



(51) International Patent Classification:

C12Q 1/70 (2006.01) C12M 1/12 (2006.01)

C12Q 1/04 (2006.01) G01N 33/497 (2006.01)

C12Q 1/24 (2006.01) G01N 1/22 (2006.01)

C12Q 1/6888 (2018.01) C12Q 1/6806 (2018.01)

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

— with sequence listing part of description (Rule 5.2(a))

(21) International Application Number:

PCT/EP2021/065932

(22) International Filing Date:

14 June 2021 (14.06.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

20382510.4 12 June 2020 (12.06.2020) EP

(71) Applicant: **CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS (CSIC)** [ES/ES]; C/ Serrano, 117, 28006 MADRID (ES).

(72) Inventors: **ALCAMI PERTEJO, Antonio**; Centro de Biología Molecular Severo Ochoa, Nicolás Cabrera, 1, Campus Cantoblanco UAM, 28049 MADRID (ES). **RASTROJO LASTRAS, Alberto**; Centro de Biología Molecular Severo Ochoa, Nicolás Cabrera, 1, Campus Cantoblanco UAM, 28049 MADRID (ES).

(74) Agent: **PONS ARIÑO, Angel**; Glorieta Rubén Darío, 4, 28010 MADRID (ES).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: DEVICE AND METHOD FOR CAPTURING AND ANALYSING AIRBORNE ORGANISMS

(57) Abstract: The present invention refers to a device comprising polytetrafluorethylene (PTFE) filters and the use of the same for collecting, detecting and identifying organisms present in air ecosystems. The invention also provides a method suitable for the capture, detection and identification of whole airborne biological particles, including viruses and other important air pathogens, which involves the use of the device. This method allows to perform organism, preferably viral, metagenomics to sequence all DNA and RNA organisms captured in the filters. This methodology may be used to detect, for instance, SARS-CoV2 particles in air samples.

DEVICE AND METHOD FOR CAPTURING AND ANALYSING AIRBORNE ORGANISMS

5 The invention belongs to the fields of microbiology, public health and methods for analysing air ecosystems, in particular to methods and devices useful for collecting air samples and analysing the airborne microbiota present in them. These methods allow to gain a global view of the airborne biological communities as well as to identify pathogens (viruses, bacteria or fungi), allergens and the like present in them.

10 BACKGROUND ART

Air pollution is known to be a major environmental risk to public health. In addition to inorganic pollution, air also carries several airborne biological particles (ABPs) containing archaea, bacteria, virus, fungi and pollen grains, coming from other environments (soil, 15 water or plant/animals microenvironments). A great effort is being made by the international community to reduce air pollution but very little is known about the biological community (airbiota) present in the outdoor air. The study of the airbiota is relevant due to its potential role in dissemination of plant, animal and human diseases, with important implications in public health, and a huge economic impact on crops and stockbreeding 20 productivity. The current knowledge about air microbial community is mainly restricted to pollen and fungal spores, studied using traditional methods such as optical microscopy or quantitative PCR, which because of its importance as allergens are daily measured and included among air quality indicators. Culture-dependent techniques were also used for the study of air microorganisms (fungi, bacteria and viruses), despite the small 25 fraction of organisms that can be cultured. Furthermore, studies on viruses are almost inexistent. Therefore, a global overview of the airborne biological community is crucial to understand air ecosystem dynamics, to identify marker organisms and to establish new air quality indicators. However, there is not a standardized method to capture the whole diversity present in the air.

30

The interest in the study of airborne microbiota is increasing to understand the connectivity of different ecosystems, the survival of microbes in the air ecosystem and the transmission of human pathogens. In spite of its relevance, the characterization of microbiota in the air is in its infancy. A variety of sampling devices are being used by the 35 research laboratories that have initiated these studies, but detailed studies on the

performance of these devices are scarce.

Some of the airborne microorganisms are the causative agent of important human diseases such as *Influenza* virus, *Rhinoviruses*, *Mycobacterium tuberculosis* or
5 *Aspergillus spp.* Others, although non-pathogenic, are known to act as allergens, such as grass pollen or *Alternaria* spores, affecting to several million people worldwide. Traditionally, air was considered mainly as a dissemination path for these microorganisms, being studied primary for its important implications in human health and livestock and agriculture productivity. However, in recent studies, the scientific
10 community has suggested that air has its own biota, and thus, must be considered as an ecosystem in itself. Nevertheless, very little is known about airborne biological community, the dynamics governing them or the potential relations among microorganisms and the influence of environmental factors on its composition.

15 Next-generation sequencing (NGS) is a promising approach to study global microbial diversity that is being applied to other ecosystems such as ocean, freshwater bodies or thermal springs (Rastrojo A., Alcamí A., 2018, Adv Virus Res, 101:39–54; Rastrojo A., Alcamí A., 2017, Virus Res, 239:87–96). NGS was also applied to the study of air communities, mainly through targeted metagenomics of marker genes, but most of the
20 studies were focused on a single group of organisms, mainly bacteria, or only on a couple of groups (fungi and bacteria) (Núñez A., *et al.*, 2016, Int Microbiol, 19:1–13). Among airborne microorganisms, viruses have received little attention. Although there are many studies on specific viruses, such as *Influenza* virus, only one report has characterized the global viral diversity in outdoor air (Whon TW., *et al.*, 2012, J Virol 86:8221–8231).
25 Moreover, to date, only a few studies have used a shotgun sequencing approach to study whole airborne community, including viruses (Cao C., *et al.*, 2014, Environmental Science and Technology, 48:1499–1507; Be N. A. *et al.*, 2015, Microb Ecol 69:346–355; Yooseph S., *et al.*, 2013, PLoS One 8:e81862).

30 A number of devices with very different capture mechanisms have been used to study the airborne community (Núñez A., *et al.*, 2016, Int Microbiol, 19:69–80). Therefore, the evaluation of different capturing methods is required to establish a standardized framework. It has been previously demonstrated that Hirst-type samplers are a good choice for monitoring the airborne biological community, by comparing classical
35 approaches such as optical microscopy of pollen and fungal spores with targeted

metagenomics, resulting in a good correlation between both techniques (Núñez A., *et al.*, 2017, Appl Environ Microbiol, 83). Although Hirst samplers can give a good description of the airborne biological community, it is far from being complete. Viruses cannot be studied by targeted metagenomics using marker genes such as 16S/18S
5 rRNA and others approaches are required. Hirst samplers are based on the inertial capture of ABPs in a vaseline covered strip that can be used for DNA extraction. A shotgun sequencing of this DNA could give a more complete view of the airborne community including DNA viruses. However, viral genomes are represented in a very low proportion of sequencing reads in a shotgun approach due to their small genomes
10 compared to cellular genomes. Viral particles should be purified to obtain a complete analysis of the viral community (both DNA and RNA viruses), and this cannot be done from the vaseline strip without introducing some important biases (Prussin AJ., *et al.*, 2014, FEMS Microbiol Lett, 357:1–9; Thurber RV., *et al.*, 2009, Nat Protoc 4:470–483). Additionally, Hirst-type samplers are expensive devices and usually not portable, so its
15 usage is limited to a restricted number of sampling sites simultaneously.

Polytetrafluorethylene (PTFE) filters have been also tested for collecting airborne bacteria, virus and other particles in the 10-900 nm size range (Nancy C. B., *et al.*, 2007, Ann. Occup. Hyg., 51: 2, 143–151), however, no efficient analysing method has been
20 proposed for identifying the whole captured biological particles.

Therefore, there is a need for developing improved capturing and analysing methods which use devices and methodologies that allow to collect and study the whole airborne biological community, including viruses and other important pathogens and allergens, in
25 an efficient way.

The document of Booth et al, "*Detection of Airborne Severe Acute Respiratory Syndrome (SARS) Coronavirus and Environmental Contamination in SARS Outbreak Units*" discloses that severe acute respiratory syndrome (SARS) is characterized by a risk of
30 nosocomial transmission; however, the risk of airborne transmission of SARS is unknown. During the Toronto outbreaks of SARS, we investigated environmental contamination in SARS units, by employing novel air sampling and conventional surface swabbing. Two polymerase chain reaction (PCR)–positive air samples were obtained from a room occupied by a patient with SARS, indicating the presence of the virus in the
35 air of the room. In addition, several PCR-positive swab samples were recovered from

frequently touched surfaces in rooms occupied by patients with SARS (a bed table and a television remote control) and in a nurses' station used by staff (a medication refrigerator door). These data provide the first experimental confirmation of viral aerosol generation by a patient with SARS, indicating the possibility of airborne droplet transmission, which emphasizes the need for adequate respiratory protection, as well as for strict surface hygiene practices. Samples were collected on a polytetrafluoroethylene (PTFE; "Teflon") membrane filter with a pore size of 0.3 μm in a closed-face, 3-piece disposable plastic cassette by use of a personal sampling pump operating at ~ 2 L/min. This resulted in the collection of samples that were dry. Samples were shipped with ice packs and refrigerated upon receipt at the laboratory. In this publication, all 28 air samples collected with PTFE filters were negative for SARS by PCR. The only 2 samples that were positive were collected with a different device (a high-resolution slit-sampler system designed by Defense Research and Development Canada).

The document of [PO YING CHIA *et al* "Detection of air and surface contamination by SARS-CoV-2 in hospital rooms of infected patients", *NATURE COMMUNICATIONS*, vol.11, no.1, 29 May 2020] discloses that Understanding the particle size distribution in the air and patterns of environmental contamination of SARS-CoV-2 is essential for infection prevention policies. Here we screen surface and air samples from hospital rooms of COVID-19 patients for SARS-CoV-2 RNA. Environmental sampling is conducted in three airborne infection isolation rooms (AIIRs) in the ICU and 27 AIIRs in the general ward. 245 surface samples are collected. 56.7% of rooms have at least one environmental surface contaminated. High touch surface contamination is shown in ten (66.7%) out of 15 patients in the first week of illness, and three (20%) beyond the first week of illness ($p = 0.01$, χ^2 test). Air sampling is performed in three of the 27 AIIRs in the general ward and detects SARS-CoV-2 PCR-positive particles of sizes $>4 \mu\text{m}$ and $1\text{--}4 \mu\text{m}$ in two rooms, despite these rooms having 12 air changes per hour. This warrants further study of the airborne transmission potential of SARS-CoV-2. In D2 it is disclosed that Six NIOSH BC 251 bioaerosol samplers were placed in each of three AIIRs in the general ward to collect air samples (set-up depicted in Supplementary Fig. 1). Particles collected with the NIOSH sampler are distributed into three size fractions. Particles $>4 \mu\text{m}$ in diameter are collected in a 15 mL centrifuge tube, particles $1\text{--}4 \mu\text{m}$ in diameter are collected in a 1.5 mL centrifuge tube, and particles $<1 \mu\text{m}$ in diameter are collected in a self-assembled filter cassette containing a 37-mm diameter, PTFE filter with 3 μm pores. All NIOSH samplers were connected to either SKC AirCheck TOUCH Pumps or

SKC Universal air sampling pumps set at a flowrate of 3.5 L/min and run for 4 h, collecting a total of 5040 L of air from each patient's room. In this publication, no SARS-CoV-2 was detected by PCR in the PTFE filters (see Table 1).

- 5 The document of [KK COLEMAN et al, "Bioaerosol Sampling for Respiratory Viruses in Singapore's Mass Rapid Transit Network", *SCIENTIFIC REPORTS*, vol.8, no.1, 30 November 2018] discloses a UV-sterilized Biological Safety Cabinet Class II and sterile consumables, and equipment were used to process samples. Using a filter-handling kit (225–8372; SKC), each polytetrafluoroethylene (PTFE) filter was removed from the
- 10 NIOSH sampler cassettes and transferred to a 50 mL Falcon tube and vortexed while dry for 5 seconds. To minimize cross-contamination, forceps were sterilized with 70% ethanol in between each filter transfer. One millilitre of 0.5% BSA solution was then added to each 50 mL Falcon tube containing a filter and vortexed again for 15 seconds. One millilitre of 0.5% Bovine Serum Albumin (BSA) solution was added to each 1.5 mL
- 15 conical tube from the NIOSH samplers and vortexed for 10 seconds. Using cryotube vials, the BSA solutions from the 50 mL Falcon tubes containing filters were pooled together with their respective 1.5 mL conical tube sample. Two millilitres of BSA solution were added to each 15 mL Falcon tube from the NIOSH samplers, vortexed for 15 seconds, and transferred to cryotube vials. In this publication, it is mentioned that
- 20 'prior to molecular analysis, the particles collected in the 1-4 μm and <1 μm size fractions were combined'. Therefore, a virus-positive results does not demonstrate that the PTFE filter (<1 μm size) was able to capture the virus, since it may have been retained in the previous tube (1-4 μm size). Also, this publication shows positive results for influenza virus, adenovirus and RSV, but not for coronavirus (see Table 1).

- 25 The document of [T.A. MYATT "BMC Public Health Airborne rhinovirus detection and effect of ultraviolet radiation on detection by a semi-nested RT-PCT assay", 13 January 2003] discloses an experiment with aerosolized rhinovirus in a small aerosol chamber. Experiments were conducted with decreasing concentrations of rhinovirus. To determine
- 30 the effect of UV irradiation on detection of rhinoviral aerosols, we also conducted experiments in which we exposed aerosols to a UV dose of 684 mJ/m². Aerosols were collected on Teflon filters and rhinovirus recovered in Qiagen AVL buffer using the Qiagen QIAamp Viral RNA Kit (Qiagen Corp., Valencia, California) followed by semi-nested RT-PCR and detection by gel electrophoresis. In this publication Teflon filters
- 35 used had a 2 μm pore, and the rhinovirus detected was generated with a nebulizer in the

laboratory-

The document of [T.A. MYATT "*Detection of Airborne rhinovirus detection and its relation to outdoor air supply in office environments*", *AMERICAN JOURNAL OF RESPIRATORY AND CRITICAL CARE MEDICINE*, vol.169, no.11, 1 June 2004] discloses an experiment wherein weekly air samples were collected in two centralized locations in each study area from 9 a.m. to 5 p.m. on workdays at 4 L/minute with closed-faced cassettes on 37-mm Teflon filters (Teflon 2.0 µm pore; Pall Gelman, Ann Arbor, MI). The filters were placed approximately 1 m from and horizontal to the floor (23). The closed-faced cassettes prevented direct contact with occupants and limited the chance that large droplets would be collected. We prepared weekly field and laboratory blanks and outdoor background samples. Filters were stored at -80°C until analysis. This publication used Teflon filters of 2µm pore size, and detected only rhinovirus, not coronavirus, in an office environment.

The document of [K. ROSARIO et al. "*Diversity of DNA and RNA viruses in Indoor Air As assessed via Metagenomic Sequencing*" *ENVIRONMENTAL SCIENCE & TECHNOLOGY*, vol.52, no.3, 23 January 2018] disclose that diverse bacterial and fungal communities inhabit human-occupied buildings and circulate in indoor air; however, viral diversity in these man-made environments remains largely unknown. Here we investigated DNA and RNA viruses circulating in the air of 12 university dormitory rooms by analysing dust accumulated over a one-year period on heating, ventilation, and air conditioning (HVAC) filters. A metagenomic sequencing approach was used to determine the identity and diversity of viral particles extracted from the HVAC filters. We detected a broad diversity of viruses associated with a range of hosts, including animals, arthropods, bacteria, fungi, humans, plants, and protists, suggesting that disparate organisms can contribute to indoor airborne viral communities. Viral community composition and the distribution of human-infecting papillomaviruses and polyomaviruses were distinct in the different dormitory rooms, indicating that airborne viral communities are variable in human-occupied spaces and appear to reflect differential rates of viral shedding from room occupants. This work significantly expands the known airborne viral diversity found indoors, enabling the design of sensitive and quantitative assays to further investigate specific viruses of interest, and providing new insight into the likely sources of viruses found in indoor air.) This publication does not use PTFE filters, but air conditioning HVAC filters.

The document of [YI YU LAI et al. "Metagenomic Human respiratory air in a hospital environment, PLOS ONE, vol.10, no.10, 2 October 2015] discloses that all samples were collected in Beijing Hospital (Google Marker location: 39°54'3.5", 116°25'0.78"), details of the room conditions were listed in Table 3. We set up an air sampling pump (SKC, PA, US) and connected to a 47mm filter holder (PALL, NY, U.S.) by TYGON tube (Saint-Gobain Corporation, US), with a 47mm,0.2um PTFE Membrane filter (PALL, NY, U.S.) inside. Fig 2A shows the tube with the holder. All filters were sterilized by autoclaving following the user's guide, then packaged in sterilized aluminium foil and stored in a sealed bag until loading into the filter holder. The holder and all the tools used for changing new filters were cleaned with 75% ethanol or autoclaved each time to avoid contamination. The air sampler was drawn at an average flow rate of 4 L/min for 24 h (4:00 PM to 4:00 PM the next day). The sampling day was December 20, a clear day without PM2.5 pollution outdoor. At that day PM2.5 = 30.3 PM10 = 42.5, lower than mean value of 2013 (PM2.5 = 89). This publication used 0,2µm PTFE filters, and after a shotgun metagenomic approach describes the identification of low amounts of viruses and archaea, with bacteria and eukaryote species exceeding 90% of the pathogens identified. This report does not use PCR to identify coronaviruses.

Finally, the document US10663454B2 discloses a handheld air sampler device for enrichment of airborne substances and/or particles, in particular microorganisms, includes: an inlet configured for air intake into a flow channel; an outlet configured for fluidic ally connecting the flow channel to an external vacuum device; a permeable backing element arranged in the flow channel, the backing element being configured to receive and support a filter element; and a sealing element configured for sealing a filter element received on the backing element such that, when a negative pressure is applied to the outlet, air entering the inlet forms an airstream passing the filter element and airborne substances and/or particles, in particular microorganisms, are enriched in the filter element. An air sampling arrangement includes a handheld air sampler device of this type. This publication describes a pipeline to process clinical samples, but it is not focused on viruses present in air samples, nor described an air sampler to capture viruses. This publication does not detect viruses (see Fig.9), only viruses in an artificial mixture (mock) generated in the laboratory (see Fig. 10).

In front of the prior art documents, the present invention improves the collection of viral

particles in the air. Consequently, the objective technical problem that the invention solves is to provide a device form improving the collection of viral particles in the air. The claimed solution comprises the synergistic effect between a PTFE filter with a pore size of 1 μm or 5 μm and coupled with a processor operable to enable the air to flow through
5 or over the substrate, thereby causing the particles in the air to be captured by the substrate, centrifugal means configured to remove cellular organisms from the substrate; buffering means configured to house the substrate, and filtration means configured to filtrate the supernatants containing the virus particles, wherein the centrifugal means are also configured to concentrate the supernatants.

10

DESCRIPTION OF THE INVENTION

The present invention refers to a device for capturing air biological particles which comprises polytetrafluorethylene (PTFE) filters and its use for collecting, detecting and
15 identifying organisms present in air ecosystems. The invention also provides a method suitable for the capture, detection and identification of whole airborne biological particles, including viruses and other important air pathogens and allergens, with the use of the device comprising the PTFE filters and preferably a sequencing shotgun metagenomic approach. This method allows to perform organism, preferably viral, metagenomics to
20 sequence all DNA and RNA organisms captured in the filters, and/or to amplify specific genomes by gene amplification methods. This methodology may be applied to detect, for instance, SARS-CoV2 particles in air samples.

The study of the airborne biological community, composed by very different organisms
25 (such as bacteria, fungi, pollen and virus), requires the use of an efficient sampling method that captures a good representation of the whole community. In the present invention, inventors have analysed the microbial community captured with seven different air samplers, based on different capturing methods (impact samplers, cyclonic, impinger and filters), to determine which is the most suitable to obtain a more complete
30 view of the airborne biological community and to establish it as a standardized method.

Thus, in the present invention different methods for collecting airborne microorganisms, including impactor samplers, liquid impingers, cyclonic samplers and different filters, have been tested. Additionally, to avoid the biases of culture-dependent methods or
35 amplicon-based analyses, the inventors have performed a next generation sequencing

(NGS) shotgun approach to uncover the whole airborne diversity. A detailed analysis of the genetic diversity of the airbiota captured using these different methodologies is presented in the Examples shown below. Therefore, the Examples of this application provide a detailed comparison of the biodiversity detected by seven different air sampling
5 devices, with the analysis of shotgun metagenomics and amplicon sequencing of prokaryotic/eukaryotic organisms to evaluate the biodiversity captured by the different samplers. The results obtained evidence that, while isolation and purification of virus particles is not possible from many devices, PTFE filters are the best option for capturing and analysing viral metagenomics and thus these filters are the methodology of choice
10 for performing the most efficient sampling and analysing method that captures the best representation of the whole airborne organisms community.

Consequently, the Examples shown below, which compare airbiota analysis from different air samplers, evidence that PTFE filters show the best efficiencies capturing
15 microbial diversity from the air (Figures 2-4).

PTFE filters have also the following advantages: they are cheaper than others samplers, replicates can be sampled using two or more filter holders, and several localizations (both outdoor and indoor) can be studied simultaneously using several devices with a
20 reduce cost. Additionally, taking into account other factors such as cost and portability, PTFE filters were the most convenient device of those tested. Finally, it has been demonstrated herein that PTFE filters allow the extraction of the viral fraction for viral metagenomics, which allows the analysis of the whole airborne biological community. Consequently, PTFE filters represent the best option for performing viral metagenomics
25 because of the easy to extract cellular organism and virus particles by shaking or sonication of the filters in a buffer.

Therefore, in a first aspect, the present invention relates to a device for capturing air biological particles, hereinafter "the device of the invention", comprising:

- 30
- a substrate configured to capture particles from flowing air thereon using at least one captor;
 - at least one processor operable to enable the air to flow through or over the substrate, thereby causing the particles in the air to be captured by the substrate;
- wherein the substrate and the at least one processor are communicatively coupled.

35

Preferably, the substrate comprises at least one filter acting as the at least one captor.

Preferably, the at least one filter is a polytetrafluoroethylene (PTFE) filter.

- 5 Preferably, the polytetrafluoroethylene (PTFE) filter is a synthetic fluoropolymer of tetrafluoroethylene filter.

In a preferred embodiment of the device, the PTFE filter has a nominal pore size between 0.2 μm and 5 μm , preferably 1 μm or 5 μm , more preferably of 1 μm . A PTFE filter with
10 a nominal pore size of 1 μm is specially advantageous for collecting viral particles from air.

Optionally, the device for capturing air biological particles further comprises an air pump operable to increase or decrease flow of air,
15 wherein, to capture the particles into the substrate, the at least one processor is operable to:
cause the air pump to increase or decrease the flow of air containing the particles toward the substrate.

- 20 Preferably, the air pump is a vacuum pump.

Another aspect refers to the use of the device of the invention for collecting or capturing or sampling, and for detecting and identifying organisms present in air (including outdoor and indoor air), preferably wherein the detection and identification are performed by
25 metagenomic analysis, more preferably by a sequencing shotgun metagenomic approach. Alternatively, by amplification of specific genomes with the use of specific primers and gene amplification methods, followed by sequencing.

The invention also refers to a system for detecting and identifying organisms present in
30 the air comprising the above explained device for capturing air biological particles.

Preferably, the system for detecting and identifying organisms present in the air further comprises centrifugal means configured to remove cellular organisms from the substrate, and buffering means configured to house the substrate.

Preferably, the system for detecting and identifying organisms present in the air further comprises filtration means configured to filtrate the supernatants containing the virus particles, wherein the centrifugal means are also configured to concentrate the supernatants.

5

Preferably, the system for detecting and identifying organisms present in the air further comprises nuclease treatment means configured to remove all non-encapsidated DNA or RNA.

10

Preferably, the system for detecting and identifying organisms present in the air further comprises gene amplification and sequencing means for the amplification of collective or specific viral genomes captured in the filters, followed by sequencing.

15 In another more preferred embodiment, the detection and identification of organisms present in air, referred to in the present invention, is performed by metagenomic analysis, more preferably by a next generation sequencing (NGS) shotgun approach.

In another more preferred embodiment, for sequencing and detection of specific viruses,
20 the viral genomes or a specific gene are amplified by PCR with specific primers and the product is identified by quantitative (qPCR) or quantitative reverse transcription PCR (RT-qPCR), or sequenced using Illumina HiSeq 200 machine or equivalent.

In another preferred embodiment the system for detecting and identifying organisms
25 present in the air comprises the device of the invention and further comprises centrifugal means configured to remove cellular organisms from the substrate, buffering means configured to house the substrate filtration means configured to filtrate the supernatants containing the virus particles, wherein the centrifugal means are also configured to concentrate the supernatants, nuclease treatment means configured to remove all non
30 encapsidated DNA or RNA and gene amplification and sequencing means for the amplification of collective or specific viral genomes captured in the filters, followed by sequencing.

The organisms present in air or in air samples will be also referred to in the present
35 invention as "airborne organisms", "airbiota" or "airborne biological community". These

organisms are, preferably, microorganisms, more preferably, the organisms are selected from the list consisting of: viruses, including DNA and RNA viruses, plants, bacteria, pollen and fungi. Even more preferably, the organisms are air pathogens, particularly viruses, preferably SARS-CoV2.

5

“Air pathogens” are those organisms present in the airbiota which are capable of trigger a disease, an infection, a clinical condition, and the like, in humans or non-human animals. Examples of “air pathogens” are, but without limitations, bacteria such as mycobacterium tuberculosis, bordetella pertussis and mycoplasma pneumoniae, and
10 viruses such as influenza virus, coronavirus, adenovirus, varicella zoster virus and respiratory syncytial virus.

Thus, another aspect of the invention refers to the use of the device of the invention for collecting, detecting and identifying air organisms, hereinafter “the use of the invention”.

15

In a preferred embodiment of the use of the invention the air organisms are selected from the list consisting of: viruses, plants, bacteria, pollen and fungi.

In another preferred embodiment of the use of the invention the organisms are air
20 pathogens, preferably virus pathogens, more preferably the virus is SARS-CoV2.

Another aspect of the invention refers to a method, hereinafter “the method of the invention”, for collecting, detecting and identifying organisms, preferably microorganisms, present in the air, or for the analysis of airbiota, wherein said method
25 comprises the steps of:

- a. sampling an air sample with the device of the invention,
- b. extracting the genetic material comprised in the device after step (a),
- c. randomly amplifying the genetic material extracted in step (b), and
30 d. sequencing, preferably by metagenomic analysis, more preferably by a sequencing shotgun metagenomic approach, the genetic material amplified in step (c).

The sampling of step (a) may be performed during at least 1 hour, preferably at least one
35 day, more preferably during at least 4 consecutive days, even more preferably during at

least 8 consecutive days, even more preferably during at least 12 hours per day.

The genetic material of step (b) may be DNA or RNA, single or double stranded, circular or lineal. This genetic material may be extracted by conventional DNA or RNA extraction
5 methods well-known in the art.

Amplification of step (c) is preferably carried out by multiple displacement amplification (MDA), more preferably using primers for 16S rRNA, for bacteria, and Internal transcribed spacer 2 (ITS2), for plants and fungi.

10

In another preferred embodiment, the metagenomics analysis of step (d) is performed by an NGS shotgun approach, more preferably using an Illumina HiSeq 2000 machine or equivalent.

15 In another preferred embodiment, the method of the invention comprises an alternative step c (c') by amplification of specific genomes with the use of specific primers and gene amplification methods, followed by sequencing.

For sequencing and detection of specific viruses of step (c'), the viral genomes or a
20 specific gene are amplified by PCR with specific primers and the product is identified by quantitative (qPCR) or quantitative reverse transcription PCR (RT-qPCR), or sequenced using Illumina HiSeq 200 machine or equivalent.

In another preferred embodiment of the method of the invention, the organisms are
25 selected from the list consisting of: viruses, including DNA and RNA viruses, plants, bacteria, pollen and fungi. Even more preferably, the organisms are air pathogens, particularly viruses, preferably the SARS-CoV2 virus.

Thus, the method of the invention is useful for collecting, detecting and identifying air
30 pathogens.

In another preferred embodiment, the method of the invention comprises an additional step (a'), between steps (a) and (b), in which cellular organisms are removed from the device, more preferably from the PTFE filter comprised in the device. Previously to this
35 step (a'), the PTFE filter comprised in the device may be optionally vortex, more

preferably at maximum speed for 1 h at 4°C.

The cellular organisms may be removed by, for instance but without limitation, resuspending the airborne particles captured in the filter in saline buffer, followed by
5 centrifugation, shaking or sonication, and finally filtration. Afterwards, a concentration step of the supernatants containing the virus particles may be performed.

Furthermore, preferably, an additional step after step (a`) and before step (b), is carried out, which comprises removing the non-encapsidated DNA or RNA, more preferably with
10 a nuclease treatment.

Methods and uses described in the present invention may be carried out in air samples collected from indoor or outdoor.

15 Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skilled in the art to which this invention belongs. Methods and materials similar or equivalent to those described herein can be used in the practice of the present invention. Throughout the description and claims the word "comprise" and its variations are not intended to exclude other technical
20 features, additives, components, or steps. Additional objects, advantages and features of the invention will become apparent to those skilled in the art upon examination of the description or may be learned by practice of the invention. The following examples and drawings are provided by way of illustration and are not intended to be limiting of the present invention.

25

DESCRIPTION OF THE DRAWINGS

Fig. 1. Contigs taxonomic classification. Contigs were first aligned using Blastn against NT database. Unassigned contigs were then aligned using Blastx against NR
30 database. Quality filtered reads were mapped against all contigs using BWA. The percentage of mapped reads is represented, considering all contigs (A) or assigned contigs (B).

Fig. 2. Bacteria metataxonomic analysis. A) Taxonomic profile at the order level.
35 Orders with abundances lower than 2% were merged. B) Alpha diversity indexes plot.

Replicates, when available, were merged for this analysis. Samples are indicated as follow: B, Burkard; DM, Burkard Multivial; BP, BioPump; I, impiger; N, Nanoparticles; S, SAS; T, PTFE. C) Beta diversity analysis. Normalized counts were used for a PCoA analysis based on weighted UNIFRAC distances. A second analysis was done removing
 5 Impinger samples (small square). D) Heatmap showing the percentage of common taxa (at family level) shared by the different samples. E) Number of OTUs with significant differences in their abundances (q-value < 0.1).

Fig. 3. Fungal metataxonomic analysis. A) Taxonomic profile at the order level. Orders
 10 with abundances lower than 2% were merged. B) Alpha diversity indexes plot. Replicates, when available were merged for this analysis. Samples are indicated as follow: B, Burkard; DM, Burkard Multivial; BP, BioPump; I, impiger; N, Nanoparticles; S, SAS; T, PTFE. C) Beta diversity analysis. Normalized counts were used for a PCoA analysis based on weighted UNIFRAC distances. D) Heatmap showing the percentage
 15 of common taxa (at family level) shared by the different samples. E) Number of OTUs with significant differences in their abundances (q-value < 0.1).

Fig. 4. Plants metataxonomic analysis. A) Taxonomic profile at the order level. Orders
 with abundances lower than 2% were merged. B) Alpha diversity indexes plot.
 20 Replicates, when available were merged for this analysis. Samples are indicated as follow: B, Burkard; DM, Burkard Multivial; BP, BioPump; I, impiger; N, Nanoparticles; S, SAS; T, PTFE. C) Beta diversity analysis. Normalized counts were used for a PCoA analysis based on weighted UNIFRAC distances. D) Heatmap showing the percentage
 of common taxa (at family level) shared by the different samples. E) Number of OTUs
 25 with significant differences in their abundances (q-value < 0.1).

Fig. 5. PTFE virome analysis. A) General overview of taxonomy. Contigs were aligned against NCBI NR database using BlastX (the percentage of reads in contigs is represented). B) Viral contigs classification (BlastX against NR viral protein database).
 30 C) Complete circular viral genomes detected. Coverage is represented as a grey histogram around genome scheme.

EXAMPLES

1. Material and methods.

Sampling

5 Several sampling devices based on different capturing mechanisms were chosen to compare airborne biological particles (ABPs) capture: three impact/slit samplers (Hirst spore trap sampler (Burkard), Surface Air System DUO 360 (SAS), and Zefon Bio-Pump® Plus (Biopump)), a cyclonic device (Burkard Multivial), a custom impinger and two different filters (Polytetrafluoroethylene (PTFE) and Glass fiber (GFC, 1.7 µm)

10 sandwich with and inner layer of nanoparticles). Hirst spore trap sampler (Burckard) is an impact sampler based on the capture of vacuum-accelerated airborne particles in a vaseline-covered strip. Surface Air System DUO 360 (SAS) was designed for the direct culture of the microorganisms captured from air as is equipped with two heads where agar plates are placed. To reduce the culture bias we replaced the agar plates by

15 vaseline-covered Petri dishes (new plates were used for each sampling day). Zefon Bio-Pump ® Plus (Biopump) uses cartridges containing a slide covered by a patented aqueous solution to capture the ABPs which is easier to recover after the sampling (1 cartridge was used per sampling day). Burkard Multivial uses vacuum suction to create a cyclon inside a 1.5 ml tube where ABPs are deposited due to the centrifugal force. A

20 new sterilized 1.5 ml tube was used every day. The custom impinger was built with gas washer bottles (100-200 µm nominal pore size) to capture ABPs by the interchange of particles between the air a buffered solution (MSM buffer: 50 mM Tris-HCl, 100 mM NaCl and 8 mM MgCl₂, pH 7.5). For each sampling day a 500 ml bottle of autoclaved buffer were used. Finally, two different filters were used: PTFE (5 µm nominal pore size) and a

25 GFC (1.7 µm, Whatman) sandwich with an inner layer of iron-coated particles (MIL-100(Fe) nanoparticles with 1-10 µm pore size). A GFC-nanoparticle sandwich was prepared directly onto the filter holders. First, a GFC filter was added before injecting an ethanol solution containing a suspension of nanoparticles (~40 mg). Then, a second GFC filter was added and the sandwiches were dried at 100°C for 24 h. Filters were

30 reused for the four samplings days. We used duplicates when enough devices were available (Table 1). Sampling was carried out in the roof of the Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid (Spain, 40.439881°N, 3.689409°W, ca 30 m over the street). All samplers were run according to device characteristics (Table 1). Air was sampled during 12 h per day during four consecutive

35 days in May of 2016. Biopump was only run during 10 h due to software limitations, and

SAS was only run during 20 min, 3 times per day, to reduce the total volume capture for comparison purpose, since its flow rate is very high (150 L/min) with respect to the other samplers.

5 Table 1. Samplers summary.

ID	Sampler	Flow (l x min ⁻¹)	Time ¹	Total air volume (m ³)
B1	Burckard	10	12 hours/day	28.8
B2	Burckard	10	12 hours/day	28.8
BM	Burckard Multivial	15	12 hours/day	43.2
BP	Biopumb	15	10 hours/day	39
I1	Impinger	6-6.5	12 hours/day	17.3
I2	Impinger	4.5-5	12 hours/day	13
N1	Nanoparticles	5.5	12 hours/day	37.3
N2	Nanoparticles	11	12 hours/day	15.8
S1	SAS	150	1 h/day, 3x20 min (9,13,17h)	43.2
S2	SAS	150	1 h/day, 3x20 min (9,13,17h)	43.2
T1	PTFE filter	11-12	12 hours/day	33
T2	PTFE filter	11-13	12 hours/day	34.5

¹ Sampling was repeated during 4 consecutive days.

DNA extraction

- 10 After sampling, total DNA was directly extracted from all devices using PowerMax® Soil DNA Isolation Kit (MO BIO Laboratories), except in the case of impingers that were subjected to tangential flow filtration (70 kDa cartridge) to reduce sample volume from 2 L to ~10 mL. Vaseline from Burkard and SAS sampler were recovered using a sterile blade and subjected to direct DNA extraction. The aqueous solution of the four Biopump
- 15 cartridges used, one per day, were recovered and mixed for DNA extraction. ABPs captured by the Burkard Multivial were resuspended in MSM buffer before DNA

extraction. Filters (PTFE and GFC) were added directly to DNA extraction tubes. The eluted DNA (5 mL) was then ethanol precipitated. Briefly, 200 μ L of 5 M NaCl, 2.5 μ L of linear polyacrylamide as carrier (LPA, 25 μ g/ μ L, Sigma) and 10.5 mL of cold absolute ethanol were added. Samples were centrifuged for 30 min at 2500 x g. Pellets were washed with cold 70% ethanol and air dried. DNA was resuspended in 100 μ L of nuclease free water (Ambion). All procedures were carried out in a UV-cabinet (BioSan UVT-B-AR) and all equipment was treated with 0.1 M HCl solution to avoid contaminations. Two mock DNA extractions were included to test for putative contaminations during extraction. Samples were also tested for the presence of microorganism by PCR using different marked genes (16S rDNA and ITS2) (Table 2). No amplification was observed in mock samples.

Table 2. Primers used for amplicon sequencing and contamination controls.

Primer	Sequence (5->3)	Target	Reference
BacF	CCTACGGGNGGCWGCAG (SEQ ID NO: 1)	16S rDNA	Takahashi S, Tomita J, Nishioka K, Hisada T, and Nishijima M. 2014. PloS One 9 (8):e105592
BacR	GACTACHVGGGTATCTAATCC (SEQ ID NO: 2)		
ITS86F	GTGAATCATCGAATCTTTGAA (SEQ ID NO: 3)	Fungal ITS2	Op De Beeck M, Lievens B, Busschaert P, Declerck S, Vangronsveld J, and Colpaert JV. 2014. PloS One 9 (6):e97629.
ITS4	TCCTCCGCTTATTGATATGC (SEQ ID NO: 4)		
ITS4-For	YGAATCTCGGCAACGGATA (SEQ ID NO: 5)	Plantae ITS2	White, TJ, Bruns TD, Lee SB, Taylor JW, and John S. 1990. In PCR Protocols, 31:315–22. Academic Press, Inc.
ITS4-Rev	TCCTCCGCTTATTGATATGC (SEQ ID NO: 6)		

Shotgun sequencing and analysis

Total DNA samples were randomly amplified using multiple displacement amplification (MDA, GenomiPhi kit, GE Healthcare). 10 μ L of each sample were used for amplification

following manufacturer's instructions. Samples were amplified for 2.5 h, except for B1 and I1 that required 3.5 h of amplification to obtain sufficient DNA for sequencing. Sample I2 was discarded because no amplification was observed after 6 h. No amplification was observed in mock samples. Library preparations and sequencing were performed at Centro Nacional de Análisis Genómico (CNAG, Barcelona, Spain). Sequencing was done using an Illumina HiSeq 2000 machine obtaining ~40 M paired-end reads (2x126 pb) for each sample. Raw reads were quality filtered using PRINSEQ (minimum read length 100 pb and minimum average quality 25). Orphan reads were discarded. Taxonomy binning of reads was carried out using Centrifuge against NCBI non-redundant nucleotide database. Only reads uniquely assigned with a score higher than 200 were considered. Reads assigned to Human or PhiX174 were excluded. Centrifuge-assigned reads were normalized using metagenomeSeq and used for beta diversity analysis using PhyloSeq. Quality filtered reads were then assembled with IDBA_UD (--pre_correction --mink 20 --maxk 120 --min_contig 500). Assembled contig with low complexity were removed using PRINSEQ (-lc_method entropy -lc_threshold 70). A hybrid-Blast approach was used for contig taxonomic classification. First, contigs were aligned using Blastn against NCBI non-redundant nucleotide database (NT, downloaded on May 2017). Only those hits with e-value < 1e-3 and score > 50 were considered, and best hit was assigned. Contigs with no hits or with hits with e-value > 1e-3 and score < 50 were then subjected to Blastx alignment against NCBI non-redundant protein database (NR, downloaded on May 2017). Again, best hit was used for taxonomic assignments (e-value < 1e-3 and score < 50). Reads were mapped to contigs using BWA MEM.

25 **Metataxonomic sequencing and analysis**

In order to obtain a taxonomic profile of the different sampling devices we used a targeted metagenomic approach. MDA-unamplified DNA samples were used as template for marker genes amplification. We used 16S rRNA for bacteria and Internal transcribed spacer 2 (ITS2) for plants and fungi (Table 2 above). PCR amplification and library preparation were carried out at Parque Científico de Madrid (Madrid, Spain). Briefly, 100 pg of DNA, quantified using Picogreen, were used for a first PCR using marker genes primers linked with Illumina adaptors (CS1 and CS2). The PCR conditions were as follow: 98°C for 30 min, followed by 26 cycles of denaturation at 98°C for 30 s, 50°C for 20 s, 72°C for 20 s, followed by a final extension at 72°C for 2 min. A second PCR of 8

cycles were performed using a 1/25-1/200 dilution of the first one using the same cycling conditions but using CS1 and CS2 primers linked to different barcodes and additional Illumina adapters also required for sequencing (p5 and p7). Q5® High-Fidelity DNA Polymerase (New England Biolabs) were used for PCRs. Amplified products were quantified and pooled before sequencing in a MiSeq machine obtaining ~200,000 reads (2x300pb). Raw reads were quality filtered as described above and paired reads were joined and adapters were removed using PANDAseq. Taxonomic assignments were performed using Qiime software. The Greengenes database (version gg_13_8 implemented in Qiime) was used for bacterias and UNITE (version no. 7.1) was used as fungal database. A custom database was used for plant classification. OTUs were defined at 97% sequence similarity and only those with at least 5 counts and present in at least 2 samples were kept for further analysis. Chloroplast and mitochondrion OTUs were removed from bacterial analysis. Although specific ITS2 primers for plants or fungi were used, some cross-amplification can be produced, and therefore fungal OTUs were removed from plant analysis and plant OTUs from fungal analysis. OTU counts were normalized using the metagenomeSeq method. Taxonomic profile, diversity indexes and beta diversity analysis were performed on normalized counts using Phyloseq package. Common family analysis were performed after merging replicates using Phyloseq. A dedicated Qiime script (differential_abundance.py) was employed to analysis the significance of OTU abundance differences using DESeq2 method.

Virus sampling, sequencing and analysis

For virus analysis we used PTFE filters (1 µm, PALL Zefluor). A vacuum pump equipped with 2 filter holders was placed on Alcobendas city (at 14 km from Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid) at 1.5 m over the ground. Filters were sampling for 8 days (from the 9th to the 17th of February of 2017) at 2.5 L/min. After sampling, filters were vortexed at maximum speed in 10 mL of MSM buffer for 1 h at 4°C. Cellular organisms were removed by centrifugation (20' at 3000xg) followed by 0.45 µm filtration of the supernatant. The supernatants containing the virus particles were then concentrated using centrifugal units (100 kDa, Amicon). Concentrated virus particles were subjected to nuclease treatment to remove all non-encapsidated DNA or RNA (200 U of DNase I, 120 U of Nuclease S7 and 10 µg of RNase A) for 1 h at RT. Viral DNA was extracted with the MinElute kit (QIAGEN), following manufacturer's instructions. Samples were tested for the presence of cellular

microorganism by PCR using different marked genes (16S rDNA and ITS2) (Table 2 above) and no amplification was observed neither in the samples nor in the mock sample (10 mL of MSM buffer that were subjected to all steps). Then, 10 μ L of each sample were used to randomly amplify all DNA using GenomiPhi kit (GE Healthcare) for 2.5 h. We
5 obtained amplification in both replicates, but not in mock sample. One of the replicates were sequenced at Parque Científico de Madrid using a MiSeq machine, obtaining 682,832 paired reads (2x300 pb). Raw reads were quality filtered using PRINSEQ (minimum read length 150 pb and minimum average quality 25). Paired quality filtered reads were used for assembly using SPAdes v3.9.0. Contigs shorter than 500 pb were
10 discarded. Contigs were aligned using BlastX against the NCBI non-redundant protein database (NR). To remove any putative non-viral sequence, only the contigs with at least 20% of viral hits (e value $<1e-3$) were considered as viral. Contigs without any hit were also considered as putative viral sequences. Viral contigs were then classified using BlastX against a non-redundant viral protein database (extracted from the NCBI NR
15 database using taxonomic id 10239). Best hits were assigned (e value $<1e-3$ and score >50). Sequencing reads were mapped against contigs using BWA MEM. Complete circular viral genomes were detected using Minimus2.

Detection of SARS-CoV2 using PTFE filters

20

Five vacuum pumps (KNF) (D1-D5) equipped with 2 filter holder each were installed in La Paz hospital (Madrid, Spain). Two consecutive rounds of sampling were done; from 20/03/2020 to 23/03/2020 (Experiment 1, 3 days) and from 23/03/2020 to 27/03/2020 (Experiment 2, 4 days) (Table 5). Polytetrafluoroethylene (PTFE) filters were used for air
25 sampling at a flow rate of 2.5 l/min. After sampling, filters were immediately immersed in 4 ml of TRIzol (Invitrogen). Three RNA extraction from 1 ml from TRIzol were performed following manufacturer's instructions. For each sample, 3 RNA pellet were resuspended together in 25 μ L of Nuclease-Free water (Ambion). For all procedures, DNA Lobind tubes were used (Eppendorf).

30

SARS-CoV2 E gene was measure using a Taqman assay (Bio-Rad) containing primers and probe designed by Charité Hospital (Berlin) and approved by World Health Organization for clinical diagnosis (Table 3). A 2-step approach was applied. Briefly, 5 μ L of RNA was retrotranscribed using the SuperScript IV First-Strand Synthesis System
35 (Invitrogen). Then, 1 μ L of cDNA was used as template in a 10 μ L qPCR reaction using

Taqman Fast Universal PCR Master Mix (Applied Biosystems). Thermal conditions were as follow: 20 seg at 95°C, 45 cycles of 5 seg at 95°C and 20 seg at 60°C. qPCR was performed in a ABI PRISM 7900HT SDS device (Applied Biosystems). Samples were evaluated using technical triplicates.

5

Table 3: Primers used for SARS-CoV2 detection. FAM: 6-carboxyfluorescein; BBQ: blackberry quencher

Primer	Sequence (5'>3')
E_Sarbeco_F1	ACAGGTACGTTAATAGTTAATAGCGT (SEQ ID NO: 7)
E_Sarbeco_R2	ATATTGCAGCAGTACGCACACA (SEQ ID NO: 8)
E_Sarbeco_P1	FAM-ACACTAGCCATCCTTACTGCGCTTCG-BBQ (FAM-SEQ ID NO: 9-BBQ)

ORF1ab and N genes were measure in a one-step approach using the Novel Coronavirus (2019-nCoV) Nucleic Acid Diagnostic Kit (Sansure). Briefly, 2.5 µl and 5 µl of RNA from Experiments 1 and 2 respectively, were used in a 25 µl qPCR reaction. Thermal conditions were as follow: 30 min at 50°C for retrotranscription, 1 minute at 95°C followed by 45 cycles of 15 seg at 95°C and 30 seg at 60°C. One step qPCR was performed in a CFX-384 device (Bio-Rad).

15

2. Results.

Shotgun sequencing approach

To obtain a global view of the airborne biological community without the restriction of targeted metagenomic approaches, that do not include viruses, we decided to sequence total DNA using a shotgun sequencing approach. However, the amount of total DNA recovered from the air samples was <1 ng/µl, and therefore we had to randomly amplify the DNA samples using Multiple Displacement Amplification (MDA) to obtain sufficient DNA for library preparation. All samples were amplified for 2.5 h, except by Burkard 1 (B1) and Impinger 1 (I1) that required 3.5 h of amplification to obtain sufficient DNA. The Impinger 2 (I2) sample was discarded because no DNA amplification was observed even after 6 h of MDA. Total DNA was directly sequenced using the Illumina technology, obtaining ~40 million paired reads for each sample. Raw reads were subjected to quality

25

filtering prior to contig assembly. We obtained on average 13,000 contigs per sample, being the impinger (I1) the sample with fewer contigs, closely followed by the Biopump (BP) and one of the Burkard replicates (B1). By contrast, SAS, Burkard Multival (BM) and the other Burkard replicate (B2) showed the highest number of contigs. Total number of assembled bases was ~18 million on average, but one of the Burkard samples (B1), the Biopump (BP), one of the GFC filters (N1) and the impinger (I1) were clearly under this average which correlated with a reduced number of contigs. However, average contig length and N50 were similar among all samples except for I1 that was almost twice with respect to all other samplers.

10

Contigs were classified using both Blastn and Blastx against the NCBI non-redundant databases (Figure 1). On average 60% of the contigs were assigned successfully, that represents ~70% of total assembled bases and ~33% of the reads (Figure 1A). Nevertheless, most of the assigned contigs were classified as belonging to *Viridiplantae* (Figure 1B). This was an unexpected result taken into account previous reports where plant sequencing reads were less abundant than for instance bacterial sequencing reads, including samples collected in spring when the highest pollen concentration is found, suggesting a higher bacterial abundance. However, in our experimental set, plant sequencing reads represented on average ~86% of total mapped reads in the assigned contigs. Excluding the I1 sample, the average percentage of plant sequencing reads increased to up to 94%, reaching up to 97% in several samples (BM, S1, T2). Furthermore, most contigs assigned as *Viridiplantae* were classified as *Pinales* (~80% on average). *Pinus* genome is one of the highest plant genomes sequenced to date (~22 Gb) and this could be the main reason of these unexpected results. One single copy of a *Pinus* genome would account for up to 99% of the reads if it was sequenced together with a single bacteria (4 Mb genome). Therefore, it seems that *Pinus* pollen have monopolized most of the sequencing reads. In order to verify this, quality filtered reads were directly classified using Centrifuge. Although sequencing reads contain less information than contigs, and therefore a small fraction of sequencing reads can be classified, the results showed a similar pattern. Most of the reads were assigned to *Viridiplantae* (~75% of assigned reads) and most of them belong to *Pinus* genus (~67%). Then, we used the assigned sequencing reads to compare different sampler through beta diversity analysis but the results are clearly dominated by plant reads. Sequencing reads assigned to different groups (Bacteria, Fungi and Viridiplantae) were analyzed independently, and the distribution observed using all reads was almost identical to the

35

one observed using only plant reads. Bacteria plot, although inverted, was also similar, which could indicate an influence of this group in the ordination of the samples. By contrast, fungal sequencing reads analysis was more dispersed, being grouped by replicates, which could indicate an influence in the capture of fungal spores depending on the sampler. However, the number of non-plant reads was very small for a shotgun analysis to be properly compared, and therefore we must be cautious. Nevertheless, we could observe a clear contamination in sample I1 (Figure 1) that, in contrast to other samples, was dominated by bacterial and metazoan reads, being the assigned reads almost 100%. Most of the bacterial assigned reads were related to *Alphaproteobacteria* (*Rhodobacteraceae*) that most likely grew in the impingers during the 12 h of sampling.

Although the results of these shotgun experiments are dominated by plant sequencing reads, mainly from *Pinus* genus, we were able to detect some viral contigs (Table 4).

15 Table 4. Taxonomic profile of the assigned viral reads. Numbers represent the number
of mapped reads in viral contigs.

[illegible]

	-	<i>Others</i>	81	0	0	0	418	0	0	0	0	0	0
Retro-transcribing viruses	-	<i>Caulimoviridae</i>	0	4946	2983	0	0	0	17	3582	3015	169	708
	-	<i>Retroviridae</i>	0	0	0	0	0	0	33	0	0	0	0
ssDNA viruses	-	<i>Circoviridae</i>	89	0	0	0	0	0	0	0	0	0	0
	-	<i>Genomoviridae</i>	0	0	0	0	0	0	0	544	0	0	0
	-	<i>Nanoviridae</i>	0	0	0	0	0	0	0	28	0	0	0
	-	<i>Others</i>	0	0	0	0	161	0	0	0	0	0	0
environmental samples	-	-	0	0	0	0	461	0	0	0	0	0	0
ssRNA viruses	-	<i>Flaviviridae</i>	0	0	0	0	0	0	204	0	0	0	0
unclassified bacterial viruses	-	-	0	0	0	0	17102	0	0	0	299	0	0
unclassified viruses	-	-	0	0	281	0	0	0	23	0	0	0	0
Total reads in contigs			732	4985	3264	0	28675	63	362	5632	7764	2941	1003

Most of them were classified as *Baculoviridae* and *Caulimoviridae*, insect and plant

viruses respectively. We were also able to detect some bacteriophage (*Caudovirales*) contigs in sample I1, which correlates with a bacterial contamination in that sample, although some of them could represent prophages rather than free virus. Among viral contigs we were able to detect 5 complete circular viral genomes: one *Circoviridae*-like
 5 contig detected in B1, one *Genomoviridae*-like virus detected in S1 and three *Caulimoviridae*-like virus detected in B2, BM and S1. Another near-complete *Caulimoviridae*-like virus was also detected in S2 sample, and many caulimo-related contigs were also detected in other samples. Moreover, using the *Caulimoviridae*-like virus detected in B2 (contig B2_191) as reference we were able to detect reads from this
 10 virus in almost all samples except in I1. Samples B1 (Burkard) and BP (BioPump) showed very few *Caulimoviridae*-like reads compared to the other samples. These *Caulimoviridae*-like viruses have been demonstrated to infect *Pinus* and could represent the first gymnosperm-infecting *Caulimovirus* described to date.

15 **Metataxonomic approach (targeted metagenomics)**

Due to the *Viridiplantae* dominance in shotgun sequencing and the impossibility of doing a convenient comparative analysis, we decided to use the unamplified DNA samples to amplify marker genes to obtain a taxonomic profile of the microorganisms captured by
 20 the different sampling devices. We used 16S rRNA for bacteria and ITS2 for plants and fungi. Amplicons were sequenced in a MiSeq device obtaining ~200,000 paired sequencing reads (2x300 pb) from each sample. Using this approach, we were able to amplify 16 rRNA from both impinger samples, however we did not obtain amplification of ITS2 (either from fungi or plants) from these two samples neither using a higher number
 25 of PCR cycles (data not shown). This result was consistent with the bacterial contamination observed in the shotgun analysis of these samples.

Bacteria analysis

30 As mentioned above, Impingers taxonomic profile were mostly composed by *Rhodobacterales* from *Paracoccus* genus arising up to ~95% of the reads in sample I1 (Figure 2A). This genus is also present in all other samples but with an average abundance of only 3%, and therefore the increased presence on Impingers is likely to be a contamination caused by the favourable growing conditions that this kind of samplers
 35 have (a liquid buffered solution). All other samples showed a similar taxonomic profile

among them (Figure 2A) with exception of the Nanoparticle filters which were the more dissimilar between replicates (N1 and N2) and among samples. The huge differences observed between N1 and N2 could be due to the air flows that were substantially different between them (Table 1 above). Both filters were attached to the same vacuum pump and due to differences in filter setting air retention was stronger on one of them (N1) and most of the vacuum pressure went to the other filter (N2). *Bacillales*, mostly composed by soil-dwelling *Alicyclobacillus* genus, was a relative abundant group observed in most of the samples (excluding Impingers). However, no or very few bacteria from this order/genus were observed in Burkard Multivial and in SAS replicates, and Burkard samples showed a great discrepancy in its relative abundance (B1 ~21.3% versus B2 ~5.6% at order level). The differences observed between Burkard replicates could be caused by air flow variations during sampling such as in Nanoparticles, as the air flow was only tested at the beginning and at the end of every sampling day (not observing significant differences). By contrast, PTFE filters showed a very similar profile between replicates in spite of a small air flow differences (Table 1).

Regarding the captured diversity, Burkard (B), SAS (S) and PTFE (T) showed the highest value in almost all indexes (Figure 2B). By contrast, impingers (I) were the worst as only a very limited number of OTUs could be detected due to the contamination. Burkard Multivial (BM), Nanoparticles (N) and BioPump (BP) showed a reduced diversity. However, in beta diversity analysis all samples clustered together except the impingers (Figure 2C). When impingers were removed from the analysis, Nanoparticles appeared far from all other samples (Figure 2C inset), as this samples showed the most different taxonomic profile. The rest of samples were clustered together being SAS and PTFE replicates the most closely ordinated between replicates, while Burkard were more dispersed. Similarly, when common families were analysed, SAS, PTFE and Burkard showed the highest percentages, followed by Nanoparticles, Burkard Multivial and Biopump (Figure 2D). Again, Impingers only shared less than half of the detected bacterial families with the other samples. This data correlated well with the number of OTUs showing significant differences (Figure 2E). SAS, PTFE and BM did not show any significant difference. Burkard and Nanoparticles showed some differences with SAS, PTFE and BM, and more between them. Finally, Impinger showed the highest number of OTUs with significant differences with all other samples.

Fungi analysis

The ITS2 region was amplified and sequenced for fungi taxonomic analysis. A general dominance of the order *Capnodiales*, most of them from *Cladosporium* genus, was observed in all samples (Figure 3A). Impingers were not included in the analysis because no amplification of ITS2 was observed even after increasing the number of PCR cycles, likely caused by the bacterial contamination. All other samples showed a similar taxonomic profile, and replicates were particularly good for SAS and PTFE. Some differences in the most abundant taxa could be observed. For instance, BM showed a highly reduced relative abundance of *Capnodiales*, and therefore a greater abundance of all others groups. This fact produced an increase in some of the diversity indexes (Shannon and Simpson) of this sample (Figure 3B). These index calculations rely on relative abundances (frequencies) that are more evenly distributed when an abundant taxa is not present or its relative abundance is substantially decreased. The second most abundant order, *Pleosporales*, where the relevant *Alternaria* genus belongs, showed a similar profile among all samples. By contrast, *Agaricales*, that were present in most samples, were almost absent in SAS replicates, in BM and in one the Burkard replicates (B1). B1, as seen in the bacterial analysis, showed a clear different profile to the B2 replicate (and with the rest of samplers), being absent some abundant taxas such as *Pucciniales* or *Erysiphales*.

However, Burkard, together with SAS and PTFE, showed high values in most diversity indexes (Figure 3B). The results of SAS in Shannon and Fisher indexes could be explained by the increased relative abundance of *Capnodiales*, that, unlike in BM, showed a reduction in these indexes. Apart from these differences in the taxonomic profiles, SAS, PTFE and Burkard (mainly B2) were clustered together when beta diversity was analysed. Nanoparticles, Burkard Multivial and BioPump were more dispersed probably due to differences in some taxa not captured by these samplers (such as *Erysiphales* or *Pucciniales*) (Figure 3C). The percentage of shared families and the number of OTUs with significant differences were consistently with beta diversity analysis (Figure 3D,E). SAS, PTFE and Burkard samples shared the highest number of families while SAS and PTFE showed a reduced number of OTUs with significant differences, being only 3 between SAS and PTFE (Figure 3E). BM, BP and N shared only ~50% of the detected families with all other samples and showed a high number of OTUs with significant differences, being nanoparticles the most different (Figure 3E).

Plants analysis

Plants were mainly capture in the samplers through pollen grain or small fragments present in the air. In general, the results of the analysis from plants showed a similar behavior than in bacterial and fungal analysis (Figure 4). Again, impingers were not included because no ITS2 amplification was observed. Most other samplers showed a dominance of *Fagales* (mainly *Quercus*) followed by *Lamiales* (*Plantago*) and *Pinales* (*Pinus*), in consonance with the pollinic profile of May in the city of Madrid. However, one of the Burkard replicates (B1) showed, such as in bacterial and fungal analysis, a clearly different profile. A single OTU assigned as *Solanum* (*Solanales*) account for more than 50% of assigned reads in B1 sample (19.178 reads) (Figure 4A). Nevertheless, this huge difference could be partially solved with the use of replicates, reducing the impact of this problem in the rest of the analysis (Figure 4). BioPump, that had no replicates due to resources availability, also had a discordant taxonomic profile, showed a divergent behavior in all analysis. It showed maximum diversity values in Shannon and Simpson parameters (Figure 4B) likely produced by a biased capture of pollen from *Fagales* and an increase on *Pinales*, that could influence the relative abundance of all other taxa, having a more even distribution increasing those diversity indexes. This phenomenon also affected partially to Burkard replicates, since the replicates average abundances are also influenced by the dominance of *Solanum* genus in B1. SAS and PTFE obtained high values in most of the diversity indexes, followed by Burkard and Burkard Multivial (Figure 4B). Consequently, these samples (SAS, PTFE, Burkard (B2) and Burkard Multivial) clustered together in the beta diversity plot (Figure 4C). BioPump, Burkard (B1) and Nanoparticles (both replicates) seemed to be more dispersed. SAS, PTFE and Burkard shared 82% of OTUs at family level and only 8 or 9 OTUs showed significant differences in their abundances among them (Figure 4D,E). Burkard Multivial (BM) showed a similar pattern to these three samplers, clustering together with them, although it showed a smaller diversity compared to them and a reduced number of shared families. However, this sampler did not show any OTUs with significant difference compared with SAS or PTFE and only 11 OTUs compared with Burkard (Figure 4E).

Virome analysis

PTFE and SAS showed the best efficiencies capturing microbial diversity from the air

(Figures 2-4). Burkard, our reference device, showed a good performance but it was less efficient and the replicates were not highly consistent. These three samplers had obtained high value in several diversity indexes, were clustered together in beta diversity analysis and shared the highest number of families while very few OTUs showed significant differences (Figure 2-4). Therefore, we can consider them as suitable for capturing the air diversity, at least regarding bacteria, fungi and plants. However, Burkard and SAS use vaseline for the capture, which is convenient to avoid desiccation, but it represents a problem to extract intact microorganisms. Our attempts to extract viral particles from the Burkard device vaseline were unsuccessful. Therefore, we decided to use PTFE filters for the analysis of viruses from the air using a similar set as that used for the comparative analysis among samplers. Air was sampled continuously for 8 days at 2.5 L/min getting a total volume of 28,8 m³. Viral particles were purified from the filters and viral DNA was extracted as described in Example 1. Viral DNA was then randomly amplified using MDA and sequenced in a MiSeq machine. After quality filtering, reads were assembled into contigs. Then, contigs were filtered to remove any putative cellular sequence before taxonomic classification. We have also included the data from Whon and Coworkers to compare our results. Most of the reads were classified as viral and only a small fraction was discarded as being putatively from cellular origin (Figure 5A). Of these, most of the contigs were classified as ssDNA viruses (Figure 5B). However, in our data *Genomoviridae* seemed to be a dominant group. *Genomoviridae* is a recently described family of viruses that were found to putatively infect plant pathogenic fungi, although the hosts of these viruses remain largely unknown. Moreover, we could detect one *Genomoviridae* complete circular genome (Figure 5C). Several other contigs were also classified as belonging to *Genomoviridae* but were not assembled completely (data not shown). Additionally, we were able to detect other complete circular genomes belonging to *Geminiviridae* and *Circoviridae* families and one similar to an unclassified ssDNA viral sequence (similar to Pacific flying fox faeces associated circular DNA virus-12, KT732830.1) containing a viral Rep gene (Figure 5C).

30 **Detection of SARS-CoV2**

For experiment 1 and 2, default setting were applied for the analysis. Only cycle threshold (Ct) below 40 with "S" shape were considered as valid. No template controls (NTC) and positive controls were included in all runs. In the table below, Ct for all three genes (E, ORF-1ab, N) in both experiments are shown in table 5.

Table 5. Location of devices: D1, neonates room; D2, medical doctors room; D3, intensive care unit 1; D4 pediatric emergencies; D5 intensive care unit 2. ND: Not determined; "-": Ct > 40 or not detected.

Device	Replicate	Experiment 1			Experiment 2		
		E	ORF-1ab	N	E	ORF-1ab	N
D1	1	-	-	38,9	35,2	32,7	34,8
		-			36,1		
		-			34,2		
	2	-	-	-	35,1	-	36,3
		-			-		
		-			-		
D2	1	-	35,3	37,0	-	32,0	34,1
		-			35,3		
		-			-		
	2	35,3	33,8	35,7	-	38,4	38,5
		-			-		
		36,3			-		
D3	1	-	-	-	-	-	-
		-			-		
		-			-		
	2	-	36,1	-	-	32,8	36,1
		-			-		
		-			-		
D4	1	ND	ND	ND	35,7	34,9	35,3
		ND			-		
		ND			-		
	2	ND	ND	ND	-	33,5	34,9
		ND			-		
		ND			-		
D5	1	36,3	-	-	-	-	37,3
		-			-		
		-			-		
	2	35,9	35,0	36,9	-	-	-

		36,8		35,3	
		36,1		-	

3. Conclusions.

- 5 Here we have presented a comparative analysis of different air samplers. Our first choice was a shotgun approach to obtain a global view of all organisms present in the air. In our experimental setting, plants seemed to dominate regardless of the sampling method used (Figure 1). It is clear that pollen grains are an important component of the airbiota, and, as plants have large genomes, they should be an abundant group in shotgun reads.
- 10 However, very few complete plant genomes are present in databases, which could introduce an important bias in reads analysis, being underestimated. In our data, *Pinus*, with one of the longest and best studied plant genomes, was the most abundant taxa in almost all samples.
- 15 It is interesting to identify the differences in the plant taxonomic profile detected using metataxonomic and shotgun approaches. *Fagales* were the most abundant taxa in the ITS2 profile (~50% on average), while *Pinales* dominated shotgun reads (~80%, Figure 1). This could be partially explained by the different nature of both approaches. For metataxonomic, only a marker gene is amplified regardless the length of the genomes
- 20 of the species analysed. By contrast, the shotgun approach is clearly influenced by the length of the genomes. Nevertheless, we can normalize the sequencing reads assigned to each taxa by genome length, when available, to be able to compare the data from both approaches. Although, the shotgun sequencing reads corrected by the median genome length results in an important reduction of *Pinales* abundance, from ~93% to
- 25 ~30% (only considering plant orders with at least one available genome), the normalized taxonomic profiles were clearly different from that obtained using the metataxonomic approach. These differences could be explained by the lack of information in databases. Plant genomes are scarce in databases, and therefore many shotgun sequences/contigs cannot be properly classified (only 60% of contigs could be classified corresponding to
- 30 33% of the sequencing reads). It is also possible that some reads assigned as *Pinus* actually belong to other plants not yet sequenced, and therefore *Pinus* abundance could be also over-estimated to some extent. By contrast, amplicon databases are more curated with a high diversity of sequences available that allow a better classification. It

is important to take into account that the most frequently used marker genes are usually multi-copy and their exact number of copies is unknown for many species, and this could also introduce a bias in targeted metagenomic approaches. Nevertheless, beta diversity analysis of shotgun reads, although dominated by plant reads, showed a similar pattern to that observed in targeted metagenomics, being most of the samplers clustered together, except the Impinger (I1) and one of the Burkard replicates (B1) (Figures 2-4). These two samples were randomly amplified for 3.5 h while all other samples were only amplified for 2.5 h. MDA is known to introduce biases especially when very small input DNA is used. This could explain the different behavior of samples I1 and B1 in shotgun analysis. However, a similar result was observed in these samples using a targeted metagenomic approach where no MDA was used, and therefore we can consider that these differences are most likely to be real than a bias introduced by MDA amplification.

Nevertheless, using a targeted metagenomic approach we were able to compare different samplers avoiding the dominance of sequencing reads from plant genomes observed by shotgun analysis. In our analysis of the outdoor air, samplers had a low relative influence on the capture of biodiversity, showing most of them a similar taxonomic profile, although some important differences were observed (Figures 1-4). Impingers were almost discarded due to a contamination of bacteria, most likely produced by the growing of bacteria in the buffered solution used. The growing of bacteria could be partially avoided by covering the bottles with aluminum foil to reduce sun incidence, but other non-photosynthetic bacteria would also be able to grow. Impingers are generally used for very short sampling times (~30 min) which reduce the possibility of bacterial growth, and therefore, we can discard this type of samplers when longer sampling times are required. GFC filter sandwiches (Nanoparticles) showed great differences between replicates and among all other samplers. These in house build filters were included to test whether iron coated nanoparticles could increase viral capture, however, they showed great differences when bacteria, fungi or plants are analyzed. As mentioned before, different air flows could be responsible for the differences between replicates, but, in spite of having a similar flow rate than other samplers, the second replicate (N2) also showed important differences in the results. Burkard Multival and BioPump also showed some important differences regarding all other samplers. BM, although showed a similar bacterial profile to most other samplers, did not capture any *Bacillales*, an abundant group (Figure 2). *Agaricales* was also absent from BM profile in fungi analysis, and a great decrease in *Capnoidiales* was observed,

which produced a more even distribution of other capture taxa (Figure 3). This even distribution caused an increase of the diversity indexes that relies on frequencies (Shannon and Simpson). By contrast, the BM plant profile was similar to that from all other samples (Figure 4), showing high diversity values. BP, although similar in bacteria or fungi analysis, showed a very different plant profile, with an important reduction in *Fagales* and an important increase in *Pinales* capture (Figure 4). Similar to BM, these differences in plant BP profile had produced an increase in Shannon and Simpson indexes due to a more even taxa distribution. Unfortunately, no replicates of BM or BP were available to determine if these differences could be caused by experimental errors or by the capture mechanism by themselves. In this sense, Burkard (B1 and B2) showed great differences between replicates in all groups. B1 showed a different profile in all analysis and was far from the main cluster in all beta diversity analysis. Differences in airflow rates during sampling could explain these capture deficiencies in B1 regarding B2. Airflows were checked at the beginning and at the end of every sampling day, and no differences were observed. However, we cannot discard fluctuations during the sampling. We have observed previously that variation on Burkard airflows produce differences in the recovered DNA and a greater dispersion of the data, especially in bacteria, and therefore, especial care should be taken to control Burkard airflow to reduce this phenomenon. By contrast, PTFE filters and SAS sampler showed a good correlation at profile level in all analysis, even with small airflow differences in PTFE samples (Table 1). These two samplers showed very few differences between them and, followed by the Burkard device, have obtained the best results regarding diversity indexes, number of shared families and a reduced number of OTUs with significant differences in their abundances (Figures 2-4). Therefore, we can conclude that the PTFE and SAS samplers are equally suited to monitor bacteria, fungi and plants on outdoor air, followed by the Burkard device. However, the cost and the portability of the different samplers are important factors to consider them. Burkard is a heavy and fixed device with a high cost, which reduces the possibilities of using this kind of devices to only a limited number of places. Burkard Multival and BioPump are also very expensive and, although BP is portable, its high cost also reduces the possibilities of sampling many places simultaneously. Impingers are the cheapest devices but as mentioned care should be taken when long sampling times are used to reduce bacterial growth, for example using azide. SAS is also expensive, but two replicates can be sampled at the same time which makes this sampler more flexible, but again sampling several localization simultaneously could be prohibitive using this device. Finally, filter samplers,

which are mainly a vacuum pump with a filter holder, are cheaper than others, replicates can be sampled using two or more filter holders, and several localization can be studied simultaneously using several devices with a reduce cost.

5 Apart from the cost, one of the major challenges for study airborne microorganisms is the amount of biomass and therefore the amount of DNA that can be obtained. This could be partially solved by the used of random amplification methods such as MDA or Sequence-Independent Single-Primer Amplification (SISPA). However, as shown here, when a shotgun approach is used, it is important to consider the presence of pollen that
10 can dominate most of the reads, although most of them could not be classified due to database deficiencies. In this sense, targeted metagenomics could be a good solution. However, viruses cannot be studied following this targeted approach. Viruses do not have any common marker gene to study them, and only a shotgun approach can be used. Using such approach, in spite of the dominance of plant reads, we were able to
15 detect viral contigs, being some of them complete genomes. Moreover, we were able to detect and describe a new viral genome (*Pinus nigra virus 1*) belonging to a new putative *Caulimoviridae* genus infecting gymnosperms. By contrast, the circo-like contig detected in B1 sample that was 99% identical to MG023129.1, a viral genome detected on marine isopods, could represent a contamination of the silica used in the DNA extraction kit
20 columns as it was previously reported.

Additionally, viruses can have DNA or RNA genomes and analysing only DNA will result in a biased view of viral diversity. Moreover, most human infecting viruses are RNA viruses such as *Influenza* or *Enterovirus* (*Rhinovirus*). Therefore, for environments with
25 very different organisms and a variety of genome lengths, the best option to study viruses is the purification of virus particles before sequencing. However, this is not an easy task for most samplers. For example, SAS and Burkard use vaseline layers to capture the airborne particles and the purification of viruses or cells from this substance, by using chloroform or other organic solvent, can destroy most of cellular organisms and some
30 viruses. As we show here, PTFE filters represent the best option for viral metagenomics because of the easy to extract cellular organism and virus particles by shaking or sonication of the filters in a buffer.

In conclusion, we have tested several sampling devices, using different capturing
35 mechanisms, for the analysis of the airborne biological community in order to establish

a standardized method. First, we tried using a shotgun sequencing approach, a method that would allow the sequencing of all microorganisms present in the samples. However, we observed that this approach is not the most convenient when the community is composed by organisms with very different genomes sizes. In our experiments, pollen grains from *Pinus*, which have a very large genome (22 Gb), have monopolized most of the sequencing reads and only a small fraction of the reads belonged to other taxonomic groups. A deeper sequencing could solve partially this problem, although the sequencing cost could also make the use of this approach prohibitive. Therefore, community diversity should be considered before using a shotgun sequencing approach. By contrast, amplicon sequencing, a simpler and cheaper method, allowed us to compare the communities captured by the different sampling devices, despite the genome size diversity. Using this approach, we showed that PTFE filters and SAS are the most suitable samplers for the analysis of bacteria, fungi spores and pollen from plants, with very little differences between them. Other samplers, such as Burkard, also showed a good performance, but the consistency between replicates was poor compared to SAS and PTFE, which is crucial in comparative studies. Additionally, taking into account other factors such as cost and portability, PTFE filters were the most convenient device. Finally, we have demonstrated that PTFE filters allow the extraction of the viral fraction for viral metagenomics, which allows the analysis of the whole airborne biological community.

CLAIMS

1. A system for detecting and identifying organism present in the air comprising:
a substrate configured to capture particles from flowing air thereon using at least
5 one polytetrafluorethylene (PTFE) filter with a nominal pore size of 1 μm or 5 μm ; and
at least one processor operable to enable the air to flow through or over the
substrate, thereby causing the particles in the air to be captured by the substrate;
wherein the substrate and the at least one processor are communicatively
coupled;
10 centrifugal means configured to remove cellular organisms from the substrate,
buffering means configured to house the substrate, and
filtration means configured to filtrate the supernatants containing the virus
particles, wherein the centrifugal means are also configured to concentrate the
supernatants.
15
2. The system according to claim 1, further comprising an air pump operable to increase
or decrease flow of air, preferably the air pump is a vacuum pump, wherein, to capture
the particles into the substrate, the at least one processor is operable to: cause the air
pump to increase or decrease the flow of air containing the particles toward the substrate.
20
3. The system of claim 1 or 2 further comprising:
nuclease treatment means configured to remove all non-encapsidated DNA or
RNA.
- 25 4. The system of claim 3 further comprising:
gene amplification means for all, or specific viral genomes and sequencing
means for identification or detection of specific virus or collective viruses
5. Use of the system according to any one of claims 1 to 4 for collecting, detecting, and
30 identifying organisms present in air, preferably the organisms are air pathogens.
6. The use according to claim 5, wherein the organisms are selected from the list
consisting of: viruses, plants, bacteria, pollen, and fungi, preferably the organisms are
viruses, more preferably the virus is SARS-CoV2.

7. A method for collecting, detecting, and identifying organisms present in the air, preferably the organisms are air pathogens executable in a system according to any of claims 1 to 4 wherein said method comprises the steps of:
- (a) sampling an air sample with:
 - 5 a substrate configured to capture particles from flowing air thereon using at least one polytetrafluorethylene (PTFE) filter with a nominal pore size of 1 μm or 5 μm ; and
 - at least one processor communicatively coupled with the substrate and operable to enable the air to flow through or over the substrate, thereby causing
 - 10 the particles in the air to be captured by the substrate,
 - (b) extracting the genetic material comprised in the device after step (a),
 - (c) randomly amplifying the genetic material extracted in step (b), and
 - (d) sequencing the genetic material amplified in step (c).
- 15 8. The method according to claim 7, wherein the organisms are selected from the list consisting of: viruses, plants, bacteria, pollen, and fungi, preferably the organisms are viruses, more preferably the virus is SARS-CoV2.
9. The method according to any of claims 7 or 8, which comprises an additional step (a'),
- 20 between steps (a) and (b), in which cellular organisms are removed.
10. The method according to any one of claims 7 to 9, wherein the amplification of step (c) is performed by multiple displacement amplification (MDA) or PCR.

Fig. 1

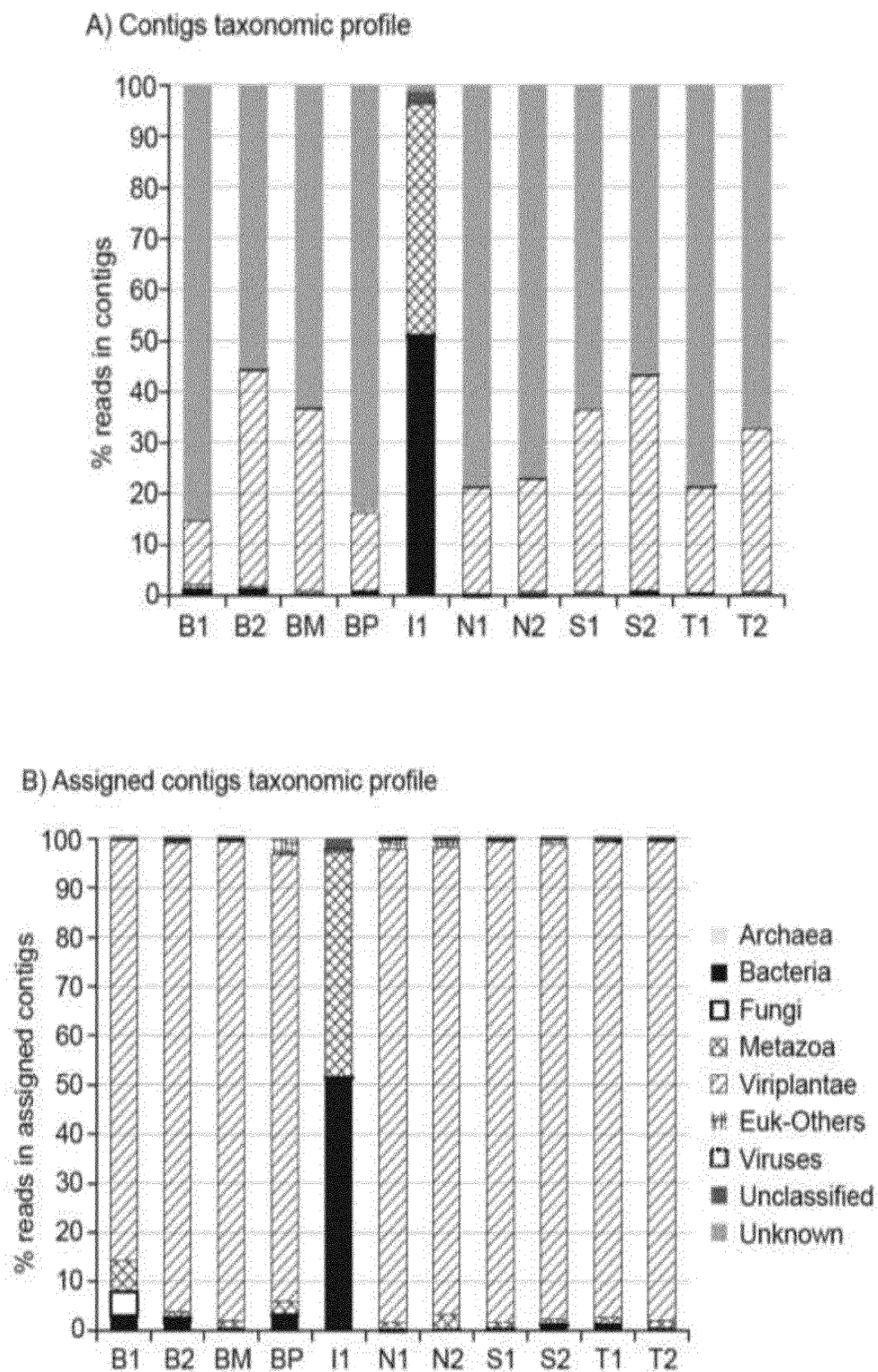


Fig. 2

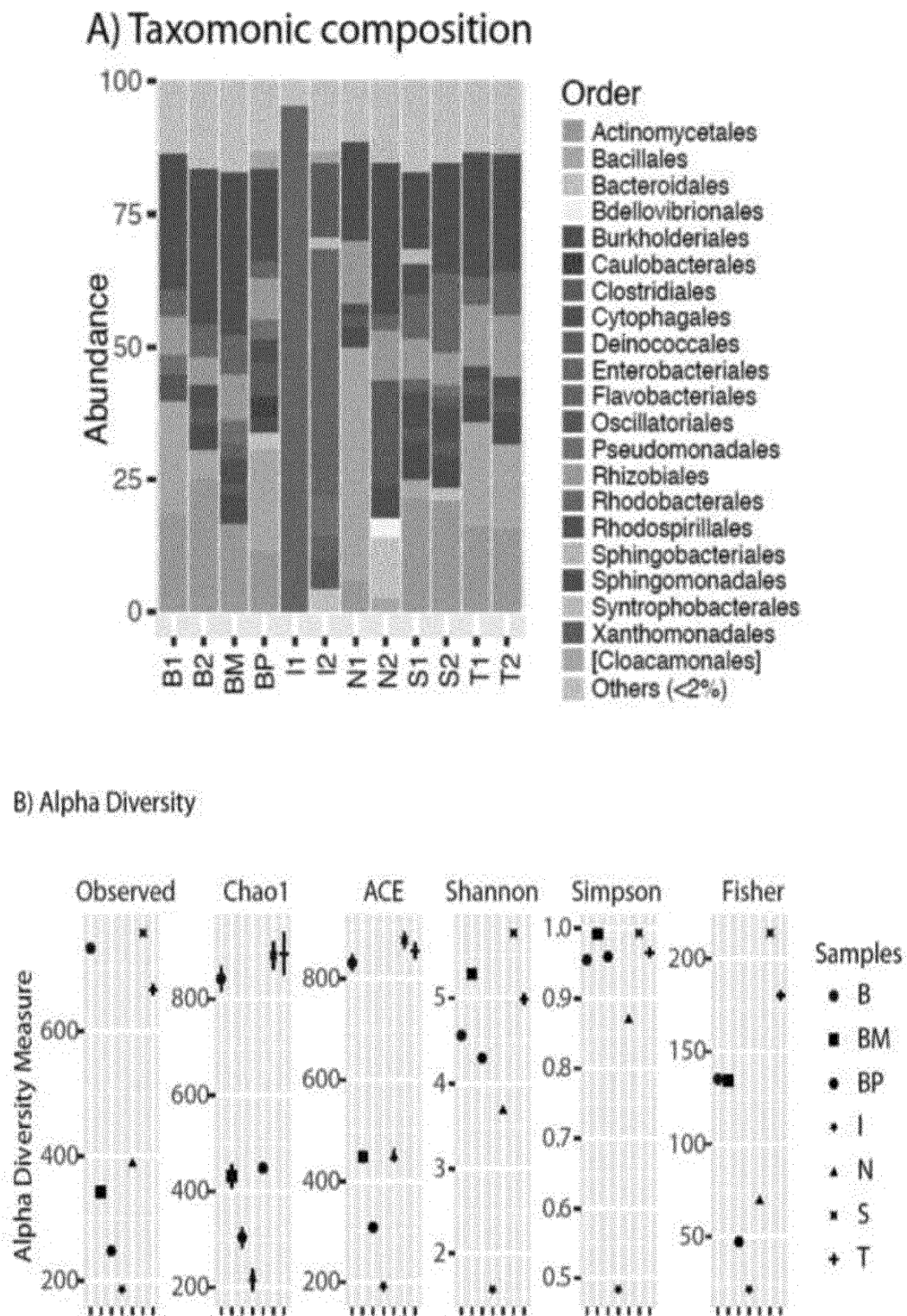


Fig. 2 (cont.)

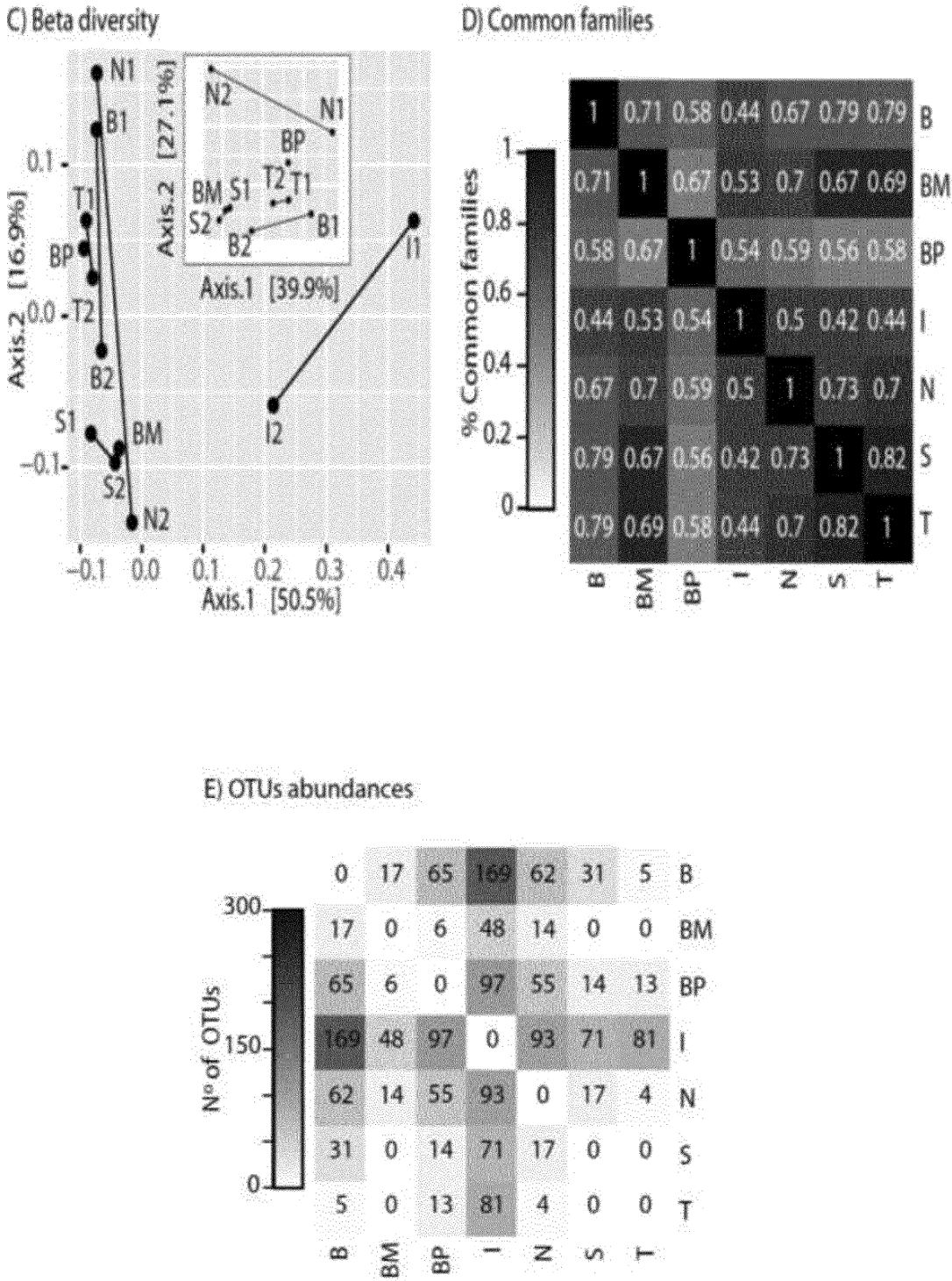
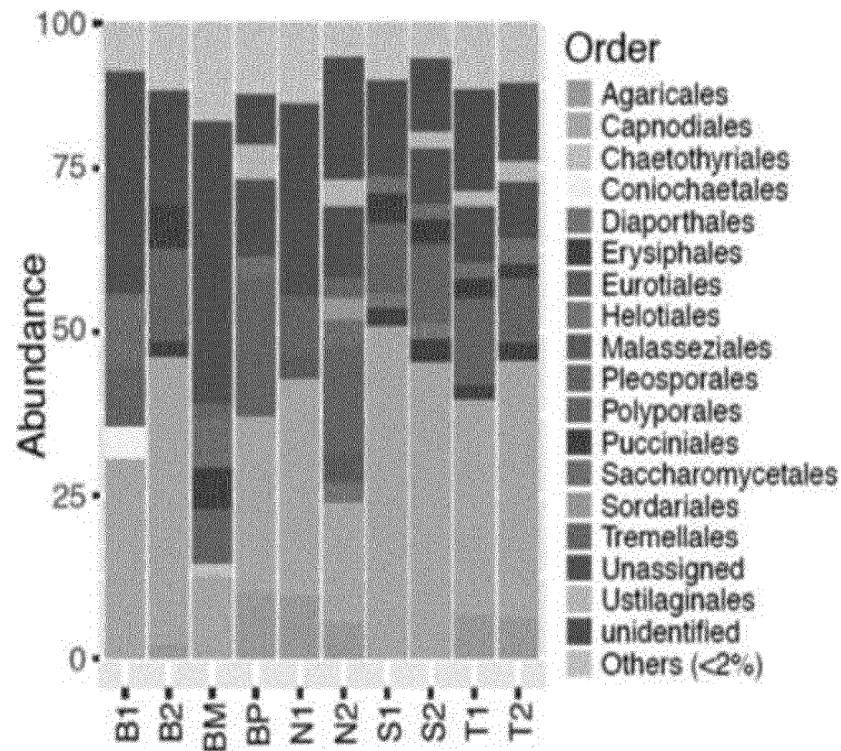


Fig. 3

A) Taxonomic composition



B) Alpha Diversity

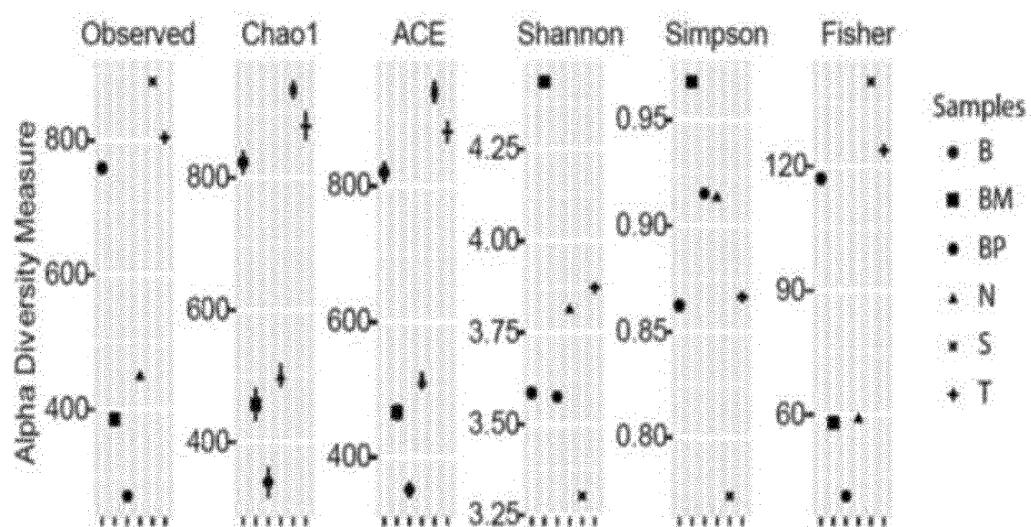


Fig. 3 (cont.)

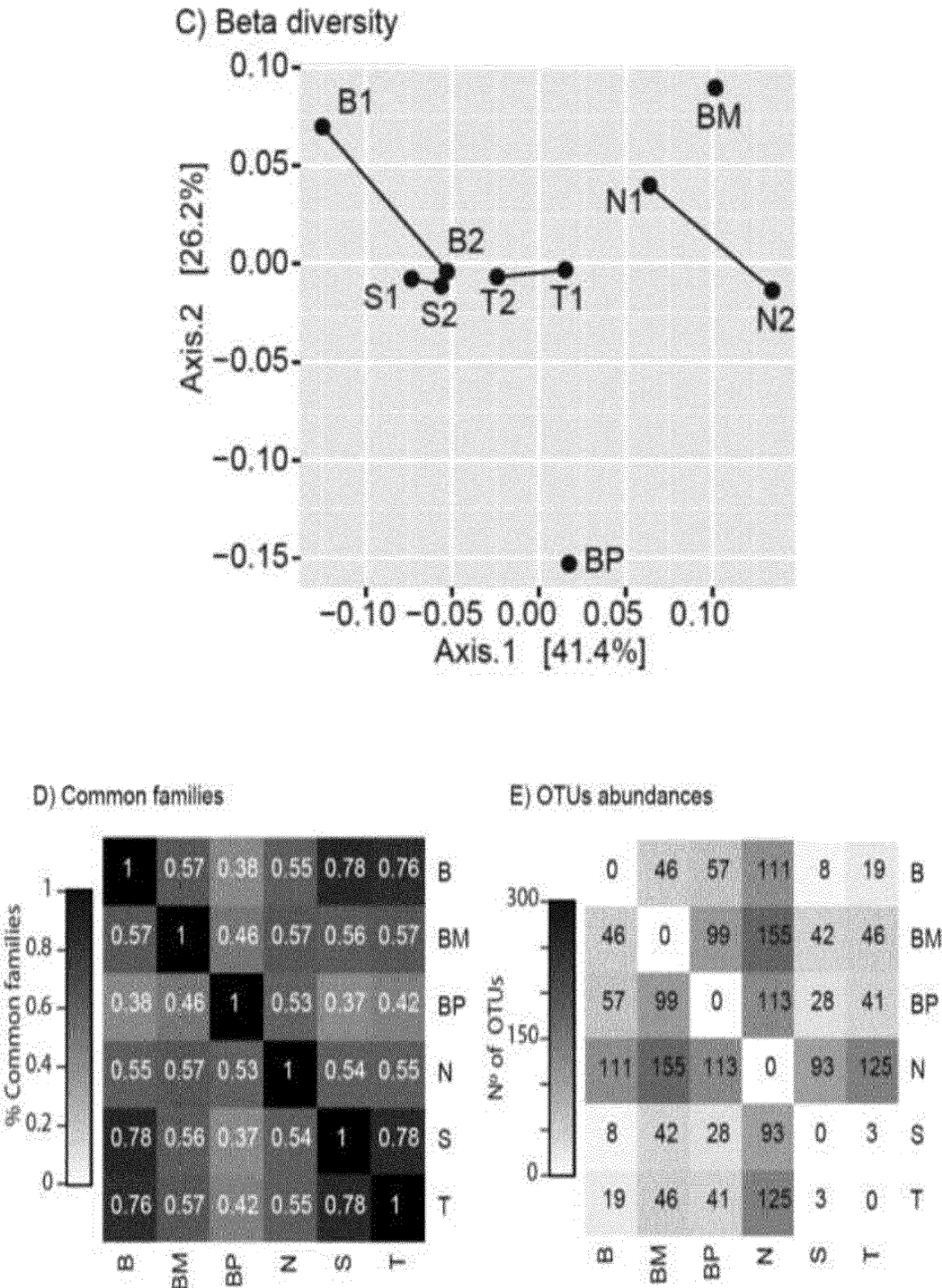


Fig. 4

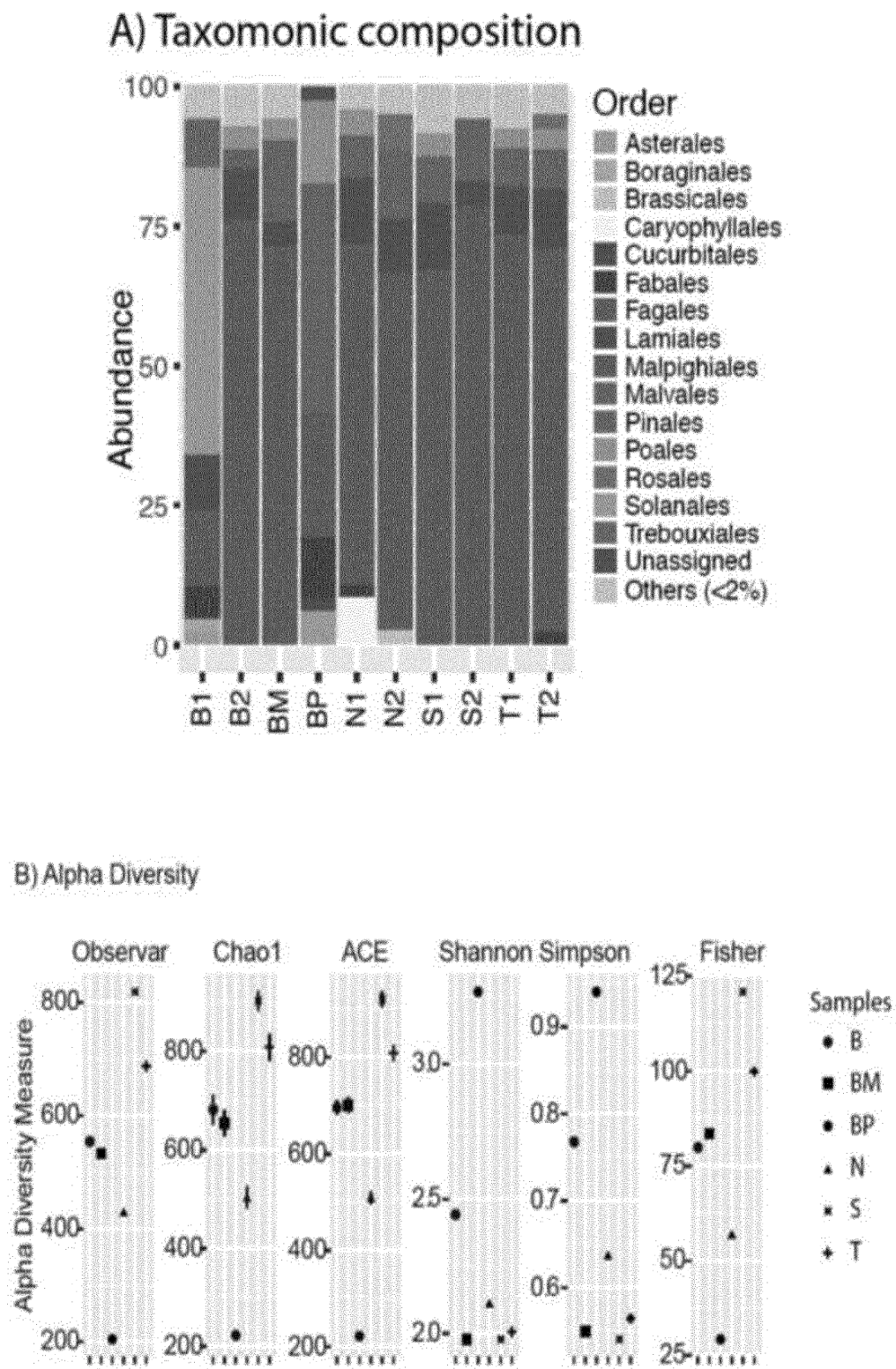


Fig. 4 (cont.)

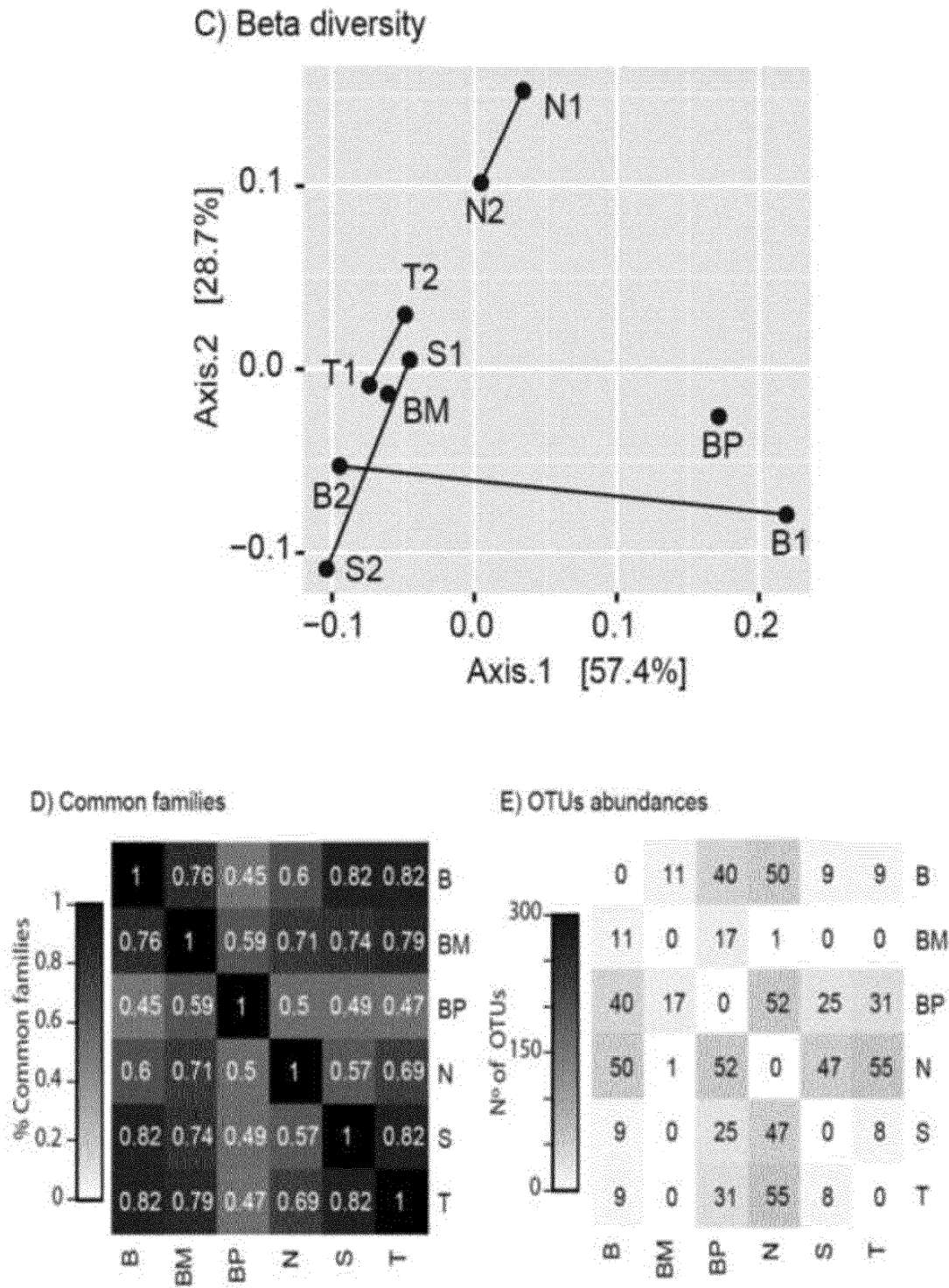
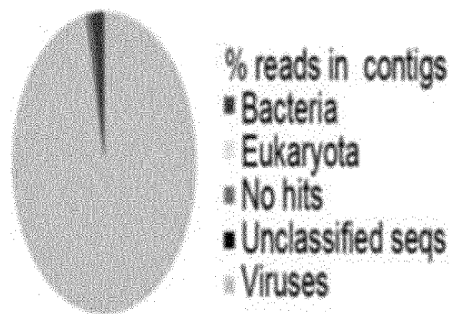
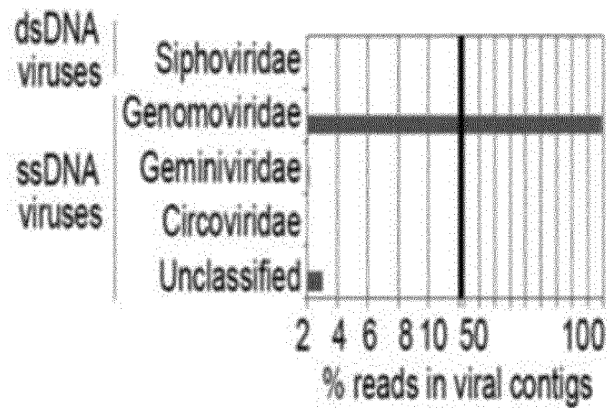


Fig. 5

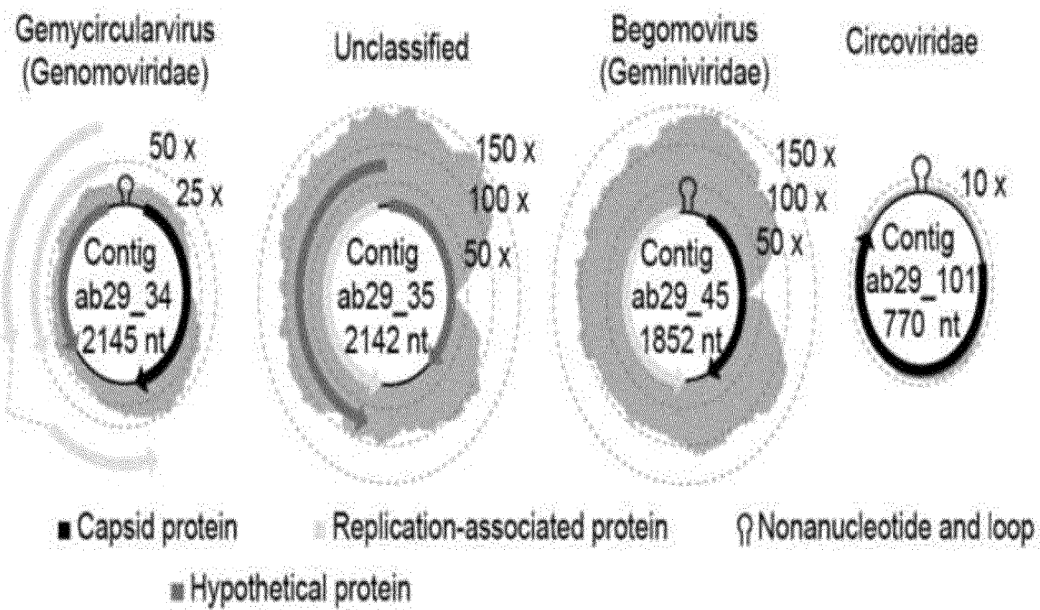
A) Taxonomic overview



B) Viral taxonomy



C) Complete circular genomes



INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2021/065932

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12Q1/70 C12Q1/04 C12Q1/24 C12Q1/6888 C12M1/12
G01N33/497 G01N1/22 C12Q1/6806

ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q C12M G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>Timothy F Booth ET AL: "472 @BULLET JID 2005:191 (1 May) @BULLET Booth et al. Detection of Airborne Severe Acute Respiratory Syndrome (SARS) Coronavirus and Environmental Contamination in SARS Outbreak Units",</p> <p>1 May 2005 (2005-05-01), XP055737310, Retrieved from the Internet: URL:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7202477/pdf/191-9-1472.pdf [retrieved on 2020-10-06] see p. 1472-1474</p> <p style="text-align: center;">----- -/-</p>	1-10

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

29 June 2021

Date of mailing of the international search report

08/07/2021

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Diez Schlereth, D

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2021/065932

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	PO YING CHIA ET AL: "Detection of air and surface contamination by SARS-CoV-2 in hospital rooms of infected patients", NATURE COMMUNICATIONS, vol. 11, no. 1, 29 May 2020 (2020-05-29), XP055737313, DOI: 10.1038/s41467-020-16670-2 see p. 1-2, 4, 6 -----	1-10
Y	KRISTEN K. COLEMAN ET AL: "Bioaerosol Sampling for Respiratory Viruses in Singapore's Mass Rapid Transit Network", SCIENTIFIC REPORTS, vol. 8, no. 1, 30 November 2018 (2018-11-30), XP055737523, DOI: 10.1038/s41598-018-35896-1 see p. 1-5, Fig. 5 -----	1-10
A	Niosh: "NIOSH personal bioaerosol cyclone sampler", 29 January 2014 (2014-01-29), XP055737594, Retrieved from the Internet: URL:youtube.com/watch?v=YCH0hw-DQTC [retrieved on 2020-10-07] the whole document -----	1-10
Y	Theodore A Myatt ET AL: "BMC Public Health Airborne rhinovirus detection and effect of ultraviolet irradiation on detection by a semi-nested RT-PCR assay", 13 January 2003 (2003-01-13), XP055737259, Retrieved from the Internet: URL:http://www.biomedcentral.com/1471-2458/3/5 [retrieved on 2020-10-06] see p. 1-4, Fig. 1 -----	1-10
Y	THEODORE A. MYATT ET AL: "Detection of Airborne Rhinovirus and Its Relation to Outdoor Air Supply in Office Environments", AMERICAN JOURNAL OF RESPIRATORY AND CRITICAL CARE MEDICINE, vol. 169, no. 11, 1 June 2004 (2004-06-01), pages 1187-1190, XP055737263, ISSN: 1073-449X, DOI: 10.1164/rccm.200306-7600C see p. 1187-1188 ----- -/--	1-10

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2021/065932

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KARYNA ROSARIO ET AL: "Diversity of DNA and RNA Viruses in Indoor Air As Assessed via Metagenomic Sequencing", ENVIRONMENTAL SCIENCE & TECHNOLOGY, vol. 52, no. 3, 23 January 2018 (2018-01-23), pages 1014-1027, XP055737248, US ISSN: 0013-936X, DOI: 10.1021/acs.est.7b04203 see p. 1014-1015 -----	1-10
A	YI YU LAI ET AL: "Metagenomic Human Respiratory Air in a Hospital Environment", PLOS ONE, vol. 10, no. 10, 2 October 2015 (2015-10-02), page e0139044, XP055737239, DOI: 10.1371/journal.pone.0139044 see p. 1, 3, 6, Fig. 1 -----	1-10
A	US 10 663 454 B2 (AIRBUS DEFENCE & SPACE GMBH [DE]) 26 May 2020 (2020-05-26) see col. 1-4, 11-12 -----	1-10
A	CN 110 093 455 A (INST PATHOGEN BIOLOGY CAMS) 6 August 2019 (2019-08-06) see abstract, [0007, 0015-0017, 0022-0032, 0054] -----	1-10
A	WO 2018/195404 A1 (THE BOARD OF THE TRUSTEES OF THE LELAND STANFORD JUNIOR UNIV [US]) 25 October 2018 (2018-10-25) the whole document -----	1-10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2021/065932

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☒ forming part of the international application as filed:
 - ☒ in the form of an Annex C/ST.25 text file.
 - ☐ on paper or in the form of an image file.
 - b. ☐ furnished together with the international application under PCT Rule 13~~ter~~.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
 - c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
 - ☐ in the form of an Annex C/ST.25 text file (Rule 13~~ter~~.1(a)).
 - ☐ on paper or in the form of an image file (Rule 13~~ter~~.1(b) and Administrative Instructions, Section 713).
2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2021/065932

Patent document cited in search report		Publication date		Patent family member(s)		Publication date
US 10663454	B2	26-05-2020	EP	3336515 A1		20-06-2018
			US	2018164283 A1		14-06-2018

CN 110093455	A	06-08-2019	NONE			

WO 2018195404	A1	25-10-2018	US	2020131509 A1		30-04-2020
			US	2020377927 A1		03-12-2020
			WO	2018195401 A1		25-10-2018
			WO	2018195404 A1		25-10-2018
