nature protocols

https://doi.org/10.1038/s41596-023-00949-x

Protocol

Check for updates

Improved sampling and DNA extraction procedures for microbiome analysis in food-processing environments

Coral Barcenilla[1,16], José F. Cobo-Díaz[1,16], Francesca De Filippis[2,3], Vincenzo Valentino[2], Raul Cabrera Rubio[4], Dominic O'Neil[5], Lisa Mahler de Sanchez[5], Federica Armanini[6], Niccolò Carlino[6], Aitor Blanco-Míguez[6], Federica Pinto[6], Inés Calvete-Torre[7,8], Carlos Sabater[7,8], Susana Delgado[7,8], Patricia Ruas-Madiedo[7,8], Narciso M. Quijada[9,10,11], Monika Dzieciol[10], Sigurlaug Skírnisdóttir[12], Stephen Knobloch[12,13], Alba Puente[1], Mercedes López[1], Miguel Prieto[1], Viggó Thór Marteinsson[12,14], Martin Wagner[9,10], Abelardo Margolles[7,8], Nicola Segata[6], Paul D. Cotter[4,15], Danilo Ercolini[2,3] & Avelino Alvarez-Ordóñez[1]✉

Abstract

Deep investigation of the microbiome of food-production and food-processing environments through whole-metagenome sequencing (WMS) can provide detailed information on the taxonomic composition and functional potential of the microbial communities that inhabit them, with huge potential benefits for environmental monitoring programs. However, certain technical challenges jeopardize the application of WMS technologies with this aim, with the most relevant one being the recovery of a sufficient amount of DNA from the frequently low-biomass samples collected from the equipment, tools and surfaces of food-processing plants. Here, we present the first complete workflow, with optimized DNA-purification methodology, to obtain high-quality WMS sequencing results from samples taken from food-production and food-processing environments and reconstruct metagenome assembled genomes (MAGs). The protocol can yield DNA loads >10 ng in >98% of samples and >500 ng in 57.1% of samples and allows the collection of, on average, 12.2 MAGs per sample (with up to 62 MAGs in a single sample) in ~1 week, including both laboratory and computational work. This markedly improves on results previously obtained in studies performing WMS of processing environments and using other protocols not specifically developed to sequence these types of sample, in which <2 MAGs per sample were obtained. The full protocol has been developed and applied in the framework of the European Union project MASTER (Microbiome applications for sustainable food systems through technologies and enterprise) in 114 food-processing facilities from different production sectors.

Key points

- This protocol outlines a procedure for sampling the microbiomes of environments with low-biomass yields such as those in a clean food-processing facility and analyzing them through WMS.

- The procedure includes an optimized DNA-extraction stage to maximize DNA yield and allow WMS-based analysis, offering a more complete analysis of the microbiome than targeted methods currently used in industry and avoiding issues of bias associated with targeted high-throughput sequencing.

Key reference

Valentino, V. et al. *Food Res. Int.* **162**, 112202 (2022): https://doi.org/10.1016/j.foodres.2022.112202

A full list of affiliations appears at the end of the paper. ✉e-mail: aalvo@unileon.es

Introduction

The composition and function of food microbiomes are of critical importance for food quality and safety, and this extends to the microbiomes present in the facilities where food is produced, processed or stored. The food-production and -processing environment can be home to many different types of microorganisms, and the composition of its microbiome depends on the specific availability of nutrients, raw materials used and external contamination sources[1,2]. The survival of microorganisms in such hostile environments is also dependent on their ability to form biofilms or tolerate routine cleaning and sanitation practices[3,4].

Considering that microorganisms in food-production and -processing environments can have a substantial impact on the quality and safety of the end products, specific microbial taxa (mainly spoilage and/or pathogenic microbes) are routinely searched for within the food industry by using target-specific (typically, traditional culture-based) approaches. However, these methodologies sometimes fail in giving a complete picture of the contamination pattern of food-production and -processing environments or in tracking the food-contamination sources, because they rely on the selective enrichment and/or isolation of specific culturable microbes, which represent only a minor part of the microbiome. High-throughput sequencing (HTS)-based analysis of metagenomic DNA has revolutionized the study of microbial communities in a wide range of fields by providing reliable means for environmental microbiome characterization and the identification of unknown or overlooked agents[2,5,6]. Compared to culture-based analyses, this approach can provide information on many different microbial contaminants in a single analysis.

Initial studies applying HTS for the characterization of the microbiome of food-production and -processing environments relied on amplicon-based approaches, in which a gene of taxonomic relevance—e.g., the *16S rRNA* gene from bacteria and archaea or ITS2 regions from fungi—is amplified by using PCR from total microbial DNA directly extracted from samples (also known as 'metataxonomics' or 'amplicon sequencing')[7,8]. However, this gives information only on the overall taxonomic composition of the microbiota in a given environment within the facility, with low discriminatory resolution for some taxa. In fact, it is not always possible to distinguish between closely related organisms, and the detection of different strains by using one or a few hypervariable regions of marker genes is challenging. Moreover, the technique can be affected by several technical biases such as the preferential amplification of some taxa and differences in the copy number of the targeted gene(s) among different taxa[2].

More recently, whole-metagenome sequencing (WMS) approaches, based on the fragmentation and sequencing of total DNA without any prior selection or amplification steps, have been explored. These techniques provide a wealth of information, including the taxonomic composition (even at the species and strain level) of prokaryotic[9,10], eukaryotic[11,12] and viral[13,14] communities; the functional potential of the global, or a specific, community[15]; the occurrence and composition of virulence genes, antimicrobial resistance genes and mobile genetic elements[16]; and the reconstruction and characterization of metagenome assembled genomes (MAGs)[9,10], allowing the detection of new taxa[9,17] or even phyla[18]. The WMS approach could therefore provide the food industry the opportunity to gain information on the environmental microbiome composition in their facilities, understand the functional potential of the microbial communities inhabiting their processing plants or identify the presence of dangerous strains or genes responsible for undesired activities. However, there are several technical challenges that might jeopardize the application of WMS technologies for mapping environmental microbiomes at food-processing facilities, with the most relevant one being the recovery of a sufficient amount of DNA from samples taken from industry equipment, tools and surfaces, which frequently harbor very low microbial loads[2]. Aspects of primary importance to improve the recovery of DNA for environmental monitoring activities in the food industry are the design of the sampling approach (the choice of samples to be collected and the sampling procedure, including the sampling kit) and the nucleic acid extraction procedure used. Current sampling procedures for food-production and -processing environments have been developed for the specific aim of isolating and enumerating microorganisms (e.g., ISO standard 18593) and are not appropriate for HTS-based approaches. In addition, most commercial DNA-purification kits available on the market have been optimized for stool, foods or soil samples rather than for low-biomass environmental samples. Therefore, there is an urgent need to develop standard procedures tailored to the particular requirements of low-biomass samples from food-production and -processing environments, especially dealing with sampling approaches, sample manipulation and storage and DNA extraction, but also covering other more unspecific aspects of microbiome analyses like library preparation, sequencing and bioinformatic analysis.

In this protocol, we present a complete workflow, with optimized sampling and DNA-purification methodology, to obtain high-quality WMS sequencing results from low-biomass environmental samples taken from food-production and -processing environments.

Development of the protocol

This protocol integrates a sampling procedure with an optimized DNA-purification approach for monitoring microbiomes at food-production and -processing environments for quality and safety purposes. The protocol aims to maximize the amount of microbial cells collected and the DNA yield, avoiding undesired contamination with exogenous matter or inhibitors that may hinder subsequent sequencing. The application of the protocol described here can yield DNA quantities ranging from ~10 ng to >500 ng (see below). This amount of DNA is sufficient for WMS on the NovaSeq platform (Illumina), and whole-genome amplification to increase the available DNA concentration is not required. This is a clear advantage because it is well documented that random whole-genome amplification might represent a source of bias[19]. A basic downstream bioinformatics workflow for reads filtering, reads assembly into contigs and contigs binning to recover MAGs is also presented.

The full protocol has been developed and applied in the framework of the European Union (EU) project MASTER (Microbiome applications for sustainable food systems through technologies and enterprise; https://www.master-h2020.eu/) by six partner institutions across 114 food-processing facilities from different production sectors (85 dairy, 19 meat, 6 fish, 3 ready-to-eat vegetables and 1 ice-cream processing). It has also been used in a recent study characterizing the microbiome of food-processing facilities handling minimally processed vegetables[20]. Other large collaborative studies within the MASTER consortium applying the protocol will follow soon.

In total, 931 samples from processing environments have been collected in the MASTER project, of which 88.7% did not fail in the library-preparation and sequencing steps and yielded >1 million reads. For those samples that failed sequencing, possible reasons were low DNA concentration (<0.1 ng/µl), failure of library preparation or sequencing that did not generate $\geq 10^6$ reads (Supplementary Fig. 1a). Only 63 samples from processing environments (54 from food contact samples and 9 from non-food contact samples), alongside 140 negative control samples, failed sequencing. In addition, of the 140 negative controls, most of the samples failing sequencing (94.1%) had <10 ng of DNA/µl (Supplementary Fig. 1b). The mean DNA concentration obtained from successfully sequenced samples was 50.87 ng/µl, with 66.6% of samples having >10 ng/µl, which allowed the generation of an average of 61,385,112 reads, 62,620.6 contigs and 12.2 MAGs per sample (with median values of 56,171,821 reads, 49,829 contigs and 10 MAGs) (Fig. 1). These results demonstrate the success of the approach for the deep characterization of the microbiome of low-biomass environments. Our protocol markedly improves on results previously obtained in studies performing WMS of processing environments and using other protocols not specifically developed to sequence these types of samples. For example, a total of 162 MAGs (10 of them with high quality) were previously obtained from 93 samples (1.7 MAGs per sample) in dairy environments[21]. Likewise, an average of 0.8 MAGs per sample were obtained from the analysis of the sequencing reads of another previous study characterizing meat-processing environments[22].

Applications

Although our focus is on swab samples from food-production and -processing environments, we envisage that the protocol will also be appropriate for microbiome-monitoring activities in other built environments, such as hospitals or households, and for analyzing other similar environmental surface samples with low microbial biomass such as those from urban

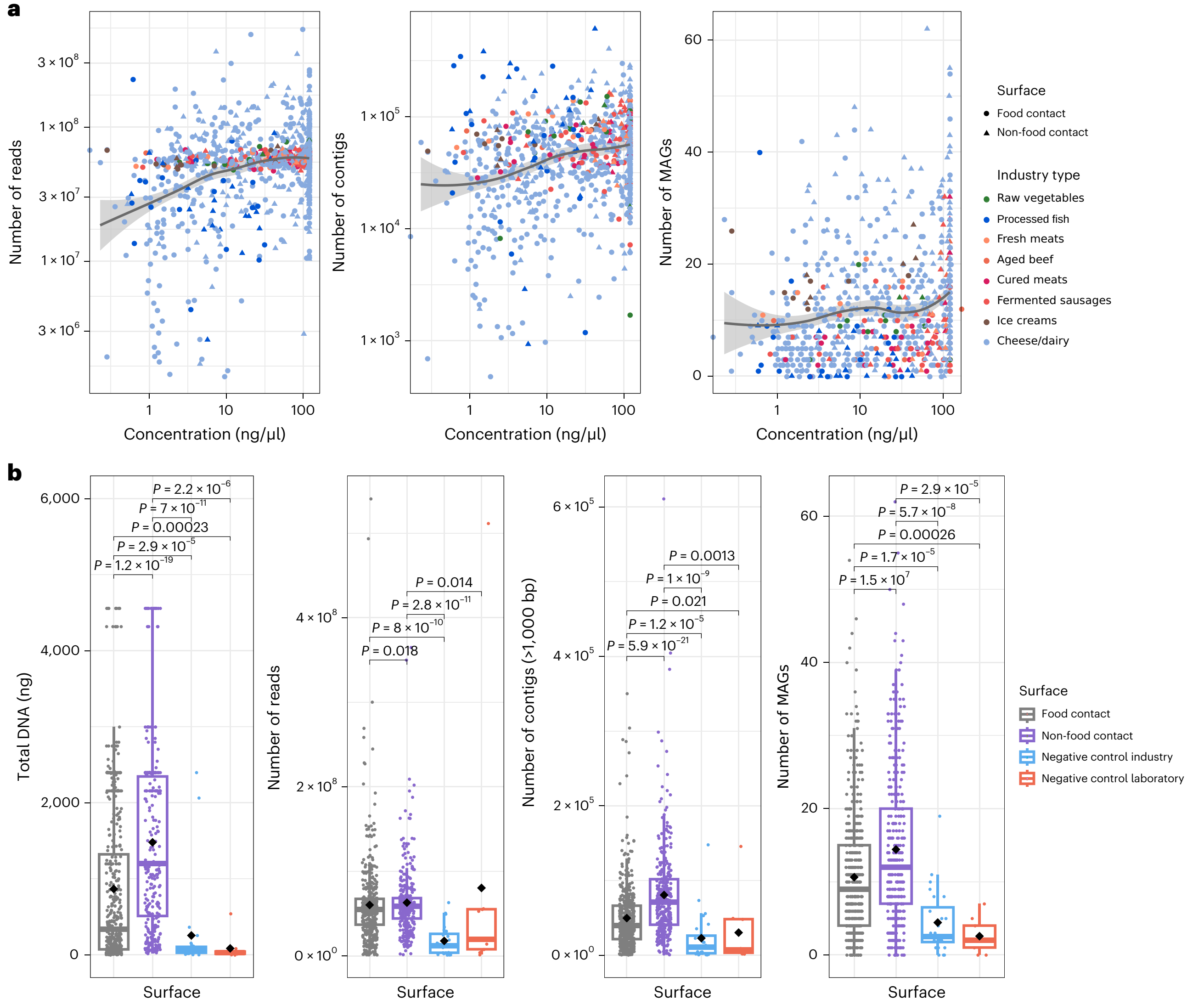

Fig. 1 | Overview of WMS results. Results after reads filtering for all the samples successfully sequenced with ≥1 million reads obtained. **a**, Total number of reads, contigs and MAGs obtained per sample as a function of the DNA yield of the sample. The type of surface is indicated by shape, and the type of industry is indicated by colors. The gray line indicates the smoothed conditional means (calculated by geom_smooth and the 'lm' method in the ggplot2 R-library), and the gray shaded area indicates the standard error of the trend line. **b**, Total DNA and number of reads, contigs and MAGs by surface type, including negative controls taken in both the food-processing site and the laboratory. Black diamonds indicate mean values, and the central lines of box plots indicate median values. The upper and lower horizontal lines indicate the first quartile (Q_1) and the third quartile (Q_3), respectively; where $Q_3 - Q_1$ = IQR (interquartile range), while the vertical lines indicate $Q_1 - 1.5 \times$ IQR and $Q_3 + 1.5 \times$ IQR, respectively, from bottom to top. Samples with DNA concentration above the limit of detection of the Qubit high-sensitivity double-stranded DNA quantification kit (120 ng/µl) are represented as having a DNA concentration equal to 120 ng/µl. Most negative controls could not be sequenced because of very low DNA yields, and the mean represented in this figure is calculated from the small proportion of negative controls that could be sequenced.

environments. Moreover, in principle, the protocol could also be used for other different sample types, such as food or water samples, although to do this, the sample-preparation step before cell lysis and DNA purification might need to be adapted, for example, by adopting different homogenization or cell-concentration methods. For these other sample types, it is recommended to review any sample-specific protocols that currently exist, for example, those for human tissues[23] or water samples[24].

The sampling and DNA-purification steps of the protocol have been validated for WMS with short-read Illumina technology. We have found that the approach can yield output DNA with fragment lengths above 10,000 bp, and therefore we believe that the procedure described here could also be appropriate, with some minor adaptations, for WMS with long-read technology (e.g., Oxford Nanopore Technology). The library-preparation steps of this protocol are specific to sequencing with the Illumina NovaSeq platform, and the bioinformatic workflow presented is also tailored to the processing and analysis of short-read outputs. These steps of the protocol would require adaptation for long-read sequencing approaches.

Advantages and limitations

The main advantage of metagenomics-based approaches over classical methods for the microbiome characterization of food-processing environments is that they are untargeted approaches capable of simultaneously detecting a vast number of microbial taxa and, in the case of WMS, gene categories (e.g., antimicrobial resistance genes or virulence genes) without the need for selective enrichment and cultivation steps, thus offering much broader information on the microbial contaminants that may be present in a given sample. The main limitations, in comparison with classical culture-dependent methodologies, are those related to the fact that sampled DNA can originate from both living and dead cells, the limited sensitivity of the technology for the detection of low-abundance microorganisms and the fact that only relative abundance data (and not absolute quantification) can be obtained from the analyses[2,5]. Interestingly, some methodologies to distinguish between viable and non-viable cells, such as the use of propidium monoazide and ethidium monoazide treatments[25], are being studied, although further research is still needed before systematically applying them for microbiome mapping. Another important limitation for some types of samples is the contamination of microbial DNA with DNA from non-microbial sources (e.g., human, animal or plant DNA), which can happen in environmental samples if surfaces have contamination from workers or traces of food or derived organic material. In addition, with WMS, it is difficult to characterize some low-abundance microbes (such as some pathogens or antimicrobial resistant microorganisms), although quasi-metagenomic approaches involving WMS, such as those that involve sequencing after the selective enrichment of a subset of specific microorganisms, can be an attractive approach for genome assembly in this case[26]. Overall, a targeted culture-dependent or qPCR approach may be more advantageous if analysis is focused on the detection and characterization of a specific microbial contaminant, whereas if the interest is in getting a more general picture of the composition of the microbial communities inhabiting the processing environment and their genetic repertoire, an untargeted metagenomics-based approach is more appropriate.

When comparing WMS with amplicon-based metataxonomic approaches, the main advantages of the former are that they can provide resolution at the species or even strain level, information on the repertoire of genetic elements (including virulence and antimicrobial resistance determinants) and the functional potential of the microbial community and the ability to reconstruct genomes from the most-dominant taxa prevailing in the given environments. On the contrary, the main limitation is that, to obtain reliable results, a higher amount of high-quality DNA is required, because no DNA amplification step is used, unlike in metataxonomy approaches[2]. In addition, the limit of detection of WMS is higher compared to that of amplicon sequencing, given that low-abundance microbial taxa may not be sequenced in taxonomically uneven samples (where a few taxa predominate) or in samples that have a relatively high concentration of non-microbial DNA. Other limitations, when compared to amplicon sequencing, are the higher monetary cost (approximately three times higher) and computational needs and the extensive knowledge in data analysis required. This is the first protocol developed with the aim of ensuring the purification of sufficient DNA (with mean DNA concentrations ranging from 43.3 ng/µl for food contact surfaces to 74 ng/µl for non-food contact surfaces) from food-processing environments, compatible with the generation through WMS of high-quality sequencing reads and the reconstruction of contigs and MAGs.

Alternative methods

Various detailed protocols are publicly available for sample collection, manipulation, storage, processing and DNA purification in microbiome-characterization studies. Many protocols are specifically tailored to particular sample types and in most cases deal with the investigation of the human microbiome (see, for instance, the protocols of the Human Microbiome Project; https://www.hmpdacc.org/hmp/resources/), although there are protocols adapted for the microbiome profiling in soil[27], air[28], plant[29] and water[30]. Such alternative protocols could in principle be used and have been used in the past, with minor adaptations regarding sample collection, for obtaining DNA for WMS of food-processing environments. However, food-processing environments are challenging samples due mainly to their low microbial biomass and possible contamination with detergents, disinfectants or residual food matrix materials that may inhibit subsequent enzymatic steps, which often result in low-quality sequencing results or even in failed library preparation for sequencing, as demonstrated by the study by Cobo-Díaz and colleagues[22], in which a low amount of reads (<200,000) was obtained on various samples from food contact surfaces. Hence, there was an obvious need to develop standard procedures to obtain high DNA yields for WMS from food-processing environments. The protocol described here successfully addresses this need, because the DNA loads, number of contigs and number of MAGs obtained with it markedly exceeds those previously described in the literature, applying different procedures to similar sample types.

Experimental design

Here, we describe our protocol for an improved sampling and extraction of DNA for WMS from food-processing environments (Fig. 2), as well as our workflow for sequencing and bioinformatic analysis. Specifically, we describe our methods for sample collection, manipulation and storage (Steps 1–13), microbial cell lysis and DNA purification (Steps 14–35), library preparation and sequencing (Steps 36 and 37) and bioinformatic analysis (Steps 38 and 39). It is important to bear in mind that this protocol might need to be slightly adapted for specific sample types and objectives, hence performing a first trial with a few representative samples is recommended. More information about steps that might be specifically modified to obtain optimal performance can be found in the troubleshooting table.

Sampling, sample manipulation and storage (Steps 1–13)

We recommend preparing a detailed sampling plan in which information on the selected sampling time, number of samples and surfaces to be sampled, among other relevant factors, is fully recorded. The most appropriate sampling time will depend on the rationale of the microbiome study. Thus, for instance, if the main objective is to characterize the resident microbiome in a food-processing facility, the ideal sampling time should be when the processing plant is clean before starting the manufacturing activities. Other sampling time points (during production, after production, before and just after cleaning and sanitation, etc.) can be more suitable for answering other biological questions. For example, investigations evaluating the efficacy of particular sanitation regimes would require restricted samplings immediately before and after the intervention is applied. To increase the microbial loads recovered from the clean surfaces, we recommend collecting and pooling at least five different samples from each given sample category. For example, for studying the microbiome of meat-cutting tables in a meat-processing plant, five ~1-m^2 surfaces from one or various cutting tables can be swabbed, and swabs should then be pooled for follow-up activities. Figure 3 provides, as an example, the types of samples recommended to characterize the resident microbiome and evaluate the impact of different sources on the microbiome of the end products in a cheese-making facility and a plant producing fermented sausages, respectively. This sampling plan is just a recommendation and can be adapted to other needs. Zoning of processing environments for sampling may be approached in different ways, for example, high-care, standard-care and low-care hygiene areas; wet and dry areas; and food contact surfaces and non-food contact surfaces. The selection of sampling points could take into account areas that are probably contaminated, such as wet areas, hard-to-reach places and difficult-to-clean equipment, and processing environments more frequently linked to persistence of specific hazardous microbes. Furthermore, sampling plans including more-intense sampling regimes could be used if assessing the effects of construction or investigating outbreaks or nonconformities in conventional microbiological analyses of foods to identify potential harborage niches in the facility and determine how far contamination has spread.

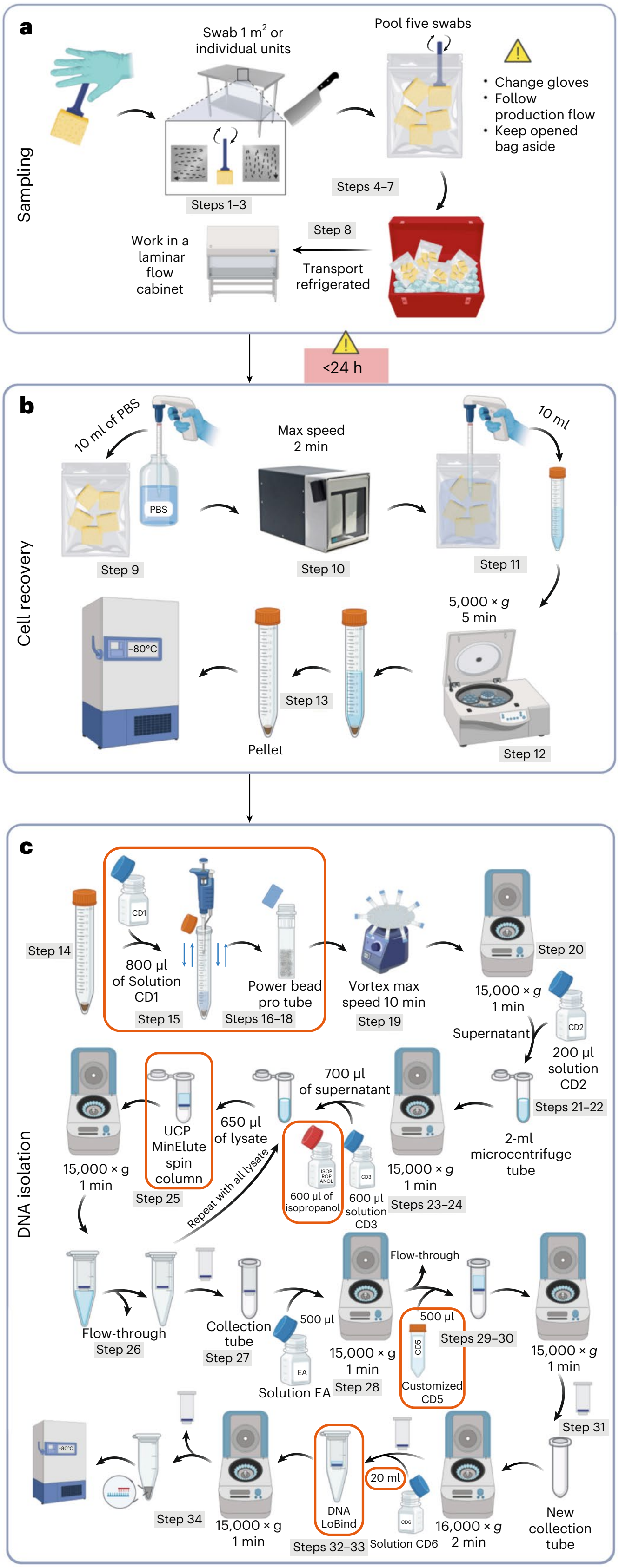

Fig. 2 | Workflow for sampling, cell recovery and DNA purification. a, Swab samples are taken from food-processing environments, by using personal protective equipment to avoid contamination, and pooled in sampling bags (five pooled swabs per sample category). **b**, PBS is added to the sampling bag, swabs are homogenized and cells are harvested through centrifugation and stored in the ultrafreezer. **c**, DNA is extracted from the cell pellet by using the tailored protocol based on the DNeasy PowerSoil Pro kit chemistry with modifications and QIAGEN's UCP MinElute spin columns. After DNA has been purified and meets the quality standards, it can be used for library preparation for Illumina sequencing. All steps of the DNA-purification protocol that deviate from that of the DNeasy PowerSoil Pro kit are indicated by orange squares on the scheme. Figure created with BioRender.com.

a

	× 5 swabs	Processing room	Cold room	Ripening room	Smoking room	Packing room
Fermented meat facility	Food contact	• Mincer/bagging machines • Knives • Tables • Chopping boards	• Trays • Shelves • Trolleys	• Trays • Shelves • Trolleys	• Trays • Shelves • Trolleys	• Slicers • Packing machine
	Non-food contact	• Drains • Floors • Walls	• Drains • Floors • Walls	• Drains • Floors • Walls	• Drains • Floors • Walls	• Drains • Floors • Walls
	Other(s)	• Meat batter (100 g) • Fresh sausage (before ripening) × 2 samples	–	–	–	• Fermented sausage (after ripening) × 2 samples
	Operators	Hands and/or aprons				

b

	× 5 swabs	Processing room	Cold room	Ripening room	Smoking room	Packing room
Cheese-producing facility	Food contact	• Cheese vats • Curd shredders • Draining tables • Molds • Molding machines	–	• Trays • Shelves	–	• Slicers • Packing machine • Knives
	Non-food contact	• Drains • Floors • Walls	–	• Drains • Floors • Walls	–	• Drains • Floors • Walls
	Other(s)	• Raw milk (50 ml) • Whey (200 ml) • Brine (200 ml) • Fresh cheese × 2 samples	–	–	–	• Ripened cheese × 2 samples
	Operators	Hands and/or aprons				

Fig. 3 | Example sampling plan. a,**b**, Sampling plan proposed for the characterization of the resident microbiome and the evaluation of the impact of different sources on the microbiome of the end products in a plant producing fermented sausages (**a**) and a cheese-making facility (**b**).

An aspect of primary importance is the choice of the type of swab and the swabbing procedure. The use of sponge swabs is recommended because these have a wider sampling surface and allow a better recovery of microbial cells than other alternatives. The most common sponge swabs in the market are cellulose derived, which have a cotton or rayon tip that is made of fibers wrapped around a plastic rod, or made of synthetic materials, such as polyester, polyurethane or nylon. Cellulose-derived swabs tend to trap bacterial cells within the fiber matrix, thus hampering the release of cells in the recovery. In addition, they can release plant DNA, thus contaminating the extracted microbial DNA[2]. On the other hand, polyurethane sponge swabs offer several distinct advantages over traditional cellulose sponges, including resistance to tearing, flaking or fraying during sample collection and improved release of organisms for more accurate test results. In addition, polyurethane's synthetic manufacturing process yields a more consistent biocide-free material without any components that may interfere with downstream test methods[31]. For these reasons, in our protocol, we recommend the use of swabs made of synthetic materials, in this case polyurethane.

When wide surfaces are sampled (e.g., floors or walls), we recommend sampling an ~1-m^2 surface by swabbing surfaces first horizontally and then vertically, turning the swab around in between swabs. For other types of surfaces, on which swabbing ~1-m^2 may not be possible (e.g., drains and knives), we recommend swabbing individual units (e.g., one drain and one knife). To sample the operators, consider swabbing the hands/gloves, aprons, caps and/or shoes

(Supplementary Video 1 and Supplementary Note). When swabbing, the bag opening should be kept to the side to decrease air-born contamination. Once the swab is taken, the air in the bag should be removed manually before sealing it.

Once samples are taken, it is important that they be kept refrigerated (for instance, using a portable cooler filled with ice packs) until processing in the laboratory, which should ideally take place <24 h after sampling. Alternatively, samples could be snap-frozen in liquid nitrogen or, if this is not possible, placed on dry ice before long-term storage in a freezer (ideally at −80 °C) until sample processing.

For cell recovery from the swabs, we recommend the addition of a small volume of sterile PBS to the sampling bag containing the pool of five swabs, followed by thorough homogenization in a Stomacher and the centrifugation of the recovered volume to obtain a cell pellet. This cell pellet will be the matrix used for cell lysis and DNA purification in the follow-up steps of the protocol. These subsequent steps can take place immediately after centrifugation, or we recommend the storage of the cell pellet until use at −80 °C. We recommend storage at −80 °C for both samples and extracted DNA because it is widely recognized that storage temperature can have a significant impact on the stability of the microbial communities and the quality of extracted nucleic acids.

Microbial cell lysis and DNA purification (Steps 14–35)

The cell pellets collected from the surfaces of food-processing facilities are expected to contain diverse, but low-abundance, microbial communities, as well as inorganic and organic contaminants from the sampled surfaces encompassing residuals of sanitizers or food matrices. Hence, the DNA extraction workflow must achieve comprehensive cell lysis and high DNA recovery rates while minimizing carryover of various contaminants. The choice of an adequate DNA extraction procedure and the specific methodology used for cell lysis and DNA purification is vital becausee the approach followed can affect the observed microbial diversity, which can be a limitation in this type of metagenomics workflow. Here, the DNeasy PowerSoil Pro kit (QIAGEN) was used as the basis for development of a modified protocol.

Lysis of microbial cells for DNA purification is usually achieved through either enzymatic or mechanical approaches. Enzymatic approaches may cause biases associated with the differential effectiveness of lytic enzymes, especially among the wide diversity of microbes expected in the sample (e.g., different degrees of lysis for Gram-positive and Gram-negative bacteria). Mechanical approaches, usually based on vigorous bead beating, can cause some DNA shearing but produce a more unbiased lysis of different bacterial species. In this protocol, cell lysis occurs through a combination of mechanical methods (bead beating in QIAGEN's PowerBead Pro Tubes) and chemical methods (lysis buffer CD1 of the DNeasy PowerSoil Pro kit, QIAGEN). After lysis, inhibitors are removed through the precipitation of non-DNA organic and inorganic material like polyphenolic and humic substances, cell debris and proteins.

To maximize the recovery of total microbial DNA (Fig. 4), the DNeasy PowerSoil Pro kit was modified as follows: the standard spin columns were replaced by QIAGEN's QIAamp UCP MinElute spin columns, which allow flexible elution volumes down to 20 µl. Elution in lower volumes increases the end concentration, which can be critical for enabling WMS workflows from low-biomass samples (Fig. 4a). Moreover, the QIAamp UCP MinElute columns are treated through a physical process to remove background microbial DNA, reducing potential contamination risks for the sequencing analysis. Besides the substitution of the silica columns included in the standard DNeasy PowerSoil Pro kit, the addition of isopropanol during DNA binding to the silica membrane improved total nucleic acid yield (Fig. 4b), although this appears to be specific for the swabs used in this protocol. Subsequent steps involve two washes to remove protein and other non-aqueous contaminants, as well as residual salt, humic acid and other contaminants from the spin column while allowing the DNA to stay bound to the silica membrane. The final elution of the purified DNA is achieved by adding a small volume (20 µl) of an elution buffer, allowing the complete release of the DNA from the spin column filter membrane (Fig. 4c,d). During optimization of the DNA-extraction protocol, a *16S rRNA* qPCR using 515F-806R primers to amplify the V4 hypervariable region was performed as described in the Supplementary Methods to quantify the *16S rRNA* gene copy numbers

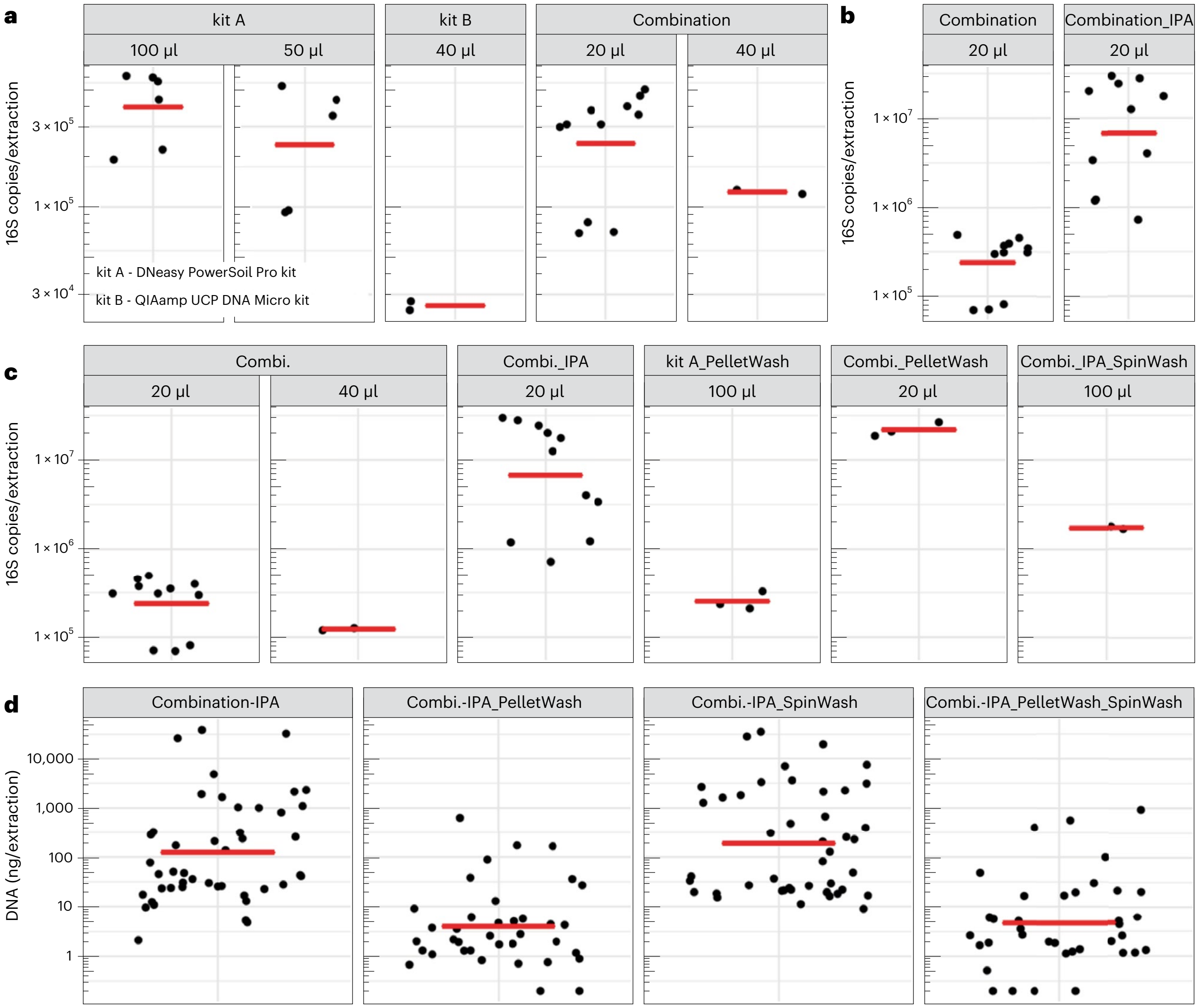

Fig. 4 | Optimization of DNA extraction from surface swabs. The cell pellet derived from pooled surface swabs was subjected to cell lysis and subsequent DNA extraction. Cell pellets were obtained by following the described surface swab sampling protocol in a standard laboratory environment. The compared conditions for the extraction workflow are indicated by the first row of graph headings. Commercial kits (Kit A: DNeasy PowerSoil Pro kit; Kit B: QIAamp UCP DNA Micro kit) with their standard protocols, a combination of kit A and spin columns of kit B or further alterations in the standard protocol of kit A were tested. The second row of graph headings denotes the elution volume, which is regulated by the choice of spin columns. **a**–**c**, Depicted are the resulting *16S rRNA* gene copy numbers obtained per individual extraction (black points) as proxy for bacterial DNA content as determined by *16S* V4 qPCR. **d**, The total DNA yield in nanograms per extraction is depicted as quantified by Qubit. Red crossbars indicate the mean of all extractions for the corresponding approach. **a**, Comparison of two commercial kits and their unaltered standard protocols and a combination of kit A with spin columns of kit B following the protocol of kit A. **b**, Comparison of the aforementioned combination of kits without (Combination) or with addition of isopropanol during binding of DNA to the silica membrane (Combination_IPA). **c**, Comparison of various alterations during the extraction protocol of kit A. 'IPA' denotes as before the addition of isopropanol during DNA binding to the silica membrane, 'PelletWash' denotes the additional washing of the swab-derived cell pellet before cell lysis and 'SpinWash' denotes the increased concentration of ethanol during spin column washing while the DNA is already bound to the silica membrane. **d**, The combination of kit A with spin columns of kit B following the protocol of kit A with the addition of isopropanol during DNA binding was used as standard for DNA extraction from surface swabs. It was compared with the inclusion of two optional steps, which are, as before, the additional washing of swab-derived cell pellets before cell lysis (PelletWash) and the increased concentration of ethanol during spin column washing while the DNA was already bound to it (SpinWash) and a combination thereof. These protocols were tested on surface swabs collected in food-processing facilities. Combi., combination.

obtained per extraction and evaluate the performance of the DNA-extraction procedures tested. This qPCR analysis can be applied to all or a subset of samples (e.g., a few representative samples from different sample categories) as a quality control process to check the amount of bacterial DNA extracted and detect possible issues linked to the preponderance of DNA from non-microbial sources.

The purified DNA sample will be the matrix used for library preparation and WMS in the follow-up steps of the protocol. These subsequent steps can take place immediately from DNA purification, or we recommend the storage of the DNA sample until use at −80 °C.

We recommend assessing the purified DNA with a Qubit fluorometer by using the Qubit high-sensitivity double-stranded DNA (dsDNA) quantification kit, which has a quantitation range from 0.1 to 120 ng/µl. The Illumina DNA prep kit requires an input of only 1 ng of DNA. However, we have found that three samples with even less DNA yields have been successfully sequenced.

Library preparation and sequencing (Steps 36 and 37)

The library preparation for Illumina NovaSeq metagenomic sequencing is based on using the Illumina DNA prep kit and following the manufacturer's protocol (available at https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/illumina-dna-prep.html). Libraries are multiplexed by using dual indexing and sequenced for 150-bp paired-end reads (average of 6.5 GB/sample) on the NovaSeq 6000 sequencing system.

Bioinformatic analysis (Steps 38 and 39)

Sequenced metagenomic reads are quality-controlled by using a pre-processing pipeline available at https://github.com/SegataLab/MASTER-WP5-pipelines/tree/master/02-Preprocessing. First, sequencing adapters, reads of low quality (Phred score <20), short reads (<75 bp) and reads with more than two ambiguous nucleotides are removed by using Trim Galore (v0.6.6) (https://github.com/FelixKrueger/TrimGalore). Then, contaminant DNA is identified by using Bowtie2 version 2.2.9 (with the *–sensitive-local* parameter)[32], removing reads from the phiX174 Illumina spike-in (National Center for Biotechnology Information (NCBI) accession number NC_001422) as well as potential human contamination (using the GRCh38.p13 human genome, NCBI accession number GCF_000001405.39). In addition, genome contamination with non-microbial DNA from other different origins (e.g., animal or plant DNA from particular host species) can be removed by following the same Bowtie2 approach, where appropriate. The remaining high-quality reads are sorted and split to create standard forward, reverse and unpaired reads files for each metagenomic sample.

To reconstruct microbial genomes, a single-sample metagenomic assembly and contig binning approach is applied (https://github.com/SegataLab/MASTER-WP5-pipelines/tree/master/05-Assembly_pipeline). Briefly, contigs are assembled from the metagenomic reads by using MEGAHIT version 1.1.1[33] with default parameters. Contigs longer than 1,000 nt are then binned by using MetaBAT2 version 2.12.1[34] with parameters *–maxP 95 –minS 60 –maxEdges 200 –unbinned –seed 0*. Finally, quality control of the MAGs is performed by using CheckM version 1.0.7[35] with default parameters. To ensure the quality of the MAGs, only medium- (completeness between 50% and 90% and contamination <5%) and high-quality (completeness >90% and contamination <5%) MAGs are kept.

To facilitate the execution of this basic bioinformatic analysis and many other more-advanced bioinformatic analyses, many tutorials are available on bioBakery at https://github.com/biobakery/biobakery.

Controls (Steps 7, 9 and 18)

Including both positive and negative control samples alongside the samples from the food-processing environments being analyzed is recommended. As a positive control, commercial mock communities, such as the ZymoBIOMICS microbial community standard, can be used. The ZymoBIOMICS standard includes three easy-to-lyse Gram-negative bacteria, five tough-to-lyse Gram-positive bacteria and two tough-to-lyse yeasts. Including different dilutions of the mock community (e.g., 10^{-6}, 10^{-4} and 10^{-2} cells/ml) is highly recommended to produce positive samples with diverse DNA concentrations and thus get more complete information on potential contaminants coming from sample manipulation and the materials used[36].

As a negative control, different type of samples can be used to understand whether the sampling materials and the environment where samples from food-processing environments are taken and/or manipulated influenced their microbiome composition. If DNA is obtained from the negative control samples so that library preparation can be completed and sequencing reads can be obtained, there exist some strategies that can be used for the in silico removal of contaminant reads from real samples, for example, by using the R-package *decontam*[37]. This tool identifies contaminants on the basis of their frequency and/or prevalence in negative control samples over 'real' samples.

In the validation of our protocol, we included as negative controls pools of five swabs left exposed for 1 min to the air of the processing plant (negative control–industry) or of the laboratory where samples were manipulated and DNA extracted (negative control–laboratory). Because of the low DNA yield obtained, only 33.3% of these negative control samples could be sequenced, the vast majority of them with a low number of reads obtained (Fig. 1).

Including negative controls for the DNA-extraction step is also recommended to check the free-DNA status of the components of the extraction kit. These can consist of empty tubes. All the negative controls from this category included in our validation of the protocol showed DNA concentrations below the detection limit of the Qubit high-sensitivity dsDNA quantification kit and failed in the library-preparation step.

Materials

Sampling materials

- Whirl-Pak B01592WA Hydrated PolyProbe sampling bags with sampling sponges and 8-inch probe, 24 ounces, sterile; 100/box (hydrated with 10 ml of HiCap neutralizing broth)
- Portable cooler
- Ice packs
- Personal protective equipment for sampling, including disposable masks, disposable coats, disposable caps, disposable shoes and disposable gloves

Laboratory reagents (sample pre-processing and DNA purification)

- PBS tablets (Sigma-Aldrich, cat. no. P4417-50TAB)
- DNeasy PowerSoil Pro kit (QIAGEN, cat. no. 47016). The following reagents from the kit will be used: Solution CD1, Solution CD2, Solution CD3, Solution C5, Solution EA and Solution C6 (10 mM Tris)

  ▲ **CAUTION** Solution EA and Solution C5 are flammable. Do not add bleach or acidic solutions directly to the sample-preparation waste. Solution CD1 and Solution CD3 contain chaotropic salts, which can form highly reactive compounds when combined with bleach. If liquid containing these buffers is spilt, clean with a suitable laboratory detergent and water. If the spilt liquid contains potentially infectious agents, clean the affected area first with laboratory detergent and water and then with 1% (vol/vol) sodium hypochlorite.
- ZymoBIOMICS microbial community standard (Zymo Research, cat. no. D6300)
- Isopropanol (e.g., Sigma-Aldrich, cat. no. I9516)
- Ethanol absolute (e.g., Sigma-Aldrich, cat. no. 1.07017)
- Qubit high-sensitivity dsDNA quantification kit (Invitrogen, cat. no. Q32851)

Laboratory reagents (library preparation)

- Illumina DNA prep kit (Illumina, cat. no. 20018705)
- Nuclease-free water

Equipment

- P1, P10, P100, P1000 and 10-ml pipettes
- 1.5-ml sterile Eppendorf tubes
- 15-ml sterile plastic tubes

- DNA LoBind tubes (e.g., Eppendorf, cat. no. 0030108051)
- DNeasy PowerSoil Pro kit (QIAGEN, cat. no. 47016); the following materials from the kit will be used: PowerBead Pro tubes and 2-ml microcentrifuge collection tubes
- QIAamp UCP DNA Micro kit (QIAGEN, cat. no. 1103588); the MinElute spin columns from this kit will be used in the procedure

  ▲ **CRITICAL** Using UCP MinElute columns is critical to reduce background DNA amounts when working with low-biomass samples.
- 96-well PCR plates
- Microseal 'B' adhesive seal
- 1.7-ml microcentrifuge tubes (e.g., Sigma-Aldrich, cat. no. CLS3620)
- Eight-strip PCR tubes
- P1, P10, P100 and P1000 pipette tips
- 20-µl multichannel pipette
- 200-µl multichannel pipette
- 96-well 0.8-ml polypropylene deepwell storage plates (midi plate) (e.g., Thermo Fisher Scientific, cat. no. AB0859)
- Microseal 'F' foil seal (e.g., Bio-Rad, cat. no. MSF1001)
- Stomacher (e.g., IUL Instruments, cat. no. 9000400)
- Vortex with adapter for 1.5- to 2-ml tubes (Vortex-Genie 2 mixer, Scientific Industries, cat. no. SI-0236); alternatively, TissueLyser II or PowerLyzer 24 homogenizer (QIAGEN, cat. no. 85300 and 13155, respectively)
- Centrifuge(s) for 1.5- and 15-ml tubes

  ▲ **CRITICAL** It is important to use a centrifuge in which the PowerBead Pro tubes rotate freely without rubbing.
- Laminar flow hood
- Ultra-freezer (−80 °C)
- Qubit fluorometer (Thermo Fisher Scientific)
- Thermal cycler (for library preparation)
- Illumina NovaSeq 60000 sequencer (Illumina)

Reagent setup

▲ **CRITICAL** All reagents should be freshly prepared before the experiment.

Customized wash buffer C5

Prepare a mix of *N* × (500 µl of Solution C5 + 333 µl of ethanol absolute), where *N* is the number of samples that will be processed for DNA extraction.

PBS

Dissolve one tablet per 200 ml of purified water and sterilize the solution at 121 °C for 15 min. Prepare ≥10 ml per sample.

Equipment setup

Sampling plan

The day before sampling, define and document the sampling plan that will be followed. See Fig. 3 for an example sampling plan.

Sampling materials

Open the boxes containing the Whirl-Pak B01592WA Hydrated PolyProbe sampling bags and organize them in groups of five bags (using the yellow strip of one of the bags to keep the five of them grouped). Label the first bag, by using a permanent marker, with the sample code to be collected. Repeat this until all five-bag groups are properly labeled. Prepare the portable cooler and personal protective equipment (disposable masks, disposable coats, disposable caps, disposable shoes and gloves). Put the ice packs in the freezer (remember to introduce them into the portable cooler on the sampling day).

Procedure

Sampling of the food-processing facility

● **TIMING 1.5 h per food-processing facility (for collecting 20 composite samples) plus travel time**

▲ **CRITICAL** To avoid airborne contamination and other sources of cross-contamination, single-use disposable protective clothing (i.e., gloves and disposable masks, coats, caps and shoes) should be worn. Gloves should be changed between samples. It is advisable to perform sampling in the order of the food chain production flow to avoid cross-contamination of the end product with raw materials or other foreign materials that the sampling procedure might bring to the facility. See Supplementary Video 1 for the detailed procedure.

1. Put on a new set of gloves and rub them with hand sanitizer before starting sampling.
2. Locate the first surface that you are going to sample and take a corresponding pre-labelled Whirl-Pak B01592WA Hydrated PolyProbe swab bag. Prepare the swab as follows:
   - Keeping the Whirl-Pak B01592WA Hydrated PolyProbe swab bag in a vertical position, open it carefully by using the marks on the top of the bag. Take care not to spill any liquid from the bag or touch any other surface with your gloves, the stick and/or the sponge.
   - Hold the swab from the stick without touching the inside of the bag with your gloves. Carefully, without taking the swab out of the bag, move the stick slowly to moisten the sponge with the liquid buffer inside the bag.
   - Once the sponge is sufficiently moistened with the liquid buffer inside the bag, take the swab out of the bag. Place the empty bag in a safe place and away from air flows. The bag will be used to store the swabs and any cross-contamination must be avoided.
3. Sample the surface as follows:
   - Rub the swab (by one of its sides) slowly on the surface to be sampled by doing horizontal movements, covering an ~1-m^2 area.
   - Rotate the swab to use the other side of the sponge and proceed to sample the same surface area again by using vertical movements.

     ◆ **TROUBLESHOOTING**
4. Once the swabbing is completed, return the swab to the plastic bag. Take care not to touch any other surface with the sponge and keep holding the stick with one hand. With the other hand, separate the stick from the sponge carefully by unscrewing and discard the stick.
5. Repeat Steps 2–4, pooling the swabs in the same bag until five swabs are collected in a single bag (the bags from the second to fifth swab can be discarded).
6. Squeeze the air from inside the bag, roll down the top of the bag and then use the yellow strips to hermetically seal the bag. Place the hermetically closed swab bag in a vertical position into the portable cooler filled with ice packs.
7. Repeat Steps 1–6 for each of the different sample categories included in the sampling plan.

   ▲ **CRITICAL** Collecting negative control samples is highly recommended. For this, expose the swabs for 1 min to the air in the food-processing facility.
8. Transport the samples to the laboratory.

Sample pre-processing

● **TIMING 1.5 h per food processing facility (for 20 samples)**

▲ **CRITICAL** Gloves should be used during sample manipulation, which ideally should take place in a laminar flow hood.

▲ **CRITICAL** Samples should be processed within the next 24 h after sampling. Alternatively, samples could be snap-frozen in liquid nitrogen or, where this is not possible, placed on dry ice before long-term storage in a freezer (ideally at −80 °C) until sample processing.

▲ **CRITICAL** At this point, collecting negative control samples in the laboratory where the samples will be pre-processed is highly recommended. To do this, expose Whirl-Pak B01592WA Hydrated PolyProbe swabs for 1 min to the air of the laboratory. Negative control swabs can be pooled and then pre-processed according to the steps detailed for the industry samples below.

9. Move the sampling bags to a laminar flow hood. In the hood, carefully open the first sampling bag, add 10 ml of sterile PBS and close it again. Repeat for each sampling bag.
10. Homogenize each bag in the Stomacher at 175 rpm for 2 min.
11. In the laminar flow hood, carefully open each sampling bag, recover 10 ml of homogenized liquid by using a pipette and transfer it to a sterile 15-ml plastic tube.

    ▲ **CRITICAL** Because the sponge swabs can retain liquid, gently squeezing the sponges from outside the sampling bag while pipetting is necessary to facilitate the release of the liquid from the sponges.
12. Centrifuge at 5,000*g* for 5 min at room temperature (20–25 °C).
13. Carefully discard the supernatant and keep the tube with the cell pellet; note that some pellets might be very small.

    ■ **PAUSE POINT** The tube with the cell pellet can be stored in the ultra-freezer at −80 °C for several months. Optionally, to save space in the ultra-freezer, the cell pellet can be resuspended in a small volume (500 µl) of sterile PBS, the liquid transferred to a 1.5-ml Eppendorf tube, the sample centrifuged at 5,000*g* for 5 min at room temperature, the supernatant discarded and the tube with the cell pellet stored at −80 °C.

DNA purification

● **TIMING 4 h (for 20 samples)**

14. Thaw the tubes with the cell pellets for 15 min at room temperature.
15. Add 800 µl of Solution CD1 to each cell pellet and resuspend by pipetting up and down.
16. Spin the PowerBead Pro tubes briefly to ensure that the beads have settled at the bottom.
17. Transfer the complete CD1 suspensions to fresh PowerBead Pro tubes.
18. At this step, adding a positive control such as the ZymoBIOMICS microbial community standard is recommended. Dilute the mock community (e.g., 10^{-6}, 10^{-4} and 10^{-2} cells/ml) and add 20 µl of each of the corresponding dilutions to PowerBead Pro tubes with 800 µl of Solution CD1. Adding a new negative control sample is also highly recommended; the negative control of the DNA-purification step can be prepared by adding 800 µl of Solution CD1 to an empty PowerBead Pro tube.
19. Secure the PowerBead Pro tubes horizontally on a vortex adapter for 1.5- to 2-ml tubes in the Vortex-Genie 2. Vortex at maximum speed for 10 min.

    ▲ **CAUTION** When using the vortex adapter for >12 preparations simultaneously, increase the vortexing time by 5 min.

    ▲ **CRITICAL** Other alternative materials may be used for bead beating. Some examples are provided in the 'Protocol: Detailed' section of QIAGEN's DNeasy PowerSoil Pro kit handbook.
20. Centrifuge the PowerBead Pro tubes at 15,000*g* for 1 min.
21. Transfer the supernatants (~500–600 µl) to clean 2-ml microcentrifuge collection tubes. The supernatants may still contain some particles.
22. Add 200 µl of Solution CD2 and vortex for 5 s.
23. Centrifuge the tubes at 15,000*g* for 1 min at room temperature. Avoiding the pellets, transfer ≤700 µl of each supernatant to clean microcentrifuge collection tubes.

    ▲ **CAUTION** The pellet contains non-DNA organic and inorganic material. For best DNA yields and quality, avoid transferring any of the pellet.
24. Add 600 µl of Solution CD3 and 600 µl of isopropanol and vortex for 5 s.
25. Load 650 µl of the lysate onto a UCP MinElute spin column and centrifuge at 15,000*g* for 1 min.
26. Discard the flow-through and repeat Step 25 by using the same UCP MinElute spin column until all of the lysate has passed through the column.
27. Carefully place the UCP MinElute spin column into a clean microcentrifuge collection tube.

    ▲ **CAUTION** Avoid splashing any flow-through onto the UCP MinElute spin column.
28. Add 500 µl of Solution EA to the UCP MinElute spin column and centrifuge at 15,000*g* for 1 min.
29. Discard the flow-through and place the UCP MinElute spin column back into the same microcentrifuge collection tube.

30. Add 500 µl of customized C5 wash buffer to the UCP MinElute spin column and centrifuge at 15,000*g* for 1 min.
31. Discard the flow-through and place the UCP MinElute spin column into a new microcentrifuge collection tube.
32. Centrifuge at 16,000*g* for 2 min. Carefully place the UCP MinElute spin column into a DNA LoBind 1.5-ml tube.
33. Carefully add 20 µl of Solution C6 to the center of the white filter membrane.

    ▲ **CRITICAL STEP** Ensure that the entire membrane is wet. This will result in a more efficient and complete elution of the DNA from the filter membrane.

    ▲ **CRITICAL STEP** DNA can be eluted in TE buffer without loss of yield, but note that the EDTA may inhibit downstream reactions such as PCR and automated sequencing. DNA may also be eluted in sterile, DNA-free PCR-grade water.
34. Centrifuge at 15,000*g* for 1 min. Discard the UCP MinElute spin column and retain the flow-through.
35. Quantify the DNA concentration of the flow-through by using a Qubit fluorometer and the Qubit high-sensitivity dsDNA quantification kit, following the manufacturer's instructions (https://tools.thermofisher.com/content/sfs/manuals/Qubit_dsDNA_HS_Assay_UG.pdf). The suggested minimum concentration is 2 ng/µl. We recommend performing qPCR according to the Supplementary Methods to check the amount of microbial DNA.

    ▲ **CAUTION** Because DNA is eluted in Solution C6 (10 mM Tris), it must be stored at −20 or −80 °C to prevent degradation.

    ■ **PAUSE POINT** The DNA is now ready for downstream applications. The tube with DNA can be stored in the ultra-freezer at −80 °C. We recommend storing these samples for ≤6 months.

    ◆ **TROUBLESHOOTING**

Library preparation

● **TIMING 6 h for 96 samples with the use of a multichannel pipette**

36. Add 2–30 µl of each DNA sample to a well of a 96-well PCR plate so that the total input amount is 100–500 ng of DNA and proceed by following the Illumina DNA prep reference guide (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/illumina_prep/illumina-dna-prep-reference-guide-1000000025416-10.pdf), with the following two modifications:
    - At the 'clean-up library step' stage, use 0.6× AMPure XP beads.
    - During the resuspension of the library pool, resuspend with one-quarter of the initial pool volume.

Sequencing

● **TIMING 2 d per sequencing run**

37. Sequence on the NovaSeq6000 sequencing system (average of 6.5 GB/sample) by following the manufacturer's instructions (https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/novaseq/1000000019358_16-novaseq-6000-system-guide.pdf). Run 384 indexed samples on four lanes of the flow cell S4.

Bioinformatic analysis

● **TIMING 5 d per food-processing facility (for 20 samples)**

▲ **CRITICAL STEP** Before performing the analysis, install scripts and software by using the following command: `conda install preprocessing -c fasnicar`

38. Pre-process the raw data as instructed in https://github.com/SegataLab/MASTER-WP5-pipelines/tree/master/02-Preprocessing. Run the pipeline through the preprocess.sh script by using the following command:

```
parallel -j NCPU 'preprocess.sh -i {} [other params]' ::: 'ls input_
folder'
```

where the `input_folder` contains the raw reads. The absolute pathway to the input folder (from /home) should be written (e.g., `/home/analysis/input_folder/`).
Some important optional parameters to use are:

- *-e*, extension of raw input files (default = '.fastq.gz')
- *-t* and *-b*, number of threads for trimgalore and bowtie2, respectively (depending on the computer or availability)
- *-x*, pathway to bowtie2 index files for the genomes to be removed from the data set, at least for the GRCh38.p13 human genome (GCF_000001405.39) and phiX174 (NC_001422).

Alternatively, trimgalore (https://github.com/FelixKrueger/TrimGalore) can be run independently of the proposed pipeline with the parameters *--nextera --stringency 5 --length 75 --quality 20 ′--max_n 2 --trim-n --dont_gzip --no_report_file --suppress_warn*. Bowtie2 can be run with the parameters *--sensitive-local --un*.

39. Run the assembly pipeline (https://github.com/SegataLab/MASTER-WP5-pipelines/tree/master/05-Assembly_pipeline) by running the command `pipeline_assembly.sh`.
▲ **CAUTION** The code assumes that inside a master folder with absolute path *pathReads=/path/${dataset_name}/reads*, there is a folder for each sample (named after the sample), which contains the files with the reads. The files are in fastq format and zipped with respective names *${samplename}_R1.fastq.bz2*, *${samplename}_R2.fastq.bz2*, *${samplename}_UN.fastq.bz2* (i.e., */path/${dataset_name}/reads/${samplename}/${samplename}_R1.fastq.bz2*).
Optionally, the six steps run automatically by `pipeline_assembly.sh` can be run independently, even adding modifications to adapt them to procedures normally used by each research group:
   - Step 1: perform assembly of reads in contigs by using MEGAHIT v1.1.124[33] with default parameters.
   - Step 2: filter contigs according to length by using the `filter_contigs.py` script (https://github.com/SegataLab/MASTER-WP5-pipelines/blob/master/05-Assembly_pipeline/filter_contigs.py), which by default removes those shorter than 1,000 bp.
   - Step 3: align filtered reads against filtered contigs by using bowtie2 v2.2.9, with *–very-sensitive-local –no-unal* parameters.
   - Step 4: find contig depth by *jgi_summarize_bam_contig_depths*, from MetaBAT2 v2.12.125[34].
   - Step 5: use MetaBAT2 v2.12.125[34] with *-m 1500 –unbinned –seed 0* parameters to compact contigs into bins/putative MAGs.
   - Step 6: use CheckM v1.0.726[35] with default parameters to verify completeness and contamination. Only high-quality (completeness >90%, contamination <5%) and medium-quality (completeness 50–90%, contamination <5%) MAGs are kept for further analysis, according to parameters previously proposed[7].
     ◆ **TROUBLESHOOTING**

Troubleshooting

Troubleshooting advice can be found in Table 1.

Table 1 | Troubleshooting table

Step	Problem	Possible reason	Solution
3	There is not enough surface to be sampled	The organization or structure of the industry is not exactly as expected. Small surfaces of special interest (such as knives or drains) are to be sampled	Where swabbing 1 m^2 is not possible (such as drains or knives), swabbing individual units (one drain or one knife) must be sufficient

Table 1 (continued) | Troubleshooting table

Step	Problem	Possible reason	Solution
35	Low concentration of eluted DNA; DNA concentration is recommended to be >2 ng/µl for optimal library preparation and sequencing	Poor cell lysis; cell wall structure of Gram-positive bacteria vary in thickness, quantity, length distribution and degree of cross-linking of the peptidoglycan, making some more difficult to lyse	After adding Solution CD1 (Step 14) and before the bead-beating step, incubate at 65 °C for 10 min and then resume the protocol from Step 18. As an alternative to the vortex adapter for the bead-based lysis, a TissueLyser II with appropriate adapter set facilitates a more comprehensive sample disruption of more samples simultaneously in a shorter time (suggested: 5 min at 25 Hz). Observe if and how the final yield is influenced for new standard samples
		An inadequate concentration of ethanol might decrease the DNA yield. Customized solution C5 (used in Step 29) is an ethanol-based solution that removes residual salts, humic acid and other contaminants while allowing the DNA to stay bound to the membrane of the column	Instead of using the customized C5 solution to wash the UCP MinElute cpin Column as described in Step 29, try to use the same volume of supplied Solution C5 or of 70% (vol/vol) ethanol. Observe if and how the final yield is influenced for new standard samples
		The eluted DNA is suspended in too great a volume of buffer	DNA may be concentrated by adding 3 µl of 3 M NaCl and flicking the tube for mixing. Next, add 20 µl of 100% cold ethanol and flick the tube for mixing. Incubate at –30 to –15 °C for 30 min and centrifuge at 10,000*g* for 5 min at room temperature. Decant all liquid. Briefly dry residual ethanol in a speed vacuum or at ambient air pressure. Avoid over-drying the pellet, or resuspension may be difficult. Resuspend precipitated DNA in the desired volume of Solution C6
39	Negative control samples show a large number of reads/contigs/MAGs with similar profiles to those of samples from the processing environments	Negative control samples might be contaminated with airborne microbes	Some bioinformatic tools can be applied for contaminant removal. For example, decontam[37] is a tool that identifies the contaminants on the basis of their frequency and/or prevalence in negative control samples over the 'real' ones. In addition, the software includes two algorithm functions, IsContaminant and IsNotContaminant, that should be applied when the real samples are high or low biomass (based on DNA yields), respectively. For the proper utilization of the tool, sequencing reads should be clustered into different features at strain level by using MetaPhlAn profiling[38]
	Samples have a high amount of 'unclassified' reads	High proportion of non-microbial reads (animal/plant host DNA)	An additional host-removal step can be performed by using the bowtie2 pipeline (Step 38) and the food animal or vegetal reference genome (e.g., *Sus scrofa* for some meat samples and *Bous taurus* for some cheese samples)

Timing

Steps 1–8, sampling: 1.5 h for 20 samples
Steps 9–13, sample pre-processing: 1.5 h for 20 samples
Steps 14–35, DNA purification: 4 h for 20 samples
Step 36, library preparation: 6 h for 96 samples
Step 37, sequencing: 2 d for each run
Steps 38–39, bioinformatic analysis: 5 d for 20 samples

Anticipated results

This protocol describes methods of sampling, DNA purification, sequencing and bioinformatic analysis for the characterization of the microbiome of food-processing environments through WMS. The sampling and DNA extraction procedures here described have been applied to many food-processing plants, with DNA concentrations of >10 ng/µl in 66.9% of sequenced samples and >0.5 ng/µl in 98.9% of sequenced samples, which is sufficient for library preparation without PCR amplification and subsequent sequencing on an Illumina Novaseq sequencer. We have been capable of generating from 0.2 to 81 Gbp of short-read data from a range of food-processing environments (not considering those samples with <1 million reads), which has allowed us to reconstruct a total of 9,564 MAGs from 807 samples (from 0 to 62 MAGs per sample, with >50% of the samples having >10 MAGs) (Fig. 1).

The sequencing reads, assembled contigs and MAGs obtained from the application of the protocol can be subjected to detailed taxonomic and functional analyses. Successful examples of the types of results that can be expected from such detailed analyses can be seen in a previous publication[20], where, among others, the results of a principal coordinates analysis of the taxonomic composition of samples, a phylogenetic tree of the reconstructed MAGs and boxplots showing the abundance of virulence factor genes in different sample categories can be observed.

Reporting summary

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

Data availability

Raw reads are available on the Sequence Read Archive of the NCBI under the BioProjects numbers PRJNA897099 (for vegetable facilities), PRJNA941197 (for an ice-cream facility), PRJNA997800 (for meat facilities), PRJNA997821 (for cheese facilities, except those located in Ireland) and PRJNA996188 for control samples. Raw reads for fish-processing factories and Irish cheese factories are available on the European Nucleotide Archive database under the accession numbers PRJEB62794 and PRJEB63604, respectively.

Code availability

The code used for raw reads filtering, assembly and binning into MAGs is available at https://github.com/SegataLab/MASTER-WP5-pipelines.

Received: 23 March 2023; Accepted: 9 November 2023;
Published online: 24 January 2024

References

1. Møretrø, T. & Langsrud, S. Residential bacteria on surfaces in the food industry and their implications for food safety and quality. *Compr. Rev. Food Sci. Food Saf.* **16**, 1022–1041 (2017).
2. De Filippis, F., Valentino, V., Alvarez-Ordóñez, A., Cotter, P. D. & Ercolini, D. Environmental microbiome mapping as a strategy to improve quality and safety in the food industry. *Curr. Opin. Food Sci.* **38**, 168–176 (2021).
3. Alvarez-Ordóñez, A., Coughlan, L. M., Briandet, R. & Cotter, P. D. Biofilms in food processing environments: challenges and opportunities. *Annu. Rev. Food Sci. Technol.* **10**, 173–195 (2019).
4. Fagerlund, A., Langsrud, S. & Møretrø, T. Microbial diversity and ecology of biofilms in food industry environments associated with *Listeria monocytogenes* persistence. *Curr. Opin. Food Sci.* **37**, 171–178 (2021).
5. Koutsoumanis, K. et al. Whole genome sequencing and metagenomics for outbreak investigation, source attribution and risk assessment of food-borne microorganisms. *EFSA J.* **17**, e05898 (2019).
6. Yap, M. et al. Next-generation food research: use of meta-omic approaches for characterizing microbial communities along the food chain. *Annu. Rev. Food Sci. Technol.* **13**, 361–384 (2022).
7. Capouya, R. D., Mitchell, T., Clark, D. I., Clark, D. L. & Bass, P. D. A survey of microbial communities on dry-aged beef in commercial meat processing facilities. *Meat Muscle Biol.* **4**, 1–11 (2020).
8. Zwirztz, B. et al. The sources and transmission routes of microbial populations throughout a meat processing facility. *NPJ Biofilms Microbiomes* **6**, 26 (2020).
9. Pasolli, E. et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* **176**, 649–662.e20 (2019).
10. Xie, F. et al. An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants. *Microbiome* **9**, 137 (2021).
11. Olm, M. R. et al. Genome-resolved metagenomics of eukaryotic populations during early colonization of premature infants and in hospital rooms. *Microbiome* **7**, 26 (2019).
12. Lind, A. L. & Pollard, K. S. Accurate and sensitive detection of microbial eukaryotes from whole metagenome shotgun sequencing. *Microbiome* **9**, 58 (2021).
13. Santos-Medellin, C. et al. Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. *ISME J.* **15**, 1956–1970 (2021).
14. Nayfach, S. et al. Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nat. Microbiol.* **6**, 960–970 (2021).
15. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* **35**, 833–844 (2017).
16. Kim, H., Kim, M., Kim, S., Lee, Y. M. & Shin, S. C. Characterization of antimicrobial resistance genes and virulence factor genes in an Arctic permafrost region revealed by metagenomics. *Environ. Pollut.* **294**, 118634 (2022).
17. Picone, N. et al. Metagenome assembled genome of a novel verrucomicrobial methanotroph from Pantelleria Island. *Front. Microbiol.* **12**, 666929 (2021).
18. Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
19. Yilmaz, S., Allgaier, M. & Hugenholtz, P. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat. Methods* **7**, 943–944 (2010).
20. Valentino, V. et al. Evidence of virulence and antibiotic resistance genes from the microbiome mapping in minimally processed vegetables producing facilities. *Food Res. Int.* **162**, 112202 (2022).
21. McHugh, A. J. et al. Microbiome-based environmental monitoring of a dairy processing facility highlights the challenges associated with low microbial-load samples. *NPJ Sci. Food* **5**, 4 (2021).
22. Cobo-Díaz, J. F. et al. Microbial colonization and resistome dynamics in food processing environments of a newly opened pork cutting industry during 1.5 years of activity. *Microbiome* **9**, 204 (2021).
23. Bruggeling, C. E. et al. Optimized bacterial DNA isolation method for microbiome analysis of human tissues. *Microbiologyopen* **10**, e1191 (2021).
24. Hinlo, R., Gleeson, D., Lintermans, M. & Furlan, E. Methods to maximise recovery of environmental DNA from water samples. *PLoS ONE* **12**, e0179251 (2017).
25. Yap, M., O'Sullivan, O., O'Toole, P. W. & Cotter, P. D. Development of sequencing-based methodologies to distinguish viable from non-viable cells in a bovine milk matrix: a pilot study. *Front. Microbiol.* **13**, 1036643 (2022).
26. Chng, K. R. et al. Cartography of opportunistic pathogens and antibiotic resistance genes in a tertiary hospital environment. *Nat. Med.* **26**, 941–951 (2020).
27. Echeverría-Beirute, F., Varela-Benavides, I., Jiménez-Madrigal, J. P., Carvajal-Chacon, M. & Guzmán-Hernández, T. eDNA extraction protocol for metagenomic studies in tropical soils. *Biotechniques* **71**, 580–586 (2021).

28. Jiang, W. et al. Optimized DNA extraction and metagenomic sequencing of airborne microbial communities. *Nat. Protoc.* **10**, 768–779 (2015).
29. Fadiji, A. E. & Babalola, O. O. Metagenomics methods for the study of plant-associated microbial communities: a review. *J. Microbiol. Methods* **170**, 105860 (2020).
30. Cabello-Yeves, P. J. et al. The microbiome of the Black Sea water column analyzed by shotgun and genome centric metagenomics. *Environ. Microbiome* **16**, 5 (2021).
31. Keeratipibul, S. et al. Effect of swabbing techniques on the efficiency of bacterial recovery from food contact surfaces. *Food Control* **77**, 139–144 (2017).
32. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
33. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
34. Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
35. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
36. Karstens, L. et al. Controlling for contaminants in low-biomass *16S rRNA* gene sequencing experiments. *mSystems* **4**, 2379–5077 (2019).
37. Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226 (2018).
38. Blanco-Míguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).

Acknowledgements

This work was funded by the European Commission under the European Union's Horizon 2020 research and innovation program under grant agreement no. 818368 (MASTER). C.B. is grateful to Junta de Castilla y León and the European Social Fund for awarding her a pre-doctoral grant (BOCYL-D-07072020-6). A.P. is grateful to Ministerio de Ciencia e Innovación for awarding her a pre-doctoral grant (PRE2021-098910). N.M.Q. is currently funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 101034371. We thank AV Star Systems for their role in creating the Supplementary Video, and M. Coakley and S. Mortensen for their help in its preparation.

Author contributions

M.L., M.P., D.O'N., V.T.M., M.W., A.M., N.S., P.D.C., D.E. and A.A.-O. conceived the study and obtained the funding. J.F.C.-D., C.B., F.D.F., V.V., R.C.R., I.C.-T., C.S., S.D., P.R.-M., N.M.Q., M.D., S.S., S.K. and A.P. performed the samplings at food-processing facilities. D.O'N. and L.M.d.S. designed and tested the improvements in the DNA-extraction protocol, and C.B., F.D.F., V.V., R.C.R. and A.P. tested the different versions of the DNA-extraction protocol for optimization. C.B., F.D.F., R.C.R., I.C.T., C.S., S.D., P.R.-M., N.M.Q., M.D., S.S., S.K. and A.P. applied the improved DNA-extraction protocol on samples from the food industry. F.A., F.P. and N.S. sequenced the extracted DNA. N.C., A.B.-M. and F.P. performed the bioinformatic analyses. J.F.C.-D., F.D.F., V.V., R.C.R., N.C., C.S. and N.M.Q. collated all the information. L.M.d.S., J.F.C.D. and C.B. prepared the figures. J.F.C.D., C.B. and A.A.-O. wrote the manuscript with input from all the authors. All authors read and approved the final manuscript.

Competing interests

D.O'N. and L.M.d.S. are employees of QIAGEN GmbH. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41596-023-00949-x.

Correspondence and requests for materials should be addressed to Avelino Alvarez-Ordóñez.

Peer review information *Nature Protocols* thanks Lena Florl, Andrea Moreno Switt, Bernard Taminiau and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Related links

Key reference using this protocol

Valentino, V. et al. *Food Res. Int.* **162**, 112202 (2022): https://doi.org/10.1016/j.foodres.2022.112202

[1]Department of Food Hygiene and Technology and Institute of Food Science and Technology, Universidad de León, León, Spain. [2]Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy. [3]Task Force on Microbiome Studies, University of Naples Federico II, Naples, Italy. [4]Teagasc Food Research Centre, Moorepark, Cork, Ireland. [5]QIAGEN GmbH, Hilden, Germany. [6]Department of Cellular, Computational and Integrative Biology, University of Trento, Trento, Italy. [7]Dairy Research Institute of Asturias, Spanish National Research Council (IPLA-CSIC), Paseo Río Linares, Villaviciosa, Asturias, Spain. [8]Health Research Institute of Asturias (ISPA), Avenida Hospital Universitario, Oviedo, Asturias, Spain. [9]Austrian Competence Centre for Feed and Food Quality, Safety and Innovation, FFoQSI GmbH, Tulln an der Donau, Austria. [10]Department for Farm Animals and Veterinary Public Health, Unit of Food Microbiology, Institute of Food Safety, Food Technology and Veterinary Public Health, University of Veterinary Medicine Vienna, Vienna, Austria. [11]Department of Microbiology and Genetics, Institute for Agribiotechnology Research (CIALE), University of Salamanca, Salamanca, Spain. [12]Microbiology Research Group, Matís ohf., Reykjavík, Iceland. [13]Senckenberg Biodiversity and Climate Research Centre, Frankfurt, Germany. [14]Faculty of Food Science and Nutrition, University of Iceland, Reykjavik, Iceland. [15]APC Microbiome Ireland and VistaMilk Research Centres, Cork, Ireland. [16]These authors contributed equally: Coral Barcenilla, José F. Cobo-Díaz.

nature portfolio

Corresponding author(s): *Double-blind peer review submissions: write DBPR and your manuscript number here instead of author names.*

Last updated by author(s): *YYYY-MM-DD*

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	No software was used
Data analysis	Trim Galore v0.6.6; Bowtie2 v2.2.9; MEGAHIT v1.1.1; MetaBAT2 v2.12.1; CheckM v1.0.7; https://github.com/SegataLab/MASTER-WP5-pipelines/02-Preprocessing; https://github.com/SegataLab/MASTER-WP5-pipelines/tree/master/05-Assembly_pipeline

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw reads are available on the Sequence Read Archive of the National Center of Biotechnology Information (NCBI) under the BioProjects numbers PRJNA897099, PRJNA941197, PRJNA997800, PRJNA997821and PRJNA996188; and on the European Nucleotide Archive database under the accession number PRJEB62794 and PRGEB63604.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

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

Field-specific reporting

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

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

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

Ecological, evolutionary & environmental sciences study design

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

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	A total of 931 samples from processing environments from 114 food facilities from different production sectors were collected.
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	Data was generated by whole metagenome sequencing of DNA extracted from samples.
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☒ Yes □ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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