 nt) contigs and scaffolds (range of N50 values: 142 to 2510, Supplementary Table 8). From these contigs and scaffolds 125 bins (8 to 20 bins per sample) could be generated. Most bins were obtained from the private house, while only a few genomes could be binned from the ICU dataset (Supplementary Table 2). In general, most markers (1389) and marker sets (369) could be binned into 39 genomes assigned to the genus *Pseudomonas* with a completeness of 93.47%, contamination of 2.51%, and heterogeneity of strains of 26.67%. Supportive 16S rRNA gene amplicon analysis for higher resolutions of respective built environment locations resulted in 225 to 37,831 sequences per sample from a total of 837,216 quality sequences and 10,814 assigned OTUs (Supplementary Table 10, Supplementary Table 11 and Supplementary Table 12).

#### Verification of bioenv results:

Associations of the microbiome with microclimate or location specific variables could not be further distinguished. MaAsLin was able to define specific taxa (distinct sets, only 6 of 82 were overlapping) for microclimate and location specific variables (e.g. microclimate: *Bauldia, Gaiella* and *Intrasporangium*; location: *Commensalibacter, Chlorocromatium*; both: *Iamia, Rubrobacter*). However, regression models using balances in gneiss showed that microclimate and location dependent variables contributed to similar proportions (~2%) to the total explained community variation (~70%). Moreover over-fitting of the model could not be ruled out (in 4 out of 6 cross-validations the prediction accuracy was higher than the within model error). Finally, linear mixed effect models were used to test if microbial composition changed over microclimate or location in response to confinement and architecture (room size). This analysis showed that microbial composition was not significantly impacted by these selected variables. Hence, we concluded that environmental variables of the microclimate and the location were confounded in our sample design and were not appropriate to tell if the microclimate or the location has a bigger impact on the microbial composition.

## **Supplementary Figures**



## Supplementary Figure 1: Diversity estimates

Diversity estimates of confined and unrestricted built environments based on 16S rRNA gene amplicon analysis.



Supplementary Figure 2: Distance estimates

Bray-Curtis distance estimates of confined and unrestricted built environments based on 16S rRNA gene amplicon analysis.



Supplementary Figure 3: Domain profile

Single reads BLASTx (rapsearch and diamond) vs. NCBI nr. superkingdom level (derived from MEGAN, excluding unassigned reads, normalized data set).





### Supplementary Figure 4: Phyla profile

Single reads BLASTx (rapsearch and diamond) vs. NCBI nr. phylum level (derived from MEGAN,

excluding unassigned reads, normalized data set).



## Supplementary Figure 5: Species profile

Space filling radial chart of taxa (species level, excluding unassigned reads, normalized, percentage) assigned (BLASTx NCBInr, diamond and rapsearch) to different built environments (MEGAN).



#### Supplementary Figure 6: Distinctive taxa of controlled built environments (CB)

LEfSe analysis (LDA effect size) on taxa (according to NCBInr database) of single reads from metagenomes of CB (ICU, gowning area, cleanroom) and UB (public buildings, public and private houses) built environments with the following parameters: per-sample normalization to 1M, factorial Kruskal-Wallis test among classes (alpha = 0.01), pairwise Wilcoxon test between subclasses (alpha = 0.01), threshold for the LDA score (1.0), strategy for multi-class analysis (all-against-all, more strict).



### Supplementary Figure 7: Core microbiome

Core OTU network based on G-tests for independence of 16S rRNA gene amplicons resolved to genus level. Edge-weighted spring embedded algorithms implemented in Cytoscape were used for visualizations. OTU abundance is reflected by node size. Edge weights by line widths and opacities. Colors refer to different sampled built environments: Cleanroom facility (blue), intensive care unit (red), public buildings, public houses, private houses (all in green).



Supplementary Figure 8: Gram positive bacteria

Phenotype prediction of Gram positive bacteria based on 16S rRNA gene amplicon analysis.



Supplementary Figure 9: Gram negative bacteria

Phenotype prediction of Gram negative bacteria based on 16S rRNA gene amplicon analysis.





Phenotype prediction of potential pathogens based on 16S rRNA gene amplicon analysis.



Supplementary Figure 11: Potential stress tolerance

Phenotype prediction of potential stress tolerant bacteria based on 16S rRNA gene amplicon analysis.



Supplementary Figure 12: Distinctive taxa based on 16S rRNA gene amplicons

LEfSe analysis (LDA effect size) on 16S rRNA gene amplicons of controlled (gowning area, cleanroom), moderate controlled (ICU) and uncontrolled (public buildings, public and private houses) built environments with the following parameters: per-sample normalization to 1M, factorial Kruskal-Wallis test among classes (alpha = 0.05), pairwise Wilcoxon test between subclasses (alpha = 0.05), threshold for the LDA score (2.0), strategy for multi-class analysis (all-against-all, more strict).



## Supplementary Figure 13: Functional profile (barchart)

Single reads BLASTx (rapsearch and diamond) vs. NCBI nr. SEED level 1 (derived from MEGAN, excluding unassigned reads, normalized data set).



### Supplementary Figure 14: Functional profile (radial chart)

Space filling radial chart of SEED annotations on level 1 (species level, excluding unassigned reads, normalized, percentage) assigned (BLASTx NCBInr, diamond and rapsearch) to different built environments (MEGAN).



#### Supplementary Figure 15: Distinctive functions

LEfSe analysis (LDA effect size) on functions (according to SEED database) of single reads from metagenomes of CB (ICU, gowning area, cleanroom) and UB (public buildings, public and private houses) built environments with the following parameters: per-sample normalization to 1M, factorial Kruskal-Wallis test among classes (alpha = 0.05), pairwise Wilcoxon test between subclasses (alpha = 0.05), threshold for the LDA score (3.0), strategy for multi-class analysis (all-against-all, more strict).



Supplementary Figure 16: Meta-analysis of organisms and functions

Comparative analysis of metagenome samples from CB and UB environments with publically available metagenome samples from plants, urban indoor air and the human microbiome project on organism and functional abundance levels visualized through MG-RAST.

top SEED function	public buildings	public houses	private houses	ICU	gowning areas	cleanrooms
Amino Acids and Derivatives						
Carbohydrates						
Cofactors, Vitamins, Prosthetic Groups, Pigments						
Membrane Transport						
Phages, Prophages, Transposable elements, Plasmids						
Protein Metabolism						
Fatty Acids, Lipids, and Isoprenoids						
Virulence, Disease and Defense						
DNA Metabolism						
Respiration						
Photosynthesis						
Stress response						
	0%		509	6		100%

## Supplementary Figure 17: Main genome functions

Relative proportions of annotated SEED functions with RAST for high quality bins from the metagenomics dataset of all sampled built environments.



Supplementary Figure 18: Resistome of genomes and plasmids

Distances of binned genomes and plasmids according to detected resistance genes (CARD database).



Supplementary Figure 19: Unrestricted buildings (UB)

Sampling map of public buildings (L) and public (P) and private houses (F) in a wildlife park in Grossenaspe, Germany (Figure was adapted from <u>https://www.wildpark-eekholt.de/besucherinformationen\_lageplan.htm</u>).



Supplementary Figure 20: Controlled built environment (CB) - Intensive Care Unit Sampling map of the intensive care unit (ICU) at the state hospital in Graz, Austria (Figure was adapted from <sup>1</sup>).



Supplementary Figure 21: **Controlled built environment (CB) - Cleanroom facility** Sampling map of the Thales Alenia space cleanroom facility in Turin, Italy (Figure was adapted from <sup>2</sup>).

## **Supplementary Tables**

Supplementary Table 1: Alpha diversity estimates from single shotgun reads of the metagenomics dataset against the NCBI nr database using the blastX algorithm.

	Тах	a	Taxa (withou	t <i>Hominoidae</i> )	KEG	G	SEED	
samples	Shannon	Simpson	Shannon	Simpson	Shannon	Simpson	Shannon	Simpson
public_buildings	8.44	68.69	8.44	68.67	11.41	1652.82	10.92	1279.98
public_buildings_PMA	8.42	61.75	8.42	61.73	11.38	1616.20	11.01	1276.63
public_houses	8.97	99.16	8.97	99.08	11.58	1814.65	11.00	1255.54
public_houses_PMA	9.00	101.20	9.00	101.12	11.58	1811.69	11.01	1279.79
private_houses	4.98	6.30	4.98	6.30	11.35	1675.57	10.81	1191.86
private_houses_PMA	5.10	6.78	5.10	6.78	11.27	1591.29	10.97	1223.26
gowning_area	7.76	35.81	8.06	42.62	11.86	1947.20	11.02	884.36
ICU	6.27	12.32	7.28	28.89	12.09	1017.25	11.12	1404.91
cleanroom	7.54	21.66	8.42	56.57	11.90	941.92	11.12	1114.67

Supplementary T	Sable 2: Summary	on binned g	genomes from	the shotgun	metagenomics	data set.
-----------------	------------------	-------------	--------------	-------------	--------------	-----------

Sample	bin No.	Marker lineage CheckM	# genomes	s # marke	ers # marker sets	0	1 2	34	S+ Completeness	Contamination	Strain heterogeneity	binID	Count_Function	Fraction_Function	KAST Function	Count_OTU	Fraction_OTU	KASIUIU	KASI subsystem information
	10	k_Bacteria (UID203)	5449	102	57	94 8	0	0 0	0 7.5	0	0	CR8_bin10	185	0.069917	Mobile element protein	160	0.171858	Pseudomonas	virulence, disease, and defense
	13	kBacteria (UID203)	5449	104	58	72 31	1	0 0	0 33.32	1.72	100	CR8_bin13	11	0.006034	Mobile element protein	184	0.516854	Acinetobacter sp. ATCC	Amino Acids and Derivatives
	16	k_Bacteria (UID1453)	901	171	117	147 21	2	1 0	0 4.17	0.31	20	CR8_bin16	5	0.049505	LSU ribosomal protein L2p (L8e),	1	0.0625	Cyanothece sp PCC 8801	photosynthesis
	21	t_Moraxellaceae (UID4680)	86	689	365	30 50	95	4 0	0 95.05	15.27	94.39	CR8_bin21	141	0.031105	Mobile element protein	43	0.078611	Acinetobacter sp. ATCC	Amino Acids and Derivatives
	2/	K_Bacteria (UID203) Sobiogramonadaler (UID2210)	36	103	57	276 17	0 15	0 0	0 0.16	0 97	40	CR8_bin2/	39 57	0.021261	Cysteine desulturase (EC 2.8.1.7), ISCS	8	0.020305	Sobiogobium co. AR49	visulance directo and deferre
ISO 8 cleanroom (CR8)	31	k Bacteria (UID203)	5449	103	57	91 12	0	0 0	0 12 2	0	0	CR8 bin31	18	0.00681	Transcriptional regulator TysB family	370	0.506849	Pseudomonas	Amino Acids and Derivatives
	33	f Xanthomonadaceae (UID4214)	55	659	290	298 27	3 76	6 1	0 56.9	16.61	45	CR8 bin33	38	0.005798	Transcriptional regulator, LysR family	1651	0.58298	Stenotrophomonas	Amino Acids and Derivatives
	36	k_Bacteria (UID203)	5449	104	58	102 2	0	0 0	0 0.31	0	0	CR8_bin36	2	0.026667	RND multidrug efflux transporter;	6	0.285714	Micavibrio	Stress response
	39	kBacteria (UID203)	5449	103	58	101 2	0	0 0	0 0.34	0	0	CR8_bin39	6	0.034483	Phage tail fiber protein	4	0.070175	Escherichia coli K12	Phage, RNA metabolism
	6	k_Bacteria (UID203)	5449	104	58	94 10	0	0 0	0 12.36	0	0	CR8_bin6	147	0.071118	Mobile element protein	35	0.095368	Acinetobacter	virulence, disease, and defense
	8	o_Actinomycetales (UID1530)	622	259	152	20 21	5 23	0 0	0 91.67	7.89	100	CR8_bin8	46	0.013154	conserved hypothetical protein	1277	0.740719	Propionibacterium	Carbonhydrates
	10	K_Bacteria (UID203) f_Enterobacteriaceae (UID5054)	322	975	305	212 51	0	0 0	0 2.59	5 27	012	F_DIN1	15	0.002705	Mobile element protein	28	0.269231	Rocuria mizophila Rantona ananatir IMG	Amino Acids and Darivatives (140)
	13	<ul> <li>Lactobarillales (LID544)</li> </ul>	293	475	267	268 18	17	1 0	0 46 32	4.01	45	F bin13	18	0.011159	Mobile element protein	867	0.81015	Enterococcus	Carbohydrates (120)
	15	k Bacteria (UID203)	5449	104	58	15 24	46	15 3	1 97.65	114.5	21.85	F bin15	18	0.003797	Oligopeptide ABC transporter.	1639	0.650139	Leuconostoc	Carbohydrates (442)
	16	f Micrococcaceae (UID1623)	39	457	218	288 14	5 23	1 0	0 37.4	5.51	15.38	F bin16	29	0.013507	Mobile element protein	149	0.20781	Micrococcus luteus NCTO	Amino Acids and Derivatives (108)
	17	k_Bacteria (UID203)	5449	104	58	56 30	14	4 0	0 25.31	7.84	61.54	F_bin17	7	0.015184	Mobile element protein	81	0.558621	Enhydrobacter	Respiration (24)
	2	k_Bacteria (UID203)	5449	104	58	43 30	20	74	0 84.28	68.1	56.92	F_bin2	89	0.012583	Mobile element protein	139	0.016931		Amino Acids and Derivatives (590)
	20	kBacteria (UID203)	5449	104	58	85 19	0	0 0	0 24.64	0	0	F_bin20	8	0.007561	Predicted transcriptional regulators	556	0.827381	Lactococcus garvieae	Carbohydrates (46)
	21	o_Lactobacillales (UID544)	293	475	267	125 22	1 116	12 1	0 71.87	30.56	13.92	F_bin21	13	0.005066	Lipid A export ATP-binding/permease	818	0.541005	Enterococcus faecalis	Carbohydrates (196)
private houses (F)	23	k_Bacteria (UID203)	5449	104	58	13 24	~ ~	1/ 15	13 94.98	229.5	30.54	F_DINZ3	28	0.002/48	memorane protein, putative	1180	0.341335	Macrococcus	Carbonydrates (1283)
	20	k_Bacteria (UID203)	5449	104	58	9/ /	0	0 0	0 8.62	0	0	F_DINZS	2	0.006667	Proposed peptidogiycan lipid li	118	0.819444	Ennydrobacter	Amino Acids and Derivatives (12)
	31	c Bacilli (UID259)	750	273	152	65 71	82	35 19	1 84.18	93.53	84.89	F bin31	47	0.005469	aretyltransferase GNAT family	4665	0.847411	Exiguobacterium	Amino Arids and Derivatives (631)
	33	k Bacteria (UID203)	5449	104	58	44 40	17	3 0	0 49.06	14.13	38.46	F bin33	12	0.007417	Oligopeptide ABC transporter.	91	0.232737	Weissella confusa LBAE	Carbohydrates (104)
	34	c_Bacilli (UID285)	586	325	181	42 22	5 57	1 0	0 93.38	20.3	90	F_bin34	6	0.00274	Lipid A export ATP-binding/permease	97	0.211329	Macrococcus	Carbohydrates (176)
	35	k_Bacteria (UID203)	5449	104	58	0 5	76	13 10	0 100	130.27	32	F_bin35	87	0.014897	Mobile element protein	1242	0.513223	Lactococcus garvieae	Carbohydrates (753)
	36	k_Bacteria (UID203)	5449	103	57	69 12	5	10 7	0 11.93	7.72	24.68	F_bin36	4	0.04	LSU ribosomal protein L18p (L5e)	2	0.222222	Bacillus coahuilensis m4	
	4	k_Bacteria (UID203)	5449	104	58	94 8	2	0 0	0 7.52	1.72	0	F_bin4	13	0.02305	Mobile element protein	17	0.084577	Leuconostoc	Phages, Prophages, Transposable
	7	o_Actinomycetales (UID1530)	622	257	151	155 99	3	0 0	0 36.89	1.99	100	F_bin7	5	0.004386	Acetyl-coenzyme A synthetase (EC	104	0.314199	Rothia mucilaginosa	Carbohydrates (47)
	8	k_Bacteria (UID203)	5449	104	58	24 33	23	14 8	2 83.54	86.21	30.08	F_bin8	139	0.016931	Mobile element protein	1853	0.538663	Staphylococcus aureus	Carbohydrates (580)
	10	<ul> <li>Lactobacillalar (UID205)</li> </ul>	471	2/10	199	5/ 0	: 65	2 0	0 4.7	22.4	79.29	F_plus_bin0	22	0.013008	Mobile element protein	1540	0.810545	Laucoportor	Amino Acids and Dariustives (359)
	12	k Bacteria (UID203)	5449	104	58	62 9	11	13 1	8 15.44	21.65	14.76	F plus bin12	8	0.024169	SSU ribosomal protein S17p (S11e)	6	0.1111111	Lactococcus garvieae	Protein Metabolism (7)
	13	k Bacteria (UID203)	5449	103	57	4 46	44	7 2	0 95.45	44.58	57.14	F plus bin13	19	0.005522	6-phospho-beta-glucosidase (EC	449	0.487514	Weissella confusa LBAE	Carbohydrates (297)
	14	o Lactobacillales (UID543)	294	472	265	34 34	81	64	0 98.24	26.26	91.06	F plus bin14	28	0.010068	Mobile element protein	493	0.644444	Streptococcus iniae SF1	Carbohydrates (414)
	16	f_Enterobacteriaceae (UID5054)	223	875	305	341 49	40	0 0	0 54.71	3.42	92.5	F_plus_bin16	14	0.004848	diguanylate	248	0.438163	Pantoea ananatis LMG	Amino Acids and Derivatives (111)
	18	k_Bacteria (UID203)	5449	103	57	55 21	22	50	0 71.93	49.12	94.59	F_plus_bin18	39	0.005431	acetyltransferase, GNAT family	4405	0.843385	Exiguobacterium	Amino Acids and Derivatives (606)
	19	kBacteria (UID203)	5449	104	58	65 34	5	0 0	0 51.1	8.62	100	F_plus_bin19	8	0.004499	Cell division protein FtsK	816	0.828426	Macrococcus	Carbohydrates (68)
	20	cGammaproteobacteria (UID4201)	1164	275	174	203 70	2	0 0	0 23.53	1.15	100	F_plus_bin20	11	0.016667	Mobile element protein	196	0.710145	Enhydrobacter	Virulence, Disease and Defense (30)
private nouses PMA (F_prus)	21	k_Bacteria (UID203)	5449	104	58	61 29	11	30	0 47.79	18.57	0	F_plus_bin21	40	0.013717	Mobile element protein	516	0.321696	Lactococcus garvieae	Carbohydrates (217)
	20	<ul> <li>Gammanroteobacteria (UID4201)</li> </ul>	1164	325	174	AD 02	05	20.7	2 91.05	14.33	67.6 55.11	F_plus_bin23	92	0.0122868	Mobile element protein	77	0.192489	Acinetobacter rp. ATCC	Amino Acids and Desivatives (440)
	32	k Bacteria (UID203)	5449	104	58	69 22	12	1 0	0 37.46	15.83	13 33	F plus bin32	17	0.009346	Mobile element protein	126	0 208264	Kocuria nalustris PFI	Carbohydrates (61)
	35	k Bacteria (UID203)	5449	103	57	98 5	0	0 0	0 0.99	0	0	F plus bin35	3	0.044776	Mobile element protein	6	0.206897	Exiguobacterium	Phages, Prophages, Transposable
	37	o_Lactobacillales (UID543)	294	469	262	69 37	29	0 0	0 86.45	6.54	96.55	F_plus_bin37	17	0.009403	Mobile element protein	963	0.833766	Lactococcus garvieae	Carbohydrates (156)
	38	k_Bacteria (UID203)	5449	104	58	83 21	0	0 0	0 19.51	0	0	F_plus_bin38	5	0.007657	Lipid A export ATP-binding/permease	363	0.862233	Enterococcus	Carbohydrates (43)
	4	k_Bacteria (UID203)	5449	104	58	21 33	24	19 5	2 84.77	89.99	29.01	F_plus_bin4	124	0.016286	Mobile element protein	1634	0.530864	Staphylococcus aureus	Carbohydrates (610)
	7	oActinomycetales (UID1530)	622	257	151	169 86	2	0 0	0 29.67	0.99	100	F_plus_bin7	5	0.005568	Methionine ABC transporter ATP-	78	0.304688	Rothia mucilaginosa	Amino Acids and Derivatives (45)
	9	k Bacteria (UID203)	5449	104	58	8 31	29	25 10	1 97.18	138.04	39.11	F plus bin9	31	0.003972	membrane protein, putative	768	0.345324	Staphylococcus aureus	Carbohydrates (955)
	1	K_Bactena (UID203)	5449	103	360	102 1	0 24	2 0	0 0.16	0	0	ICU_bin1	1	0.2	transposase Mehilo alamant antala	700	0 202000	Desudements	Anima Anida and Daskathura (620)
	19	root (UID1)	5656	56	24	55 1	0	0 0	0 1.04	0	0	ICU bin19	3	0.022388	Plasmin protease 501 233	4	0.057143	Pseudomonas	RNA Metabolism (1)
	20	k Bacteria (UID203)	5449	103	58	97 6	ō	0 0	0 2.59	0	0	ICU bin20	2	0.071429	DNA-directed RNA polymerase beta	1	0.25	Leuconostoc	(4)
ICU	23	k Bacteria (UID203)	5449	104	58	100 4	ō	0 0	0 2.19	0	0	ICU bin23	10	0.042017	Mobile element protein	9	0.125	Pseudomonas	Membrane Transport (10)
	24	f_Moraxellaceae (UID4680)	86	689	365	18 66	2 9	0 0	0 97.59	0.59	0	ICU_bin24	217	0.058808	Mobile element protein	52	0.106776	Acinetobacter	Amino Acids and Derivatives (336)
	3	oActinomycetales (UID1530)	622	258	151	17 20	37	0 0	0 92.19	14.29	100	ICU_bin3	46	0.013699	conserved hypothetical protein	1230	0.750916	Propionibacterium	Amino Acids and Derivatives (255)
	7	k_Bacteria (UID1453)	901	171	117	148 21	2	0 0	0 3.74	0.52	0	ICU_bin7	3	0.032609	photosystem I assembly related	1	0.166667	Prevotella sp. oral taxon	Photosynthesis (17)
	0	f_Micrococcaceae (UID1623)	39	457	218	35 35	5 64	20	0 94.1	14.79	92.86	L_bin0	15	0.003896	Aldehyde dehydrogenase (EC 1.2.1.3)	506	0.397799	Arthrobacter arilaitensis	Amino Acids and Derivatives (389)
	1	c_Bacilli (UID259)	750	2/3	152	210 62	1	0 0	0 6.85	0.06	100	L_DIN1	3	0.017241	Phage terminase, large subunit	48	0.648649	Exiguobacterium	Phages, Prophages, Transposable
	12	k Bacteria (UID202)	500	104	101	5 JU 74 JP	2	0 0	0 95.13	2.45	100	L bin12	12	0.002725	ABC transporter ATB-binding protein	1054	0.197872	Arthrobacter aurercent	Carbobudrater (224)
	14	k Bacteria (UID203)	5449	104	57	82 17	2	2 0	0 43.05	0.96	12.5	L bin14	9	0.003131	Mobile element protein	125	0.354138	Brachybacterium	Membrane Transport (21)
	15	k Bacteria (UID203)	5449	104	58	92 8	4	0 0	0 2.45	0.63	0	L bin15	5	0.032895	Mobile element protein	8	0.145455	Bacillus anthracis str.	Phages, Prophages, Transposable
public buildings (L_clean)	16	f Micrococcaceae (UID1623)	39	457	218	220 23	8 4	0 0	0 52.37	1.16	100	L bin16	18	0.005513	Aldehyde dehydrogenase (EC 1.2.1.3)	97	0.117291	Arthrobacter aurescens	Amino Acids and Derivatives (134)
	17	c_Bacilli (UID259)	750	273	152	75 10	68	22 3	0 85.8	57.95	87.5	L_bin17	30	0.005201	acetyltransferase, GNAT family	3553	0.842942	Exiguobacterium	Amino Acids and Derivatives (514)
	18	f_Micrococcaceae (UID1631)	31	367	171	263 90	13	1 0	0 25.45	3.96	50	L_bin18	17	0.006474	Beta-galactosidase (EC 3.2.1.23)	741	0.607875	Arthrobacter aurescens	Cofactors, Vitamins, Prosthetic
	4	kBacteria (UID203)	5449	104	58	74 19	9	2 0	0 43.1	20.69	93.33	L_bin4	47	0.006359	Mobile element protein	1784	0.547743	Jani bacter sp. HTCC2649	Carbohydrates (342)
	6	oActinomycetales (UID1593)	69	400	198	320 79	1	0 0	0 18.37	0.51	100	L_bin6	4	0.006126	3-oxoacyl-[acyl-carrier protein]	10	0.064516	Burkholderia	Membrane Transport (20)
	7	k_Bacteria (UID203)	5449	104	58	21 37	20	16 9	1 71.26	43.71	11.36	L_bin7	37	0.011534	Mobile element protein	557	0.359587	Enterococcus	Carbohydrates (146)
	10	T_MICROCOCCACEAE (UID1623) k Bacteria (UID202)	39	45/	218	2 90	22	0 0	0 64.25	1.62	85.33	L plus_bin10	18	0.002226	Aldenyde denydrogenase (EC 1.2.1.3) Mobile element protein	108	0.115/56	Arthrobacter aurescens Marroroccur	Amino Acids and Derivatives (211) Protein Metabolism (245)
	13	o Lactobacillales (LID544)	293	475	267	301 15	14	8 1	0 32 93	3 51	18 18	L plus bin13	12	0.009119	Mobile element protein	530	0 703851	Enterococcus	Carbobydrates (36)
	14	k Bacteria (UID203)	5449	104	58	85 16	3	0 0	0 29.31	5.17	100	L plus bin14	15	0.003315	Butyryl-CoA dehydrogenase (EC	1029	0.513217	Janibacter sp. HTCC2649	Carbohydrates (223)
	16	c Bacilli (UID259)	750	273	152	39 13	86	15 2	0 93.41	56.03	92.31	L plus bin16	23	0.004423	acetyltransferase, GNAT family	3258	0.852656	Exiguobacterium	Amino Acids and Derivatives (422)
public building: RMA (Lolur)	17	k_Bacteria (UID203)	5449	104	58	59 41	4	0 0	0 63.11	5.17	100	L_plus_bin17	20	0.003972	Beta-glucosidase (EC 3.2.1.21)	1378	0.59835	Arthrobacter aurescens	Carbohydrates (338)
public buildings ( http://doi.org/	18	k_Bacteria (UID203)	5449	103	57	71 14	13	50	0 12.6	5.98	0	L_plus_bin18	6	0.017442	Replicative DNA helicase (EC 3.6.1);	16	0.140351	Bacillus anthracis str.	Phages, Prophages, Transposable
	4	oActinomycetales (UID1593)	69	400	198	283 11	6	0 0	0 26.37	1.19	50	L_plus_bin4	6	0.004946	Alpha-glucosidase (EC 3.2.1.20)	22	0.071429	Microbacterium	Carbohydrates (30)
	6	f_Micrococcaceae (UID1623)	39	457	218	54 35	5 47	1 0	0 89.88	8.8	78	L_plus_bin6	19	0.005278	Aldehyde dehydrogenase (EC 1.2.1.3)	468	0.397621	Arthrobacter arilaitensis	Amino Acids and Derivatives (380)
	<i>.</i>	k_Bacteria (UID203)	5449	104	58	/8 23	3		0 31.11	3.61	35.33	L_plus_bin/	13	0.004779	FIGUU761799: memorane protein	544	0.503238	Brachybacterium	Carbonydrates (76)
	°	f_Microscoccaceae (UD1621)	21	267	171	45 20	12	1 0	0 35.09	37.06	24.07	L plus_bin8	40	0.004357	Mobile element protein	1430	0.458403	Arthrobacter sp. HTCC2845	Carbohudrater (62)
	0	root (UID1)	5656	56	24	55 1	0	0 0	0 417	0	0	P bin0	5	0.006935	Beta-plurosidase (FC 3 2 1 21)	13	0.068783	Janiharter sn HTCC2649	Carbohydrates (10)
	1	k_Bacteria (UID203)	5449	104	58	100 3	0	1 0	0 6.9	3.45	33.33	P_bin1	65	0.05893	Mobile element protein	53	0.161585	Kocuria rhizophila	Virulence, Disease and Defense (33)
	10	oActinomycetales (UID1530)	622	257	152	27 21	2 16	2 0	0 91.89	10.2	100	P_bin10	50	0.01197	Mobile element protein	457	0.322512	Kocuria rhizophila	Amino Acids and Derivatives (321)
	13	kBacteria (UID203)	5449	104	58	53 36	13	1 1	0 44.37	10.19	68.18	P_bin13	18	0.00542	Methyl-accepting chemotaxis protein I	247	0.346424	Pantoea ananatis LMG	Carbohydrates (113)
	14	c_Bacilli (UID259)	750	273	152	13 12	108	28 0	0 96.05	70.95	94.27	P_bin14	38	0.00598	acetyltransferase, GNAT family	3900	0.844704	Exiguobacterium	Amino Acids and Derivatives (541)
	15	f_Enterobacteriaceae (UIDSUS4)	223	8/4	303	122 42	284	4/1	0 87.38	44.72	90.81	P_bin15	10	0.003578	Methyl-accepting chemotaxis protein i	213	0.184896	Pantoea ananatis LMG	Carbonydrates (430)
public houses (P)	18	<ul> <li>Actinomycetales (UID1530)</li> </ul>	622	257	152	19 20	7 26	5 0	0 93 33	18 55	95.12	P bin18	19	0.004781	Mobile element protein	505	0 385791	Kocuria nalustris PFI	Amino Acids and Derivatives (305)
	19	o Lactobacillales (UID375)	177	350	163	76 20	64	3 0	0 73.49	20.72	8.22	P bin19	15	0.007845	Mobile element protein	143	0.282051	Aerococcus viridans	Carbohydrates (133)
	20	k Bacteria (UID203)	5449	104	58	56 18	16	5 9	0 24.03	14.02	24.71	P bin20	5	0.021834	LSU ribosomal protein L14p (L23e)	20	0.4	Enhydrobacter	Fatty Acids, Lipids, and Isoprenoids (3)
	21	k_Bacteria (UID203)	5449	104	58	5 79	16	4 0	0 93.1	19.28	100	P_bin21	8	0.003521	Lipid A export ATP-binding/permease	98	0.207188	Macrococcus	Protein Metabolism (209)
	22	k_Bacteria (UID203)	5449	104	58	101 2	1	0 0	0 3.45	0.86	100	P_bin22	38	0.064189	Mobile element protein	39	0.171806	Exiguobacterium	DNA Metabolism (26)
	30	root (UID1)	5656	56	24	55 0	1	0 0	0 0.38	0.38	0	P_bin30	3	0.05	NADH-ubiquinone oxidoreductase	3	0.142857	Synechococcus sp. WH	
	33	root (UID1)	5656	56	24	54 2	0	0 0	0 8.33	0	0	P_bin33	5	0.017921	Mobile element protein	27	0.442623	Acinetobacter sp. ATCC	Membrane Transport (6)
	54	k_Bacteria (UID203)	5449	104	58	86 11 00 F	-	0 0	0 4.58	11	28.5/	P_bin34	3	0.0011628	Inymidylate synthase (EC.2.1.1.45)	9	0.088235	Arthrobacter aniaitensis	Phages, Prophages, Transposable
	10	c Recilli (UD295)	5940	225	191	19 22	1 27	0 0	0 8.62	12.57	04 50	P_plus_bin10	,	0.008052	Mobile element protein	10	0.066546	Marracer sp. HTCC2849	Carbonydrates (24) Protein Metabolicm (177)
	12	<ul> <li>Actinomycettler (UID1520)</li> </ul>	500	325	161	12 21	3 3/	0 0	0 95.09	10.99	94.39	P_plus_bin12	0 21	0.005197	Mobile element protein	571	0.19181	Kocuria palurtrir PEI	Amino Acids and Darivativas (217)
	16	o Actinomycetales (UID1530)	622	257	152	24 22	1 12	0 0	0 90.73	4.5	91.67	P plus bin16	33	0.008623	Mobile element protein	443	0.352426	Kocuria rhizophila	Amino Acids and Derivatives (312)
	21	f_Micrococcaceae (UID1623)	39	457	218	305 14	3 3	1 0	0 32.84	1.65	33.33	P_plus_bin21	18	0.010664	Mobile element protein	203	0.373162	Arthrobacter arilaitensis	Carbohydrates (77)
public bourser (P. pture)	23	kBacteria (UID203)	5449	104	58	94 10	0	0 0	0 7.84	0	0	P_plus_bin23	57	0.051677	Mobile element protein	42	0.124629	Kocuria rhizophila	Virulence, Disease and Defense (35)
puone nouses (P_pius)	3	k_Bacteria (UID203)	5449	103	57	3 64	31	50	0 95.61	55.48	91.3	P_plus_bin3	36	0.005802	acetyltransferase, GNAT family	3810	0.846291	Exiguobacterium	Carbohydrates (529)
	4	k_Bacteria (UID203)	5449	104	58	99 3	2	0 0	0 0.78	0.31	0	P_plus_bin4	2	0.011765	Thymidylate synthase (EC 2.1.1.45)	4	0.047619	Pantoea ananatis LMG	Phages, Prophages, Transposable
	5	f_Enterobacteriaceae (UID5054)	223	876	305	111 45	276	30 2	0 90.68	37.88	45.77	P_plus_bin5	26	0.004153	Methyl-accepting chemotaxis protein I	206	0.175021	Pantoea ananatis LMG	Carbohydrates (502)
	6	o_Lactobacillales (UID355)	490	327	182	216 10	2 9	0 0	0 30.13	2.21	22.22	P_plus_bin6	10	0.01443	Mobile element protein	143	0.60084	Aerococcus viridans	Carbohydrates (29)
	8	<pre>k_sacteria (UID203) root (UID1)</pre>	5449	104	58	53 29	18	40	U 37.4	12.85	/b.67	r_plus_bin8	16	0.004799	rranscriptional regulator, LysR family Mobile element protein	280 21	u.389972	Pantoea ananatis LMG	Carbohydrates (118) Rhaner, Broohater, Transport-bi-
	0	f Elavoharteriaceae (UED2917)	3030	511	24	JS 2	1	0 0	0 683	3.53	55	pros_DIN9	12	0.004847	Two-component system response	16	0.035011	Flavobacterium	Amino Arids and Derivatives
	16	<ul> <li>Pseudomonadales (UID4488)</li> </ul>	185	813	308	465 27	3 64	92	0 44.24	11.59	38.83	UR bin16	62	0.008713	Transcriptional regulator, LysR family	892	0.46775	Pseudomonas	Amino Acids and Derivatives
	17	o Sphingomonadales (UID3310)	26	569	293	311 23	27	1 0	0 43.13	5.28	16.67	UR bin17	73	0.016276	Mobile element protein	123	0.097156	Novosphingobium sp.	Carbonhydrates
	18	o_Sphingomonadales (UID3310)	26	569	293	37 49	35	0 0	0 92.7	9.33	85.71	UR_bin18	28	0.00546	Mobile element protein	320	0.255387	Sphingobium sp. AP49	Amino Acids and Derivatives (371)
	19	root (UID1)	5656	56	24	54 1	1	0 0	0 8.33	4.17	0	UR_bin19	231	0.053696	Mobile element protein	391	0.253238	Pseudomonas	Virulence, Disease and Defense (129)
	21	kBacteria (UID203)	5449	104	58	64 25	12	3 0	0 56.72	26.55	76.19	UR_bin21	66	0.005966	Transcriptional regulator, LysR family	2738	0.552796	Stenotrophomonas	Amino Acids and Derivatives
and the formation	24	oActinomycetales (UID1530)	622	259	152	34 19	1 31	3 0	0 85.8	13.18	92.5	UR_bin24	44	0.012185	conserved hypothetical protein	1290	0.746096	Propionibacterium	Carbonhydrates
gowning area (UR)	26	r_knizoblaceae (UID3564)	/8	840	354	/ 71	, 102	137	1 99.01	18.85	34.2 100	UR_bin26	242	0.022524	Mobile element protein	1846	0.456028	Agrobacterium	Carbohydrates (832)
	32	pouterorderes (UID2605)	350 564	340	210	27 27	1 10	0 0	0 20.00	3.41	50	UR bic??	10	0.01433	diguarylate	-0 83	0.148479	-y-mgouacterium Micavibrio	Protein Metabolism (167)
	33	f Moraxellaceae (IIIn4680)	86	689	365	40 50	5 330	11 2	0 92.75	24.59	84	UR bin33	256	0.038694	Mobile element protein	79	0.089569	Adjactor on ATCC	Amino Acids and Derivatives (414)
	39	k Bacteria (UID203)	5449	103	57	84 17	2	0 0	0 31.58	3.51	50	UR bin39	94	0.014971	Glyoxylate reductase (EC 1.1.1.7%) /	1497	0.650587	Pseudomonas	Amino Acids and Derivatives
	41	k_Bacteria (UID203)	5449	104	58	20 42	21	18 2	1 78.15	62.15	45.36	UR_bin41	33	0.004112	Inosine-uridine preferring nucleoside	1917	0.599437	Stenotrophomonas	Amino Acids and Derivatives
	7	k_Bacteria (UID203)	5449	104	58	39 52	10	3 0	0 52.32	7.21	10.53	UR_bin7	23	0.005852	Mobile element protein	1145	0.582994	Caulobacter crescentus	Membrane Transport (94)
	9	k_Bacteria (UID1453)	901	171	117	140 25	4	2 0	0 8.04	1.66	20	UR_bin9	22	0.043738	Mobile element protein	12	0.090909	Acinetobacter	Phages, Prophages, Transposable

Supplementary Table 3: Pan- and Core genome analysis of different built environments and species.

pan-genomes	of differe	ent built environments								
genome bin	CDS (	CDS (without artefact fam.)	Pan CDS	Core CDS \	ar CDS	Strain specific CDS	Core CDS (%)	Var CDS (%)	Strain spe. CDS (%)	Excluded CDS (%)
CR8 bin_21	4642	4642	4642	5	4637	4600	0.108	99.892	99.095	0
CR8 bin_33 CR8 bin_8	5238 3070	5238	5238 3070	3	5235 3067	3063	0.057	99.943	99.275	0
F bin_15	5169	5169	5169	4	5165	712	0.077	99.923	13.774	0
F bin_23	10627	10627	10627	7	10620	2551	0.066	99.934	24.005	0
F bin_31 F bin_34	7567 2183	7567	7567 2183	2	7565 2181	1563	0.026	99.974	20.655	0
F bin_35	6345	6345	6345	5	6340	2926	0.079	99.921	46.115	0
F_plus bin_10	2695	2695	2695	2	2693	468	0.074	99.926	17.365	0
F_plus bin_13 F_plus bin_14	4170 2762	4170	4170 2762	1	4169 2759	1563	0.024	99.976 99.891	37.482	0
F_plus bin_18	7064	7064	7064	3	7061	1381	0.042	99.958	19.55	0
F_plus bin_23	2222	2222	2222	3	2219	205	0.135	99.865	9.226	0
F_plus bin_30 F_plus bin_4	6136 9563	6136	6136 9563	1	6135 9557	6042	0.016	99.984	98.468	0
F_plus bin_9	8301	8301	8301	2	8299	1626	0.024	99.976	19.588	0
ICU bin_18	6265	6265	6265	8	6257	6167	0.128	99.872	98.436	0
ICU bin_24	3951 2826	3951	3951	8	3943	3861	0.202	99.798	97.722	0
Lbin_0	3892	3892	3892	5	3887	711	0.128	99.872	18.268	0
Lbin_10	2265	2265	2265	5	2260	156	0.221	99.779	6.887	0
Lbin_17 L plus bin_11	5420 2544	5419 2544	5419 2544	3	5416 2539	1181	0.055	99.945	21.794	0
L_plus bin_16	4973	4971	4971	4	4967	954	0.08	99.92	19.191	0
L_plus bin_6	3608	3608	3608	4	3604	523	0.111	99.889	14.496	0
Pbin_10 Pbin_14	3488 6196	3488	3488 6196	4	3484 6192	901 1000	0.115	99.885	25.831	0
P bin_15	4987	4987	4987	6	4981	1473	0.12	99.88	29.537	0
P bin_18	2975	2975	2975	5	2970	709	0.168	99.832	23.832	0
P bin_21 P plus hin 10	2281 2109	2281	2281 2109	5	2276 2105	202	0.219	99.781 99.81	8.856	0
P_plus bin_10	3026	3026	3026	4	3022	755	0.13	99.868	24.95	0
P_plus bin_16	3130	3130	3130	4	3126	568	0.128	99.872	18.147	0
P_plus bin_3 P_plus bin_5	6021 5086	6021	6021 5086	5 6	6016 5080	864 1626	0.083	99.917 99.927	14.35	0
UR bin_18	4216	4216	4216	0	4216	4096	0.118	100	97.154	0
UR bin_24	3332	3331	3331	0	3331	3325	0	100	99.82	0
UR bin_26	10242 2545	10242	10242 2545	0	10242 2545	10061	0	100	98.233	0
UR bin_33	7639	7639	7639	0	7639	7504	0	100	98.233	0
UR bin_41	6116	6116	6116	0	6116	6037	0	100	98.708	0
UR bin_9	846	843	843	0	843	793	0	100	94.069	0
pan-genome o	f Acineto	obacter								
genome bin	CDS (	CDS (without artefact fam.)	Pan CDS	Core CDS \	ar CDS	Strain specific CDS	Core CDS (%)	Var CDS (%)	Strain spe. CDS (%)	Excluded CDS (%)
CR8 bin 21	4642	4642	4642	1542	2409	784 366	39.028	63.033	19.843	0
F_plus bin_30	6136	6136	6136	2209	3927	3157	36.001	63.999	51.45	0
UR bin_33	7639	7639	7639	1802	5837	2857	23.589	76.411	37.4	0
pan-genome o	of confine	d built environments								
and a state later.										
genome bin	CDS (	CDS (without artefact fam.)	Pan CDS	Core CDS	ar CDS	Strain specific CDS	Core CDS (%)	Var CDS (%)	Strain spe. CDS (%)	Excluded CDS (%)
CR8 bin_8 CR8 bin_21	CDS 0 3070 4642	CDS (without artefact fam.) 3070 4642	Pan CDS 3070 4642	Core CDS \ 0 0	/ar CDS 3070 4642	Strain specific CDS 559 410	Core CDS (%) 0 0	Var CDS (%) 100 100	Strain spe. CDS (%) 18.208 8.832	Excluded CDS (%) 0 0
CR8 bin_8 CR8 bin_21 CR8 bin_33	CDS 0 3070 4642 5238	CDS (without artefact fam.) 3070 4642 5238	Pan CDS 3070 4642 5238	Core CDS \ 0 0 0	/ar CDS 3070 4642 5238	Strain specific CDS 559 410 3575	Core CDS (%) 0 0 0	Var CDS (%) 100 100 100	Strain spe. CDS (%) 18.208 8.832 68.251	Excluded CDS (%) 0 0 0
CR8 bin_8 CR8 bin_21 CR8 bin_33 UR bin_9	CDS 0 3070 4642 5238 846	CDS (without artefact fam.) 3070 4642 5238 843	Pan CDS 3070 4642 5238 843	Core CDS \ 0 0 0 0	/ar CDS 3070 4642 5238 843	Strain specific CDS 559 410 3575 766	Core CDS (%) 0 0 0 0 0	Var CDS (%) 100 100 100 100	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 27.002	Excluded CDS (%) 0 0 0
CR8 bin_8 CR8 bin_21 CR8 bin_21 CR8 bin_33 UR bin_9 UR bin_18 UR bin_24	CDS 0 3070 4642 5238 846 4216 3332	CDS (without artefact fam.) 3070 4642 5238 843 4216 3331	Pan CDS 3070 4642 5238 843 4216 3331	Core CDS V 0 0 0 0 0 0	/ar CDS 3070 4642 5238 843 4216 3331	Strain specific CDS 559 410 3575 766 4093 970	Core CDS (%) 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.12	Excluded CDS (%) 0 0 0 0 0 0
CR8 bin_8 CR8 bin_21 CR8 bin_21 CR8 bin_33 UR bin_9 UR bin_18 UR bin_24 UR bin_26	CDS 0 3070 4642 5238 846 4216 3332 10242	CDS (without artefact fam.) 3070 4642 5238 843 4216 3331 10242	Pan CDS 3070 4642 5238 843 4216 3331 10242	Core CDS V 0 0 0 0 0 0 0 0 0	/ar CDS 3070 4642 5238 843 4216 3331 10242	Strain specific CDS 559 410 3575 766 4093 970 10001	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.12 97.647	Excluded CDS (%) 0 0 0 0 0 0 0 0 0
genome bin CR8 bin_8 CR8 bin_21 CR8 bin_33 UR bin_9 UR bin_18 UR bin_24 UR bin_26 UR bin_32 UR bin_32	CDS 0 3070 4642 5238 846 4216 3332 10242 2545 700	CDS (without artefact fam.) 3070 4642 5238 843 4216 3331 10242 2545 7270	Pan CDS 3070 4642 5238 843 4216 3331 10242 2545 7620	Core CDS V 0 0 0 0 0 0 0 0 0	/ar CDS 3070 4642 5238 843 4216 3331 10242 2545	Strain specific CDS 559 410 3375 766 4093 970 10001 2503	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 100	Strain spe. CDS (%) 18.208 8.832 68.251 90.868 97.083 29.12 97.647 98.33 0.015 0.01	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0
genome bin CR8 bin_8 CR8 bin_21 CR8 bin_33 UR bin_9 UR bin_9 UR bin_18 UR bin_24 UR bin_26 UR bin_32 UR bin_33 UR bin_41	CDS 0 3070 4642 5238 846 4216 3332 10242 2545 7639 6116	CDS (without artefact fam.) 3070 4642 5238 843 4216 3331 10242 2545 7639 6116	Pan CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116	Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	/ar CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116	Strain specific CDS 559 410 3575 766 4093 970 10001 2503 2988 4189	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 100 100	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.12 97.647 98.35 39.115 68.492	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome bin           CR8 bin_8           CR8 bin_21           CR8 bin_33           UR bin_33           UR bin_18           UR bin_24           UR bin_23           UR bin_33           UR bin_41           ICU bin_3	CDS 0 3070 4642 5238 846 4216 3332 10242 2545 7639 6116 2826	CDS (without artefact fam.) 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2525 2545 7639 6116	Pan CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2826	Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	/ar CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2826	Strain specific CDS           559           410           3575           766           4093           970           10001           2503           2988           4189           478	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.12 97.647 98.35 39.115 68.492 16.914	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome bin         genome bin           CR8 bin_8         CR8 bin_21           CR8 bin_21         CR8 bin_33           UR bin_18         UR bin_24           UR bin_24         UR bin_24           UR bin_32         UR bin_33           UR bin_33         UR bin_33           UR bin_33         ICU bin_33           ICU bin_33         ICU bin_24	CDS 0 3070 4642 5238 846 4216 3332 10242 2545 7639 6116 2826 6266 6265	CDS (without artefact fam.) 30707 4642 5238 843 4216 3331 10242 2545 10242 2545 10242 2545 10242 2545 10242 2545 10242 2545 10242 2545 10242 1024 1024	Pan CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2826 6265	Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	/ar CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2826 6265 6265	Strain specific CDS 559 4100 3575 766 4093 9700 10001 2503 2988 4189 2988 4189 5979	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.12 97.647 98.35 39.115 68.492 16.914 95.435 54.95 55.55	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome bin           CR8 bin_8           CR8 bin_21           CR8 bin_31           UR bin_9           UR bin_24           UR bin_226           UR bin_32           UR bin_33           UR bin_33           UR bin_33           UC bin_3           ICU bin_3           ICU bin_3           ICU bin_24	CDS 0 3070 4642 5238 846 4216 3332 10242 2545 7639 6116 2826 6265 3951	CDS (without artefact fam.) 30707 4642 5238 843 4216 3331 4216 3331 4216 3331 4216 3331 4216 3331 4216 3331 4216 3331 4256 4256 5626 3351	Pan CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2826 6265 3951	Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	/ar CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2826 6265 3951	Strain specific CDS 559 410 3575 766 4093 970 10001 2503 2503 2503 2508 4189 4189 4189 5979 853	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.822 68.251 99.866 97.083 29.12 97.647 98.35 39.115 66.492 16.914 95.435 21.589	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome bin           CR8 bin_8           CR8 bin_12           CR8 bin_33           UR bin_19           UR bin_124           UR bin_32           UR bin_33           UR bin_33           UR bin_18           ICU bin_18           ICU bin_24	CDS 0 3070 4642 5238 846 4216 3332 10242 2545 7639 6116 2826 6265 3951 of unrestr	CDS (without artefact fam.) 30707 4642 5238 843 4216 3331 10242 2545 7633 6116 2626 6265 3351 icted built environments	Pan CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2826 6265 3951	Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	/ar CDS 3070 4642 5238 4216 3331 10242 2545 7639 6116 2826 6265 3951	Strain specific CDS         S59           410         3575           766         4003           970         10001           2503         2988           4189         4198           5979         853	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.12 97.647 98.35 33.115 68.492 16.914 95.435 21.589	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome bin CR8 bin_8 CR8 bin_21 CR8 bin_33 UR bin_9 UR bin_18 UR bin_24 UR bin_32 UR bin_33 UR bin_33 UR bin_33 ICU bin_18 ICU bin_18 ICU bin_24 pan-genome o genome bin	CDS 0 3070 4642 5238 846 4216 3332 10242 2545 7639 6116 2826 6265 3951 of unrestr	CDS (without artefact fam.) 30707 4642 5233 943 4216 3331 10242 2545 10242 2545 10242 2545 10242 2545 10242 2545 10242 2545 10242 2545 10242 2545 10242 1020	Pan CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2826 6265 3951 Pan CDS	Core CDS \ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	/ar CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2826 6265 3951 /ar CDS	Strain specific CDS         559           410         3375           766         4033           970         10001           2503         2988           4189         478           5979         853           Strain specific CDS         Strain specific CDS	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.12 97.647 98.35 33.115 68.492 16.914 95.435 21.589 Strain spe. CDS (%)	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome bin         26           CR8 bin_21         27           CR8 bin_33         UR bin_33           UR bin_124         UR bin_24           UR bin_32         UR bin_33           UR bin_41         100           ICU bin_31         100           ICU bin_18         100           UCU bin_124         pan-genome o           genome bin_15         5           E bin_32         2	CDS 0 3070 4642 5238 846 4216 3332 10242 2545 7639 6116 2826 6265 3951 if unrestr CDS 0 5169 10637	CDS (without artefact fam.) 30707 4642 5238 443 4216 3331 10242 2545 7639 6116 2256 3951 1cted built environments CDS (without artefact fam.) 5169	Pan CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2826 6265 3951 Pan CDS 5169 10627	Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	(ar CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2826 6265 3951 (ar CDS 5169 10227	Strain specific CDS         S59           410         3575           766         4093           9707         10001           2503         2988           4189         478           5979         853           Strain specific CDS         712           712         712	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.12 97.647 98.33 30.115 68.492 16.914 95.455 21.589 Strain spe. CDS (%) 13.747 23.767	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
yenome on CR8 bin, 2 CR8 bin, 21 CR8 bin, 23 UR bin, 21 UR bin, 23 UR bin, 24 UR bin, 24 UR bin, 24 UR bin, 24 UR bin, 24 ICU bin, 24 pan-genome on F bin, 15 F bin, 23 F bin, 23	CDS 0 3070 4642 5238 846 4216 3332 10242 2545 7639 6116 2826 6265 3951 0 f unrestr CDS 0 5169 10627 7567	CDS (without artefact fam.) 3070 4642 5238 4843 4843 4216 3331 10242 2545 7639 6116 2626 6265 6265 10222 CDS (without artefact fam.) 5169 10627 7566	Pan CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2826 6265 3951 Pan CDS 5169 1067 7566	Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	(ar CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 6166 6265 3951 (ar CDS 5169 10627 7566	Strain specific CDS SS9 410 3575 766 4003 970 10001 2503 2988 4189 478 5979 853 Strain specific CDS 712 2539 973	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.12 97.647 98.35 39.115 66.492 16.914 95.435 21.589 Strain spe. CDS (%) 13.774 23.872 12.595	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome on CRB bin, 3 CRB bin, 21 CRB bin, 21 CRB bin, 21 UR bin, 24 UR bin, 24 UR bin, 24 UR bin, 24 UR bin, 24 UR bin, 31 ICU bin, 31 ICU bin, 31 ICU bin, 31 ICU bin, 31 ICU bin, 28 ICU bin, 28 F bin, 23 F bin, 24	CDS 0 3070 4642 5238 846 4216 3332 10242 2545 7639 6116 2826 6265 3951 0 funrestr CDS 0 5169 10627 7557 2183	CDS (without artefact fam.) 3070 4642 5283 843 4216 3331 10242 2545 47633 6116 2626 6265 3951 icted built environments CDS (without artefact fam.) 10627 7566 10627 7566 2183	Pan CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2826 6265 3951 Pan CDS 5169 10627 7566 2183	Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	/ar CDS 3070 4642 5238 843 4216 33311 10242 2545 7639 6116 6265 3951 4/ar CDS 5169 5169 5169 7566 2183	Strain specific CDS 559 410 3575 766 4003 970 10001 2503 2988 4189 478 5979 853 Strain specific CDS 5train specific CDS 712 2539 953 103	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.102 97.647 98.35 33.115 68.492 16.914 95.435 21.589 Strain spe. CDS (%) Strain spe. CDS (%) 13.774 23.892 12.586 4.718	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome on CR8 bin, 3 CR8 bin, 21 CR8 bin, 21 CR8 bin, 21 CR8 bin, 24 UR bin, 32 UR bin, 31 F bin, 31 F bin, 31 F bin, 31 F bin, 31	CDS 0 3070 4642 5238 846 4216 3332 10242 2545 7639 6116 2826 6265 3951 f unrestr CDS 0 5169 10627 7567 2183 6345 0757	CDS (without artefact fam.) 30707 4642 5233 943 4216 3331 10242 2545 4216 3331 10242 2545 4256 3351 10242 2545 10242 2545 10242 10242 10242 1025 1057 10627 106777 106777 106777 106777 106777 106777 106777 106	Pan CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2826 6265 3951 Pan CDS 5169 10627 7566 2183 6345 5457 5569 10627 7566 2183 6345 5567 5577 5567 5577 5567 5577 5567 5577 5577 5567 5577 5567 55777 55777 55777 5577 5577 5577 5577 55777 5577 5577	Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	/ar CDS 3070 4642 5238 843 4216 33311 10242 2545 7639 6116 6265 3951 4ar CDS 5169 5169 5169 7566 62183 6345 2823	Strain specific CDS 559 410 3375 766 4003 970 10001 2503 2988 4189 478 5979 853 5170 712 2539 953 103 2539 953 103 2991	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.200 8.832 68.251 90.866 97.083 29.12 97.647 98.35 33.115 68.492 16.914 95.435 21.589 Strain spe. CDS (%) 13.774 23.892 12.596 4.718 64.718 64.718 55.718	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome bin (R8 bin, 3 (R8 bin, 21 (R8 bin, 21 (R8 bin, 21 (R8 bin, 24 UR bin, 24 UR bin, 24 UR bin, 24 UR bin, 24 UR bin, 32 (UU bin, 33 ICU bin, 31 F bin, 23 F bin, 23 F bin, 31 F bin, 35 F bin, 35 F plus bin, 4 F plus bin, 4	CDS 0 3070 4642 5238 846 4216 3332 22545 7639 6116 2826 6625 3951 funrestr CDS 0 5169 7567 7567 7567 2183 6445 9563 8401	CDS (without artefact fam.) 30707 4642 5238 843 4216 3331 10242 2545 7639 6116 2826 6265 3951 icted built environments CDS (without artefact fam.) 5160 10622 7566 2035 10622 7566 3953 3062 3062 3062 3062 3075 3062 3075 3062 3075 307	Pan CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 2826 6265 3951 Pan CDS 5169 10627 7566 2183 6345 9563 8301	Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Aar CDS 3070 4642 5238 843 3331 10242 2545 7639 6116 2826 6265 3951 0627 5169 10527 5169 10527 5169 2183 6345 9563 8301	Strain specific CDS S59 410 3575 766 4003 970 10001 2533 2588 4189 478 5979 853 5579 853 5772 2539 953 103 2953 103 2953 103 2953 103 2953 103 2953 105 105 105 105 105 105 105 105	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.12 97.647 98.35 33.115 68.492 16.914 95.435 21.589 Strain spe. CDS (%) 13.774 23.892 12.596 4.718 40.036 62.146 19.516	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome on CR8 bin, 3 CR8 bin, 21 CR8 bin, 21 CR8 bin, 21 UR bin, 24 UR bin, 24 UR bin, 24 UR bin, 24 UR bin, 32 UR bin, 32 UR bin, 31 ICU bin, 31 ICU bin, 31 ICU bin, 31 ICU bin, 32 ICU bin, 34 F bin, 35 F	CDS 0 3070 0 4642 2 5238 846 4216 2 3332 10242 2 545 5 7639 0 10242 2 545 5 7639 0 10242 1 2826 6 665 3 3951 1 6116 6 10627 7 566 3 10627 7 567 3 10267 1 2833 6 445 2 516 9 10627 1 2833 6 10627 1 2833 6 10627 1 2835 6 10707 1 2855 6 10707 1 2007 1000 10000 100000000000000000000000	CDS (without artefact fam.) 3070 4642 5238 843 4216 3331 10242 2545 4216 2545 4216 2545 4216 2555 4216 2555 4216 2555 4217 5169 10627 2556 2188 6345 5455 2558 8301 1057 2655 2188 8455 2655 2188 8455 2185 2185 2195	Pan CDS         3070           4642         5238           843         4216           3331         10242           2545         7639           6116         2826           6265         3951           Pan CDS         5169           10627         7566           6345         9563           6345         2183           6345         2695	Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	rar CDS 3070 4642 5238 843 3331 10242 2545 7639 6116 2826 6265 3951 7639 5169 10627 7566 2183 6445 9563 6445 9563 8301 2695	Strain specific CDS SS9 410 3575 766 4003 970 10001 2503 2988 4189 478 5979 853 Strain specific CDS 712 2539 103 2933 103 2933 103 2931 1620 458 459 458 458 458 458 458 458 458 458	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.12 97.647 98.453 33.115 66.492 15.914 95.435 21.589 Strain spe. CDS (%) 13.774 23.892 12.596 4.718 46.066 62.146 15.516 16.994	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome on CR8 bin, 3 CR8 bin, 21 CR8 bin, 21 CR8 bin, 21 UR bin, 24 UR bin, 24 UR bin, 24 UR bin, 24 UR bin, 24 UR bin, 32 UR bin, 32 UR bin, 31 ICU bin, 24 ICU bin, 24 ICU bin, 24 pan-genome on genome bin F bin, 35 F bin, 35	CDS         4           3070         4           3072         4           3070         4           4642         5238           846         4216           3332         10242           10242         2545           10242         2545           6116         6265           3951         10627           7567         2183           9563         8301           4100         2695           4100         2702	CDS (without artefact fam.) 3070 4642 5283 843 4216 3331 10242 2545 47633 6116 2626 6265 3951 icted built environments CDS (without artefact fam.) 10627 7566 10627 7566 10627 7566 3634 3645 3645 3645 36555 3655 3655 3655 36555 36555 36555 365555 36555 36555	Pan CDS 3070 4642 5238 843 4216 4331 10242 2545 7639 6116 62255 5169 5169 10627 7566 6345 9563 8301 10624 21756 6345 9563 8301 10627 7566 6345 9563 8301 10627 7566 6345 9563 8301 10627 7566 6345 9563 8301 10627 7566 6345 9563 8301 10627 7566 6345 9563 8301 10627 7566 7578 8301 10627 7566 75788 7578 7578 75788 75788 7578 7	Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	ar CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 6255 5169	Strain specific CDS 559 410 3575 766 4003 970 10001 2503 2988 4189 478 5979 853 Strain specific CDS 5239 953 103 2981 103 2981 103 2981 103 2981 103 2981 103 2981 103 2981 103 2981 103 2981 103 2981 103 103 103 103 103 103 103 10	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.12 97.647 98.35 33.115 68.492 16.914 95.435 21.589 Strain spe. CDS (%) Strain spe. CDS (%) 12.596 4.718 4.036 62.146 15.166 16.994 37.482 24.257 24.257 25.2588 25.2588 25.2588 25.2588 25.2588	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
<u>genome bin</u> CR8 bin, 32 CR8 bin, 21 CR8 bin, 21 CR8 bin, 24 UR bin, 32 UR bin, 31 F bin, 35 F bin, 31 F bin, 31 F bin, 35 F pilus bin, 47 F pilu	CDS         4           3070         4           4522         5           3332         10242           10242         2545           10242         2545           3050         6116           6126         5029           3951         10027           7557         7557           7556         9563           8301         20595           4170         2762           4170         2762           7064         7064	CDS (without artefact fam.) 3070 4642 5283 843 4216 3331 10242 2545 4216 3331 10242 2545 (3351 10242 2545 (3351 10242 10242 10242 1025 1027 1067 1067 1067 1067 1067 1067 1067 1067 1067 1067 1067 1067 1067 1067 1067 1067 1067 107 107 107 107 107 107 107 10	Pan CDS 3070 4642 8238 8433 4216 3331 10242 2565 5169 51	Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	ar CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 6255 3951 10627 5169 10627 7566 6345 5169 10627 7566 8301 2695 4170 2762 4170 7664 4170 7664 7664 7664 7664 7664 7664 7664 7664 7664 7664 7664 7664 7664 7664 7664 7666 7	Strain specific CDS         S59           3575         3576           4003         3575           7666         4093           970         10001           2503         2988           4189         478           5979         853           7112         2539           953         103           2921         5539           1032         2984           1620         488           1563         1620           488         1563           289         878	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 99.0866 97.083 29.12 97.647 98.35 33.115 66.492 16.914 95.435 21.589 Strain spe. CDS (%) 13.774 23.892 12.596 4.718 46.036 62.146 19.516 15.944 37.482 19.516 16.944 37.482 10.463 11.549 10.463 11.549 10.463 10.464 10.463 10.464	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome on           CR8 bin, 21           UR bin, 24           UR bin, 24           UR bin, 24           UR bin, 24           UR bin, 32           UR bin, 31           ICU bin, 31           ICU bin, 35           F bin, 31           F bin, 35           F bin, 35           F bin, 35           F bin, 35           F plus bin, 31           F plus bin, 35           F plus bin, 36           F plus bin, 37	CDS         0           3070         0           3072         0           4642         5238           846         3322           10242         5238           2545         7639           6116         6265           33951         1           funrestr         10627           5169         10627           7567         7567           2183         8301           2695         38301           2695         38301           2762         7064           2170         2222	CDS (without artefact fam.) 30707 4642 5238 843 4216 3331 10242 2545 4216 3331 10242 2545 4216 3331 10242 2545 4265 3351 10222 10627 1070 10707 107	Pan CDS 3070 4642 5238 843 4216 3331 10242 2565 3331 10242 2565 3331 10242 2565 3351 10242 2555 3550 10277 2757 2752 2755 2755 2755 2755 2755 2755 2755 2755 2755 2755 27	Core CD5 V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 62826 6265 3351 10627 7566 2183 6455 9563 8301 10627 7566 2183 8301 10627 7566 2183 8301 10627 7566 2183 8301 10627 7566 2183 2065 2183 2065 2183 2065 2183 2065 2183 2065 2183 2070 207	Strain specific CDS S59 410 3575 766 4003 970 10001 2503 2988 4189 478 5979 853 5576 5712 2539 953 103 2921 5943 1620 4588 1620 4588 1620 878 1620 878 1620 878 1620 878 1620 878 1620 878 1620 878 1620 878 1620 878 1620 878 1620 878 1620 878 1620	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 1000	Strain spe. CDS (%) 18.208 8.832 68.251 99.684 99.083 29.12 99.647 98.35 30.115 68.492 16.914 95.455 21.589 Strain spe. CDS (%) 13.774 23.892 12.596 4.718 46.036 62.146 19.516 16.994 37.482 10.463	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome on CR8 bin, 3 CR8 bin, 21 CR8 bin, 21 CR8 bin, 21 UR bin, 24 UR bin, 24 UR bin, 24 UR bin, 24 UR bin, 24 UR bin, 32 UR bin, 32 UR bin, 31 ICU bin, 33 ICU bin, 33 ICU bin, 33 ICU bin, 33 ICU bin, 33 ICU bin, 33 ICU bin, 33 F bin, 33 F bin, 34 F bin, 35 F bin, 34 F bin, 34 F bin, 35 F bin,	CDS         0           30700         30700           30700         4642           5238         846           3322         5238           3426         5238           3321         10042           5238         7639           6116         6265           6265         3351           10622         5268           5169         10627           10757         2183           6345         9563           8301         26953           20542         7064           4170         2762           7064         2222           6136         2010	CDS (without artefact fam.) 3070 4642 5238 843 4216 3331 10242 2545 4763 6166 2826 6265 6265 6265 6265 6265 10627 7566 2183 6345 1062 2183 6345 2183 2185 2183 2185	Pan CDS 30700 46422 5238 8 8433 4216 33311 10242 5455 7639 6116 62255 39511 10242 5169 77566 6245 5169 10627 77566 6345 5169 9563 4170 2183 8301 2655 4170 2183 8301 2655 4170 2183 8301 2655 4170 2183 8301 2055 4170 2183 8301 2055 4170 2183 8301 2055 4170 2183 8301 2055 4170 2183 8301 2055 2188 8301 2055 2188 2185 2165 2175 2185 2185 2165 2175 2185 2185 2165 2175 2185 2185 2175 2185 2185 2175 2185 2185 2185 2185 2185 2175 2185 2185 2185 2185 2185 2175 2175 2175 2185 2185 2185 2185 2185 2175 2175 2175 2185 2185 2185 217	Core CDS V / 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Ar CDS 3070 4642 5238 843 4216 3331 10242 2545 7639 6116 62826 6265 3951 5169 516	Strain specific CDS SS9 410 3575 766 4003 970 10001 2503 4189 4189 478 5979 853 Strain specific CDS 527 2539 103 2953 103 2951 1620 4584 1162 2898 878 1162 289 878 1162 289 878 1162 289 1662 16	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 99.686 99.083 29.121 97.647 98.635 29.647 97.647 98.635 21.589 Strain spe. CDS (%) 51.774 23.892 12.596 4.718 4.036 12.596 4.718 4.036 15.166 15.944 13.516 15.944 13.516 15.944 13.545 15.944 13.545 15.944 13.744 13.545 15.944 13.545 15.944 13.744 13.545 15.944 13.744 13.545 15.944 13.744 13.545 15.944 13.744 13.545 15.944 13.744 13.545 15.944 13.744 13.545 14.945 15.945 1	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome bin           CR8 bin, 31           CR8 bin, 21           CR8 bin, 21           CR8 bin, 23           UR bin, 24           ICU bin, 31           F bin, 13           F bin, 35           F plus bin, 30           F plus bin, 31           P plus bin, 33           P plus bin, 30           P plus bin, 31	CDS         0           30700         0           30700         4642           5238         846           25238         846           3322         10242           2332         2545           7639         351           2826         6166           2826         6265           5169         3051           10627         7567           2183         6345           95631         2695           41700         2762           7064         22762           3488         6136           3488         6136           3488         6136	CDS (without artefact fam.) 3070 4642 5283 843 4216 3331 10242 2545 4763 4216 4255 4255 4255 4255 4255 4255 10627 7566 10627 7566 10627 7566 10627 7566 10627 7566 4170 2285 4170 2782 4170 2782 4170 2785 4170 4170 2785 4170 2785 4170 2785 4170 2785 4170 2785 4170 2785 4170 2785 4170 2785 4170 2785 4170 2785 4170 2785 4170 2785 4170 41	Pan CDS 30700 46422 8433 4216 5238 8433 4216 5238 8434 7639 5169 5169 5169 7566 6345 5169 10627 7566 6345 5169 2183 3051 10627 7566 6345 5169 2183 3051 10627 7566 6345 2183 3051 10627 7566 6345 2183 3051 10627 7566 6345 2183 3051 10627 7566 6345 2183 3051 10627 7566 6345 2183 3051 10627 7566 6345 2183 3051 10627 7566 6345 2183 3051 10627 7566 6345 2183 3051 10627 7566 6345 2183 3051 10627 7566 6345 2183 3051 10627 7566 6345 2183 3051 10627 7566 6345 2183 3051 10627 7566 6345 2183 3051 10627 7566 6345 2183 3051 10627 7566 6345 2183 3051 10627 7566 6345 2183 3051 10627 7766 6345 2183 3051 10627 7766 6345 2183 3051 10627 7766 6345 2183 3051 10627 7766 6345 2762 7766 6345 7766 6345 77766 6345 7776 6345 77766 6345 77766 6345 77766 6345 77766 6345 77766 6345 77766 6345 77766 6345 77766 6345 7776 7766 7766 7766 7766 77766 7766 7766 7776 7766 7766 77766 7776 7766 7766 7766 7776 7766 7766 7776 7766 7776 7766 77766 7776 7766 7776 7766 7776 7766 7776 7766 7776 7776 7766 7776 7766 7776 7766 7776 7766 7776 7766 7776 7766 7776 7776 7766 7776 7776 7766 7776 7776 7767 7776 7767 7777 7767 7777 7767 7777 7767 7777 7767 77777 77777 7767 777777 7767 77777777	Core CDS V / 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	ar CDS           3070           4642           5238           843           4216           33311           10242           2545           7639           3351           10242           2545           3331           10242           5169           10627           7566           2183           8301           2695           9563           9563           8301           2695           7064           4170           2762           7064           6136           3488	Strain specific CDS 559 410 3575 766 4003 970 10001 2503 2988 4189 478 5979 853 5train specific CDS 5train specific CDS 712 723 933 103 2991 103 2991 103 2991 105 289 878 111 5965 886 886 886 886 886 886 886 8	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.102 97.647 98.35 33.115 68.492 16.914 95.435 21.589 Strain spe. CDS (%) Strain spe. CDS (%) 13.774 23.892 12.596 4.718 46.036 62.146 15.166 16.994 37.482 37	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
<u>genome on</u> CR8 bin, 3 CR8 bin, 21 CR8 bin, 21 CR8 bin, 24 UR bin, 24 ICU bin, 3 ICU bin, 3 ICU bin, 3 ICU bin, 3 ICU bin, 3 ICU bin, 3 F bin, 35 F bin, 31 F bin, 35 F bin, 35 F bin, 31 F bin, 35 F bin, 31 F bin, 35 F bin, 31 F bin, 35 F bin, 31 F bin, 35 F bin, 35 F bin, 35 F bin, 35 F plus bin, 40 F bin, 35 F plus bin, 14 F plus bin, 24 F pl	CDS         4           30700         4         2           30700         4         2           52288         8         6           3320         10242         5           3321         10242         2           3051         10242         5           3051         6         6           5169         6         6           5169         6         6           5169         6         6           5169         10627         7           7567         7         7           7062         2         2           41700         2         7           7064         2         2           3488         6         196           6         196         3488           6         196         1987	CDS (without artefact fam.) 30707 4642 5238 843 4216 3331 10242 2545 4216 3331 10242 2545 4256 4256 4256 4256 4256 4256 4256 4256 4256 4256 4256 4256 4256 4257 4256 4257 4256 4257 4256 4257 4256 4257 4256 4257 4256 4257 4256 4257 4256 4257 4257 4256 42577 42577 4257777 425777 425777777 42577777777777777	Pan CDS 30700 46422 5238 8433 4216 33311 10242 2545	Core CDS V (0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0	ar CDS 3070 4642 5238 843 4216 3331 10242 2545 3351 10242 2545 3551 10242 2545 3551 10242 2545 3551 10242 2545 3551 10242 2545 3551 10242 2545 3551 10242 2545 3551 10242 2545 3552 10247 2545 3510 10247 2545 3510 10247 2545 3510 10247 2545 3510 10247 2545 3510 10247 2545 3510 1027 27566 3431 2695 3454 3454 2762 3645	Strain specific CDS         S59           559         410           3575         766           4033         970           10001         2503           2598         4189           419         478           5979         853           712         2539           953         103           2921         5539           1620         488           1563         1620           488         1563           1263         2891           1458         1563           1458         1458	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%)           18.208         8.832           68.251         90.866           97.083         29.12           97.647         98.35           33.115         66.492           16.914         95.435           91.647         21.589           Strain spe. CDS (%)         13.774           23.892         12.586           4.036         62.146           19.516         19.974           13.974         23.892           12.596         4.718           40.036         62.146           19.516         19.494           37.482         10.463           11.429         4.995           97.213         25.688           8.862         29.236	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome on           CR8 bin, 21           UR bin, 24           UR bin, 24           UR bin, 24           UR bin, 24           UR bin, 32           UR bin, 31           ICU bin, 31           ICU bin, 34           genome bin           F bin, 13           F bin, 31           F bin, 35           F bin, 35           F bin, 35           F plus bin, 34           F plus P plus bin, 34           P plus bin, 34           P bin, 10           P bin, 14           P bin, 14           P bin, 15	CDS         0           30700         3           30700         4           25238         8           4642         2           5338         6           3320         6           616         6           6265         3951           funrestr         5           7639         9           9633         3645           2653         3951           10627         7           7567         7           7567         7           9563         30615           2635         30615           2636         30616           27064         4170           2702         6136           6199         3488           6196         3408           6196         3408           6196         3408           6196         3488           6196         3488           6196         3488           6196         3488           6196         3488           6196         3488           6196         3488           6196         5498	CDS (without artefact fam.) 3070 4642 5238 4741 3331 10242 2545 4741 3331 10242 2545 4763 4776 4776 4776 4776 4776 4776 4777 4776 4777 4776 47777 4777 4777 4777 4777 4777 4777 4777 4777 4777	Pan CDS 30700 4642 4216 42	Core CD5 V (2000) 0 0 0 0 0	ar CDS           3070           4642           5238           843           3331           10242           2545           3631           10242           2545           3631           10242           2545           3631           6265           3951           10627           7566           5169           10627           7566           2695           4170           2762           7064           22762           7064           22762           7064           2495           4170           3608           4197	Strain specific CDS S59 410 3575 766 4003 970 10001 2503 2988 4189 478 5979 853 547 547 547 547 543 103 2921 543 1053 2921 543 1053 2921 543 1055 8056 548 549 544 544 544 544 544 544 544	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 1000	Strain spe. CDS (%) 18.208 8.832 68.251 97.683 29.12 97.647 98.35 33.115 68.492 16.914 95.435 21.589 Strain spe. CDS (%) 13.774 23.892 12.596 4.718 46.036 62.146 13.464 13.464 14.455 16.994 37.482 10.463 12.429 4.955 37.688 8.862 29.236 8.862 29.236 14.246	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome on           CR8 bin, 3           CR8 bin, 32           CR8 bin, 32           UR bin, 24           UR bin, 31           UR bin, 32           UR bin, 31           ICU bin, 24           pan-genome on           genome bin           F bin, 35           F, bin, 36           F, bin, 37           F, bin, 38           F, plus bin, 34           F, plus bin, 35           L, plus bin, 54           L, plus bin, 54           L, plus bin, 54 <td>CDS         0           3070         0           3070         0           3070         0           4642         5338           4642         3322           3332         10242           2545         3332           6116         6265           3351         10627           7567         7567           7567         7567           7064         4100           20222         636           3488         3488           4170         2222           6366         3488           3488         6196           61926         4367           2544         3468           3488         2544</td> <td>CDS (without artefact fam.) 3070 4642 5238 4716 3331 10242 2545 4763 4775</td> <td>Pan CDS 3000 46422 5238 843 4216 3331 10242 2545 3651 10242 2545 3951 10242 2545 3951 10242 2545 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2155 3051 10242 2155 3051 10242 2155 3051 10242 2155 3051 1027 2155 3051 1027 2155 3051 1027 2155 3051 1027 2155 3051 1027 2155 3051 3051 3051 3052 3052 3055 3056 3055 3056 305</td> <td>Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td> <td>Arr CDS           3070           4642           5238           8333           4216           3331           10242           2548           2538           3311           10242           2563           3951           5169           10627           7556           2183           6345           9563           8301           2642           9563           8301           2642           9563           8301           2642           9563           8301           2642           9563           8301           2762           6136           9563           8301           2762           8488           6195           4987           4987           3608           2544           407**</td> <td>Strain specific CDS SS9 410 3575 766 4003 970 100011 2503 4189 4189 4189 478 5979 853 5979 853 5979 853 1000 100011 2508 478 478 478 478 478 478 478 47</td> <td>Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td> <td>Var CDS (%) 100 100 100 100 100 100 100 10</td> <td>Strain spe. CDS (%) 18.208 8.832 68.251 97.643 97.643 97.647 98.65 29.7647 97.647 98.657 97.647 97</td> <td>Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>	CDS         0           3070         0           3070         0           3070         0           4642         5338           4642         3322           3332         10242           2545         3332           6116         6265           3351         10627           7567         7567           7567         7567           7064         4100           20222         636           3488         3488           4170         2222           6366         3488           3488         6196           61926         4367           2544         3468           3488         2544	CDS (without artefact fam.) 3070 4642 5238 4716 3331 10242 2545 4763 4775	Pan CDS 3000 46422 5238 843 4216 3331 10242 2545 3651 10242 2545 3951 10242 2545 3951 10242 2545 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3951 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2154 3051 10242 2155 3051 10242 2155 3051 10242 2155 3051 10242 2155 3051 1027 2155 3051 1027 2155 3051 1027 2155 3051 1027 2155 3051 1027 2155 3051 3051 3051 3052 3052 3055 3056 3055 3056 305	Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Arr CDS           3070           4642           5238           8333           4216           3331           10242           2548           2538           3311           10242           2563           3951           5169           10627           7556           2183           6345           9563           8301           2642           9563           8301           2642           9563           8301           2642           9563           8301           2642           9563           8301           2762           6136           9563           8301           2762           8488           6195           4987           4987           3608           2544           407**	Strain specific CDS SS9 410 3575 766 4003 970 100011 2503 4189 4189 4189 478 5979 853 5979 853 5979 853 1000 100011 2508 478 478 478 478 478 478 478 47	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 97.643 97.643 97.647 98.65 29.7647 97.647 98.657 97.647 97	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome bin           CR8 bin, 31           CR8 bin, 21           CR8 bin, 21           CR8 bin, 24           UR bin, 24           pan-genome bin           F bin, 15           F bin, 31           F bin, 35           F plus bin, 35           F plus bin, 35           F plus bin, 35           F plus bin, 36           F plus bin, 37           F plus bin, 38           F plus bin, 30           F plus bin, 30           F plus bin, 31           F plus bin, 30           F plus bin, 31           Publis bin, 30           P plus bin, 31           Publis bin, 31           Publis bin, 31           Publis bin, 51           L plus bin, 51           L plus bin, 51           Publis bin, 51           Publis, 16           Publis, 16	CDS         0           3070         3           3070         4642           5338         846           4642         5338           3332         10242           2545         3332           6116         2266           3351         1           funrestr         6166           5169         10627           7567         7567           9563         8301           2022         6136           3483         8301           2702         2222           6136         3488           4987         3608           4987         3608           4987         3608	CDS (without artefact fam.) 3070 4642 5283 843 4216 3331 10242 2545 4216 3331 10242 2545 4256 4256 4256 4256 4256 4256 10627 7566 10627 7566 10627 7566 4170 2282 4170 2275 4170 2265 4170 4170 2265 4170 417	Pan CDS 30700 4642 5238 8433 4216 4217 4216 4216 4217 42	Core CDS V / 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Arr CDS           3070           4642           5238           8333           4216           3331           10242           2548           339           6116           2863           5169           5169           5169           2683           8301           26953           8301           26953           8301           2695           7064           21222           6136           4953           3488           6195           3488           2544           4971           3608           2544           3975	Strain specific CDS S59 410 3575 766 4003 970 10001 2503 2988 4189 478 5979 853 5172 2539 933 103 2991 103 2991 103 2991 115 2094 1620 488 111 5965 886 549 1458 811 1555 849 1458 845 845 845 845 845 845 845	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 97.647 97.647 98.35 33.115 68.492 15.944 95.435 21.589 Strain spe. CDS (%) Strain spe. CDS (%) 12.596 4.718 46.036 62.146 15.166 16.944 37.482 12.429 4.955 97.213 22.568 8.862 22.255 8.791 2.3788	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome on           CR8 bin, 31           CR8 bin, 21           CR8 bin, 21           CR8 bin, 24           UR bin, 24           ICU bin, 31           F bin, 35           F bin, 35           F bin, 31           F bin, 31           F bin, 35           F bin, 35           F bin, 35           F bin, 35           F bin, 36           F bin, 37           F bin, 38           F plus bin, 10           F plus bin, 13           F plus bin, 14           P bin, 12	CDS         0           30700         30700           30700         4642           5338         4642           5338         4642           5338         4642           4216         3322           3322         2545           4216         6265           5169         6265           5169         6166           10627         7567           7587         7583           6301         10627           7567         2183           6345         5436           4170         2262           6136         61366           61366         61366           61366         52544           4973         3088           2975         3264	CDS (without artefact fam.) 3070 4642 5283 843 4216 3331 10242 2545 4216 3331 10242 2545 4216 2826 6255 3351 10242 10242 10242 10242 10242 1025 1027 10677 10677 10677 10677 10677 10677 10647 106777 106777 106777 106777 106777 106777 106777 106777 106777 1067777 1067777 1067777 1067777 1067777 10677777 106777777 106777777777777777777777777777777777777	Pan CDS 30700 4642 5238 8433 4216 63331 10242 25454 7639 6116 6255 5169 7639 6116 6255 5169 7639 6126 5159 7639 6126 2183 63454 8301 2057 2183 63454 8301 2057 2183 63454 8301 2057 2183 63454 8301 2057 2183 63454 8301 2057 2183 8301 2057	Core CDS V (0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0	Arr CDS           3000           4642           5238           843           3331           10242           2545           7639           6166           62253           5169           9563           10627           7566           2183           6345           2762           28301           3088           4997           3608           2544           4971           2975           2281	Strain specific CDS         S59           357         359           410         3575           766         4033           4033         970           10011         2503           2533         970           10011         2503           2538         4189           478         5979           853         103           953         103           953         103           953         1620           953         1620           488         1563           828         1863           1111         5965           549         1488           544         514           545         437           708         54	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 99.686 97.083 29.12 97.647 98.35 33.115 68.492 16.914 95.435 21.589 Strain spe. CDS (%) 13.774 23.892 12.596 4.718 46.036 62.146 15.954 4.718 4.036 62.146 15.954 4.718 4.036 62.146 15.954 4.718 4.036 62.146 15.954 4.718 4.036 62.146 15.956 8.862 2.255 8.862 2.2568 8.862 2.2568 8.8791 2.2568 8.7911 2.2568 8.25711 2.2568 8.25711 2.2568 8.25711 2.2567 8.25711 2.2568 8.25711 2.2567	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
<u>genome bin</u> (R8 bin, 3 (R8 bin, 2 (R8 bin, 2 (R8 bin, 2 (R8 bin, 2 (R8 bin, 2 (R8 bin, 2 (R bin, 2 (C bin, 3 (C bin	CDS         Q           30700         4642           5338         4642           5338         846           4216         5332           3032         2545           3032         2545           3032         2545           3032         2616           6265         5169           5169         5169           5169         5169           7557         7557           7564         2222           6136         5169           2645         5169           36345         2264           3645         3645           27064         4388           6196         5488           6196         3488           6196         3488           6196         3488           6196         3488           6196         3488           6196         3488           6197         3608           2544         4973           2975         2181           3001         3272           3058         3497	CDS (without artefact fam.) 3070 4642 5238 4741 3331 10242 2545 4763 4773 4763 4763 4763 4763 4763 4763 4763 4763 4763 4763 4763 4763 4763 4763 4773 4773 4774	Pan CDS 30700 4642 5238 8433 4216 3331 10242 2545 4256 6265 5169 5175 5169 5169 5175 5169 5175 5169 5175 5169 5175 5169 5175 5169 5175 5169 5175 5169 5175 5169 5175 5169 5175 5	Core CDS V V 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Arr CDS           3000           4642           5238           843           10242           2545           7639           2545           3551           3616           2545           3551           3616           2545           3616           2545           3616           2545           3635           2183           6365           2563           9563           36365           26126           2183           83015           2563           36365           26136           3488           6195           3488           2544           4971           2975           2081           2094           4971           2075           2081           2095           2014           2025           2036           2037           2038           2039           2044 </td <td>Strain specific CDS S59 410 3575 766 4003 970 10001 2503 2988 4189 478 5979 853 5979 853 1020 2539 1020 478 5933 1020 478 5933 1020 478 878 1156 593 1020 458 878 1156 593 1020 458 1156 593 1020 458 1156 1595 593 1020 458 1156 1155</td> <td>Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td> <td>Var CDS (%) 100 100 100 100 100 100 100 10</td> <td>Strain spe. CDS (%) 18.208 8.832 68.251 97.683 29.12 97.647 98.35 33.115 68.492 16.914 95.435 21.589 Strain spe. CDS (%) 13.774 23.892 12.596 4.718 46.063 14.746 62.146</td> <td>Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>	Strain specific CDS S59 410 3575 766 4003 970 10001 2503 2988 4189 478 5979 853 5979 853 1020 2539 1020 478 5933 1020 478 5933 1020 478 878 1156 593 1020 458 878 1156 593 1020 458 1156 593 1020 458 1156 1595 593 1020 458 1156 1155	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 97.683 29.12 97.647 98.35 33.115 68.492 16.914 95.435 21.589 Strain spe. CDS (%) 13.774 23.892 12.596 4.718 46.063 14.746 62.146	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
<u>genome bin</u> CR8 bin, 3 CR8 bin, 21 CR8 bin, 21 CR8 bin, 21 CR8 bin, 21 CR8 bin, 24 UR bin, 24 ICU bin, 24 ICU bin, 24 F bin, 15 F bin, 24 F bin, 25 F bin, 24 F bin, 25 F bin, 24 F bin, 25 F bin, 24 F bin, 25 F bin, 20 F bin, 20 F bin, 10 F bin, 10 F bin, 10 F bin, 10 F bin, 10 F bin, 10 F bin, 12 F bin, 21 P bin, 12 P bin, 12 P bin, 12 P bin, 12 P bin, 13 P bin, 21 P bin, 21 P bin, 24 P bin, 25 P bin, 25 P bin, 25 P bin, 24 P bin, 25 P bin, 24 P bin, 25 P bin, 24 P bin, 25 P bin, 24 P bin, 25 P bin, 25 P bin, 24 P bin, 25 P bin, 24 P bin, 25 P bin, 24 P bin, 24 P bin, 24 P bin, 24 P bin, 24 P bin, 24 P bin, 25 P bin, 24 P bin, 24 P bin, 25 P bin, 24 P bin, 24	CDS         O           30700         30700           30701         30700           4642         5338           5332         5332           10242         2545           7639         6166           6265         3951           10627         2183           6365         5169           10627         7567           7567         7567           7567         7567           7567         2183           6345         9563           3048         6345           9563         3488           3448         2544           22625         7064           43975         2681           5080         2544           2975         2281           5080         2109	CDS (without artefact fam.) 3070 4642 5283 843 4412 5284 843 4416 10242 2545 4763 6165 6255 6265 6255 6265 6265 6265 6265 6265 6265 6265 6265 6265 6265 6265 6255 6265 6255 6265 6255 6265 62555 6255 6255 62555 62555 62555 625555 625555 625555555 6	Pan CDS 30700 4642 5238 8433 4642 5238 8433 4626 5439 10242 25454 6255 6255 5169 10627 7566 6265 5169 10627 77566 6345 8301 10627 77566 6345 8301 20762 7752 2083 8301 2095 4170 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 4197 77664 6195 7762 77664 6195 7762 77664 7775 7767 77664 7775 7767 77664 7775 7767 7767 7767 7767 7767 7767 7775 777	Core CDS V 2000 0000000000000000000000000000000	Arr CDS           3000           4642           5238           843           10242           2545           7639           6166           22545           33311           10242           2545           33951           10627           7566           10627           7566           9563           8301           2695           9563           8301           2695           9563           8301           2695           9563           9563           8301           2762           6136           4470           27064           4997           3608           2544           4971           2955           20612           5060           5021           5020           6121           5026           6121           5026           6121           5026           612	Strain specific CDS SS9 410 33575 7666 4003 970 100011 2503 4189 4189 4189 478 4189 478 478 5979 853 933 933 933 933 933 933 933 9	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 97.683 29.12 97.647 97.	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
genome on           CR8 bin, 32           CR8 bin, 21           UR bin, 24           Dan-genome bin           F bin, 15           F bin, 24           F bin, 24           F bin, 31           F bin, 35           F plus bin, 35           F plus bin, 35           F plus bin, 36           F plus bin, 37           P bin, 15           L plus bin, 30           P bin, 15           L plus bin, 31           P bin, 15           P bin, 21           P plus bin, 32           P plus bin, 31           P plus bin, 31           P bin, 33           P plus bin, 31           P bin, 15           L plus bin, 31           P bin, 33           P plus bin, 31           P bin, 32           P plus bin, 33           P plus bin, 31           P bin, 32 <t< td=""><td>CDS         O           30700         30700           30700         4642           5338         4642           5338         4642           5338         4642           2545         5763           6126         6265           6126         6265           5169         10627           7557         7557           7283         6345           41700         27624           2205         6136           4388         6146           4987         7064           4987         3608           4987         3608           4987         3608           2021         5244           4973         3268           2021         524           4021         5264           4021         5264           4021         5264           4021         5264           4021         5264           4021         5264           4021         5264           4021         5086           2109         5026</td><td>CDS (without artefact fam.) 3070 4642 5283 843 4216 3331 10242 2545 47633 6116 2626 6265 3351 10627 1064 10627 1067 1</td><td>Pan CDS 30700 4642 5238 8433 4216 4217 42</td><td>Core CDS V / 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td><td>Arr CDS         3000           30200         4642           34216         5238           843         343           10242         2545           7639         6116           2826         3351           5169         5169           10627         7566           2183         8301           2762         6365           9563         8301           2762         7646           9563         8301           2762         7664           9563         3488           91027         7666           2095         34100           3408         2695           34100         2762           2025         3608           3408         2554           4977         2281           3608         2109           3608         2109           3026         2162</td><td>Strain specific CDS SS9 410 3575 766 4003 970 10001 2503 2988 4189 478 5979 853 5172 2539 933 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2097 103 2097</td><td>Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td><td>Var CDS (%) 100 100 100 100 100 100 100 10</td><td>Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.102 97.647 98.35 33.115 68.492 15.944 95.435 21.589 Strain spe. CDS (%) Strain spe. CDS (%) 12.596 4.718 46.036 62.146 15.166 16.944 37.482 37.482 37.482 37.482 37.482 22.255 8.791 12.255 8.791 23.788 8.862 22.255 8.791 23.788 8.862 22.255 8.791 23.788 8.862 22.255 8.791 23.788 8.862 22.255 8.791 23.788 3.1931 3.277 7.889 3.1931 3.277 24.818</td><td>Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td></t<>	CDS         O           30700         30700           30700         4642           5338         4642           5338         4642           5338         4642           2545         5763           6126         6265           6126         6265           5169         10627           7557         7557           7283         6345           41700         27624           2205         6136           4388         6146           4987         7064           4987         3608           4987         3608           4987         3608           2021         5244           4973         3268           2021         524           4021         5264           4021         5264           4021         5264           4021         5264           4021         5264           4021         5264           4021         5264           4021         5086           2109         5026	CDS (without artefact fam.) 3070 4642 5283 843 4216 3331 10242 2545 47633 6116 2626 6265 3351 10627 1064 10627 1067 1	Pan CDS 30700 4642 5238 8433 4216 4217 42	Core CDS V / 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Arr CDS         3000           30200         4642           34216         5238           843         343           10242         2545           7639         6116           2826         3351           5169         5169           10627         7566           2183         8301           2762         6365           9563         8301           2762         7646           9563         8301           2762         7664           9563         3488           91027         7666           2095         34100           3408         2695           34100         2762           2025         3608           3408         2554           4977         2281           3608         2109           3608         2109           3026         2162	Strain specific CDS SS9 410 3575 766 4003 970 10001 2503 2988 4189 478 5979 853 5172 2539 933 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2991 103 103 2097 103 2097	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 90.866 97.083 29.102 97.647 98.35 33.115 68.492 15.944 95.435 21.589 Strain spe. CDS (%) Strain spe. CDS (%) 12.596 4.718 46.036 62.146 15.166 16.944 37.482 37.482 37.482 37.482 37.482 22.255 8.791 12.255 8.791 23.788 8.862 22.255 8.791 23.788 8.862 22.255 8.791 23.788 8.862 22.255 8.791 23.788 8.862 22.255 8.791 23.788 3.1931 3.277 7.889 3.1931 3.277 24.818	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Jerome on CR8 bin, 3 CR8 bin, 21 CR8 bin, 21 CR8 bin, 21 CR8 bin, 24 UR bin, 24 ICU bin, 3 ICU bin, 3 F bin, 33 F bin, 31 F bin, 32 F bin, 30 F bin, 32 F bin, 30 F bin, 32 F bin, 30 F bin, 14 F bin, 14 F bin, 14 F bin, 12 F plus bin, 24 F bin, 14 F bin, 14 F bin, 13 F plus bin, 24 F bin, 14 F bin, 14 F bin, 12 F plus bin, 32 F bin, 31 F bin, 14 F bin, 14 F bin, 13 F plus bin, 24 F bin, 31 F bin, 14 F bin, 14 F bin, 14 F bin, 15 F plus, 10 F bin, 14 F bin, 12 F plus bin, 32 F plus bin, 32	CDS         0           3070         3070           3070         3070           3070         3070           3070         3070           3070         3462           5338         346           4422         538           4126         245           7639         6116           5129         5136           10027         2183           5169         7567           7567         7567           7564         30301           2695         3130           2695         3130           2695         4170           2762         7368           4097         3488           6196         2544           4097         3608           44973         32675           20109         2222           16021         6026           3130         32675	CDS (without artefact fam.) 3070 4642 5283 843 4216 3331 10242 2545 4216 3331 10242 2545 4216 3351 10242 2545 4255	Pan CDS 30700 4642 5238 8433 4216 63331 10242 25454 7639 6116 6265 5169 7639 6116 6265 5169 7639 6126 210627 77566 6265 5169 7559 6126 2183 8301 2655 2183 2655 2183 2655 2183 2655 2183 2655 2183 2655 2183 2055 2184 2055 2184 2055 2184 2055 2184 2055 2184 2055 2184 2055 2184 2055 2184 2055 2195 2105 21	Core CDS V ( 10 - 10 - 10 - 10 - 10 - 10 - 10 - 10	Arr CDS           30700           4642           5238           843           843           10242           2545           7639           5169           10627           7566           2183           63351           7566           2183           6345           2655           2163           38301           2062           7064           2275           2636           3480           20702           7064           2072           2065           4100           20702           2014           20172           2018           6195           2014           20172           2016           3100           20172           2018           6021           30206           3100           3100	Strain specific CDS S59 410 3575 766 4003 970 10001 2533 4189 4189 478 478 5979 853 547 478 5973 5172 2539 478 478 5973 1020 478 478 478 478 478 478 478 478	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 99.687 99.687 99.637 99.647 98.35 21.598 5train spe. CDS (%) 13.774 23.892 12.596 4.718 46.036 62.146 15.546 16.944 15.956 4.718 4.036 62.146 15.956 4.718 4.036 62.146 15.956 4.718 4.036 62.146 15.956 4.718 4.036 62.146 15.956 8.862 20.256 8.862 20.256 8.862 20.256 8.8791 12.259 4.788 8.862 20.256 8.791 12.259 8.791 13.774 13.774 14.255 14.256 8.862 20.256 8.862 20.256 8.791 13.277 7.889 3.1931 3.272 2.367 3.1931 3.272 2.367 3.1931 3.272 2.368 3.1931 3.272 2.368 3.1931 3.272 2.368 3.1931 3.272 2.368 3.1931 3.272 2.368 3.1931 3.272 2.368 3.278 3.278 3.277 2.368 3.278 3.277 2.368 3.278 3.277 2.368 3.278 3.277 2.368 3.278 3.277 2.368 3.277 2.368 3.278 3.277 2.368 3.277 2.367 3.277 2.368 3.277 2.368 3.278 3.277 2.368 3.278 3.278 3.277 2.368 3.278 3.278 3.277 2.368 3.278 3.277 2.368 3.278 3.278 3.277 2.368 3.278 3.277 2.368 3.278 3.277 2.368 3.278 3.277 2.368 3.278 3.277 2.368 3.278 3.277 2.368 3.278 3.277 2.368 3.278 3.277 2.368 3.278 3.277 3.277 2.388 3.277 3.277 2.388 3.277	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Jenome on CR8 bin, 3 CR8 bin, 21 CR8 bin, 21 CR8 bin, 21 CR8 bin, 21 UR bin, 24 UR bin, 32 UR bin, 31 ICU bin, 3 ICU bin, 3 ICU bin, 3 ICU bin, 3 ICU bin, 3 ICU bin, 3 ICU bin, 3 F bin, 14 F bin, 18 F bin, 10 F	CDS         O           3070         3070           3070         3070           3070         3070           3070         3070           3070         3462           5238         846           4216         342           4332         332           2332         2545           7639         616           10627         2565           3551         3651           10627         7567           7567         7567           7567         7567           7564         5469           9563         8301           10627         7664           9563         8301           20752         7664           3265         3488           8301         2222           6136         3488           6196         4987           32076         2281           6120         3130           3281         3130           3282         3130	CDS (without artefact fam.) 3070 4642 5238 4741 3331 10242 2545 4763 4774	Pan CDS 30700 4642 5238 8433 4216 33311 10242 2545 4256 6265 5169	Core CDS V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Aar CDS 30700 4642 5238 8433 4216 3331 10242 2545 7639 6116 6265 3951 10242 2545 7639 6126 6265 3951 10242 2545 2183 6345 4170 2183 8301 2062 7766 6345 2183 8301 2062 7766 6345 2183 8301 2062 7766 6345 2183 8301 2062 7766 2183 8308 2192 2062 2193 2062 2193 2062 2193 2062 2193 2062 2193 2062 2193 2062 2193 2062 2193 2062 2193 2062 2193 2062 2193 2062 2193 2062 2193 2062 2193 2062 2193 2062 2076 2022 2061 2062 2076 2022 2061 2052	Strain specific CDS SS9 410 3575 766 4003 970 10001 2503 2988 4189 478 5979 853 5979 853 5979 853 1020 478 5933 1020 478 5933 1020 478 5933 1020 478 5933 1020 478 5933 1020 478 5943 1020 478 5943 1020 478 5943 1020 478 5943 1020 478 5943 1020 478 5943 1020 100	Core CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Var CDS (%) 100 100 100 100 100 100 100 10	Strain spe. CDS (%) 18.208 8.832 68.251 97.63 29.12 97.647 98.35 33.115 68.492 16.914 95.435 21.589 Strain spe. CDS (%) 54.35 21.589 47.08 47.08 47.08 46.2146 19.516 16.944 37.482 12.596 4.718 46.02 10.463 12.429 4.995 97.213 25.688 8.862 23.268 8.862 23.268 8.862 23.268 8.862 23.277 7.889 3.191 3.277 7.889 3.191 3.277 2.888 2.376 7.889 3.191 3.277 2.888 2.376 7.889 3.191 3.277 2.888 2.376 7.889 3.391 3.277 2.888 2.376 7.889 3.391 3.277 2.888 2.376 7.889 3.191 3.277 2.888 2.376 7.889 3.191 3.277 2.888 2.376 7.889 3.191 3.277 2.888 2.376 7.889 3.191 3.277 2.888 3.191 3.277 2.888 3.191 3.277 2.888 3.191 3.277 2.888 3.191 3.277 2.888 3.277 2.888 3.277 2.888 3.191 3.277 2.888 3.277 2.888 3.277 2.888 3.277 2.888 3.277 2.888 3.277 2.366 3.277 2.888 3.277 2.888 3.277 2.366 3.277 2.366 3.277 2.366 3.277 2.367 3.277 2.366 3.277 2.367 3.277 2.368 3.277 2.366 3.277 2.368 3.277 2.388 3.277 2.388 3.277 2.388 3.277 2.388 3.277 2.388 3.277 2.388 3.277 3.086 3.086 3.086 3.086 3.277 3.086 3.086 3.086 3.086 3.097 3.086 3.086 3.086 3.086 3.087 3.086 3.086 3.086 3.087 3.086 3.086 3.086 3.087 3.086	Excluded CDS (%) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

Supplementary Table 4: A list of cleaning and disinfection reagents applied for various surfaces and purposes in the sampled built environments.

						built environmen	π .				
	public buildings (L)	publi	c houses (P)	р	rivate houses (F)	intensive care unit (M	CU)	gowning ar	ea (UR)	cleanroor	n (CR)
floor	s mechanically (broom)	floors	natural soaps	floors	all-purpose cleaners		Desderman pure		isopropanol 70%		isopropanol 70%
							Desmanol pure	surfaces	JAMINAL PLUS	surfaces	JAMINAL PLUS
						hand design at an	Descoderm		KLERCIDE-CR		KLERCIDE-CR
						hand desinfection	Skinman Soft Protect	devices and products	isopropanol 70%		isopropanol 70%
							Sterilium classic pure			devices and products	vapor phase H <sub>2</sub> O <sub>2</sub>
							Sterillium LSG				autoclaving
							Kodan forte (colored)				-
						ckin anticontic	Kodan forte (colorless)				
						skiriantiseptic	Betaseptic				
							Betaisodona standardized solution				
						museral anticentic	Octenisept solution				
						indcosai antiseptic	Betaisodona standardized solution				
						surface desinfection	Incidin Plus				
						rapid desinfection (containing alcohol)	Incidin Liquid				
						rapid desinfection (nonalcobolic)	Acryl-des desinfection tissues				
						Taple desinection (nonaconone)	Acrylan				
							Gigasept Instru AF				
						desinfectant for instruments (manually preparation)	Sekusept plus				
1							Sekusept active				

Supplementary Table 5: A list of cleaning and disinfection reagents including the exposure time applied for certain cases in the ICU at the state hospital in Graz, Austria.

	range of application		
disease / germ	hand desinfection	surface de	esinfection
	hand desiniection	rapid desinfection	routine desinfection
Norovirus	Desderman pure (30 sec)	Incidin liquid (10 min)	Incidin active 1% (1 br)
Norovinus	Desmanol pure (30 sec)	incluin inquiti (10 min)	meran active 1/0 (1111)
Adenovirus	Desderman pure (30 sec)	Incidin liquid (10 min)	Incidin plus 1% (1 hr)
	Desderman pure (30 sec)		
Rotavirus	Desmanol pure (30 sec)	Incidin liquid (5 min)	Incidin plus 0.5% (1 hr)
	Descoderm (30 sec)		
Clostridium difficile	standard desinfectants according to cleaning and desinfection protocols	Incidin active 2% (15 min)	Incidin active 1% (30 min)
gas gangrene	standard desinfectants according to cleaning and desinfection p	rotocols (in-depth cleaning)	(
tuberculosis	standard desinfectants applied 2 times according to cleaning and desinfection protocols	Incidin liquid (5 min)	Incidin plus 0.5% (1 hr)
MRE (MRSA, ESBL etc.)			
hepatitides, HIV			
influenza A, B			
measles, mumps			
Meningococci	standard desinfectants according to cleaning and desir	nfection protocols	
pertussis			
Salmonellae			
SARS			
varicella			

Supplementary Table 6: Complete list of all primers used in the study.

primer name	target region	application	primer sequence
515f		16S amplicons	GTGCCAGCMGCCGCGGTAA
806r	V4 region of the 165 rPNA gene	105 amplicons	GGACTACHVGGGTWTCTAAT
515F	v4 region of the 103 rank gene	aDCD	GTGCCAGCAGCCGC
927R		yr Ch	CCCGTCAATTYMTTTGAGTT

samples	sequences (per each read)	sequence length	%GC	read	Phred Score (average quality per read; seq/q)					
CPS	6 71E±07		11	1	>2.5E+07/37					
CNO	0.712+07			2	>2.0E+07/37					
LIR	5 35E+07		/18	1	>1.8E+07/37					
ON	5.552+07		40	2	>1.4E+07/37					
ICU	3 30E+07		12	1	>1.6E+07/37					
	5.502+07		42	2	>1.4E+07/37					
	3 03E+07		60	1	>8.0E+06/36					
L .	5.05L+07		00	2	>8.0E+06/36					
L plus	3 17E+07	50-150	62	1	>9.0E+06/36					
L_pius	5.172.07	50-150	02	2	>8.0E+06/36					
D	3 335702		60	1	>1.0E+07/36					
r	J.J2L+07							00	2	>7.0E+06/36
P nlus	3 28F+07		61	1	>9.0E+06/36					
r_plus	5.201+07		60	2	>7.0E+06/36					
F	3 97F±07		45	1	>1.4E+07/37					
•	5.572+07		-J	2	>1.2E+07/37					
E plus	3 605+07		16	1	>1.2E+07/37					
r_plus	5.09ET07		40	2	>1.2E+07/37					

Supplementary Table 7: Summary on all quality reads of the shotgun metagenomics dataset.

Supplementary Table 8: Summary of all contigs and scaffolds after assembly of the shotgun metagenomics dataset.

Contigs >= 100 nt	Number	<b>Total length</b>	Average	N50	Median	Largest
CR8	8603965	2010442497	233	251	173	88313
UR	6781605	1285900173	189	169	148	77951
ICU	5846714	1194191177	204	204	162	152957
L	3719862	616578790	165	144	140	112186
L_plus	3595693	606816993	168	146	141	192834
Р	3609403	583352985	161	144	140	260735
P_plus	3529598	569566003	161	145	140	260738
F	3013516	528935438	175	143	138	128864
F_plus	3103321	529882363	170	142	138	98570
Contigs >= 500 nt	Number	Total length	Average	N50	Median	Largest
ICU	98994	73962331	747	625	574	152957
CR8	468389	338833151	723	667	613	88313
UR	103191	129616359	1256	1523	713	77951
L_plus	50525	64061830	1267	1449	771	192834
Р	31304	47448008	1515	2299	766	260735
P_plus	29900	44378695	1484	2206	748	260738
F	59104	91192370	1542	2084	870	128864
F_plus	54071	82752023	1530	2032	867	98570
L	50240	60661672	1207	1324	750	112186
		1				1
Scaffolds >= 100 nt	Number	Total length	Average	N50	Median	Largest
Scaffolds >= 100 nt ICU	<b>Number</b> 5846470	<b>Total length</b> 1194230266	Average 204	<b>N50</b> 204	Median 162	Largest 152957
Scaffolds >= 100 nt ICU CR8	Number 5846470 8602774	<b>Total length</b> 1194230266 2010603824	<b>Average</b> 204 233	<b>N50</b> 204 251	<b>Median</b> 162 173	Largest 152957 88313
Scaffolds >= 100 nt ICU CR8 UR	Number 5846470 8602774 6779560	<b>Total length</b> 1194230266 2010603824 1286163090	Average 204 233 189	<b>N50</b> 204 251 169	<b>Median</b> 162 173 148	Largest 152957 88313 101382
<u>Scaffolds &gt;= 100 nt</u> ICU CR8 UR L_plus	Number 5846470 8602774 6779560 3594500	Total length 1194230266 2010603824 1286163090 606967359	Average 204 233 189 168	<b>N50</b> 204 251 169 146	<b>Median</b> 162 173 148 141	Largest 152957 88313 101382 258085
Scaffolds >= 100 nt ICU CR8 UR L_plus P	Number 5846470 8602774 6779560 3594500 3608497	Total length 1194230266 2010603824 1286163090 606967359 583483719	Average 204 233 189 168 161	<b>N50</b> 204 251 169 146 144	<b>Median</b> 162 173 148 141 140	Largest 152957 88313 101382 258085 260735
Scaffolds >= 100 nt ICU CR8 UR L_plus P P_plus	Number 5846470 8602774 6779560 3594500 3608497 3528733	Total length 1194230266 2010603824 1286163090 606967359 583483719 569693266	Average 204 233 189 168 161 161	<b>N50</b> 204 251 169 146 144 145	Median 162 173 148 141 140 140	Largest 152957 88313 101382 258085 260735 260738
Scaffolds >= 100 nt ICU CR8 UR L_plus P P_plus F	Number 5846470 8602774 6779560 3594500 3608497 3528733 3011212	Total length 1194230266 2010603824 1286163090 606967359 583483719 569693266 529221167	Average 204 233 189 168 161 161 175	<b>N50</b> 204 251 169 146 144 145 143	Median 162 173 148 141 140 140 138	Largest 152957 88313 101382 258085 260735 260738 128864
Scaffolds >= 100 nt ICU CR8 UR L_plus P P_plus F F_plus	Number 5846470 8602774 6779560 3594500 3608497 3528733 3011212 3101251	Total length 1194230266 2010603824 1286163090 606967359 583483719 569693266 529221167 530124777	Average 204 233 189 168 161 161 161 175 170	N50 204 251 169 146 144 145 143 142	Median 162 173 148 141 140 140 138 138	Largest 152957 88313 101382 258085 260735 260738 128864 98570
Scaffolds >= 100 nt ICU CR8 UR L_plus P P_plus F F_plus L	Number 5846470 8602774 6779560 3594500 3608497 3528733 3011212 3101251 3718903	Total length 1194230266 2010603824 1286163090 606967359 583483719 569693266 529221167 530124777 616698401	Average 204 233 189 168 161 161 175 170 165	<b>N50</b> 204 251 169 146 144 145 143 142 144	Median 162 173 148 141 140 140 138 138 138 140	Largest 152957 88313 101382 258085 260735 260738 128864 98570 155114
Scaffolds >= 100 nt ICU CR8 UR L_plus P P_plus F F_plus L	Number 5846470 8602774 6779560 3594500 3608497 3528733 3011212 3101251 3718903	Total length           1194230266           2010603824           1286163090           606967359           583483719           569693266           529221167           530124777           616698401	Average 204 233 189 168 161 161 175 170 165	<b>N50</b> 204 251 169 146 144 145 143 142 144	Median 162 173 148 141 140 140 138 138 140	Largest 152957 88313 101382 258085 260735 260735 128864 98570 155114
Scaffolds >= 100 nt ICU CR8 UR L_plus P P_plus F F_plus L	Number 5846470 8602774 6779560 3594500 3608497 3528733 3011212 3101251 3718903	Total length 1194230266 2010603824 1286163090 606967359 583483719 569693266 529221167 530124777 616698401	Average 204 233 189 168 161 161 175 170 165	<b>N50</b> 204 251 169 146 144 145 143 142 144	Median 162 173 148 141 140 140 138 138 140	Largest 152957 88313 101382 258085 260735 260738 128864 98570 155114
Scaffolds >= 100 nt ICU CR8 UR L_plus P P_plus F F_plus L Scaffolds >= 500 nt	Number           5846470           8602774           6779560           3594500           3608497           3528733           3011212           3101251           3718903           Number	Total length 1194230266 2010603824 1286163090 606967359 583483719 569693266 529221167 530124777 616698401 Total length	Average 204 233 189 168 161 161 175 170 165 Average	<b>N50</b> 204 251 169 146 144 145 143 142 144 <b>N50</b>	Median 162 173 148 141 140 140 138 138 140 Median	Largest 152957 88313 101382 258085 260735 260738 128864 98570 155114 Largest
Scaffolds >= 100 nt ICU CR8 UR L_plus P P_plus F F_plus L Scaffolds >= 500 nt ICU	Number           5846470           8602774           6779560           3594500           3608497           3528733           3011212           3101251           3718903           Number           98764	Total length 1194230266 2010603824 1286163090 606967359 583483719 569693266 529221167 530124777 616698401 Total length 74008162	Average 204 233 189 168 161 175 170 165 <b>Average</b> 749	<b>N50</b> 204 251 169 146 144 145 143 142 144 <b>N50</b> 626	Median 162 173 148 141 140 138 138 140 <b>Median</b> 574	Largest 152957 88313 101382 258085 260735 260738 128864 98570 155114 Largest 152957
Scaffolds >= 100 nt ICU CR8 UR L_plus P P_plus F F_plus L Scaffolds >= 500 nt ICU CR8	Number           5846470           8602774           6779560           3594500           3608497           3011212           3011251           3718903           Number           98764           467522	Total length 1194230266 2010603824 1286163090 606967359 583483719 569693266 529221167 530124777 616698401 Total length 74008162 339066479	Average 204 233 189 168 161 175 170 165 <b>Average</b> 749 725	<b>N50</b> 204 251 169 146 144 145 143 142 144 <b>N50</b> 626 668	Median 162 173 148 141 140 138 138 140 <b>Median</b> 574 612	Largest 152957 88313 101382 258085 260735 12864 98570 155114 Largest 152957 88313
Scaffolds >= 100 nt ICU CR8 UR L_plus P P_plus F F_plus L Scaffolds >= 500 nt ICU CR8 UR	Number           5846470           8602774           6779560           3594500           3608497           3528733           3011212           3101251           3718903           Number           98764           467352           101444	Total length 1194230266 2010603824 1286163090 606967359 583483719 569693266 529221167 530124777 616698401 Total length 74008162 339066479 130012967	Average 204 233 189 168 161 175 170 165 <b>Average</b> 749 725 1281	<b>N50</b> 204 251 169 146 144 145 143 142 144 <b>N50</b> 626 668 1592	Median 162 173 148 141 140 138 138 140 <b>Median</b> 574 612 711	Largest 152957 88313 101382 258085 260735 12864 98570 155114 Largest 152957 88313 101382
Scaffolds >= 100 nt ICU CR8 UR L_plus P P_plus F F_plus L Scaffolds >= 500 nt ICU CR8 UR L_plus	Number           5846470           8602774           6779560           3594500           3608497           3528733           3011212           3101251           3718903           Number           98764           467352           101444           49546           324754	Total length 1194230266 2010603824 1286163090 606967359 583483719 569693266 529221167 530124777 616698401 Total length 74008162 339066479 130012967 64308268	Average 204 233 189 168 161 175 170 165 <b>Average</b> 749 725 1281 1297	<b>N50</b> 204 251 169 146 144 145 143 142 144 <b>N50</b> 626 668 1592 1520	Median 162 173 148 141 140 138 138 140 <b>Median</b> 574 612 711 772	Largest 152957 88313 101382 258085 260735 12864 98570 155114 Largest 152957 88313 101382 258085 258085
Scaffolds >= 100 nt ICU CR8 UR L_plus P P_plus F F_plus L Scaffolds >= 500 nt ICU CR8 UR L_plus P	Number           5846470           8602774           6779560           3594500           3608497           3528733           3011212           3101251           3718903           Number           98764           467352           101444           49546           30472	Total length 1194230266 2010603824 1286163090 606967359 583483719 569693266 529221167 530124777 616698401 Total length 74008162 339066479 130012967 64308268 47613011	Average 204 233 189 168 161 175 170 165 <b>Average</b> 749 725 1281 1297 1562	<b>N50</b> 204 251 169 146 144 145 143 142 144 <b>N50</b> 626 668 1592 1520 2510	Median 162 173 148 141 140 138 138 140 <b>Median</b> 574 612 711 772 763	Largest 152957 88313 101382 258085 260735 12864 98570 155114 155114 Largest 152957 88313 101382 258085 260735
Scaffolds >= 100 nt ICU CR8 UR L_plus P P_plus F F_plus L Scaffolds >= 500 nt ICU CR8 UR L_plus P P_plus	Number           5846470           8602774           6779560           3594500           3608497           3528733           3011212           3101251           3718903           Number           98764           467352           101444           49546           30472           29106	Total length 1194230266 2010603824 1286163090 606967359 583483719 569693266 529221167 530124777 616698401 Total length 74008162 339066479 130012967 64308268 47613011 44538821	Average 204 233 189 168 161 175 170 165 <b>Average</b> 749 725 1281 1297 1562 1530	N50           204           251           169           144           145           143           142           144           150           626           668           1592           1520           2510           2400	Median 162 173 148 141 140 138 138 140 <b>Median</b> 574 612 711 772 763 744	Largest 152957 88313 101382 258085 260735 12864 98570 155114 152957 88313 101382 258085 260735 260738
Scaffolds >= 100 nt ICU CR8 UR L_plus P P_plus F F_plus L Scaffolds >= 500 nt ICU CR8 UR L_plus P P_plus F	Number           5846470           8602774           6779560           3594500           3608497           3528733           3011212           3101251           3718903           Number           98764           467352           101444           49546           30472           29106           57173	Total length 1194230266 2010603824 1286163090 606967359 583483719 569693266 529221167 530124777 616698401 Total length 74008162 339066479 130012967 64308268 47613011 44538821 91638271	Average 204 233 189 168 161 161 175 170 165 749 749 725 1281 1297 1562 1530 1602	<b>N50</b> 204 251 169 146 144 145 143 142 144 <b>N50</b> 626 668 1592 1520 2510 2402 2264	Median 162 173 148 141 140 138 138 140 <b>Median</b> 574 612 711 772 763 744 872	Largest 152957 88313 101382 258085 260735 12864 98570 155114 152957 88313 101382 258085 260735 260738 128864
Scaffolds >= 100 nt ICU CR8 UR L_plus P P_plus F F_plus L Scaffolds >= 500 nt ICU CR8 UR L_plus P P_plus F F_plus F F_plus	Number           5846470           8602774           6779560           3594500           3608497           3528733           3011212           3101251           3718903 <b>Number</b> 98764           467352           101444           49546           30472           29106           57173           52403	Total length 1194230266 2010603824 1286163090 606967359 583483719 569693266 529221167 530124777 616698401 Total length 74008162 339066479 130012967 64308268 47613011 44538821 91638271 83164916	Average 204 233 189 168 161 161 175 170 165 <b>Average</b> 749 725 1281 1297 1562 1530 1602 1530	<b>N50</b> 204 251 169 146 144 145 143 142 144 <b>N50</b> 626 668 1592 1520 2510 2402 2264 2194	Median 162 173 148 141 140 140 138 138 140 Median 574 612 711 772 763 744 872 871	Largest 152957 88313 101382 258085 260735 260738 128864 98570 155114 152957 88313 101382 258085 260735 260738 128864 98570

software tool	application	settings
		MIN_LENGTH=50, TAILCROP=0, HEADCROP=0,
Trimmomatic-0.32	data filtering	ILLUMINACLIP:TruSeq3-PE-
		2.fa:2:30:10:4:true, SLIDINGWINDOW: 5:20
		files: paired_reads=yes,
		max_matches_per_read=100; GI Maps:
		gi_taxid-March2015X.bin, gi2kegg-
MEGAN 5	single read visualization	Feb2015X.bin, gi2seed.map; LCA Params:
		min_score=50, max_expected=0.01,
		top_percent=10, min_support_percent=0,
		min_support=1, embed_in_main_file
Dev Mate 2.2.1		K=31, MINLEN=1000, MINCOV=5,
Ray Meta-2.3.1	assembly	READLEN=150
		clusters=400, epsilon=1e-06, iterations=500,
	hinning	kmer_length=4, length_threshold=1000,
CUNCUCI-0.4.0	ginning	pca_components=0.9, read_length=100,
		<pre>seed=1, total_percentage_pca=90</pre>
MaxBin-1.4.2	binning	min_contig_length=1000, marker genes 107,
		support factor= $(2/3.0)$ min confidence=0.8
		MIN_MARKERPERCENTAGE=0.2
Amphora2	taxonomic assignment	NIM BACT MARKERS=31
		NUM ARCH MARKERS=104
		MASH sketch=1000, P ani=0.9,
		S algorithm='ANIn', S ani=0.99,
		SkipMash=False, SkipSecondary=False,
		cent index=None. clusterAlg='average'.
dRep-0.5.7	genome comparisons	cov thresh=0.1. coverage method='larger'.
<b>p</b>	8	dry=False. n PRESET='normal'.
		operation='compare_wf'.overwrite=False.
		processors=6 rup tax=Ealse warn aln=0.25
		warn dist=0.25 warn sim=0.98
		wam_ast=0.25, wam_shii=0.50
		Line rate= 6, Lines per side= 1, Offset rate= 4,
		FTable chars= 10, Strings= unpacked, len
iRep-1.1	replication rates of genomes	divisor= 4, Difference-cover sample period=
		1024, Endianness= little, Actual local
		endianness= little, Random seed= 0, Sizeofs=
		void*=8, int=4, long=8, size_t=8

Supplementary Table 9: Settings for selected bioinformatic tools.

# Supplementary Table 10: Summary of applied statistics on the 16S rRNA gene amplicon dataset.

#### summary of applied statistics (16S rRNA gene amplicons)

Test	grouping category	metric	significance	details
	environmental confinement		0.001	A: 0.1618, delta 0.4224 of 0.504
MDDD	sampling location / surface	weighted unifrac	0.001	A: 0.1583, delta 0.4242 of 0.504
IVINPP	environmental confinement	unwoighted unifrag	0.001	A: 0.1392, delta 0.6759 of 0.7851
	sampling location	unweighten unmac	0.001	A: 0.1252, delta 0.6868 of 0.7851
	environmental confinement		0.001	R2: 0.30668
adonic	type of built environment		0.001	R2: 0.51265
auonis	sampling location / surface		0.001	R2: 0.30079
	surface material		0.001	R2: 0.41641
	environmental confinement		0.001	R: 0.633738272592444
	type of built environment		0.001	R: 0.607909664422414
ANOSIN	sampling location / surface		0.001	R: 0.603552220205951
	surface material		0.001	R: 0.380902465808126
two-sided	environmental confinement		distances	controlled vs. moderate controlled 0.6 > moderate controlled vs. uncontrolled 0.59 > controlled vs. Uncontrolled 0.4
two sample	type of built environment		(maximum to	ICU_floor_vspublic_buildings_wp 0.65; public_house_vspublic_buildings_wp 0.36
Student's t-	sampling location / surface	weighted unifrac	minimum)	table_vsfloor 0.6; device_vstable 0.37
tests	surface material	weighted dimite		metal_polymer_vsconcrete_tiles 0.68; polymer_tiles_vsstone 0.21
	humidity (H)			
	temperature (°C)			
	room volume (vol)			
	surface area (surf)			
Mantel	room area (room)			
correlogram	surface material (m)			
	geographic latitude (la)			
	geographic longitude (lo)			
	sea level (sl)			
	room height (h)			
	environmental variables			
	sample		0.9773	
	sample, lo		0.9586	
	sample, la, lo		0.9705	
	sample, la, lo, sl		0.9425	
BEST	sample, vol, la, lo, sl	Bray-Curtis	0.9043	
(BioEnv)	sample, vol, H, Ia, Io, sl	2.2, 64.45	0.8846	
	sample, surf, vol, H, la, lo, sl		0.8659	
	sample, surf, h, vol, H, la, lo, sl		0.8355	
	sample, surf, room, h, vol, H, la, lo, sl		0.8044	
	sample, surf, room, h, vol, H, C, la, lo, sl		0.7738	
	sample, surf, room, h, vol, m, H, C, la, lo, sl		0.7518	

Supplementary Table 11: Summary of applied statistics on predicted functions from the 16S rRNA gene amplicon dataset with PICRUSt.

#### summary of applied statistics (PICRUSt predictions of 16S rRNA gene amplicons)

Test	grouping category	metric	significance	details
MRPP	environmental confinement		0.001	A: 0.181, delta 0.06969 of 0.08509
	type of built environment		0.001	A: 0.3074, delta 0.05893 of 0.08509
	sampling location / surface		0.001	A: 0.2373, delta 0.0649 of 0.08509
	surface material		0.001	A: 0.2896, delta 0.06045 of 0.08509
adonis	environmental confinement		0.001	R2: 0.32734
	type of built environment		0.001	R2: 0.55937
	sampling location / surface		0.001	R2: 0.43945
	surface material		0.001	R2: 0.51541
	environmental confinement		0.001	R: 0.482880658436213
	type of built environment		0.001	R: 0.508214886053013
ANUSIN	sampling location / surface		0.001	R: 0.740354908750047
	surface material		0.001	R: 0.395628422683383
two-sided	environmental confinement		distances	moderate controlled vs. uncontrolled 0.1 > moderate controlled vs. controlled 0.1 > controlled vs. uncontrolled 0.06
two sample	type of built environment		uistances	ICU_workplace_vsprivate_house 0.14; gowning_area_vspublic_house 0.04
Student's t-	sampling location / surface		(maximum to	table_vsfloor 0.12; device_vstable 0.06
tests	surface material		minimum	polymer_tiles_vsfurnished_wood 0.15; polymer_tiles_vsstone 0.04
	room volume (vol)	Bray-Curtis		
	humidity (H)			
	temperature (°C)			
	room area (room)			
Mantel	surface area (surf)			
correlogram	surface material (m)			
	geographic latitude (la)			
	geographic longitude (lo)			
	sea level (sl)			
	room height (h)			
	environmental variables			
	sample		0.8918	
BEST (BioEnv)	sample, vol		0.8996	
	sample, h, vol		0.8847	
	sample, surf, h, vol		0.8645	
	sample, h, vol, la, lo		0.8448	
	sample, surf, h, vol, la, lo		0.835	
	sample, surf, h, vol, la, lo, sl		0.8123	
	sample, surf, h, vol, H, Ia, Io, sI		0.7924	
	sample, surf, room, h, vol, H, Ia, Io, sl		0.7543	
	sample, surf, room, h, vol, m, H, Ia, Io, sl		0.7262	
	sample, surf, room, h, vol, m, H, C, Ia, Io, s		0.7024	

# Supplementary Table 12: Read statistics of the 16S rRNA gene amplicon dataset.

16S rRNA gene amplicons					
Number of samples	61				
Number of OTUs (operational taxonomic units)	10814				
Number of sequences	837216				

Number of sequences per sample		Number of sequences per sample	
L5	225	ICU.d.14	10228
L1.PMA	254	ICU.d.10	10305
ICU.f.1	2052	ICU.w.19	10504
ICU.f.5	2888	F3.PMA	10846
CR.7a	2928	ICU.f.3	10877
ICU.f.4	4424	L1	10943
L8.PMA	4576	ICU.d.16	11185
F2.PMA	5227	ICU.d.12	11408
ICU.d.6	5308	L6.PMA	12149
F2	6444	ICU.w.21	13147
ICU.d.11	6635	P3.PMA	13454
ICU.w.23	6778	ICU.w.20	14835
ICU.w.18	7057	L3	14896
ICU.f.2	7168	L9	15899
UR.7	7255	L7.PMA	16261
L3.PMA	7307	P1.PMA	18562
L4.PMA	7543	CR.7b	18888
ICU.w.17	7721	P2.PMA	19242
ICU.d.15	7841	L7	19359
P3	7920	L6	21036
L5.PMA	7965	P1	21321
ICU.d.9	8088	L4	24486
ICU.w.22	8221	F1	25455
ICU.d.13	8446	F1.PMA	25670
F3	8616	L2	26425
ICU.d.7	9243	UR	26913
ICU.w.24	9627	P2	27974
ICU.d.8	9628	L8	28344
		FO	28566
		L9.PMA	30456
		F0.PMA	35436

CR.8

L2.PMA

summary of reads	
Minimum	225
Maximum	37831
Median	10504
Mean	13724.852
Std. dev.	9289.725

36930

37831

#### **Supplementary References**

- 1. Oberauner, L. *et al.* The ignored diversity: complex bacterial communities in intensive care units revealed by 16S pyrosequencing. *Sci. Rep.* **3**, 1413 (2013).
- 2. Koskinen, K. *et al.* Microbial biodiversity assessment of the European Space Agency's ExoMars 2016 mission. *Microbiome* **5**, 143 (2017).
- Sinha, R. *et al.* Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nat. Biotechnol.* 35, 1077– 1086 (2017).
- 4. Adams, R. I., Bateman, A. C., Bik, H. M. & Meadow, J. F. Microbiota of the indoor environment: a meta-analysis. *Microbiome* **3**, 49 (2015).
- 5. Chase, J. *et al.* Geography and Location Are the Primary Drivers of Office Microbiome Composition. *mSystems* **1**, 1–18 (2016).
- Nicholson, W. L., Schuerger, A. C. & Race, M. S. Migrating microbes and planetary protection. *Trends Microbiol.* 17, 389–392 (2009).
- Rummel, J. D., Stabekis, P. D., Devincenzi, D. L. & Barengoltz, J. B. COSPAR'S planetary protection policy: A consolidated draft. *Adv. Sp. Res.* 30, 1567–1571 (2002).
- 8. Rozov, R. *et al.* Recycler: an algorithm for detecting plasmids from *de novo* assembly graphs. *Bioinformatics* btw651 (2016). doi:10.1093/bioinformatics/btw651
- Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* 19, 455–477 (2012).
- Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–95 (2010).
- Zhu, W., Lomsadze, A. & Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* 38, 1–15 (2010).
- 12. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
- Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30 (2000).
- Bairoch, A. *et al.* The Universal Protein Resource (UniProt). *Nucleic Acids Res.* 33, 154–159 (2005).

- 15. Jia, B. *et al.* CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).
- Vallenet, D. *et al.* MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res.* 34, 53–65 (2006).
- 17. Chen, L., Zheng, D., Liu, B., Yang, J. & Jin, Q. VFDB 2016: Hierarchical and refined dataset for big data analysis 10 years on. *Nucleic Acids Res.* **44**, D694–D697 (2016).
- Joensen, K. G. *et al.* Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *J. Clin. Microbiol.* 52, 1501–1510 (2014).
- 19. Miele, V., Penel, S. & Duret, L. Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics* **12**, 116 (2011).
- Cury, J., Jové, T., Touchon, M., Néron, B. & Rocha, E. P. Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic Acids Res.* 44, 4539–4550 (2016).
- Vernikos, G. S. & Parkhill, J. Interpolated variable order motifs for identification of horizontally acquired DNA: Revisiting the Salmonella pathogenicity islands. *Bioinformatics* 22, 2196–2203 (2006).
- 22. Waack, S. *et al.* Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* **7**, 1–12 (2006).
- 23. Caporaso, J. G. *et al.* QIIME allows analysis of high- throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
- 24. Bolyen, E. *et al.* QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ Prepr.* (2018). doi:10.7287/peerj.preprints.27295v1
- Schwendner, P. *et al.* Preparing for the manned Mars journey: Microbiome dynamics in the confined Mars500 habitat during simulated Mars flight and landing. *Microbiome, Submitt.* 5, 129 (2017).
- Mahnert, A., Haratani, M., Schmuck, M. & Berg, G. Enriching beneficial microbial diversity of indoor plants and their surrounding built environment with biostimulants. *Front. Microbiol.* (2018). doi:10.3389/fmicb.2018.02985
- Schloss, P. D. *et al.* Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).

- 28. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–1 (2010).
- Langille, M. G. I. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–21 (2013).
- Ward, T. *et al.* BugBase Predicts Organism Level Microbiome Phenotypes. *bioRxiv* 1–19 (2017). doi:10.1101/133462
- 31. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–504 (2003).
- Moissl-Eichinger, C. *et al.* Quo vadis? Microbial profiling revealed strong effects of cleanroom maintenance and routes of contamination in indoor environments. *Sci. Rep.* 5, 9156 (2015).
- 33. Bokulich, N. A. *et al.* q2-sample-classifier: machine-learning tools for microbiome classification and regression. *bioRxiv* (2018). doi:10.1101/306167
- Clarke, K. R. & Ainsworth, M. A method of linking multivariate community structure to environmental variables. *Mar. Ecol. Prog. Ser.* 92, 205–219 (1993).
- 35. Morgan, X. C. *et al.* Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* **13**, R79 (2012).
- Morton, J. T. *et al.* Balance Trees Reveal Microbial Niche Differentiation. *mSystems* 2, 1–11 (2017).
- Bokulich, N. A. *et al.* q2-longitudinal : Longitudinal and Paired-Sample Analyses of Microbiome Data. *mSystems* 3, 1–9 (2018).