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

Coral Barcenilla[1,16*], Jose F. Cobo-Diaz[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†,142], Martin Wagner[9,10], Abelardo Margolles[7,8], Nicola Segata[6], Paul D. Cotter[4,153], Danilo Ercolini[2,3], Avelino Alvarez-Ordóñez[1] ✉

[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, Napoli, 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]Current affiliation: Department of Microbiology and Genetics, Institute for Agribiotechnology Research (CIALE), University of Salamanca, Salamanca, Spain

[12†]Microbiology Research Group, Matís ohf., Reykjavík, Iceland

[13Present address:] Senckenberg Biodiversity and Climate Research Centre, Frankfurt, Germany

[1,2]Faculty of Food Science and Nutrition, University of Iceland, Reykjavik, Iceland

[1,5]APC Microbiome Ireland and VistaMilk Research Centres, Ireland

[1,6*] These authors contributed equally

[†]Current affiliation: Department of Microbiology and Genetics, Institute for Agribiotechnology Research (CIALE), University of Salamanca, Salamanca, Spain

[††]Present address: Senckenberg Biodiversity and Climate Research Centre, Frankfurt, Germany
[*] These authors contributed equally

✉ e-mail: aalvo@unileon.es

Editorial summary

This protocol describes a method for sampling the microbiome of food processing facilities and analyzing it using whole metagenome sequencing. The protocol includes sampling, and DNA extraction and purification steps optimized for low-biomass samples.

Proposed tweet

#NewNProt for sampling and analysing the microbiome of food processing facilities, tailored to low-biomass samples. @MetaResistantB @MASTER_IA_H2020

Proposed teaser

Mapping the microbiome of food processing sites

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 analysing them through whole metagenomic sequencing (WMS).

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

[H1] Abstract

The deep investigation of the microbiome of food production and 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 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 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 approximately 5 days including both laboratory and computational work. This ~~significantly~~ markedly improves on results previously obtained in studies performing WMS of processing environments using other protocols not specifically developed to sequence these types of samples, where less than 2 MAGs were obtained per sample. The full protocol has been developed and applied in the frame of the EU project MASTER (Microbiome applications for sustainable food systems through technologies and enterprise) in 114 food facilities from different production sectors.

[H1] 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 employed 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 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 as they rely on the selective enrichment and/or isolation of specific culturable microbes, which only represent a minority 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 characterisation, and for 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 characterisation of the microbiome of food production and processing environments relied on amplicon-based approaches, where a gene of taxonomic relevance — for example, the *16S rRNA* gene from bacteria and archaea or ITS2 regions from fungi — is amplified using PCR from total microbial DNA directly extracted from samples (also known as "metataxonomics" or "amplicon sequencing")[7,8]. However, this only gives information 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 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 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 whole metagenome sequencing 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 harbour 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.

[H2] 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 around 10 ng to more than 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 as 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 EU project MASTER (Microbiome applications for sustainable food systems through technologies and enterprise; https://www.master-h2020.eu/) by six partner institutions ~~in~~ to 114 food ~~companies~~ processing environments from different production sectors (86 dairy, 19 meat, 6 fish, 3 ready-to-eat vegetables and 1 ice-cream processing facilities). It has been also used in a recent study characterizing the microbiome of ~~industries companies~~sites processing 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 more than 1 million reads. For those samples that failed sequencing, possible reasons were low

DNA concentration (<0.1 ng/µL), library preparation failed; or sequencing did not generate at least 10^6 reads (**Supplementary figure 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, mMost of these samples failing sequencing (94.1%) had < 10 ng/µL (**Supplementary figure 1B**). The mean DNA concentration obtained from successfully sequenced samples was of 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) (**Figure 1**). These results demonstrate the success of the approach for the deep characterization of the microbiome of low biomass environments. Our protocol significantly markedly improves on results previously obtained in studies performing WMS of processing environments 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 characterising meat processing environments[22].

[H2] Applications

Although our focus is on swab samples from food production and processing environments, we envisage that the protocol will be also appropriate for microbiome monitoring activities in other built environments, such as hospitals or households, and also for analyzing other similar environmental surface samples with low microbial biomass such as those from urban environments. Moreover, in principle, the protocol could also be used for other different sample types, such as food or water samples, but 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 be also 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 ~~as well~~.

[H2] 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 are being studied, such as the use of propidium monoazide (PMA) and ethidium monoazide (EMA) treatments[25], 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 ~~to a certain extent~~ in ~~food~~ environmental samples if the surfaces ~~still have some rests~~ have contamination from workers or traces of food or derived organic material. Additionally, 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 ~~some of the~~a subset of specific microorganisms ~~present~~, 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 on 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 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, in order to obtain reliable results, a higher amount of high-quality DNA is required, as no DNA amplification step is used, unlike in metataxonomy approaches[2]. Additionally, 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 (~~around~~ approximately 3 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.

[H2] 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] or 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 resulted in low-quality sequencing results or even in failed library preparation for sequencing, as demonstrated by the study by Cobo-Diaz and colleagues[22], where a low amount of reads (less than 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, as the DNA loads, number of contigs and number of MAGs obtained with it significantly exceeds those previously described in the literature, applying different procedures to similar sample types.

[H2] Experimental design

Here, we describe our protocol for an improved sampling and extraction of DNA for WMS from food processing environments (**Figure 2**), as well as our workflow for sequencing and bioinformatic analysis. Specifically, we describe our methods for sample collection, manipulation and storage (Steps 1-1322), microbial cell lysis and DNA purification (Steps 1423-3544), library preparation and sequencing (Steps 3645 and 3746) and bioinformatic analysis (Steps 3847 and 3948).

[H3] Sampling, sample manipulation and storage (Steps 1-1322)

We recommend preparing a detailed sampling plan where 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 timepoints (during production, after production, before and just after cleaning and sanitation, etc) can be more suitable to answer other biological questions. ~~Thus, for instance~~For example, investigations evaluating the efficacy of particular sanitation regimes would require restricted samplings immediately before and after the intervention is applied. In order 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 be then 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; food contact surfaces and non-food contact surfaces. The selection of sampling points could take into account areas that are likely to be contaminated, such as wet areas, hard to reach places, and ~~poorly cleanable~~ 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, ~~, in the case of special events (e.g., construction),~~ outbreaks ~~investigations,~~ or ~~following~~ non-conformities in conventional microbiological analyses of foods ~~, a specific sampling plan could be developed, including intensified~~

samplings, to investigate the potential presence of harbourage niches in the facility or to assess how far the contamination is has spread.

An aspect of primary importance is the choice of the type of swab and the swabbing procedure. The use of sponge swabs is recommended as 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 a rayon tip that is made of fibres wrapped around a plastic rod, or those made of synthetic materials, such as polyester, polyurethane or nylon. Cellulose-derived swabs tend to trap bacterial cells within the fibre matrix, thus hampering the release of the 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. Additionally, 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, walls, etc), we recommend sampling a ~1 m^2 surface, by swabbing surfaces first horizontally and then vertically, turning the swab around in between. For other types of surfaces, where swabbing ~1 m^2 may not be possible (e.g. drains, knives), we recommend swabbing individual units (e.g. 1 drain, 1 knife). To sample the operators, consider swabbing the hands/gloves, aprons, caps and/or shoes (**Supplementary Video 1; 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 taken, it is important that samples are kept refrigerated (for instance using a portable cooler filled with ice packs) until processing in the laboratory, which should ideally take place

~~within the next~~less than 24 hours after sampling. Alternatively, samples could be snap-frozen in liquid nitrogen or, where this is not possible, placed on dry ice prior to long-term storage frozen (ideally at -80ºC) until sample processing.

For cell recovery from the swabs, we recommend the addition of a small volume of sterile phosphate buffered saline (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, alternatively, we recommend the storage of the cell pellet until use at -80ºC. We recommend storage at -80ºC for both samples and/or extracted DNA since 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.

[H3] Microbial cell lysis and DNA purification (Step 14~~23~~-35~~44~~)

The cell pellets collected from ~~food company~~industry the surfaces of food processing sites ~~is~~ 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 as the approach followed can impact the observed microbial diversity, which can be a limitation in this type of metagenomics workflow. Here, the DNeasy PowerSoil Pro kit (QIAGEN, Hilden, Germany) was used as the basis for development of a modified protocol.

Lysis of microbial cells for DNA purification is usually achieved either through 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)~~-methods~~. Post 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 (**Figure 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 (**Figure 4A**). Moreover, the QIAamp UCP MinElute columns are treated through a physical process in order 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 (**Figure 4B**), though 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 (**Figure 4C**, 4**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 obtained per extraction and evaluate the performance of the DNA extraction procedures tested.

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, alternatively, we recommend the storage of the DNA sample until use at -80ºC.

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 DNA. However, we have found that three samples with even less DNA yields have been successfully sequenced.

[H3] Library preparation and sequencing (Step 3645-3746)

The library preparation for Illumina NovaSeq metagenomic sequencing is based on the Illumina DNA Prep Kit 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 using dual indexing and sequenced for 150 bp paired-end reads (average of 6.5 GB/sample) on the NovaSeq 6000 Sequencing System.

[H3] Bioinformatic analysis (Step 3847-3948)

Sequenced metagenomic reads are quality-controlled using a pre-processing pipeline available at https://github.com/SegataLab/MASTER-WP5-pipelines/tree/master/02-Preprocessing. Firstly, sequencing adapters, reads of low quality (Phred score < 20), short reads (<75 bp), and reads with more than 2 ambiguous nucleotides are removed using Trim Galore (v0.6.6) (https://github.com/FelixKrueger/TrimGalore). Then, contaminant DNA is identified using Bowtie2 version 2.2.9 (with --sensitive-local parameter)[32], removing reads from the phiX174 Illumina spike-in (NCBI accession number NC_001422) as well as potential human contamination (using the GRCh38.p13 human genome, NCBI accession number GCF_000001405.39). Additionally, genome contamination with non-microbial DNA from other different origins (e.g. animal or plant DNA from particular host species) can be removed 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.

In order 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). ~~Shortly~~Briefly, contigs are assembled from the metagenomic reads using MEGAHIT version 1.1.1[33] with default parameters. Contigs longer than 1000 nt are then binned 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 using CheckM version 1.0.7[35] with default parameters. In order 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, and many other more advanced, bioinformatic analyses, many tutorials are available on bioBakery at https://github.com/biobakery/biobakery.

[H3] Controls (Steps ~~7~~16, ~~9~~18 and 18~~27~~)

It is recommended to include both positive and negative control samples alongside the samples from food processing environments being analysed. As 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. It is highly recommended to include different dilutions of the mock community (e.g. 10^{-6}, 10^{-4} and 10^{-2} cells/mL) in order to produce positive samples with diverse DNA concentration and thus get more ~~interesting~~ complete information on potential contaminants coming from sample manipulation and materials used[36]. As 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, library preparation can be completed and sequencing reads are 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 based on 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 one minute to the air of the processing plant (negative control – industry) or of the laboratory where samples were manipulated and DNA extracted (negative control – laboratory). Due to 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 (**Figure 1**).

It is also recommended to include negative controls for the DNA extraction step 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.

[H1] Materials

[H2] Sampling materials

- Whirl-Pak B01592WA Hydrated PolyProbe™ Sampling Bags with Sampling Sponges and 8" probe, 24 oz, sterile; 100/box (hydrated with 10 mL of HiCap™ Neutralizing Broth)
- Portable cooler
- Ice packs
- Personal protective equipment (PPE) for sampling, including disposable masks, disposable coats, disposable caps, disposable shoes and gloves

[H2] Laboratory reagents (samples preprocessing and DNA purification)

- Phosphate buffered saline (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, 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% (v/v) sodium hypochlorite.

- ZymoBIOMICS Microbial Community Standard (Zymo Research, Cat. No. D6300)
- Isopropanol (for example: Sigma-Aldrich, Cat. No. I9516)
- Ethanol 100% (for example: Sigma-Aldrich, Cat. No. 1.07017)
- Qubit High Sensitivity double-stranded DNA (dsDNA) quantification kit (Invitrogen, Cat. No. Q32851)

[H2] Laboratory reagents (library preparation)

- Illumina DNA Prep Kit (Illumina, Cat. No. 20018705).
- Nuclease-free water

[H2] Equipment

- P1, P10, P100, P1000 and 10 mL pipettes
- 1.5 mL sterile eppendorf tubes
- 15 mL sterile plastic tubes
- DNA LoBind Tubes (for example: Eppendorf, Cat. No. 0030108051)
- DNeasy PowerSoil Pro Kit (Qiagen, Cat. No. 47016). The following materials from the kit will be used: PowerBead Pro Tubes, 2 mL microcentrifuge collection tubes
- QIAamp UCP DNA Micro Kit ~~Min Elute spin columns, (~~Qiagen, Cat. No. 1103588). The MinElute spin columns from this kit will be used in the procedure.

▲ **CRITICAL STEP** Using UCP MinElute columns is critical in order to reduce background DNA amounts when working with low biomass samples.

- 96-well PCR plates
- Microseal 'B' adhesive seal
- 1.7 ml microcentrifuge tubes (for example: Sigma-Aldrich, Cat. No. CLS3620)
- 8-~~PCR~~strip PCR tube~~s strip~~
- 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) (for example: Thermo Fisher Scientific, Cat. No. AB0859)
- Microseal 'F' foil seal (for example: Bio-Rad, Cat. No. MSF1001)
- Stomacher (for example: IUL Instruments, Cat. No. 9000400)
- Vortex with adapter for 1.5-2 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) can be used.
- Centrifuge(s) for 1.5ml and 15 mL tubes
- Laminar flow hood
- Ultra-freezer (-80ºC)
- Qubit Fluorometer (Thermo Fisher Scientific, Carlsbad, California, United States)
- Thermal cycler (for library preparation)
- Illumina NovaSeq 60000 sequencer (Illumina, Inc., San Diego, California, United States)

[H2] Reagent setup

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

[H3] Customized wash buffer C5

Prepare a mix of N x (500 µL solution C5 + 333 µL EtOH(100%)), w~~W~~here N is the number of samples ~~(cell pellets) to be~~that will be processed for DNA extraction ~~at a time, prepare a mix of N x (500 µL solution C5 + 333 µL EtOH(100%))~~.

[H3] PBS

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

[H2] Equipment setup

[H3] Sampling plan

The day before sampling, ~~it is important to~~ define and document the sampling plan that will be followed. See Figure 3 for an example sampling plan.

[H3] Sampling materials

~~and to perform the following preparatory work:~~

Open the boxes containing the Whirl-Pak B01592WA Hydrated PolyProbe™ Sampling Bags and organize them in groups of 5 bags (using the yellow strip of one of the bags to keep the 5 of them grouped). Label the first bag, using a permanent marker, with the sample code to be collected. Repeat this until all 5 bags-groups are properly labelled.

Prepare the portable cooler and PPE (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).

[H1] Protocol

[H2] Sampling of the food processing facility ● Timing 1.5 h per food processing facility (for collecting 20 composite samples) (plus travel time) ~~(for collecting 20 composite samples)~~

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

1. Put on a new set of gloves and rub them with hand sanitizer before starting sampling.

2. Locate the first surface you are going to sample and take ~~the~~ a corresponding pre-labelled Whirl-Pak B01592WA Hydrated PolyProbe™ swab bag. Prepare the swab as follows:

• ~~3.~~ Keeping the Whirl-Pak B01592WA Hydrated PolyProbe™ swab bag in a vertical position, ~~and~~ 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. ~~Take care not to drop the liquid from the bag.~~

• ~~4. Take~~ 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.

• ~~5.~~ 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. ~~6. Place the sponge over the~~Sample the surface ~~to be sampled and slide the swab vigorously following horizontal, and vertical movements, as explained under the points below, to cover ~1 m^2 of surface~~as follows:~~.~~

~~? TROUBLESHOOTING~~

• ~~7.~~ Rub the swab (by one of its sides) slowly on the surface to be sampled by doing horizontal movements, covering a ~1 m^2 area.~~.~~

• ~~8.~~ Rotate the swab in order to use the other side of the sponge and proceed to sample the same surface area again using. ~~The movements will be~~ vertical ~~this time~~movements.

? TROUBLESHOOTING

4~~9~~. Once the swabbing is completed, return the swab to the plastic bag. Take care not to touch any other surface with the sponge and. ~~Introduce the sponge into the bag.~~ k~~K~~eep holding the stick with one hand.

10. With the other hand, separate the stick from the sponge carefully, by unscrewing and .

11. dDiscard the stick.

512. Repeat steps 12-4 to 11 pooling the swabs in the same bag until until you pool 5 swabs are collected in one a single bag (the bags from the second to fifth swab can be discarded).

613. Close the bag with the five swabs hermetically. For this, first rSqueezemove the air from inside the bag, manually. Then, roll down the top of the bag and then use the yellow strips to hermetically close seal the bag.

14.KeepPlace the hermetically closed swab bag in a vertical position inside into the portable cooler filled with ice packs.

15.Discard the gloves.

176. Repeat steps 1-615 for the each of the different sample categories included in the sampling plan.

▲ CRITICAL: It is highly recommended to collect negative control samples. For this, expose the swabs for 1 minute to the air in the food processing facility.

817. Introduce all pooled sample bags (containing 5 sponge-swabs per bag) into Transport the portable cooler filled with ice packs for transportsamples to the laboratory.

[H2] Sample pre-processing ● Timing 1.5 h per food processing facility (for 20 samples)

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

▲ CRITICAL: Samples should be processed within the next 24 hours after sampling. Alternatively, samples could be snap-frozen in liquid nitrogen or, where this is not possible, placed on dry ice prior to long-term storage frozen (ideally at -80ºC) until sample processing.

▲ CRITICAL At this point, it is highly recommended to collect negative control samples in the laboratory where the samples will be pre-processed. To do this, expose Whirl-Pak B01592WA Hydrated PolyProbe™ swabs for 1 minute to the air of the laboratory. Subsequently Negative control swabs can be pooled and then pre-processed s explained below for the industry samplesaccording to the steps detailed for the industry samples below.

189. Move the samplings bags to a laminar flow hood. In the hood, carefully open the first sampling bag, add 10ml of sterile PBS, and close it again. Repeat for each sampling bag.

1910. Homogenize each bag in the stomacher at 175 rpm for 2 minutes.

2011. In the laminar flow hood, carefully open each sampling bag, recover 10 mL of homogenized liquid using a pipette, and transfer it to a sterile 15 mL plastic tube.

▲ CRITICAL Since the sponge swabs can retain liquid, it is necessary to gently squeeze the sponges from outside the sampling bag while pipetting to facilitate the release of the liquid from the sponges.

2112. Centrifuge at 5,000 x g for 5 min at room temperature (20-25ºC).

22213. Carefully discard the supernatant and keep the tube with the cell pellet;. Bear in mindnote that some pellets might be very small.

■ **PAUSE POINT** The tube with the cell pellet can be stored in the ultra-freezer at -80ºC 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 x g for 5 min at room temperature, the supernatant discarded and the tube with the cell pellet stored at -80ºC.

[H2] DNA purification ● Timing 4 h (for 20 samples)

~~23~~14. Thaw the tubes with the cell pellets for 15 min at room temperature.

~~24~~15. Add 800 µL of Solution CD1 to ~~the~~ each cell pellet and resuspend ~~it~~ by pipetting up and down.

~~25~~16. Spin the PowerBead Pro tubes briefly to ensure that the beads have settled at the bottom.

! CAUTION It is important to use a centrifuge where the PowerBead Pro tubes rotate freely without rubbing.

~~26~~17. Transfer the complete CD1 suspensions to ~~a~~ fresh PowerBead Pro Tubes.

~~27~~18. At this step, adding a positive control~~,~~ such as the ZymoBIOMICS Microbial Community Standard~~, is highly recommended~~. ~~The mock community should be diluted~~Diluted the mock community (e.g., 10^{-6}, 10^{-4} and 10^{-2} cells/mL) and add 20 µL of ~~the~~ each of the corresponding dilutions ~~can be added respectively~~ to PowerBead Pro Tubes with 800 µL of Solution CD1. ~~Also, a~~Adding a new negative control sample is also highly recommended~~:~~. ~~T~~The negative control of the DNA purification step can be prepared by adding 800 µL of Solution CD1 to an empty PowerBead Pro Tube.

~~28~~19. Secure the PowerBead Pro Tubes horizontally on a Vortex Adapter for 1.5–2 mL tubes in the Vortex-Genie 2. Vortex at maximum speed for 10 min.

! CAUTION When using the Vortex Adapter for more than 12 preps 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.

~~29~~20. Centrifuge the PowerBead Pro Tubes at 15,000 x g for 1 min.

~~30~~21. Transfer the supernatants (~500–600 µL) to ~~a~~ clean 2 mL microcentrifuge collection tubes. The supernatants may still contain some particles.

~~31~~22. Add 200 µL of Solution CD2 and vortex for 5 s.

~~32~~23. Centrifuge tubes at 15,000 x g for 1 min at room temperature. Avoiding the pellets, transfer up to 700 µL of each supernatant to ~~a~~ 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.

~~33~~24. Add 600 µL of Solution CD3 and 600 µL of 100% isopropanol and vortex for 5 s.

~~34~~25. Load 650 µL of the lysate onto an UCP Min Elute Spin Column and centrifuge at 15,000 x g for 1 min.

~~35~~26. Discard the flow-through and repeat step 2~~5~~ using the same UCP Min Elute Spin Column, until all of the lysate has passed through the column.

697~~36~~27. Carefully place the UCP Min Elute Spin Column into a clean microcentrifuge collection
698tube.

699**! CAUTION** Avoid splashing any flow-through onto the UCP Min Elute Spin Column.

700~~37~~28. Add 500 µL of Solution EA to the UCP Min Elute Spin Column and centrifuge at 15,000
701x g for 1 min.

702~~38~~29. Discard the flow-through and place the UCP Min Elute Spin Column back into the same
703microcentrifuge collection tube.

704~~39~~30. Add 500 µL of customized C5 wash buffer to the UCP Min Elute Spin Column and
705centrifuge at 15,000 x g for 1 min.
706
707~~40~~31. Discard the flow-through and place the UCP Min Elute Spin Column into a new
708microcentrifuge collection tube.

709~~41~~32. Centrifuge at 16,000 x g for 2 min. Carefully place the UCP Min Elute Spin Column into
710a DNA LoBind 1.5 mL Tube.

711~~42~~33. Carefully add 20 µL of Solution C6 to the center of the white filter membrane.

712**▲ CRITICAL STEP** ~~It is important to visually make~~En~~-~~sure the entire membrane is wet. This
713will result in a more efficient and complete ~~release~~ elution of the DNA from ~~of~~ the filter
714membrane.

715**▲ CRITICAL STEP**: DNA can be eluted in TE buffer without loss of yield, but note that the
716EDTA may inhibit downstream reactions such as PCR and automated sequencing. DNA may
717also be eluted in sterile, DNA-free PCR-grade water.

718~~43~~34. Centrifuge at 15,000 x g for 1 min. Discard the UCP Min Elute Spin Column and retain
719the flow-through.

720~~44~~35. Quantify the DNA concentration of the flow-through by using a Qubit Fluorometer and
721the Qubit High Sensitivity double-stranded DNA (dsDNA) quantification kit, following the
722manufacturer~~'~~'s instructions
723(https://tools.thermofisher.com/content/sfs/manuals/Qubit_dsDNA_HS_Assay_UG.pdf). The
724suggested minimum concentration is 2 ng/µL. ~~In addition,~~We recommend performing ~~a~~ qPCR
725according to the **Supplementary Methods** ~~is recommended~~ to check the amount of microbial
726DNA.

727**? TROUBLESHOOTING**

728**■ PAUSE POINT** The DNA is now ready for downstream applications. The tube with DNA
729can be stored in the ultra-freezer at -80ºC. We recommend storing these samples no longer than
7306 months.
731
732**! CAUTION** As DNA is eluted in Solution C6 (10 mM Tris), it must be stored at -20 ºC or -
73380ºC to prevent degradation.
734
735**[H2] Library preparation ● Timing 6 h for 96 samples using a multichannel pipette**

736~~45~~36. Add 2–30 µL of each DNA sample to a well of a 96-well PCR plate so that the total input
737amount is 100–500 ng DNA and proceed following the Illumina DNA Prep reference guide
738(https://support.illumina.com/content/dam/illumina-support/documents/documentation/
739chemistry_documentation/illumina_prep/illumina-dna-prep-reference-guide-1000000025416-
74010.pdf), with the following two modifications:

- **- At the “clean-up library step” stage, use 0.6x AMPure XP beads**

- **–During the resuspension of the library pool, re-suspend with ¼ of the initial pool volume.**

[H2] Sequencing ● Timing 2 days per sequencing /run

4637. Sequence on the NovaSeq6000 Sequencing System (average of 6.5GB/sample) 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 4 lanes of the flow cell S4.

[H2] Bioinformatic analysis ● Timing 2 days per /sample

4738. Pre-process the raw data as instructed in https://github.com/SegataLab/MASTER-WP5-pipelines/tree/master/02-Preprocessing. Run the pipeline throught the preprocess.sh script ~~by typing~~using the following command:

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

~~.~~Where the input folder should contain the raw reads and the absolute pathway (from /home) should be written. 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 indexes 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)

! CAUTION Previously, you need to install scripts and software by: *conda install preprocessing -c fasnicar*

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`. ~~parameters and B~~bowtie2 ~~with~~ can be run with the parameters `--sensitive-local --un` ~~can be run independently of the proposed pipeline~~.

4839. Run the assembly pipeline (https://github.com/SegataLab/MASTER-WP5-pipelines/tree/master/05-Assembly_pipeline) by ~~typing~~ 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 name ${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 6 steps run automatically by *pipeline_assembly.sh* can be run independently, even adding modifications to adapt them to procedures normally employed by each research group:

- StepTEP 1: perform assembly of reads in contigs using MEGAHIT v1.1.124[33] with default parametners

- StepTEP 2: filter contigs according to length using *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

- StepTEP 3: align filtered reads against filtered contigs using bowtie2 v2.2.9, with *--very-sensitive-local --no-unal* parameters

- StepTEP 4: find contigs depth by *jgi_summarize_bam_contig_depths*, from MetaBAT2 v2.12.125[34]

- StepTEP 5: use MetaBAT2 v2.12.125[34] with *-m 1500 --unbinned --seed 0* parameters to compact contigs into bins/putative MAGs

- StepTEP 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].

[H1] Troubleshooting

Troubleshooting advice can be found in Table 1

Table 1. Troubleshooting table

Step	Problem	Possible reason	Solution
36	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 are to be sampled (i.e. such as knives or, drains)	Where swabbing 1 m^2 is not possible (e.g.such as drains or, knivfes), swabbing individual units (e.g. 1 drain, 1 knife) must be sufficient.
3544	Low concentration of eluted DNA. DNA concentration is recommended to be >2 ng/µl for optimal library preparation and sequencing.	Cells are difficult to lysePoor cell lysis (step 28). : Cell wall structure of gram-positive bacteria vary in thickness, quantity, length distribution and degree of crosslinking of the peptidoglycan, making them some more difficult to be lysed.	After adding Solution CD1 (Step 14) and prior to the bead-beating step, incubate at 65°C for 10 min, then, rResume the protocol from step 128. 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	Instead of using the customized C5 solution to wash the UCP Min Elute Spin Column as described in step 239, try to use the same volume of supplied

		C5 (used in step 239) is an ethanol-based solution that removes residual salts, humic acid and other contaminants, while allowings the DNA to stay bound to the membrane of the column.	Solution C5 or of 70% (v/v) Ethanol. Observe if and how the final yield is influenced for new standard samples.
		The eluted DNA is suspended in too much great a volume of buffer.	The 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 x g for 5 min at room temperature. Decant all liquid. Briefly dry residual ethanol in a speed vac or at ambient air. Avoid over-drying the pellet or resuspension may be difficult. Resuspend precipitated DNA in the desired volume of Solution C6.
3948	Negative control samples show a large number of reads/contigs/MAGs with similar profiles to that those of some samples from the processing environments	Negative control samples might be contaminated with airborne microbes.	Some bioinformatic tools can be applied for contaminants removal. For example, decontam[37] is a tool that identifies the contaminants based on their frequency and/or prevalence in negative control samples over the "real" ones. Additionally, 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 using MetaPhlAn profiling[38].
	Samples have a high amount of "unclassified" reads	Could be related to a hHigh proportion of non-microbial reads (animal/plant host DNA).	An additional host removal step can be performed by using the bowtie2 pipeline (step 347-348) and the food animal or vegetal reference genome, i.e. *Sus scrofa* for some meat samples, *Bous taurus* for some cheese samples, etc.

802
803

804**[H1] Timing**

805Steps 1-817, sampling: 1.5 h for 20 samples

806Steps 918-1322, sample pre-processing: 1.5 h 20 samples

807Steps 1423-3544, DNA purification: 4 h for 20 samples

808Step 3645, library preparation: 6 h for 96 samples

809Step 3746, sequencing: 2 days for each run

Step 3847-3948, bioinformatic analysis: 2 days for 1 sample

[H1] 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 less than 1 million reads), which has allowed to reconstruct a total of 9,564 MAGs from 807 samples (from 0 to 62 MAGs per sample, with >50% of the samples having more than 10 MAGs).

The sequencing reads, assembled contigs and MAGs obtained from the application the protocol can be subjected to detailed taxonomic and functional analyses. Successful examples of the type 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, or boxplots showing the abundance of virulence factor genes in different sample categories, can be observed.

[H1] Data and code availability

The code employed for raw reads filtering, assembly and binning into MAGs is available at https://github.com/SegataLab/MASTER-WP5-pipelines. Raw reads are available on the Sequence Read Archive of the National Center of Biotechnology Information (NCBI) under the BioProjects numbers PRJNA897099 for vegetable facilities, PRJNA941197 (for ice-cream

836facility), PRJNA997800 (for meat facilities), PRJNA997821 (for cheese facilities, except those
837located in Ireland) and PRJNA996188 for control samples. Raw reads for fish processing
838factories and Irish cheese factories are available on the European Nucleotide Archive database
839under the accession numbers PRJEB62794 and PRGEB63604, respectively.

840

841 **[H1] References:**

842 1. Møretrø, T. & Langsrud, S. Residential Bacteria on Surfaces in the Food Industry and
843 Their Implications for Food Safety and Quality. *Compr. Rev. Food Sci. Food Saf.* **16**,
844 1022–1041 (2017).

845 2. De Filippis, F., Valentino, V., Alvarez-Ordóñez, A., Cotter, P. D. & Ercolini, D.
846 Environmental microbiome mapping as a strategy to improve quality and safety in the
847 food industry. *Curr. Opin. Food Sci.* **38**, 168–176 (2021).

848 3. Alvarez-Ordóñez, A., Coughlan, L. M., Briandet, R. & Cotter, P. D. Biofilms in Food
849 Processing Environments: Challenges and Opportunities. *Annu. Rev. Food Sci. Technol.*
850 **10**, 173–195 (2019).

851 4. Fagerlund, A., Langsrud, S. & Møretrø, T. Microbial diversity and ecology of biofilms
852 in food industry environments associated with Listeria monocytogenes persistence.
853 *Curr. Opin. Food Sci.* **37**, 171–178 (2021).

854 5. Koutsoumanis, K. *et al.* Whole genome sequencing and metagenomics for outbreak
855 investigation, source attribution and risk assessment of food-borne microorganisms.
856 *EFSA J.* **17**, (2019).

857 6. Yap, M. *et al.* Next-Generation Food Research: Use of Meta-Omic Approaches for
858 Characterizing Microbial Communities Along the Food Chain. *Annu. Rev. Food Sci.*
859 *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 and Muscle Biology* **4**, 1-11 (2020).

8. Zwirztz, B. *et al*. The sources and transmission routes of microbial populations throughout a meat processing facility. *npj Biofilms and Microbiomes* **6** (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** (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., Sequino, G., Cobo-Díaz, J.F., Álvarez-Ordóñez, A., De Filippis, F. & Ercolini, D. Evidence of virulence and antibiotic resistance genes from the microbiome mapping in minimally processed vegetables producing facilities. *Food Res. Int.* **162** (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).

907 26. Chng, K. R. *et al*. Cartography of opportunistic pathogens and antibiotic resistance
908 genes in a tertiary hospital environment. *Nat. Med.* **26**, 941–951 (2020).

909 27. Echeverría-Beirute, F., Varela-Benavides, I., Jiménez-Madrigal, J.P., Carvajal-Chacon,
910 M. & Guzmán-Hernández, T. eDNA extraction protocol for metagenomic studies in
911 tropical soils. *Biotechniques* **71**, 580–586 (2021).

912 28. Jiang, W. *et al.* Optimized DNA extraction and metagenomic sequencing of airborne
913 microbial communities. *Nat. Protoc.* **10**, 768–779 (2015).

914 29. Fadiji, A. E. & Babalola, O. O. Metagenomics methods for the study of plant-associated
915 microbial communities: A review. *J. Microbiol. Methods* **170**, 105860 (2020).

916 30. Cabello-Yeves, P. J. *et al.* The microbiome of the Black Sea water column analyzed by
917 shotgun and genome centric metagenomics. *Environ. Microbiome* **16**, 5 (2021).

918 31. Keeratipibul, S. *et al*. Effect of swabbing techniques on the efficiency of bacterial
919 recovery from food contact surfaces. *Food Control* **77**, 139-144 (2017).

920 32. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat.*
921 *Methods* **9**, 357–359 (2012).

922 33. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast
923 single-node solution for large and complex metagenomics assembly via succinct de
924 Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).

925 34. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient
926 genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).

927 35. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:
928 assessing the quality of microbial genomes recovered from isolates, single cells, and
929 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.* (2023).

[H1] Acknowledgments

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 would like to thank Mairead Coakley and Samuel Mortensen for their help in the preparation of the Supplementary Video.

[H1] Author contributions

M.L., M.P., D.O., 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 sites~~companies~~. D.O. and L.M.S. designed and tested the improvements in the DNA extraction protocol, while 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., R.C.R., 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.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.

[H1] Competing interests

D. O'Neil and L. Mahler are employees of QIAGEN GmbH. All other authors declare no competing interests.

[H1] Additional information

[H2] Supplementary information

Supplementary Methods. Methodology followed for the 16S V4 qPCR.

Supplementary Figure 1: Number of reads compared to DNA concentration on those samples from the MASTER program failing sequencing. Dot color indicates the surface where the sample was taken (food contact surfaces, non-food contact surfaces, negative control samples taking in food companies, or negative control samples taken in the lab where sample pre-processing took place). Those samples with 0 reads were not successful on library preparation. B) Zoom overview of the blue rectangle in A).

Supplementary Video 1: Microbiome mapping in the food industry: detailed visual procedure on how to prepare the materials and take the samples at a food facility environment. Also, the steps that should be followed in the laboratory for sample pre-processing are shown.

Supplementary Note: Detailed information related to the Supplementary Video.

[H2] Key references using this protocol

977Valentino, V., Sequino, G., Cobo-Díaz, J.F., Álvarez-Ordóñez, A., De Filippis, F. & Ercolini,

978D. Food Research International 162 (2022): https://doi.org/10.1016/j.foodres.2022.112202

979

980
981**[H1] Figure legends**

982**Figure 1. Overview of whole metagenome sequencing results.** Results after reads filtering for
983all the samples successfully sequenced with at least 1 million reads obtained. A) Total number
984of reads, contigs and metagenome assembled genomes (MAGs) obtained per sample as a
985function of the DNA yield of the sample. Type of surface is indicated by shape while type of
986industry is indicated by colors. The grey line indicates the smoothed conditional means
987(calculated by geom_smooth and 'lm' method in ggplot2 R-library) while the grey shadowed
988area indicates the standard error of the trend line. B) Total DNA, number of reads, contigs and
989MAGs by surface type, including negative controls taken ~~on~~ in both the food ~~industry~~
990processing site and laboratory. Black diamonds indicate mean values while the central lines of
991boxplots indicate median values. Samples with DNA concentration above the limit of detection
992of the Qubit High Sensitivity double-stranded DNA (dsDNA) quantification kit (120 ng/µL) are
993represented as having a DNA concentration equal to 120 ng/µL.

994**Figure 2. Workflow for sampling, cell recovery and DNA purification.** (A) Swab samples
995are taken from food processing environments, using personal protective equipment to avoid
996contamination, and pooled in sampling bags (5 pooled swabs per sample category). (B) PBS is
997added to the sampling bag, swabs are homogenised and cells are harvested through
998centrifugation and stored at the ultrafreezer. (C) DNA is extracted from the cell pellet using the
999tailored protocol based on the DNeasy PowerSoil Pro kit chemistry with modifications and the
1000Qiagen´s UCP MinElute Spin Columns. After DNA has been purified and meets the quality
1001standards it can be used for library preparation for Illumina sequencing. All steps of the DNA
1002purification protocol that deviate from that of the DNeasy PowerSoil Pro kit are indicated by
1003orange squares on the scheme.

1004**Figure 3. Example sampling plan.** Sampling plan proposed for the characterisation of the
1005resident microbiome and the evaluation of the impact of different sources on the microbiome of
1006the end products in A) a plant producing fermented sausages and B) a cheese making facility.

1007**Figure 4. Optimization of DNA extraction from surface swabs.** The cell pellet derived from
1008pooled surface swabs was subjected to cell lysis and subsequent DNA extraction. Cell pellets
1009were obtained by following the described surface swab sampling protocol in a standard
1010laboratory environment. The compared conditions for the extraction workflow are indicated by
1011the first row of graph headings. Either commercial kits (Kit A - DNeasy PowerSoil Pro Kit; Kit
1012B - QIAamp UCP DNA Micro Kit) with their standard protocols, a combination of kit A and
1013spin columns of kit B, or further alterations in the standard protocol of kit A were tested. The
1014second row of graph headings denotes the elution volume, which is regulated by the choice of
1015spin columns. Depicted are the resulting 16S rRNA gene copy numbers obtained per individual
1016extraction (black points) as proxy for bacterial DNA content *as* determined by 16S V4 qPCR for

panels A-C*. For panel D the total DNA yield in ng 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 silica membrane (=Comparison_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; 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 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 pellet*s* before cell lysis (=PelletWash) and the increased concentration of Ethanol during spin column washing while the DNA is already bound to it (= SpinWash) and a combination thereof. These protocols were tested on surface swabs collected in food processing sites.