UNIVERSITY OF MILANO - BICOCCA

Department of Biotechnologies and Biosciences

PhD in Biology

XXVIII cycle

Drinking water microbiota:

from the source to the tap.

Tutor: **Maurizio Casiraghi**

Coordinator: **Paolo Tortora**

PhD Candidate: **BRUNO Antonia**

2016, February 15th

INDEX

ABSTRACT

Drinking water quality is a public health concern worldwide. Growing evidences depict drinking water as a complex matrix, in which a wide diversity of microorganisms interacts in a dynamic network. Dealing with environmental samples, the “great plate count anomaly” must be taken into account: only a minimal portion of bacteria can grow on cultured media. Molecular techniques can give a deeper knowledge, going beyond the limit of culture-dependent methods, even if the living/death distinction is a traditional limitation.

We collected water samples during an extended monitoring campaign of drinking water treatment plants (DWTPs) located in Milan (Italy). We analysed the microbial community at different steps of the potabilization processes, from the source to the tap: i) raw water from the groundwater, ii) after the passage across granular activated carbon filters and iii) after chlorination.

We first evaluated and standardized a new experimental workflow for microorganisms concentration, environmental DNA extraction and amplification, suitable for molecular analysis and optimized for High-Throughput DNA Sequencing (HTS) approaches.

Since molecular techniques are unable to differentiate between viable and nonviable microorganisms, live/dead ratio of microorganisms for each sampling point was estimated using fluorescent staining coupled with microscopy visualization. From our observations chlorination does not exert a full-scale effect.

We further analysed the presence and the relative abundance of microorganisms across the DWTP through Real Time PCR. The occurrence of resistance genes was detected across the entire DWTP, highlighting the presence of native resistance genes in groundwater and their permanence after potabilization processes. The presence of resistance genes in water is becoming an issue of great interest as the mobile

resistome (i. e. the collection of all the resistance genes of an ecosystem) can easily spread among species.

Recent studies revealed that drinking water treatment process can affect the microbiota structure. We evidenced that carbon filters play a key role in shaping the bacterial community. Likely filters harbour a microbial community that seeds and shapes water microbiota downstream, a variation that is visible even after chlorination. These evidences can help to unravel the dynamics underlying water microbiota changes.

We reported for the first time the presence of the so-called nanobacteria in the entire DWTP, even after chlorination. Nanobacteria showed a differential distribution across the DWTP, emphasising the role of carbon filters in shaping the nano-microbial community downstream.

It is clear that DWTP is not an inert system, but an ecosystem: complex biological processes take place between the source and the tap. A better knowledge of these networks is crucial to improve the management of drinking water facilities.

1. INTRODUCTION

"Water is essential to sustain life, and a satisfactory (adequate, safe and accessible) supply must be available to all." This is the definition of "safe drinking-water", according to the Guidelines for Drinking-water Quality (Fourth Edition, WHO), is nowadays an issue of public health concern.

Water covers about 71% of Earth's surface and just 2.5% of it is fresh water (U.S. Geological Survey*).*

Drinking water sources can be mainly

- ground sources, such as aquifers, groundwater, springs
- surface waters, such as lakes, rivers, glaciers
- precipitation, including rain, snow,…
- desalinated seawater

In most cases waters derived from these sources must be treated in order to be safely consumed. Tap water refers to water delivered to homes from water treatment plants.

WHO Guidelines intend to support the development and the improvement of drinking water safety through the control of hazardous components that could be present in water. Moreover is clearly underlined the importance of the implementation and enforcement of the drinking water quality standards.

There is a shared effort to lead to universally accepted national standards and regulations protective of public health. The importance of the implementation and enforcement of the drinking water quality standards is clearly underlined, in order to maintain drinking-water quality at the highest possible level. A fundamental concept is the awareness that "priorities set to remedy the most urgent problems (e. g. protection from pathogens) may be linked to long-term targets of further water

quality improvements (e.g. improvements in the acceptability of drinking-water in terms of its taste, odor and appearance)". This suggests that the preferable approach should take into account every aspect of drinking water quality, from the source through the drinking water potabilization processes, to the consumer. Such *holistic* approach should imply not only a systematic monitoring, but also the development of methods based on innovative techniques, but well standardized protocols.

1.1. DRINKING WATER QUALITY STANDARDS

As discussed before, still there are no universally recognized and accepted international standards for drinking water.

The **Safe Drinking Water Act (SDWA)** is the principal federal law in the U.S. that establishes the standards for a safe drinking water (since 1974). Environmental Protection Agency (EPA) is required to set standards for drinking water quality following SDWA. Drinking water standards from the EPA specify the levels of contaminants, disinfection agents, and disinfection by-products that are allowed in drinking water, for a total of 90 parameters monitored.

In Europe, the **European Drinking Water Directive (DWD)** (Council Directive 98/83/EC of 3 November 1998 on the quality of water intended for human consumption) establishes the essential quality standards which water intended for human consumption must meet. The European Drinking Water Directive emphasizes both human health, as well as the environment. Under the guidance of the DWD, member states are required to regularly test and monitor a total of 48 parameters (two of them microbiological, 26 chemical and 20 indicators) in the drinking water. These parameters are derived from the World Health Organization's standards (EPA, 2014. Drinking Water Parameters. Microbiological, Chemical and Indicator Parameters in the 2014 Drinking Water Regulations).

For countries without a legislative or administrative framework for such standards, the World Health Organisation publishes guidelines on the standards that must be achieved (Guidelines for Drinking-water Quality, Fourth Edition; World Health Organisation; 2011). China adopted its own drinking water standard for surface water in 2002 (GB3838-2002 (Type II)).

Most of the drinking water quality standards are expressed as guidelines rather than requirements. Few water standards have legal basis or are subjected to enforcement. Two exceptions are the European Drinking Water Directive and the Safe Drinking Water Act, which include requires of legal compliance with specific standards.

1.1.1. THE ITALIAN DRINKING WATER STANDARDS

THE ITALIAN LEGISLATIVE DECREE N. 31 OF 2 FEBRUARY 2001

The Italian Legislative Decree No. 31 of 2 February 2001 implements European directive 98/83/EC on the quality of water intended for human consumption. The Decree, replacing and partly amending the Presidential Decree 236/88, eliminates the concept of guide levels (GL) and maximum admissible concentrations (MAC), and introduces quality and health parameters. It also establishes new microbiological and chemical parameters. The new legal limits are below the ones set by the Presidential Decree 236/88, resulting in improved water quality.

THE ITALIAN LEGISLATIVE DECREE N. 152 OF APRIL 2006

The Italian Legislative Decree No. 152 of April 2006 concerns environmental laws. The Decree governs the procedures for Strategic Environmental Assessment (SEA), Environmental Impact Assessment (EIA) and Integrated Pollution Prevention and Control (IPPC); it sets the standards for soil protection, battle against desertification, water protection against pollution and management of the water resources; it regulates waste management and reclamation of contaminated sites, air protection

and reduction of air emissions as well as compensated protection against environmental damage. Its primary goal is to promote the quality level of human life through safeguarding and improving the conditions of the environment and by responsible and rational use of natural resources.

Metropolitana Milanese S. p. A. (MM), the drinking water company of Milan, refers to the D. Lgs. n. 31 of 2 February 2001 (law since 25 December 2003), implementing the European Directive 98/83/CEE. The drinking water monitoring programme had been defined with the Milan health authority (ASL) and was approved by the Health Institute and by the internationally recognized bodies, in compliance with the AATO provisions. More than 180,000 parameters are tested in a year to guarantee drinking water quality standards.

In particular the Laboratory normally carries out:

- microbiological tests: coliform bacteria at 37 °C, *Escherichia coli*, *Enterococcus spp.*, *Pseudomonas aeruginosa*, total mesophilic bacteria at 37 °C, total thermophilic bacteria at 22 °C;
- chemical-physical tests: total dissolved solids, hydraulic conductivity, colour, turbidity, hardness and pH level;
- chemical tests: residual free chlorine, pesticides, volatile organic compounds (VOC), aromatic hydrocarbons, metals, polycyclic aromatic hydrocarbons, chromium, anions and cations.

In addition, the laboratory constantly monitors the flow of micro-organic and inorganic pollutants not required by the law that may be found in the aquifer. If the parameter level exceeds 60% then they are internally reported as a warning status, if it exceeds 80% then it is upgraded to the alarm status. Following these tests, preventive action are undertaken to reduce the possibility of degradation of the water quality.

1.2. GROUNDWATER AND AQUIFERS

EPA defines “groundwater” as the water beneath Earth’s surface in soil pore spaces and in the fractures of rock formations and “an aquifer” as a natural underground layer of porous, water-bearing materials (sand, gravel) usually capable of yielding a large amount or supply of water. When water can flow directly between the surface and the saturated zone of an aquifer, the aquifer is defined as *unconfined*. The deeper parts of unconfined aquifers are usually more saturated since gravity causes water to flow downward. On the contrary, a *confined aquifer* is an aquifer that is overlain by a relatively impermeable layer of rock.

Surface naturally recharges groundwater and groundwater often flows to the surface, originating springs and wetlands. Groundwater is not composed only by water, but also by soil moisture, permafrost, and geothermal water.

Because of its confinement, it is less vulnerable to pollution than surface water and it is commonly used for public water supplies. Thanks to the construction of extraction wells, groundwater is withdrawn for municipal, agricultural and industrial use.

Nevertheless, the contamination of groundwater is less visible and more difficult to remove than pollution in rivers and lakes. Groundwater pollution often depends on improper disposal of wastes on land, excessive fertilizers and pesticides, industrial waste, leaking oil storages, sewage sludge.

1.2.1. GROUNDWATER AS AN ECOSYSTEM

Is it possible to consider groundwater as an ecosystem?

Groundwater constitutes the largest terrestrial freshwater biome (Griebler et al., 2014a). The groundwater realm still belongs to the least explored habitats on earth.

If today the assessment of the ecological status of surface waters is routine and regulated by national and international directives (i. e. European Water Framework Directive), in the case of groundwater an ecological approach is still missing (Griebler

et al., 2014b). Groundwater monitoring and management regards exclusively chemical and physical parameters, without taking into account the microbiological aspects.

This habitat is extreme in its environmental conditions. From a macroscale viewpoint, fluctuations of environmental conditions (i.e. temperature, water chemistry, flow parameters) are gradually diminishing with depth, becoming stable ecosystems (Griebler and Lueders, 2009). Water infiltrating into the subsurface is continuously depleted in carbon during the passage through soil and sediment (Tufenkji et al. 2002), and as a result, ground water is typically poor in dissolved organic carbon (DOC). Respiration (O_2 consumption) and biomass production (growth rates) are extremely low. However, groundwater ecosystems contribute significantly to the turnover of carbon thanks to the huge volume of aquifers and the comparably long residence times of organic matter in the subsurface (Griebler and Avramov, 2015).

Nevertheless, groundwater systems are colonized by a heterogeneous population, characterized by prokaryotes, microeukaryotes, viruses, and, in the shallow aquifer, meiofauna and macrofauna can be found (Larentis et al., 2015).

A key issue in the survival of subsurface microorganisms is the adaptation to the oligotrophic conditions. Cell densities often decrease logarithmically with depth, although abundances can vary significantly, depending on the sampling strategy and the specific geochemical characteristics of the studied site (McMahon and Parnell, 2014). Many of the bacterial and especially archaeal taxa observed in deep subsurface systems by molecular approaches are distantly related to known prokaryotes or not characterised (Auguet et al., 2010).

Thus groundwater systems are ecosystems harbouring diverse communities of microorganisms. The link between biodiversity and ecosystem dynamics is still not evaluated.

1.2.2. GROUNDWATER – MILAN

In Italy the 85% of drinking water derives from groundwater, whereas only 15% from surface water sources (Onorati et al., 2006).

Groundwater has the great advantage of being filtered by soil permeable layers and therefore it is generally "naturally purified". Deeper and confined aquifers are more protected by contaminations. In some cases water could be potable even without any potabilization treatment.

The groundwater area of Milan is about 2000 km^2 and is contained completely within the Po plain area where both agricultural and industrial activities are widespread. This area has a complex hydrogeological setting. The subsoil is characterized by Pliocene-Pleistocene sediments of fluvial-glacial origin. Aboveground the main lithotypes are constituted by sands and gravels. Going deeper, grain sizes lower and permeability decreases.

This condition determines the presence of unconfined and semi-confined aquifers in the permeable superficial sediments. These sediments reach about 100 m of depth and constitute the so-called "Traditional Aquifer", exploited by municipal water supplies with the majority of captation wells (Fig. 1.1) (Masetti et al., 2007).

Stratigraphic units of Milan underground are permeable horizons (aquifers) separated by impermeable horizons (aquicludes, formed by clay and silt) (Table 1.1):

1. First aquifer: high permeability sediments (gravel and sand with fractions of silt). It is the superficial aquifer, with a maximum of depth of 40 m.
2. Second aquifer: high or intermediate permeability sediments (gravel and sand with fine fractions and horizons composed by sandy conglomerates), goin from 30/40 m of depth to 100 m; aquifer can be confined or unconfined, connected with the first layer.

3. Third aquifer: low or intermediate permeability sediments (silt and clay, with fractions of sand); it is the deeper aquifer, from 100 m of depth, and it is confined.

Aquifer	Depth	Hydraulic conductivity (m^2/s)	Qualitative features
I aquifer	0 - 40 m	$10^{-3}/10^{-4}$	highly vulnerable; it can be subjected to microbiological and chemical contaminations
II aquifer	40 - 100 m	$10^{-4}/10^{-5}$	chemical contaminants can be recorded, in particular where the II aquifer is connected with the aquifer above
III aquifer	100 - 200 m and more	$10^{-4}/10^{-6}$	possible presence of H_2S and, at deeper level, of brackish water

Table 1.1. Aquifers and main characteristics.

The history of the Milan groundwater passed through a phase of intense exploitation, which triggered land subsidence phenomena and pollutants inflow, such as organohalogenated compounds (Beretta et al., 1992), followed by a phase in which there was the interference of groundwater level with underground structures, due to decreased well withdrawal (Beretta et al., 2004). In the last years there was a continuous piezometric level increase, exposing the area of Milan to the problem of subterranean erosion which can occur during foundations for buildings (Colombo et al., 2015).

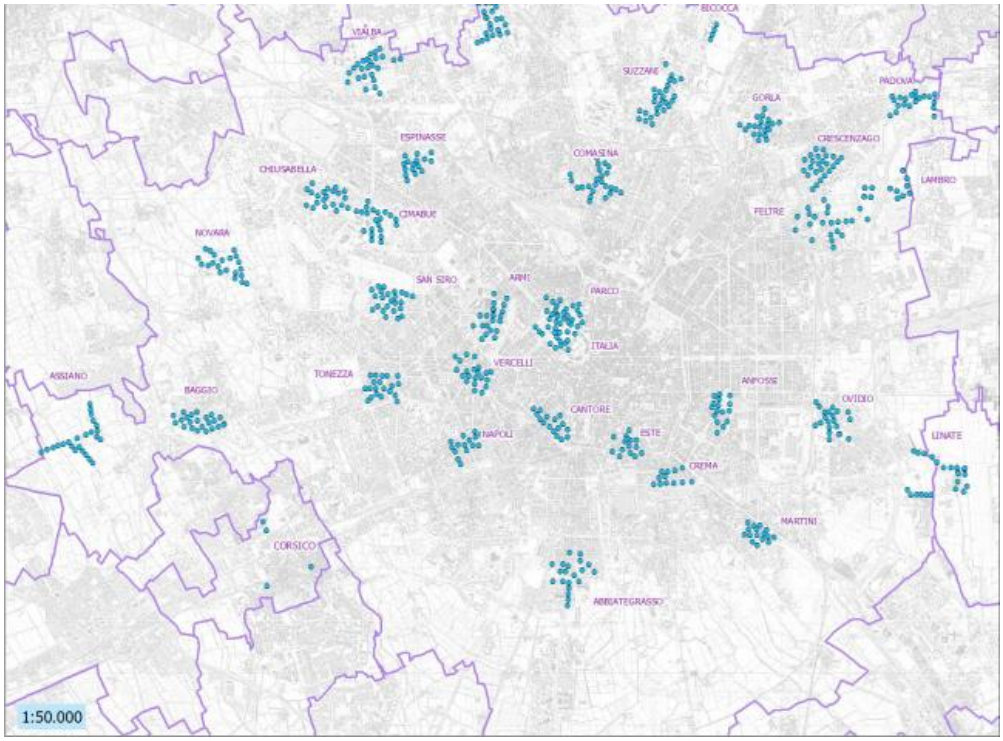


Fig. 1.1. Localization of capitation wells and drinking water treatment plant in the area of Milan, Italy. Image provided by Metropolitana Milanese S. p. A.

1.3. MILAN DRINKING WATER TREATMENT PLANT (DWTP)

Water samples for this study are collected in collaboration with Milan drinking water treatment plant (DWTP) Metropolitana Milanese S. p. A. (MM).

1.3.1. CAPTATION

Water captation is made through captation wells, spread in the entire area of the city of Milan and connected with the treatment plants.

There are 29 DWTPs and each is served by about 12-24 wells, to a total of 400 wells operating in the city.

Water is collected through wells, thanks to two-stage electric pumps: water is drawn to a collection reservoir and therefore to the drinking water pipeline.

Collection basins are used as reservoirs where sand eventually present can sediment.

Electric pump are located at a depth of 40-50 m and have a flow rate of 25-35 L/s.

Wells are activated following water demand, that can varies during the day and depending on seasonality.

The 29 DWTPs operate by remote control thanks to a complex system of telemetry. In this way water capitation is monitored and managed.

1.3.2. POTABILIZATION

GRANULAR ACTIVATED CARBON FILTERS

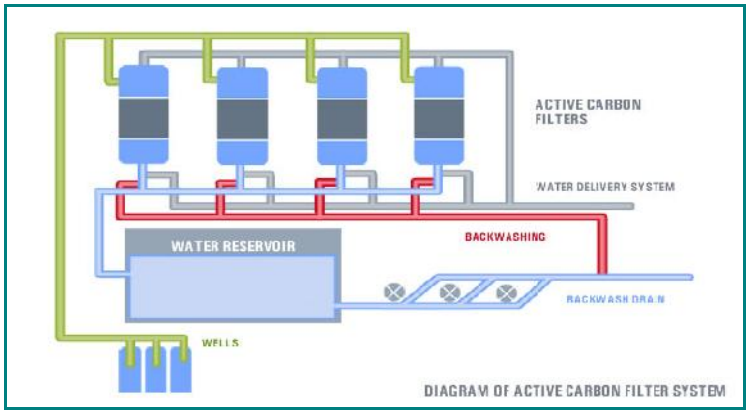
Chemical contaminants are mainly removed through granular activated carbon filters (Fig. 1.2).

Activated carbons entrap in their pores molecules and ions through the chemical process of absorption. Each particle of carbon provides a large surface area: 450 g of activated carbon have a surface of about 40 hectares (Fig. 1.3). This increases the capacity to remove chemical contaminants. Active carbon filters are most affective at removing chlorine, sediments, volatile organic compounds, pesticides, taste and odour from water. They are ineffective at removing mineral, salts and dissolved inorganic compounds. Typical particle size that can be removed by carbon filters are big molecules and range from 0.5 to 50 µm.

Granular activated carbons are obtained from mineral carbons or vegetal carbons heated at high temperatures (900 °C). The efficacy of carbon filters decrease gradually with their saturation by contaminants. After exhaustion due to the use, carbons can be regenerated with a specific heating process. Carbon filers used in Milan DWTPs

become exhausted after 10-24 months of use, depending on contaminants concentration. After several regeneration cycles, the carbon must be replaced.

Aeration towers for air stripping are used in addition or as an alternative to granular activated carbon filters for the chemical pollution removal (Fig. 1.4).



a)


b)

Fig. 1.2. a) Diagram of active carbon filter system. Image from www.metropolitanamilanese.it. b) Carbon filter tanks in Feltre DWTP (Milan).

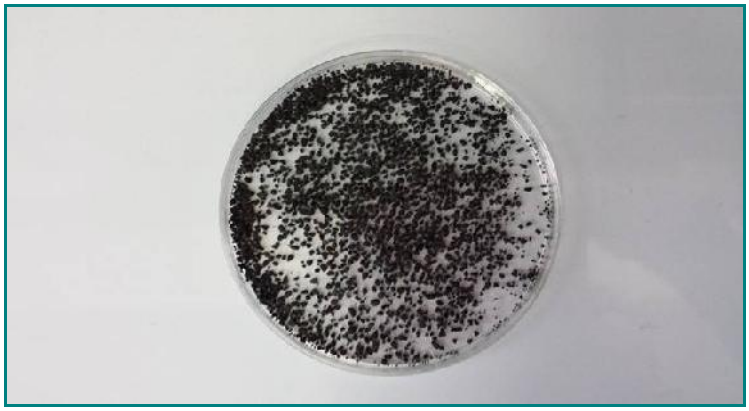


Fig. 1.3. Sample of mineral carbons present in the filters.

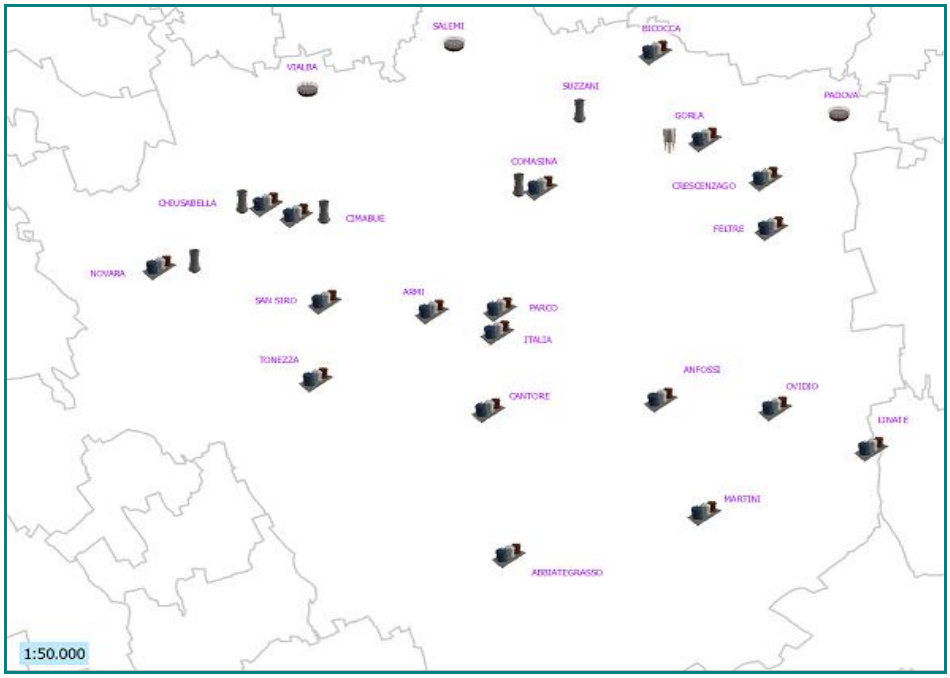


Fig. 1.4. Distribution of drinking water treatment processes to remove chemical contaminants in the city of Milan. carbon filters; aeration towers. Image provided by Metropolitana Milanese S. p. A.

CHLORINATION

Chlorination is one of the most widely used disinfection processes in drinking water treatment plants to reduce pathogenic microorganisms. In order to obtain a water of high microbiological safety, chlorine (Cl_2) or hypochlorite is added to water. This potabilization process is used to destroy pathogen microorganisms and maintain the safety along the entire pipeline. As a halogen, chlorine is a strong oxidizing agent, killing cells via the oxidation of organic molecules (Calderon, 2000). In particular, the mechanism is triggered by the penetration of chlorine and the hydrolysis product hypochlorous acid across the layer of the cell membrane. These molecules not only disintegrate the lipids that compose the cell wall, but also react with intracellular enzymes and proteins, inactivating them. In this way, chlorine exerts its bactericidal or bacteriostatic action (Kleijnen, 2011).

Chlorine is added to public water supplies to kill disease-causing pathogens, such as bacteria, viruses and protozoans, preventing the spread of waterborne disease (i.e. cholera, dysentery, and typhoid fever).

Milan DWTPs uses sodium hypochlorite to disinfect the water and guarantee that it is perfectly sanitary, from the pumping station to the tap. In Milan DWTPs chlorination is carried out at low dosage, downstream the other potabilization processes, since groundwater has typically a low bacterial load.

DISTRIBUTION

The drinking water distribution network total length is about 2300 km (e. g. the distance from Milan to Cairo), reaching more than 50000 end users. The web-like network follows the street layout of the city and consists of pipes with diameters ranging from 80 to 1200 mm. The network is made solely of metals such as steel (15 %), used mainly for large pipes which, since the ‘80s, has partly replaced grey iron (65 %) and ductile iron (20 %), materials with good corrosion resistance and a wide range of special pieces readily available on the market. The drinking water pipes are laid on

average about 1.5 m below ground, to protect the water from changes in temperature: water temperature remains constant at 14-15 °C to the point where it is delivered to the houses.

1.4. CULTIVATE OR NOT?

1.4.1. CULTURE-DEPENDENT METHODS

Standard microbiological tests typically imply an enrichment step in order to increase the level of the target/s: it is the case of classical culture-dependent methods.

In accordance with regulatory requirements, water companies routinely use culture-dependent methods to detect and enumerate specific waterborne pathogens.

In particular total and faecal coliforms are easily screened to assess risk of faecal contamination using specific media or enzymatic reaction (Colilert®, IDEXX Laboratories and a modified version, Quantity-Tray). Another good indicator of faecal contamination is the sulphite-reducing anaerobe bacterium Clostridium perfringens, since spores formed by this bacterium are mainly of faecal origin and can survive disinfection (Ashbolt et al., 2001). Culture conditions are anaerobic, on specific media.

Heterotrophic bacteria able to grow on specific solid media are counted to provide a general estimation of the bacteriological load in the water samples. Even if heterotrophic plate count yields information only about a limited fraction of the whole microbial community in a sample, it is widely considered a convenient tool for water utilities to assess the efficiency of water treatment and to infer regrowth of microorganism in the network (WHO, 2003). Advantages are the low cost, relative simplicity, wide acceptance and long history of the methods.

Culture-dependent methods have mainly two great disadvantages: first, they require an incubation step, extending the analysis time. Second, the majority of bacteria are unable to grow on cultured media. Indeed this discrepancy was called "*the great plate count anomaly*" (Staley and Konopka, 1985; Hugenholtz, 2002) and it has been well

documented for several types of samples. For example, in the case of bacteria inhabiting soil or aquatic environment, only 0.1-1% of them is able to grow on common media under standard conditions (Torsvik et al., 1990; Torsvik and Øvreås, 2002; Kogure et al., 1979; Connon and Giovannoni, 2002). Moreover, containment facilities for cultivating pathogens are required. Another great detriment is the necessity of selective media for specific bacteria, impeding the simultaneous investigation of different microorganisms that characterize complex matrices. As a consequence, at the present state of the art, **there is no standard methodology available for the simultaneous detection of different microorganisms deriving from complex matrices**.

1.4.2. CULTURE-INDEPENDENT METHODS

Despite the culture-dependent techniques represent a convenient tool for water companies, they provide limited information about the total microbial community (encompassing <1% of the diversity) and the variation that can occur in the community composition. The application of culture-independent techniques overcomes these limitations. Culture–independent techniques, such as molecular techniques, can be a suitable tool to analyze drinking water bacterial communities in depth, improving not only the knowledge about this uncovered world, but also the management of water source.

Molecular techniques, such as **qPCR** and **high-throughput DNA sequencing (HTS)**, can provide great advantages in bacterial community analyses (Galimberti et al., 2015).

The adoption of these techniques as standard method to uncover microbial diversity is slow, due to some practical reasons: there is the requirement of specialized and trained employees and specific equipment; even if the prices for the analyses are dropping, they are still more expensive than the culture dependent methods.

However, the quantity and quality of information that can be achieved using these methods is incredible higher than those obtained with other methodologies.

One of the main, and unsolved, *caveat* is the impossibility to distinguish live from dead microorganisms. So molecular techniques should be integrated with other assay, such as live/dead staining or RNA analyses, to overcome this issue (Yu and Zhang, 2012).

1.5. CONCENTRATION TECHNIQUES

Groundwater is characterized by low concentrated and often uncultivable microorganisms (Douterelo et al., 2014).

Although there is a range of high sensitivity molecular techniques available to detect bacteria, in many cases the key step of enrichment is still needed to increase the concentration of the target/s, to be compatible with the sensitivity of such techniques.

There are several different methods for simultaneous concentration of multiple microorganisms from aqueous solutions, for example filtration on membrane discs, precipitation and centrifugation (Block and Rolland, 1979; Payment et al., 1989; Polaczyk et al., 2007; Hill et al., 2005; Hill et al., 2007). Some of these methods were successfully applied in metagenomic studies for microbes concentration from environmental samples (Cai et al., 2015; Furtak, 2015; Khaler et al., 2005). Nevertheless, new improved techniques are indispensable in order to tackle new challenges: dealing with heterogeneous samples with low amount of DNA and presence of inhibitors, coupled with the necessity to abandon culture-dependent techniques, increasing the sensibility and decreasing the response time are the main issues.

Normal orthogonal flow filtration and culture-dependent methods are the standard water testing of American Public Health Association (APHA, 1992) and of European

directives (Drinking Water Directive, 98/83/EC), in which the microbes of water flowing are trapped on membrane disc filters and then plated.

Unlike orthogonal filtration, **tangential flow filtration (TFF)** does not rely on capturing microbes in the filter, avoiding filter clogging when cells and other abiotic components are trapped in the filter maze. In TFF microorganisms and other particles tend to remain in the bulk water samples during the filtration process, recirculating in the system. Ultrafilters have pore sizes that are rated by molecular weight cutoff (MWCO, typically reported in Daltons), allowing the selection of size particles concentrated.

TFF has been extensively used in the biotechnology industry to recover proteins, metabolic products, plasmids, enzymes (van Reis and Zydney, 2001), and in very few studies it was applied to microbial concentration. Recently TFF was used in some relevant papers with the aim to concentrate microorganisms, from endospores to viruses to pathogen bacteria, in different liquid matrices (Fu et al., 2005; Liu et al., 2012). Tangential flow filtration may be the most appropriate method for the concentration of aqueous solutions derived from environmental samples or liquid samples characterized by heterogeneous composition in species.

The ability of concentrate at once diverse microbes is pivotal. In the case of drinking water contaminated by an array of unknown microorganisms or more generally in environmental monitoring of water and other aqueous matrices TFF could be considered the best choice.

Overall there is also a lack of information about how filtration affects microorganisms viability. Complex communities of microorganisms, composed by both live and dead bacteria, could show a not predictable behaviour.

1.6. MOLECULAR ANALYSES

Experimental workflow must be addressed in order to obtain samples suitable for PCR and high-throughput sequencing: these are the ideal analyses to perform in the case of sensitive quantification (qPCR) and wide qualitative composition estimation (HTS).

Reviews in the current literature mainly focus on DNA extraction from bacteria cultures or human specimens (McOrist et al., 2002; Bali et al., 2014; Yuan et al., 2012). However, the application of standard methods can often be ineffective for samples of environmental origin and derived from complex matrices.

For this reason the identification of an effective metagenomic DNA extraction method for complex matrices such as environmental samples must be the goal of experimental plans that strive for subsequent high-throughput sequencing analyses.

The experimental procedure must be prone to sensitivity and must not be inclined to contamination with exogenous DNA. Moreover, the biases deriving from the involvement of several different microorganisms, with different characteristics, must be minimized. Not last, a rapid and standardized protocol, with reasonable costs is recommended.

1.7. NANOBACTERIA

The existence of ultra-small biodiversity in aquatic environments is much closer to us than previously expected. Few months ago, nanobacteria have been surprisingly isolated in groundwater on Earth (Luef et al, 2015; Brown et al., 2015).

Nanobacteria show dimensions under the minimal predicted sizes (Board, 1999) and they are currently defined as a candidate taxon, including at least 35 phyla, with no representatives isolated in culture (Brown et al. 2015). The smallest cell and genome sizes ever documented for nanobacteria reach the minimum values of 195 nm x 149 nm and with a genome of about 1 Mb (Kantor et al., 2013; Luef et al., 2015). For a

comparison the “giant” *Escherichia coli* has a cell size of 2000 nm x 250 nm and a genome size of about 4.5 Mb (Pierucci, 1978; Blattner et al., 1997). Clues on the unexpected diversity of nanobacteria started to appear just after the introduction of two technical advancements: stringent filtration of water and High-Throughput DNA Sequencing techniques.

A prominent feature of the bacterial domain is a radiation of major lineages that are defined as candidate phyla because they lack isolated representatives (Brown et al., 2015) (Fig. 1.5).

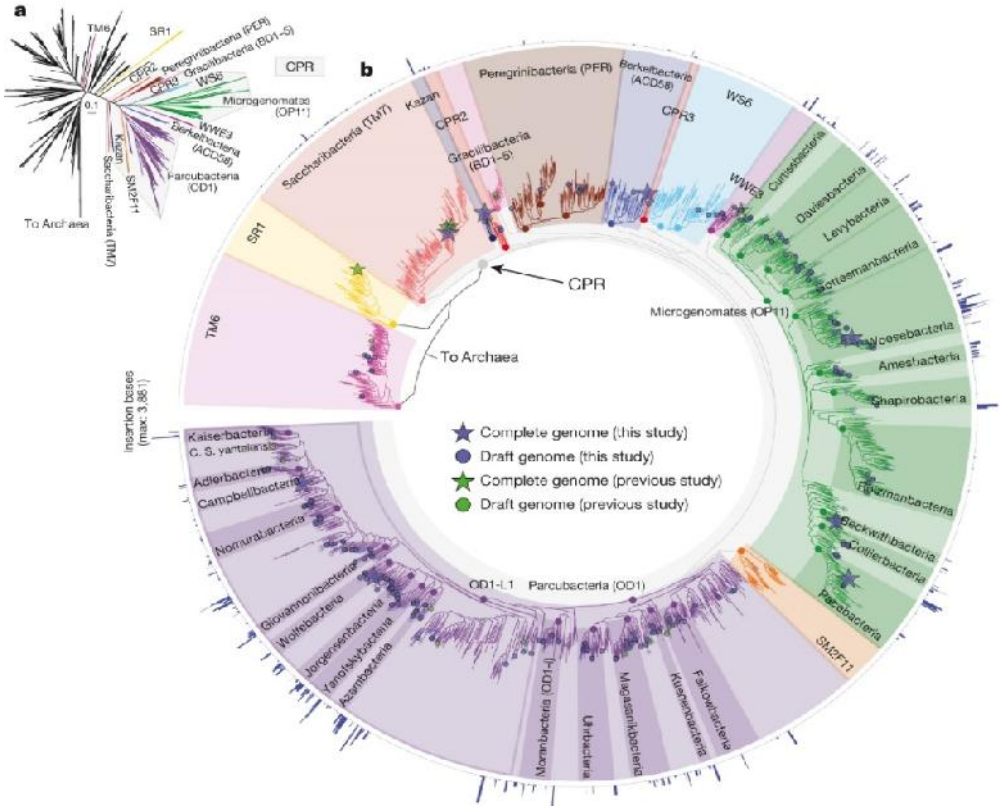


Fig. 1.5. Phylogeny and genomic sampling of the nanobacteria candidate phyla radiations (CPR). **a,b**, Subsets of a maximum-likelihood 16S rRNA gene phylogeny showing the CPR, a monophyletic radiation of candidate phyla (a), and genomic sampling of candidate phyla (b). Many CPR 16S rRNA genes encode insertions (length shown by blue bars, combined length for multiple insertions) (from Brown *et al. Nature* 000, 1-4 (2015)).

1.8. RESISTANCE GENES

As discussed before, water ecosystem is a source of highly biodiverse microorganisms, but it can be also considered an important reservoir of resistance mechanisms.

In the recent few years arose the great interest in antibiotic resistance genes and in the related mechanisms beside the spread of antibiotic resistance.

Several researches demonstrated the correlation between ecosystems and the cycling of resistance in nature, either because resistance mechanisms can originate in environmental bacteria or because human and animal commensals and pathogens can contaminate the environment (Baquero et al., 2008; Vaz-Moreira et al., 2014).

Antibiotic resistance is a natural property of bacteria, occurring in environments with reduced or null anthropogenic impacts, such as wild life or remote Earth zones. It is ancient in nature, with more than 10^6–10^9 years (D'Costa et al., 2011). Functions are as diverse as molecular signalling, transcription activation, enhanced gene transfer, stimulation of bacterial adhesion, increased mutation frequency or virulence suppression (Davies et al., 2006; Davies and Davies, 2010).

Anyway, in the presence of selective pressures, such as antimicrobial residues, bacterial lineages with acquired antibiotic resistance will have an improved fitness (i.e. a better capacity to survive and reproduce in comparison with bacteria without acquired resistance), becoming more prevalent in the community (Vaz-Moreira et al., 2014). Shi (2013) demonstrated that Proteobacteria were the main antibiotic resistant bacteria dominating in the drinking water and that chlorination caused enrichment of ampicillin, aminoglycoside, β-lactam, tetracyclines and erythromycin resistance genes.

In our research we take into account the concept of "resistance" *sensu latu*, meaning all the ways and mechanisms that bacteria can carry or acquire in order to survive under stress conditions, in order to better describe resistance dynamics in water ecosystem. Thus not only from antibiotic point of view but also in terms of the

interplay between ecology, evolutionary dynamics and natural selection (Gillings, 2013).

For example, Chao and colleagues (2013) demonstrated, through metagenomic analysis, that microorganisms in treated water (after chlorination) contained higher protective genes responding to the selective pressure of chlorination, such as glutathione related genes. Glutathione has been proven to directly increase bacterial resistance to chlorine compounds (Chesney et al., 1996) and is also indirectly implicated in the regulation of other oxidation resistant systems, such as OxyR, SoxR and SOS systems (Saby et al., 1999).

1.9. AIM OF THE STUDY

Our first intent was to evaluate and standardize a new pipeline for microorganisms concentration from an heterogeneous matrix, environmental DNA extraction and amplification, suitable for molecular analysis and optimized for High-Throughput DNA Sequencing (HTS) approaches.

While safe and of high quality, drinking water is far from being sterile.

Several questions rise from this statement: first of all, which types of microorganisms are present and if they have effects on human health. Following this, how they are abundant. Can they have effects on human health and on the environment? Can the environment have effects on them, i.e. shaping the microbial communities?

Drinking water bacterial composition is a world largely unknown and poorly investigated, but probably harbours a microbial community highly heterogeneous. Water potabilization processes can affect microbial community structure (Pinto et al., 2012), but as already stated drinking water is not sterile.

This research project aims to uncover microbial biodiversity from the source (groundwater), to the tap (drinking water). A better knowledge of drinking water

microbiota, and how it can be modified by potabilization, is crucial not only from an ecological point of view, but also because it can help managing water treatments by water companies.

2. METHODS

2.1. METHODS FOR THE DETECTION OF ENVIRONMENTAL BACTERIA FOR HIGH-THROUGHPUT BASED STUDIES: TESTS AND STANDARDIZATION

2.1.1. BACTERIAL STRAINS

We used different bacterial strains with different cell wall properties, because these features can affect the cell lysis treatment. Therefore, we evaluated quality and quantity of DNA extracted from monocultures and mock cultures of the following bacteria: the Gram negative *Salmonella choleraesuis* ATCC 7001, *Escherichia coli* ATCC 10536, *Pseudomonas aeruginosa* ATCC 15442, and *Legionella pneumophila* ATCC 33152 and the Gram positive *Clostridium perfringens* ATCC 13124, *Staphylococcus aureus* ATCC 6538 and *Enterococcus hirae* ATCC 10541. Furthermore, a selection of damaged cultures was involved in order to test how bacteria with cell membrane compromised can be concentrated. The species tested were the following (cultivated in our department): *Lactobacillus rhamnosus*, *Lactobacillus plantarum*, *Lactobacillus reuteri*, *Bifidobacterium lactis*, *Bifidobacterium longum*.

Serial ten-fold dilutions were prepared and CFU of each live bacterium was estimated by plating on selective media.

Optical densities (OD_{600}) and/or CFU of each monoculture and of each mix are listed in Table 2.1. Only optical densities were reported for damage cultures.

BACTERIA	GRAM staining	OD_{600}	CFUlog5
***Escherichia coli* ATCC 10536**	GRAM -	1.8	367
***Salmonella choleraesuis* ATCC 7001**	GRAM -	2.4	506
***Legionella pneumophila* ATCC 33152**	GRAM -	0.13	315
***Clostridium perfrigens* ATCC 13124**	GRAM +	NA	NA
***Staphylococcus aureus* ATCC 6538**	GRAM +	2.3	784
***Pseudomonas aeruginosa* ATCC 15442**	GRAM -	2.8	427
***Enterococcus hirae* ATCC 10541**	GRAM +	NA	NA
Lactobacillus rhamnosus	GRAM +	3.1	NA
Bifidobacterium longum	GRAM +	2.7	NA
Bifidobacterium lactis	GRAM +	0.08	NA
Lactobacillus plantarum	GRAM +	2.8	NA
Lactobacillus reuteri	GRAM +	3.1	NA

Table 2.1. Optical densities (OD_{600}) and/or CFU of each monoculture and of each mix. Only optical densities were reported for damage cultures.

2.1.2. ARTIFICIALLY CONTAMINATED SAMPLES

Artificially contaminated samples were prepared in order to test efficiency of recovery and nucleic acid extraction and to show biases typical of heterogeneous samples.

One litre of sterile Milli-Q water has been spiked with known amounts of the species taken into account.

Two experiments were set, with different bacteria composition (Table 2.2).

In the first experiment four different species, two of these alive and two dead, were used to contaminate one litre of Milli-Q water. Three identical spiked one litre samples were created to estimate intra-assay repeatability.

In the second experiment the number of species tested was increased, and, in addition, a second DNA extraction method was included.

2.1.3. ARTIFICIALLY CONTAMINATED ENVIRONMENTAL SAMPLES

Three samples of drinking water (one litres each), were artificially contaminated with the bacterial mix depicted in table 2.2, experiment #3, in order to evaluate the effect of environmental samples characteristics on the method.

2.1.4. ARTIFICIALLY CONTAMINATED ENVIRONMENTAL SAMPLES WITH DEGRADED DNA

Three samples of drinking water (one litres each), were artificially contaminated with the bacterial mix described in table 2.2, experiment #4. Contaminated samples were left for fifteen days at room temperature before concentration, resulting in partial cell lyses and the subsequent release of eDNA.

Exp #1 Artificially contaminated samples	Exp #2 Artificially contaminated samples	Exp #3 Artificially contaminated environmental samples	Exp #4 Artificially contaminated environmental samples in condition of degradation
	E. coli		
	S. choleraesuis		
L. pneumophila	*L. pneumophila*	*L. pneumophila*	*L. pneumophila*
C. perfringens	*C. perfringens*	*C. perfringens*	*C. perfringens*
	S. aureus		
	P. aeruginosa		
	E. hirae		
L. rhamnosus	*L. rhamnosus*	*L. rhamnosus*	*L. rhamnosus*
	B. longum		
	B. lactis		
L. plantarum	*L. plantarum*	*L. plantarum*	*L. plantarum*
	L. reuteri		

Table 2.2. Bacteria composition of artificially contaminated samples for each experiment. In green live bacteria, in red dead bacteria.

2.1.5. ENVIRONMENTAL (DRINKING WATER) SAMPLES

Three samples of drinking water (seven litres each) from a water treatment plant in Milan were additionally tested, without artificial contamination, to definitely verify the applicability in the case of environmental samples. Two different DNA extraction methods were tested.

2.1.6. TANGENTIAL FLOW FILTRATION

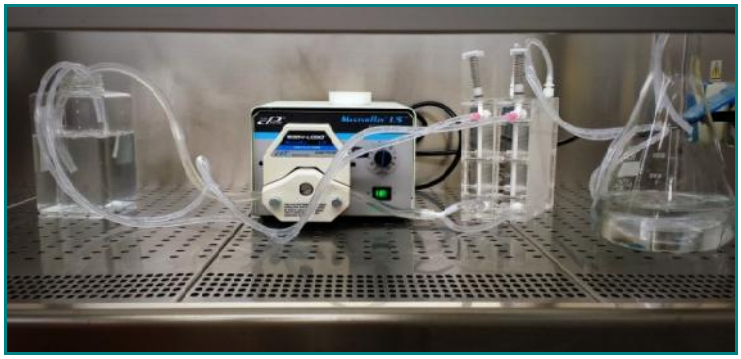


Fig. 2.1. Tangential flow filtration system.

In order to reduce the volume of the samples and therefore concentrate the bacteria we used a tangential flow filtration (TFF) system. The system involves a peristaltic pump (Masterflex L/S Economy Drive), Tygon® tubing, sterile reservoirs and filtration modules.

The tangential flow filter used was a VivaFlow 200 cassette (Sartorius) composed of polyethersulfone (PES) with a nominal pore rating of 10000 MWCO and a surface area of 200 cm^2. The system was scaled up with an additional unit connected in parallel to increase the filtration surface area and the flow speed.

All tubing, tubing connections and containers were sterilized with sodium hypochlorite or autoclaved prior to each experiment. Every step was conducted in the laminar flow cabinet (Fig. 2.1).

The TFF system was run at a transmembrane pressure of max 1.5 bar.

The initial one litre solution artificially contaminated was concentrate to a final retentate volume of 100 mL, for each of the three replicates.

Three aliquots of filtrate (that should not contain bacteria) were conserved for further tests.

The same procedure was followed for environmental samples: seven litres of drinking water was concentrated to 100 mL for each of the three replicates.

2.1.7. MICROSCOPY VISUALIZATION

All bacterial monocultures, the pre-filtration mix of bacteria, the spiked water, the concentrated water and the filtrated were visualized at an epifluorescent microscope (Nikon Y-FL) at 100x and 60x magnification.

Twenty µL of each samples was stained with 20 µL of 2X solution of SYTO9/propidium iodide (BacLight Bacterial Viability kit, Invitrogen) and incubate in the dark, at 4 °C, for 15 minutes.

SYTO 9 is a dye with similar properties of SYBR GREEN I, allowing live cell staining. Otherwise, propidium iodide penetrates only damaged cell membrane, quenching SYTO 9 fluorescence and giving red coloration to the cells. The excitation/emission wavelength is 480/500 nm for SYTO 9 stain and 490/635 nm for propidium iodide.

Live/dead ratio was estimated, with particular attention for pre-concentration and post-concentration samples (Patel et al. 2007, with modifications). All counts were made in triplicate.

2.1.8. DNA EXTRACTION

Genomic DNA (gDNA) extraction was carried out on the samples before the concentration process (the solution used for the contamination of the sample) and after the concentration process. DNA from filtrate was extracted too.

Genomic DNA extraction was carried out with two different methods: one-step rapid DNA extraction with Instagene Matrix (Bio-Rad) and an automated nucleic acid extraction using NucliSens® EasyMAG™ system (Biomerieux).

Instagene procedure one-step nucleic acid extraction utilizes a lysis solution (Chelex ®). Manufacture instructions were followed. A volume of 200 µL of supernatant containing DNA was stored at -80 °C.

For nucleic acid extraction with EasyMAG system, the specific protocol for the increase of DNA yield was used. The nucleic acids were eluted in a final volume of 50 µL and stored at -80 °C.

2.1.9. REAL TIME PCR

Quantitative Real Time PCR (qPCR) assays were performed with AB 7500 (Applied Biosystem) using species-specific primers.

Samples before the concentration process (called "pre") and after (called "post") were tested. Dilutions were used.

qPCR conditions included an initial denaturation at 95 °C for 10 min, followed by 40 cycles of denaturation at 95 °C for 15 s and annealing-elongation for 1 min. Primer sequences, targets, annealing temperatures and references are given in Table 2.3. Standard curves were generated using tenfold serial dilutions of positive controls. All samples and standards were run in triplicate. Negative controls were tested in triplicate too for each amplification. All the assays are followed by a dissociation stage and melting curves were obtained.

Amplification data were collected and analyzed with the SDS 7500 Real-Time PCR System Software (Applied Biosystems).

Bacteria	primers	sequence	gene	Tannealing	Reference
E. coli	ColiF	CATGCCGCGTGTATGAAGAA	16S	60 °C	Huijsdens et al., 2002
	ColiR	CGGGTAACGTCAATGAGCAAA			
S. choleraesuis	SAL1410f	GGTCTGCTGTACTCCACCTTCAG	bipA	60 °C	Calvò et al., 2008
	SAL1494r	TTGGAGATCAGTACGCCGTTCT			
L. pneumophila	JFP	AGGGTTGATAGGTTAAGAGC	16S-23S	60 °C	Devos et al., 2005
	JRP	CCAACAGCTAGTTGACATCG			
C. perfringens	ClperF	GCATGAGTCATAGTTGGGATGATT	plc	60 °C	Shannon et al., 2007
	ClperR	CCTGCTGTTCCTTTTTGAGAGTTAG			

S. aureus	S.aur_F	GCGATTGATGGTGATACGGTT	nucA	60 °C	Brakstad et al., 1992
	S.aur_R	AGCCAAGCCTTGACGAACTAAAGC			
P. aeruginosa	P.aer_F	AATTCGGCAAATTTGCTGCG	oprL	60 °C	Wong et al., 2014
	P.aer_R	GGAGCTGTCGTACTCGAAGT			
E. hirae	EC_F	AGAAATTCCAAACGAACTTG	23S	55 °C	He et al., 2005
	EC_R	CAGTGCTCTACCTCCATCATT			
L. rhamnosus	LrhamF	TGCTTGCATCTTGATTTAATTTTG	16S	60 °C	Zhang et al., 2012
	LrhamR	GGTTCTTGGATYTATGCGGTATTAG			
B. longum	BloF	CAGTTGATCGCATGGTCTT	16S	60 °C	Ramirez-Farias *et al*., 2009
	BloR	TACCCGTCGAAGCCAC			
B. lactis	BlacF	GCACGGTTTTGTGGCTGG	16S	60 °C	Our lab
	BlacR	GACCTGGGGGACACACTG			
L. plantarum	LplF	CCAGTTGGTTCCATAAGTTG	16S-23S	60 °C	Our lab
	LplR	GTTTCAATGACGACTAACGTC			
L. reuteri	LreuF	GGAACCTACACATCGAAG	16S	60 °C	Our lab
	LreuR	CAAATAACGCGGTGTTCTC			
panbacterial	340F	CCTACGGGNGGCWGCAG	16S	55 °C	16S Metagenomic Sequencing Library Preparation (Illumina sequencing protocol)
	806R	GACTACHVGGGTATCTAATCC			

Table 2.3. Primers used for the quantification of target bacteria through qPCR.

2.1.10. qPCR REPRODUCIBILITY

The reproducibility of the qPCR assays was assessed by intra-assay repeatability and inter-assay reproducibility, using six 1:10 serial dilutions of DNA of positive controls. Three replicates of each sample in the same experiment were tested.

2.1.11. qPCR EFFICIENCY

For each reaction six 1:10 serial dilutions of DNA of positive controls were amplified. Three replicates of each sample in the same experiment were tested and standard curve was calculated.

qPCR amplification efficiencies (E) were based on the following equation (1):

$E = 10^{(-1/slope)}-1$ (1)

and R^2 values (linearity) were 0.99.

2.1.12. RECOVERY RATES

C_T value represents the amplification cycle in which the reaction is at the exponential phase (Threshold Cycle), so they are used for the quantification.

C_T values obtained were transformed in **counts**, applying the equation (2)

Counts $=(2^{(C_{max}-C_T)}*100)/E$ (2)

Where E is the efficiency, specific of each target fragment, calculated with equation (1) and C_{max} is the highest C_T recorded (i. e. the lower limit of detection).

Recoveries (R) were calculated using equation (3)

$R = \{[counts(f)/ counts(i)] * 100\}/F$ (3)

where counts(f) is the counts value corresponding to the quantity of microorganisms DNA extracted after concentration, counts(i) before concentration and F is the factor of concentration. Values were expressed as percentages.

2.1.13. DETECTION OF INHIBITION

For each sample, 1:10 dilutions of the DNA extracts were made in sterile nuclease-free water to test for possible inhibition of DNA polymerase. The expected C_T increase after dilutions ensures that the recovery efficiency was not impaired by PCR inhibition (McKee et al., 2015). When inhibition existed diluted samples were used and counts obtained multiplied for the factor of dilution.

2.1.14. STATISTICAL ANALYSIS

Geometric means were calculated. To determine whether recovery efficiency varied, a one-way analysis of variance (ANOVA) was used. The Tukey-Kramer HSD post hoc test was then used to perform a pairwise comparison between mean recovery efficiency.

2.2. TESTS ON DRINKING WATER SAMPLES

2.2.1. CLONING AND RFLP ANALYSIS TO TEST HETEROGENEITY IN BACTERIAL COMPOSITION

DNA extraction (Biomerieux) was performed on samples 00GW, 00CF and 00CHL, collected in Dec-2013 in Feltre DWTP (Site 1). Amplification of 16S rDNA V3-V4 region was carried out with panbacterial primer pairs 340F (CCTACGGGAGGCAGCAG) - 907R (CCGTCAATTCCTTTGAGTTT). Purified fragments were cloned with *pGEM-T Easy Vector System (PROMEGA)*. Ligation was performed following the protocol (Protocol for Ligations Using the pGEM®-T and pGEM®-T Easy Vectors and the 2X Rapid Ligation Buffer; Table 2.4).

Ligation	
2X Rapid Ligation Buffer, T4 DNA Ligase	5 µl
pGEM®-T or pGEM®-T Easy Vector (50 ng)	1 µl
PCR product	X µl*
T4 DNA Ligase (3 Weiss units/µl)	1 µl
nuclease-free water	To reach 10 µl

Table 2.4. Ligation reaction. *PCR product volume is calculated on the base of the length of the fragment and of its concentration. We used 3 µl, the maximum volume allowed.

Plasmids obtained were cloned in *E. coli* competent cells and plated on specific media. DNA extracts from each colony (correctly transformed) were amplified with the primer pair M13F (5'-TGTAAAACGACGGCCAGT-3') and M13R (5'-CAGGAAACAGCTATGAC-3'). Thermal cycling were conducted using the following protocol: 94 °C for 3 min, followed by 35 cycles of 94 °C for 5 s, annealing at 54 °C for 60 s and elongation at 72 °C for 60 s.

Amplification products were digested with restriction enzymes MseI (10000 U/mL) and MspI (20000 U/mL). 5 µl of template were incubated for 2 h at 37 °C and at 65 °C for 20 min, in a 2X mix, composed by the two restriction enzymes (0.1 µl each), 1 µl of buffer and 3.8 µl of Milli-Q water. The restriction products were visualized with gel electrophoresis.

In this way, different sequences amplified for each fragment purified can be discriminated.

2.2.2. DNA EXTRACTION TESTS

Samples 00GW, 00CF and 00CHL were used for an initial comparison of several DNA extraction protocols: a) InstaGene™ Matrix (Bio-Rad) plus a purification-concentration treatment with Genomic DNA Clean & Concentrator™ kit (Zymo Research); b) NucliSens® EasyMAGTM system (Biomerieux); c) PowerWater® DNA Isolation Kit (MO BIO); d) PowerSoil® DNA Isolation Kit (MO BIO); e) QIAamp DNA Mini Kit (Qiagen); f) Invisorb® Spin Universal Kit (Invitek).

Samples obtained were then quantified through qPCR, using panbacterial primers 340F-518R. C_T values were transformed to counts following equation (2). To determine significant differences, a one-way analysis of variance (ANOVA) was used. The Tukey-Kramer HSD *post hoc* test was then used to perform a pairwise comparison. A 95% confidence level were used to determine if there are statistically significant differences among the samples analyzed.

2.3. SAMPLING CAMPAIGN

Samples were obtained from two drinking water treatment plants located in Milan, Italy (Fig. 2.2).

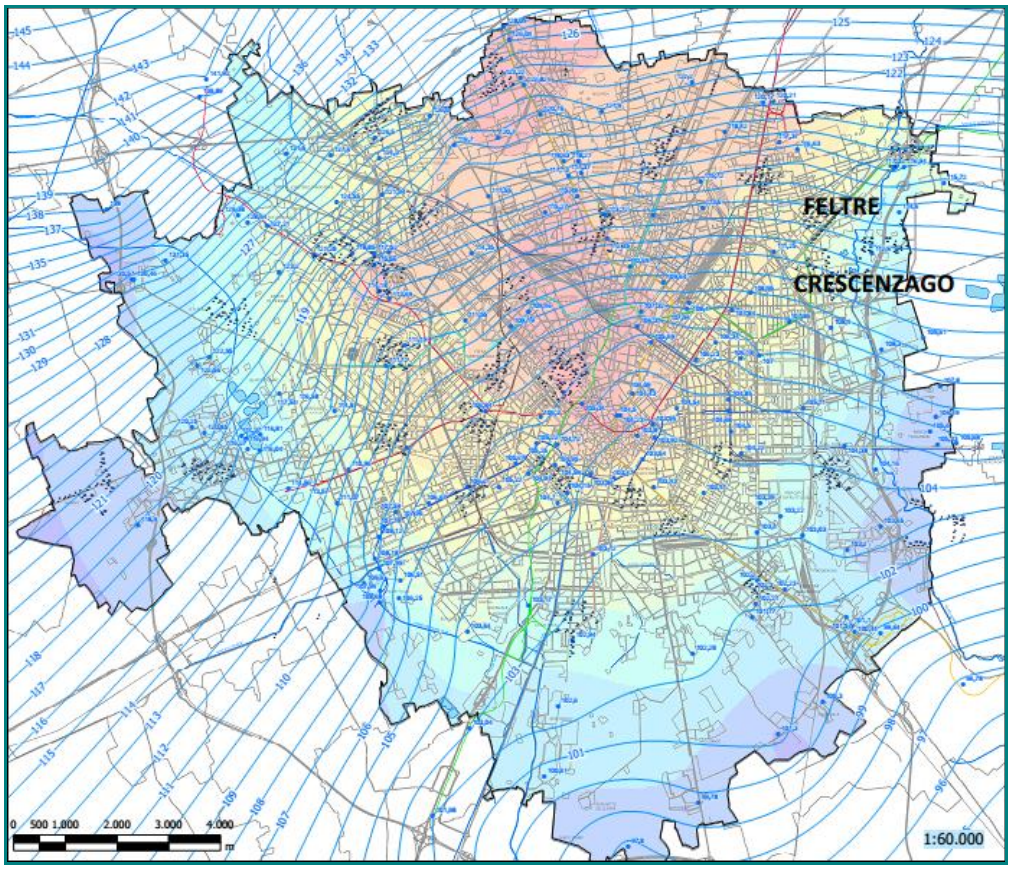


Fig. 2.2. Map of Milan (Italy), with the two DWTPs indicated. Feltre is denominated Site 1, whereas Crescenzago is denominated Site 2.

We collected water samples from different steps of the potabilization processes: i) from groundwater (GW), ii) after the granular activated carbon filters (CF) and iii) after chlorination (CHL) (Fig. 2.3).

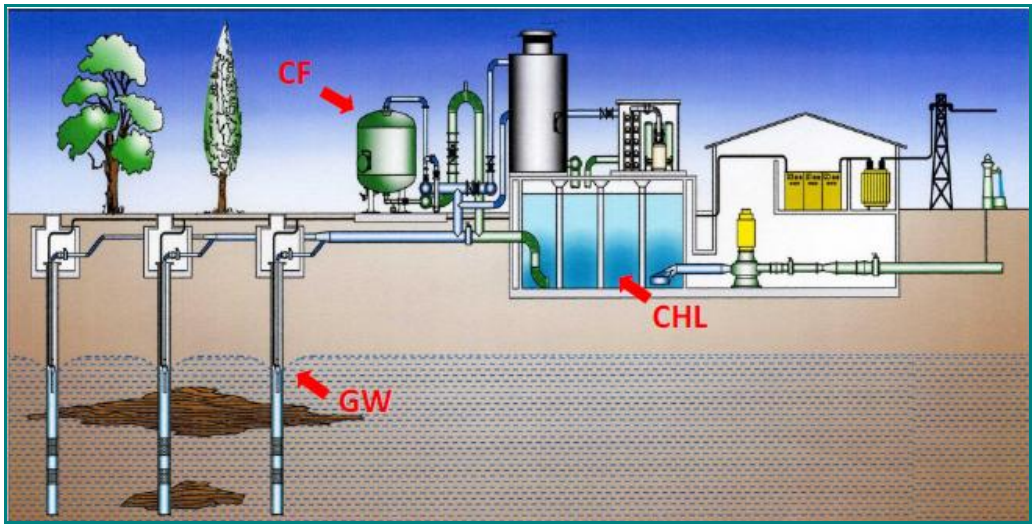


Fig. 2.3. Schematic representation of the DWTP. GW: groundwater; CF: granular activated carbon filters; CHL: chlorination basin.

The sampling campaign lasted one year, from December 2013 to November 2014. In total we collected 42 samples, listed in table 2.5. Several environmental variables were measured (e. g. weather, temperature, humidity) and are reported in table A.2 in Appendix. Moreover MM conducted chemical, physical and microbiological tests (table A.3 in Appendix).

#sample	DWTP	sampling point	date
#01	Site 1	GW	Dec-13
#02	Site 1	CF	Dec-13
#03	Site 1	CHL	Dec-13
#04	Site 1	GW	Jan-14
#05	Site 1	CF	Jan-14
#06	Site 1	CHL	Jan-14
#07	Site 1	GW	Feb-14
#08	Site 1	CF	Feb-14
#09	Site 1	CHL	Feb-14

#10	Site 1	GW	Mar-14
#11	Site 1	CF	Mar-14
#12	Site 1	CHL	Mar-14
#13	Site 1	GW	Apr-14
#14	Site 1	CF	Apr-14
#15	Site 1	CHL	Apr-14
#16	Site 1	GW	May-14
#17	Site 1	CF	May-14
#18	Site 1	CHL	May-14
#19	Site 1	GW	Jun-14
#20	Site 1	CF	Jun-14
#21	Site 1	CHL	Jun-14
#22	Site 1	GW	Jul-14
#23	Site 1	CF	Jul-14
#24	Site 1	CHL	Jul-14
#25	Site 1	GW	Aug-14
#26	Site 1	CF	Aug-14
#27	Site 1	CHL	Aug-14
#28	Site 1	GW	Sep-14
#29	Site 1	CF	Sep-14
#30	Site 1	CHL	Sep-14
#31	Site 1	GW	Oct-14
#32	Site 1	CF	Oct-14
#33	Site 1	CHL	Oct-14
#34	Site 1	GW	Nov-14
#35	Site 1	CF	Nov-14
#36	Site 1	CHL	Nov-14
#37	Site 2	GW	Oct-14
#38	Site 2	CF	Oct-14
#39	Site 2	CHL	Oct-14
#40	Site 2	GW	Nov-14

#41	Site 2	CF	Nov-14
#42	Site 2	CHL	Nov-14

Table 2.5. List of samples collected. GW: groundwater. CF: Feltre is denominated Site 1, whereas Crescenzago is denominated Site 2.

2.4. TANGENTIAL FLOW CONCENTRATION

In order to reduce the volume of the samples and therefore concentrate the bacteria we used a tangential flow filtration (TFF) system, as described in section 2.1.6.

TFF experiments were carried out as soon as practical after sampling.

For each sampling points, seven liters of water were concentrated to obtain 100 mL.

2.5. PLATE COUNT

1 ml of the concentrated water sample was transferred to a sterile Petri dish and mixed with about 25 mL warm agar-based media (composition listed in Table 2.6), incubated at 37 °C for 48 h and at 22 °C for 5 days and then colony forming units (CFU) were counted manually. All measurements were done in triplicate.

Reagents	
Yeast	0.75 g
Triptone	1.75 g
Agar	3.75 g
H_2O	250 mL

Table 2.6. Composition of plate count medium for water samples.

2.6. LIVE/DEAD STAINING

Concentrated samples were stained with fluorescent dyes SYTO9/propidium iodide (*Live/Dead BacLight Viability Kit, Molecular Probes*) to measure membrane-intact

cells, as described in section 2.1.7. All observations were made in triplicate and values are expressed as live/dead ratio.

2.7. ENVIRONMENTAL DNA EXTRACTION

Environmental DNA extraction for qPCR on target genes was performed using InstaGene™ Matrix (Bio-Rad) plus a purification-concentration treatment with Genomic DNA Clean & Concentrator™ kit (Zymo Research) DNA was eluted in a final volume of 50 µL and stored at -80 °C.

In order to increase DNA yield and avoid contaminations with exogenous DNA, environmental DNA extraction for High-Through Sequencing was carried out with an automated nucleic acid extraction (NucliSens® EasyMAG™ system, Biomerieux), based on magnetic beads. Specific protocol for liquid matrices was used. Starting from 1 mL of samples, the nucleic acids were eluted in a final volume of 50 µL and stored at -80 °C.

2.8. qPCR FOR TARGET GENES

Relative quantification of target genes was performed for all the samples.

For bacteria quantification, primer pairs on V3-V4 regions of 16S rDNA (340F-806R) were used. In the case of resistance genes primers are listed in table 2.7.

In brief, all qPCRs were performed in 10 µL reaction mix containing 5 µL of 2X SsoFast Evagreen with Low ROX supermix (Bio-Rad), 0.1 µL of primer forward 10 µM, 0.1 µL of primer reverse 10 µM, 2 µL of DNA extract and 3 µL of sterile Milli-Q water. In each reaction no template controls (negative) and 10-fold serial dilution of positive controls were included to calculate efficiencies. All samples were tested in triplicates. For each DNA extract, amplification inhibition was evaluated measuring C_T values of 1:10 dilution. Melt curve analysis was performed for each amplification. C_T values

were transformed to counts following equation (2). To determine significant differences, a one-way analysis of variance (ANOVA) was used. The Tukey-Kramer HSD post hoc test was then used to perform a pairwise comparison. A 95% confidence level was used to determine if there are statistically significant differences between the samples analyzed.

Target	primers	sequence	T annealing	Reference
ampicillin resistance	AmpCF	CCTCTTGCTCCACATTTGC	55 °C	Shi et al., 2013
	AmpCR	ACAACGTTTGCTGTGTGACG		
β-lactam resistance	blaTEM1F	CATTTTCGTGTCGCCCTTAT	55 °C	Shi et al., 2013
	blaTEM1R	GGGCGAAAACTCTCAAGGAT		
glutamate-cysteine ligase	gshAF	GGCGGCGAAGCGTATCAGAAA	59 °C	Chao et al., 2013
	gshAR	AATGCTTTGCCTGTTCCGCCA		
glutathione synthase	gshBF	CGTGATTGCCGAAACCCTGA	58 °C	Chao et al., 2013
	gshBR	GCCAGATTGCCACGGGTTTC		

Table 2.7. Primers used for the quantification of resistance genes through qPCR.

2.9. HIGH-THROUGHPUT DNA SEQUENCING

Illumina MiSeq 16S (V3-V4 regions of 16S rRNA gene) libraries were generated following standard protocol (16S Metagenomic Sequencing Library Preparation, Part # 15044223 Rev. B) with modifications, due to the low DNA concentrations. DNA extracts were normalized on C_T values of Real Time PCR with the same primer pairs, instead of measuring the total amount of microbial DNA with fluorometric/spectrophotometric methods.

Amplicon PCR was performed using the primer pairs

5’ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG 3’

5’ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC 3’

at initial concentration of [10 μM], aiming to increase the volume of DNA in the reaction.

PCR-clean up step after amplicon PCR was modified in the final resuspension volume, with a two-fold increase of sample concentration.

Libraries prepared were quantified with 2100 Bioanalyzer (Agilent Technologies).

Samples were sequenced using the 2x300 paired-end chemistry (MiSeq Reagent Kit v3). Technical replicates of each sample were included in order to verify the sequencing reproducibility (84 samples in total).

2.10. 16S rDNA SEQUENCE PROCESSING AND OPERATIONAL TAXONOMIC UNIT (OTU) SELECTION

Illumina reads were paired and pre-processed using USEARCH script (Edgar, 2010). During Quality filter step reads were filtered out if: 1) ambiguous bases were detected, 2) lengths were outside the bounds of 250 bp and/or 3) average quality scores over a sliding window of 40 bp dropped below 25. Reads were then processed by Vsearch 1.1.3 software version (doi 10.5281/zenodo.16153), which removed noise and chimeras prior to performing de novo clustering into OTUs at 97% sequence identity discarding those OTUs with < 75 sequences.

A representative sequence was selected randomly for each OTU and classified with the RDP (Ribosomal Database Project) classifier v2.2 (Cardenas et al., 2009) using the Silva reference set (119 release) (Quast et al., 2012). The taxonomic assignment of each sequence was obtained with a confidence score of at least 0.8.

For statistical analyses QIIME (Caporaso et al., 2010) and Phyloseq Bioconductor package (McMurdie and Holmes, 2012) were used.

A beta-diversity measure, Weighted UniFrac distance matrix, which measures the pairwise difference in microbial diversity among samples, and Bray-Curtis distance matrix were calculated (Lozupone et al., 2011). To provide visualization of the sample

distribution patterns, a principal component analysis (PCA) was applied to transform the distance matrices into principal coordinates.

2.10.1. MICROBIAL COMPOSITION AND COMMUNITY STRUCTURE ANALYSIS OF NANOBACTERIA

OTUs assigned to OD1, OP11, TM6 and TM7 (reported as nanobacteria in Luef et al., 2015 and Brown et al., 2015) were identified in our analysis. Moreover we decided to include also OP3, OP1, BRC1 and WS3, even if there are no information about dimensions, since they are not cultivable bacteria and still candidate phyla (not characterized). The OTU representative sequence set was aligned to the Silva set using MOTHUR (Schloss et al 2009). Based on the alignment of OTU representative sequences, a phylogenetic tree was then built using RAxML version 7.4.2 (Stamatakis 2006) with the GTRGAMMA model, bootstrapping (1,000 replicates), best maximum likelihood tree inference, and displayed with iTol (http://itol.embl.de/) representing the output of PhyloH analysis with the nanobacteria OTUs count as multibarplots.

2.10.2. PHYLOH ANALYSIS OF NANOBACTERIAL COMMUNITY

The measurement of the variety of sequences found in the different samples was done within the overall analysis framework defined in Sandionigi et al. (2014), where the ecological concept of gamma diversity $D\gamma$ and alpha diversity $D\alpha$ are identified as the exponential of the phylogenetic generalization of Shannon (Hp) proposed by Chao et al. (2010) and the exponential of the mean Hp per group of observations, respectively.

$$H_p = H_p(T) = H_{p\gamma} = -\sum_{i \in B_T} \frac{L_i}{\tau_T} p_i \log(p_i)$$

$$\tau_T = \sum_{i \in B_T} L_i p_i$$

$$H_{p\alpha} = H_p(T|G) = \sum_{g \in G} p_g \left(-\sum_{i \in B_{T_g}} \frac{L_i}{\tau_{T_g}} p_{ig} \log(p_{ig}) \right)$$

$$\tau_{T_g} = \sum_{i \subset B_{T_g}} L_i p_{ig}$$

where pi are the frequency of observation of organisms descendant of node i and Li is the length of the branch of the node i over the phylogenetic tree T. Subdividing observation in groups is possible to define pig that is the frequency of observations of organisms descendant of node i and belonging to group g. The ecological concept of beta diversity is identified with exponential of mutual information between the species observation and the grouping (I(Obs,G) as proposed by Jost (2007) and we applied this concept to the phylogenetic generalization of Shannon

$$I_p(T, G) = H_p(T) - H_p(T|G) = H_{p\beta} = H_{p\gamma} - H_{p\alpha}$$

This phylogenetic generalization of mutual information describes the information shared between the lineage and the grouping at which a given observation belong. Modifying the order of the summation is possible to extract the contribution of each branch/lineage to mutual information (Hpβi)

$$H_{p\beta i} = -\frac{L_i}{\tau_t} p_i \log(p_i) - \sum_{g \in G} -p_g \frac{L_i}{\tau_{tg}} p_{ig} \log(p_{ig})$$

$$H_{p\beta} = \sum_{i \in B_t} (H_{p\beta i})$$

Following Chao (2010), we report in our work the exponential of Hpγ or Hpγ such that the unit of measure is equivalent number of equi-abundant independent lineage, meaning the number of branch of a star tree in which each terminal taxon is equally abundant and that would produce the same level of diversity than in the actually

observed sample. As summary of the differentiation of communities, the beta diversity, we preferred not to use the exponential of Hβ that would produce estimates in equivalent number of sample, a measure quite ambiguous when sample have different number of observations. So we normalized the Hβ by its maximum possible values given the experimental design. Mutual information shared across two variables cannot be bigger than the entropy of the least entropic variable. Given Ip(T,G) the number of group is fixed, while T is unknown prior data observation, so the maximum value that mutual information could take is H(G) and so mutual information were normalized between 1 and zero using this value. This measure was defined as "***turnover***" and in a case of two groups is the percentage of observations belonging to a not shared lineage.

Discussing the results we noticed that the runs of treated water had in average many more read than the runs from the groundwater. We considered that this difference is due to factor linked to the library preparation and do not reflect difference in the actual effort to explore the three environments. So we modified the calculation of pi and pig such that each runs would contribute equally to those estimates and not proportionally to its number of reads. These changes do not obscure our capacity to correctly assay if the mutual information is different from zero, given that as in Sandionigi et al. (2014), significance was obtained by comparing original dataset with results from permuted data set in which grouping labels were randomly re-assigned to observations.

3. RESULTS

3.1. TESTS AND STANDARDIZATION FOR THE DETECTION OF ENVIRONMENTAL BACTERIA FOR HIGH-THROUGHPUT BASED STUDIES

3.1.1 LIVE/DEAD RATIO VARIATIONS

Single-species cultures were checked at the epifluorescence microscope and live/dead ratios were reported in table 3.1.

Live/dead ratio was estimated for samples from spiked solutions and after the concentration process. No differences were shown in bacteria viability after concentration process.

Moreover damaged cell membranes can stand the pressure exercised by the peristaltic pump, as we can notice in the samples after tangential flow filtration.

No cells were detected in filtrate, for each sample tested.

Bacteria	% Live bacteria pre	% Live bacteria post	Bacteria in filtrate
E. coli	100	100	0
S. choleraesuis	100	100	0
L. pneumophila	100	100	0
C. perfringens	100	100	0

S. aureus	100	100	0
P. aeruginosa	100	100	0
E. hirae	100	100	0
L. rhamnosus	0	0	0
B. longum	0	0	0
B. lactis	0	0	0
L. plantarum	0	0	0
L. reuteri	0	0	0

Table 3.1. Live/dead ratios of spiked samples (named “pre”, as they are observed before TTF) and samples after TTF (named “post”). Moreover the presence of bacteria in filtrate samples was verified.

3.1.2. REPEATABILITY OF EXPERIMENTAL PROCEDURE

In the first experiment intra-assay repeatability was estimated. The three identical spiked samples show no significant differences across all the replicates of the pre- and across all the replicates of the post-concentration samples ($p>0.05$), demonstrating the repeatability of the procedure, from the filtration to the DNA extraction (Fig. 3.1, Table 3.2).

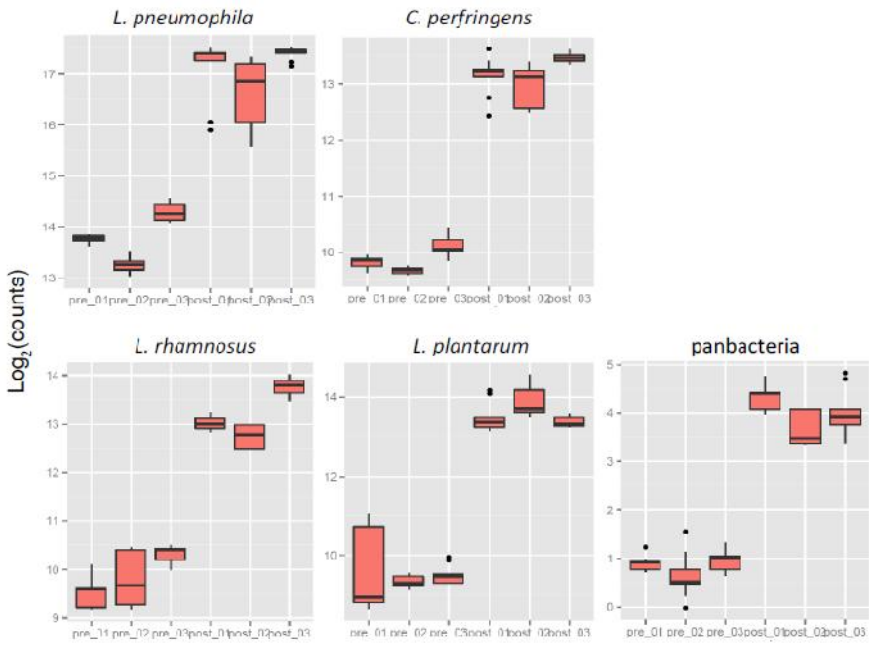


Fig. 3.1. Boxplot of log_2(counts) for each bacteria tested, comparing pre-concentration samples and post-concentration samples. The three replicates tested are shown.

3.1.3. RECOVERY IN ARTIFICIALLY CONTAMINATED SAMPLES

After tangential flow filtration all the bacteria were successfully recovered in all the experiments.

According to equation (3), in the first experiment the overall recoveries for *L. pneumophila, C. perfringens*, *L. rhamnosus* and *L. plantarum* were respectively 124%, 121%, 92%, 113%. When the amplification of all the bacteria present in the samples was carried out with panbacterial primers (16S rDNA V3-V4 regions) the recovery efficiency reach 120%. When diluting the concentrated samples of 10-fold factor and recalculating the counts, recovery efficiencies (indicated with δ) exceeding 100% were no more measured, with the only exception of *L. plantarum*. No significant differences

existed in recoveries between Gram positive and Gram negative bacteria. Moreover, no significance differences were measured between live and dead bacteria ($p<0.01$), suggesting that a damaged cell membrane can fully stand the pressure exercised by tangential flow filtration (Table 3.2, Fig. 3.2).

Target	Recovery efficiency 1.r (%)	s.d.	Recovery efficiency 2.r (%)	s.d.	Recovery efficiency 3.r (%)	s.d.	Mean recovery efficiencies (%)	s.d.	Mean recovery efficiencies δ (%)	s.d.
L. pneumophila	133	0.28	182	0.32	87	0.22	124	0.48	95	3.16
C. perfringens	122	0.25	138	0.1	98	0.2	121	0.34	100	2.6
L. rhamnosus	108	0.34	78	0.58	108	0.25	92	0.72	99	2.9
L. plantarum	112	1.12	156	0.41	104	0.28	113	1.23	113	3.4
panbacterial	192	0.18	142	0.48	119	0.33	120	0.61	90	3

Table 3.2. Recovery efficiencies for experiment #1. s.d.: standard deviation. δ: Recovery efficiencies obtained using dilutions of concentrated samples.

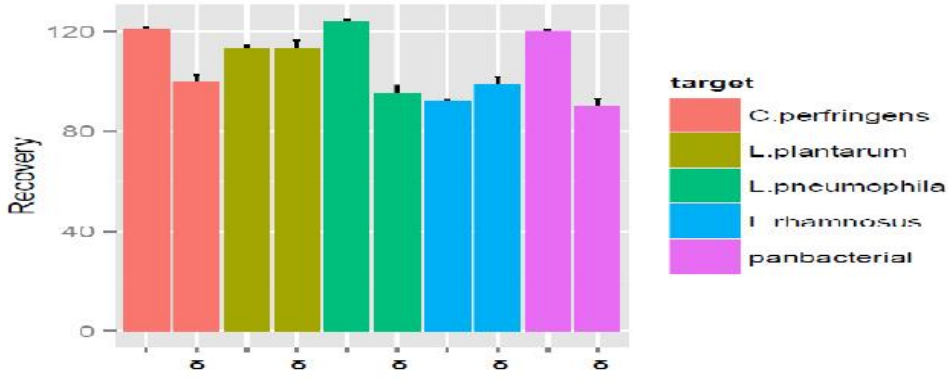


Fig. 3.2. Recovery of artificial contaminated samples in experiment #1. δ: Recovery efficiencies obtained using dilutions of concentrated samples.

When, in the second experiment, the number of species tested was increased to 12, recovery efficiencies were all greater than 61% and there is always a statistically significant difference ($p<0.05$) between samples before and after concentration, for each target (Table 3.3, Fig. 3.3). No significant differences existed in recoveries between Gram positive and Gram negative bacteria. Moreover, no significance differences were measured between live and dead bacteria ($p<0.01$). The only exception was represented by *C. perfringens*, which was not detected in the samples before concentration process, but only after concentration. This was probably due to the low starting amount of target DNA. For this reason the estimation of recovery efficiency was not possible in this case.

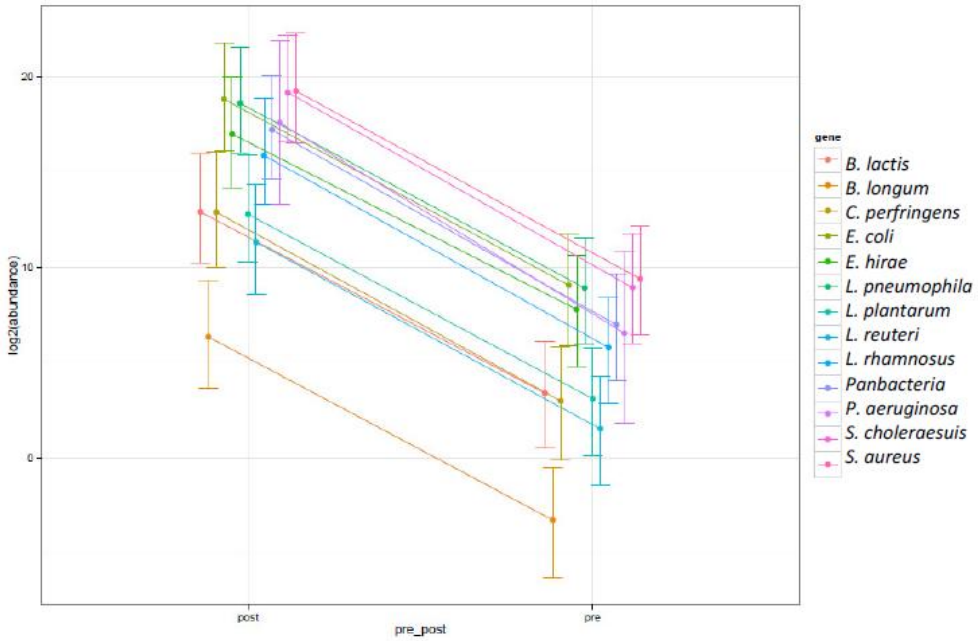


Fig. 3.3. Plot of log_2(counts) for each bacteria tested, comparing pre-concentration samples and post-concentration samples of experiment #2.

In the second experiment a different DNA extraction method was tested on the same samples: instead of the one-step extraction protocol, a more complex, but automated, method was used, in order to prove if the presence of inhibitors could impair the results. First of all, *C. perfringens* was detected even in the sample before concentration, showing an improved sensibility in the extraction method. Recovery was only 49%. In general Ct values of samples after concentration were lower than those with the first DNA extraction method. This indicates that there is a higher sensitivity (Table 3.3, Fig. 3.4, 3.5).

Target	Recovery efficiency A (%)	s.d.	Recovery efficiency B (%)	s.d.
E. coli	95	0.02	131	0.91
S. choleraesuis	117	0.04	100	0.83
L. pneumophila	97	0.13	88	0.01
C. perfringens	NA	NA	49	2.34
S. aureus	135	0.14	94	0.38
P. aeruginosa	113	0.01	106	0.1
E. hirae	71	0.01	63	0.05
L. rhamnosus	180	0.17	151	1.2
B. longum	62	0.44	79	0.33
B. lactis	61	0.04	80	0.11
L. plantarum	88	0.09	41	1.39
L. reuteri	89	0.11	93	0.01
panbacterial	65	0.04	83	0.3

Table 3.3. Recovery efficiencies for experiment #2. A: Chelex DNA extraction. B: Biomerieux DNA extraction. s.d.: standard deviation.

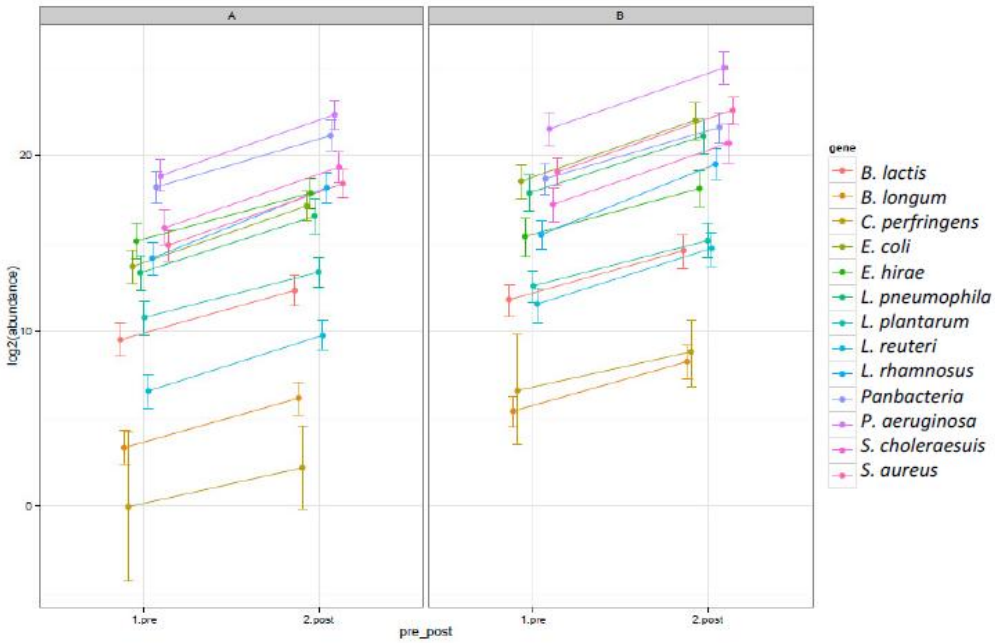


Fig. 3.4. Plot of log_2(counts) for each bacteria tested, comparing pre-concentration samples and post-concentration samples of experiment #2 and two different DNA extraction methods, called A (Chelex®) and B (Biomerieux™).

3.1.4. REPRODUCIBILITY

Inter-assay reproducibility was estimated across the experiments of artificially contaminated samples. There were no significant differences across all the experiments considering recovery efficiencies (ANOVA, Tukey *post hoc test*: p>0.05), demonstrating the reproducibility of the procedure.

3.1.5. RECOVERY IN ENVIRONMENTAL SAMPLES

To verify the feasibility of the method even in real conditions, environmental samples (drinking water samples) were artificially contaminated with the mix described in table 2.2. Our results showed that the recovery efficiencies were not impaired

(ANOVA, Tukey *post hoc test*: p>0.05) in case of environmental samples that can be characterized by the presence of inhibitors of amplification (Table 3.4, Fig. 3.5).

Target	Recovery efficiency (%)	sd
L. pneumophila	98	1.57
C. perfringens	107	1.64
L. rhamnosus	82	1.55
L. plantarum	123	9.65
panbacterial	77	1.56

Table 3.4. Recovery efficiencies for artificially contaminated drinking water samples.

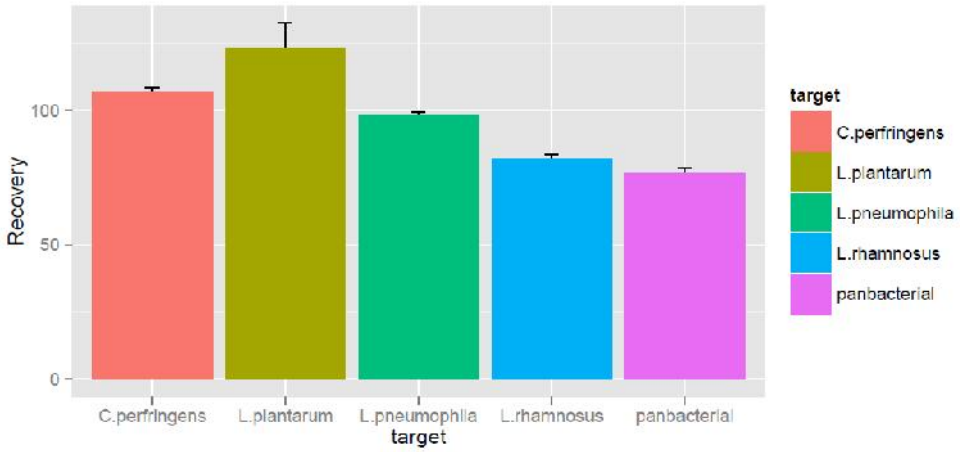


Fig. 3.5. Recovery efficiencies of artificially contaminated drinking water samples.

3.1.6. RECOVERY IN CONTAMINATED ENVIRONMENTAL SAMPLES EXPOSED TO "DNA DEGRADATION" CONDITIONS

Processing the samples immediately after collection is highly suggested, as depicted by our results. If the samples were left for 15 days at room temperature, the degradation of DNA impaired the recovery efficiency for all the bacteria tested. In the case of the Gram negative bacteria *L. pneumophila* the detection was even not possible (Table 3.5, Fig. 3.6).

Target	Recovery efficiency (%)	sd
L. pneumophila	NA	NA
C. perfringens	13	4.1
L. rhamnosus	15	3.1
L. plantarum	13	4.7
panbacterial	16	2.8

Table 3.5. Recovery efficiencies for artificially contaminated drinking water, processed after 15 days of RT incubation.

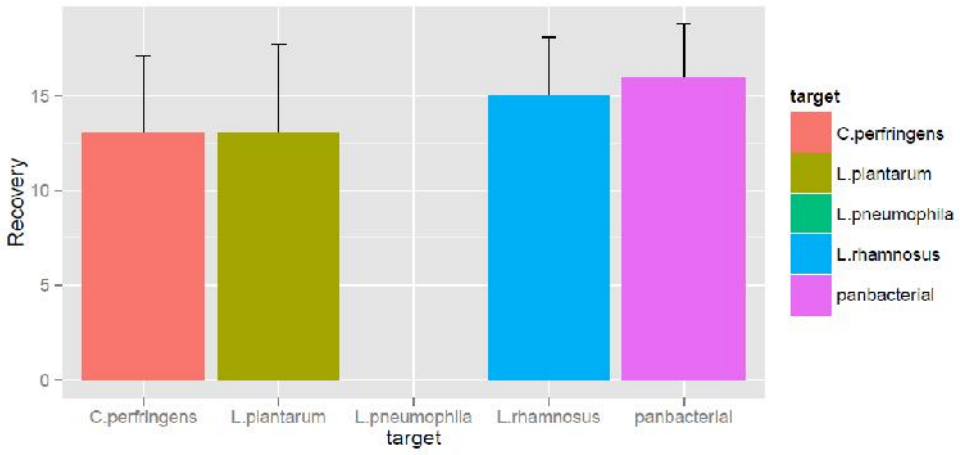


Fig. 3.6. Recovery efficiencies of artificially contaminated drinking water samples, processed after 15 days of RT incubation.

3.1.7. RECOVERY IN DRINKING WATER SAMPLES

Three samples of drinking water (seven litres each) from a water treatment plant in Milan were additionally tested, to definitely verify the applicability in the case of environmental samples. Two different DNA extraction methods were tested. Considering the samples before tangential flow concentration, it was not possible to detect DNA molecules. After concentration, it was possible to measure environmental bacteria DNA in all the samples. Biomerieux DNA extraction was significantly more efficient than Chelex® DNA extraction (ANOVA, Tukey *post hoc test*: $p<0.05$) (Table 3.6, Fig. 3.7).

Sample name	DNA extraction	target	log_2(count) pre	log_2(count) post	sd
drink03	Biomerieux	panbacterial	NA	14.19	0.73
drink04	Biomerieux	panbacterial	NA	11.02	4.35
drink09	Biomerieux	panbacterial	NA	8.11	3.87
drink03	Chelex	panbacterial	NA	5.8	0.62
drink04	Chelex	panbacterial	NA	6.09	0.45
drink09	Chelex	panbacterial	NA	6.37	0.01

Table 3.6. Counts, expressed as log_2 values, for drinking water samples, pre and post concentration, comparing two different DNA extraction methods.

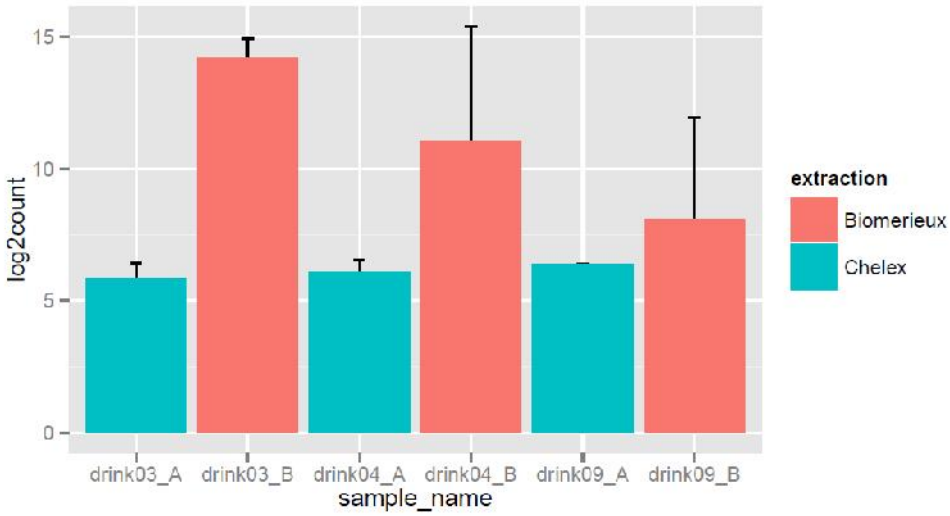


Fig. 3.7. Counts, expressed as log_2 values, for drinking water samples post concentration, comparing two different DNA extraction methods.

3.2. TESTS ON DRINKING WATER SAMPLES

3.2.1. HETEROGENEITY IN BACTERIA COMPOSITION

To verify that the heterogeneity in bacteria composition of drinking water samples was preserved from sampling through all the experimental procedures, we performed a test on water samples collected in the three sampling points in the DWTP. After cloning, ten colonies were isolated from sample 00.GW, nineteen from 00.CF and nineteen from 00.CHL. After amplification of colonies DNA with panbacterial primers and RFLP analysis the panel of restriction fragments obtained varied across samples and across colonies deriving from the same sampling points. This is the first evidence that there is a certain level of heterogeneity in bacterial composition intra and inter-samples (Fig. 3.8).

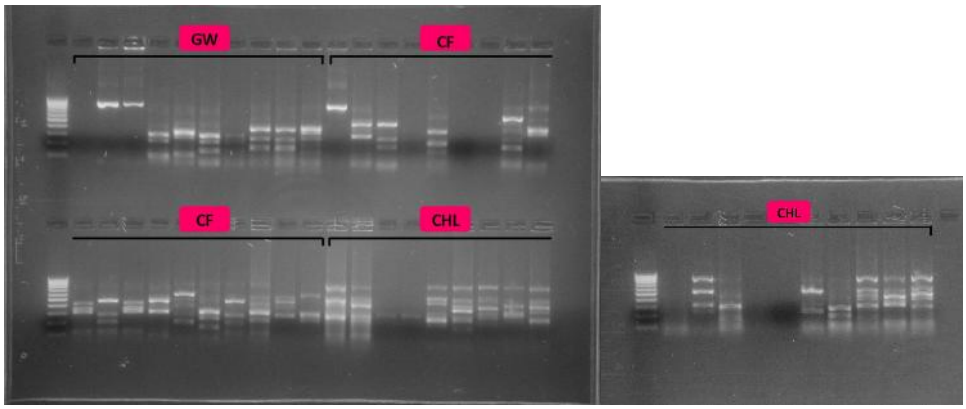


Fig. 3.8. Gel electrophoresis of RFLP analysis on samples 00.GW, 00.CF and 00.CHL. 100 bp ladder in the first well. GW: groundwater, CF: carbon filters, CHL: chlorination basin.

DNA fragments were sequenced and confirmed the high level of heterogeneity composition, revealing the presence of environmental (in particular aquatic and of soil) bacteria, most of them uncultured (see Appendix Table A.1).

3.2.2. DNA EXTRACTION TESTS

We decided to further compare additional DNA extraction kits and protocols on the concentrated water samples 00GW, 00CF, 00CHL, as described in section 2.2.2.

NucliSens® EasyMAGTM system (Biomerieux) was confirmed to be the most efficient DNA extraction kit: all the CT values were lower than the CT values obtained with the other kits, meaning that the product of the amplification was high. Thus starting DNA template was increased in Biomerieux DNA extraction. All the sample processed using this kit were significantly different compared with samples processed with other kits (ANOVA, Tukey *post hoc test*: $p<0.01$). No significant differences ($p>0.05$) were found between samples 00GW_Chelex and 00GW_Invitek; 00GW_PowerSoil and 00GW_PowerWater; 00CF_PowerSoil and 00CF_Chelex; 00CF_Qiagen and 00CF_PowerWater; 00CHL_Invitek and 00CHL_Biomerieux; 00CHL_Invitek and 00CHL_PowerWater; 00CHL_Qiagen and 00CHL_PowerSoil (Fig. 3.9, Table 3.7).

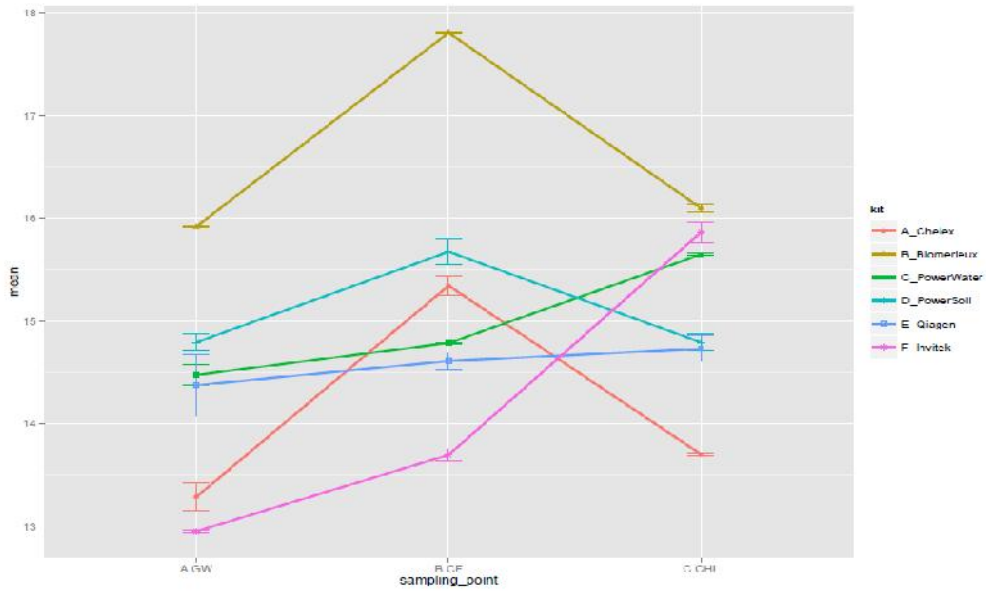


Fig. 3.9. Plot of $\log_2$(counts) for each sample tested, comparing different DNA extraction methods.

Sample name	Sampling point	kit	mean log_2(count)	sd
00GW	GW	A_Chelex	13.285	0.27
00GW	GW	B_Biomerieux	15.920	0.01
00GW	GW	C_PowerWater	14.475	0.20
00GW	GW	D_PowerSoil	14.793	0.16
00GW	GW	E_Qiagen	14.375	0.60
00GW	GW	F_Invitek	12.950	0.02
00CF	CF	A_Chelex	15.345	0.19
00CF	CF	B_Biomerieux	17.810	0.02
00CF	CF	C_PowerWater	14.785	0.01
00CF	CF	D_PowerSoil	15.675	0.2
00CF	CF	E_Qiagen	14.610	0.16
00CF	CF	F_Invitek	13.695	0.12
00CHL	CHL	A_Chelex	13.700	0.02
00CHL	CHL	B_Biomerieux	16.100	0.08
00CHL	CHL	C_PowerWater	15.650	0.01
00CHL	CHL	D_PowerSoil	14.793	0.16
00CHL	CHL	E_Qiagen	14.730	0.25403412
00CHL	CHL	F_Invitek	15.865	0.20207259

Table 3.7. Counts, expressed as log_2 values, for each sample tested, comparing different DNA extraction methods.

3.3. SAMPLING CAMPAIGN

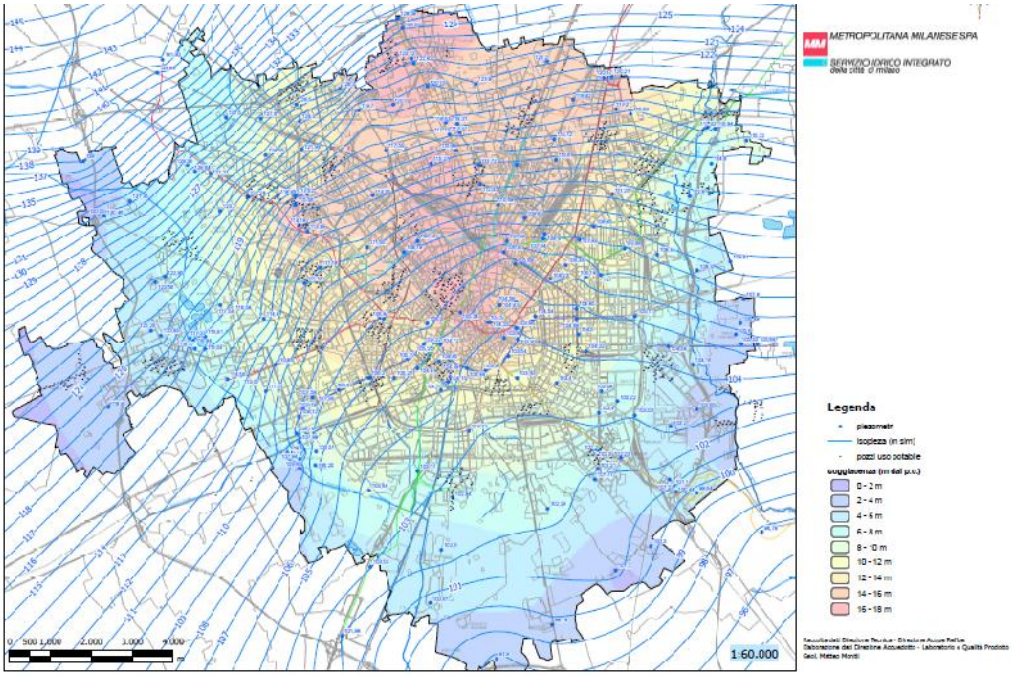


Fig. 3.10. Level of the aquifer in March 2014 and localization of wells.

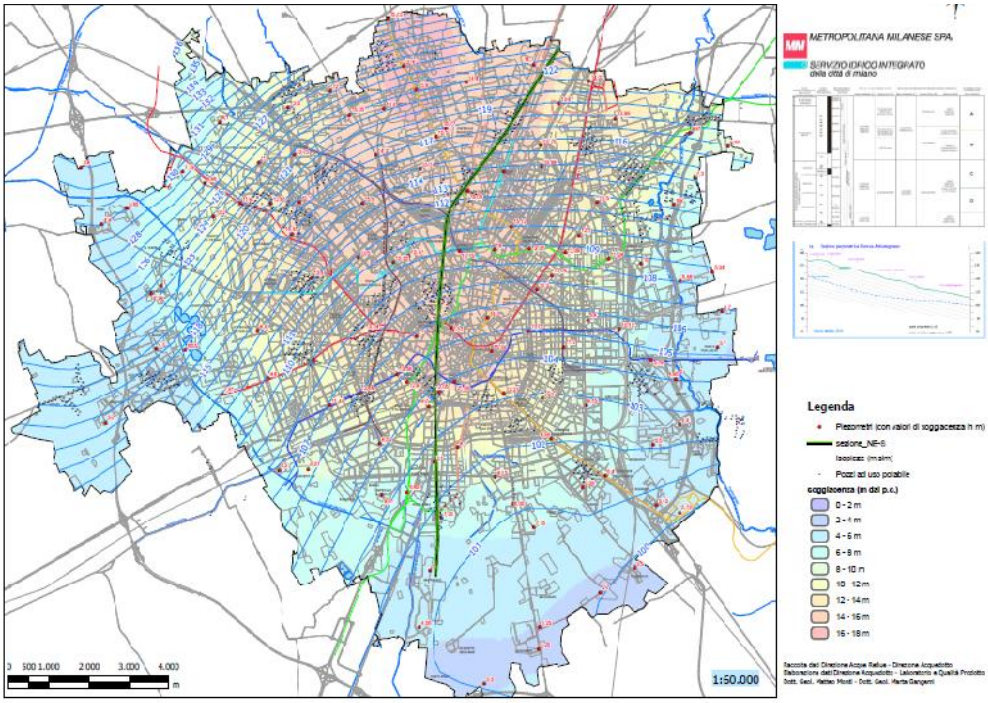


Fig. 3.11. Level of the aquifer in October 2014 and localization of wells.

As already described in the methods, the sampling campaign lasted one year and involved two DWTPs. The main chemical, physical and microbial analyses conducted by Metropolitana Milanese S. p. A. are reported in table A.3 in Appendix. In Fig. 3.10 and Fig. 3.11 aquifer levels in the area of Milan are reported.

Two events occurred during the sampling campaign are noteworthy. In Feltre DWTP (Site 1) granular activated carbon filters were renewed during October 2014, leading to new carbon filters for the sampling of November 2014. In November 2014 the rivers Seveso and Lambro overflowed, due to the high abundance of rainfall. The level of the aquifer increased drastically.

3.4. CULTIVABLE BACTERIA

Randomly, concentrated samples were plated on culture media and colonies growth at 37 and 22 °C were counted, when observed. Only in few cases colonies grew, less frequently in samples belonging to GW. DNA was extract from each colony and, after amplification with 16S rDNA panbacterial primers (27F-1061R), sequenced. All the bacteria identified are typical of aquatic environments or unclassified (the complete list is reported in table A.4 in Appendix).

MM did not detected in CHL sampling point *C. perfringens*, coliforms, enterococci, *E. coli*, *P. aeruginosa*, and pathogenic *Staphylococcus* spp. during all the sampling campaign.

3.5. LIVE/DEAD RATIO

Live/dead ratio of bacteria was quantified for each sample through live/dead staining coupled with microscopy visualization. On average, in GW samples the 70% of bacteria resulted live. A similar result was recorded for bacteria visualized in CF samples. Sampling point CHL significantly differed from sampling point GW and CF (ANOVA, Tukey *post hoc test*: $p<0.05$), showing a decrease in live bacteria percentage (50%). Noteworthy, after chlorination (CHL samples) live/dead ratio is not equal to 0/100 (Fig. 3.12).

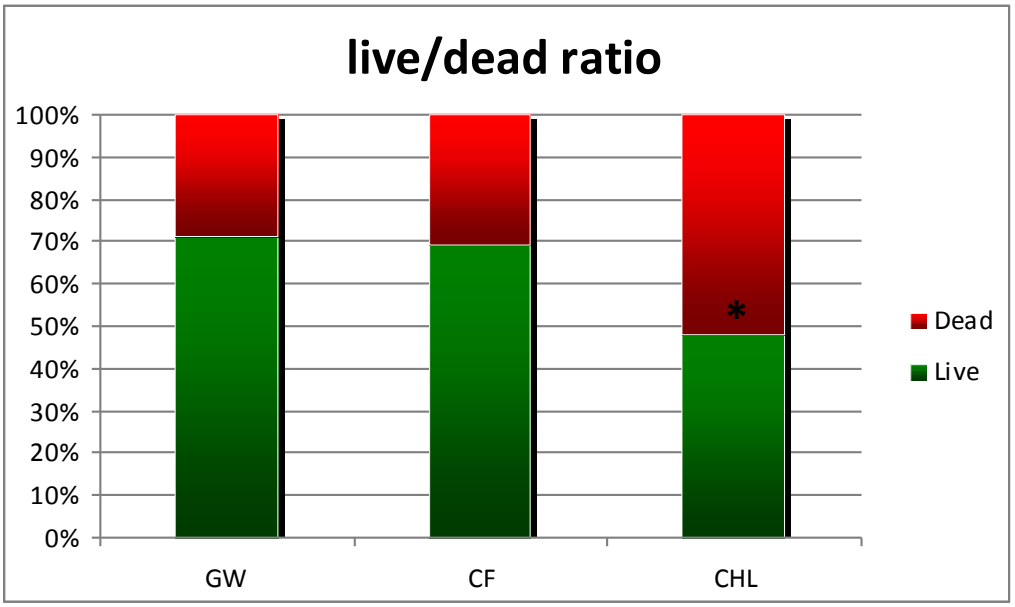


Figure 3.12. Live/dead ratio for sampling points (GW, CF, CHL) of Site 1 (Feltre DWTP). Values are expressed as percentages. CHL differs significantly from GW and CF ($p<0.05$).

3.6. BACTERIA qPCR QUANTIFICATION

Bacteria were quantified for each sample through qPCR amplification of V3-V4 regions of 16 rDNA gene (Fig. 3.13, Table A.5 in Appendix).

On average, it was possible to measure a 3.5 fold increase in bacterial load in sampling point CF in respect of sampling point GW (Fig. 3.14).

Sampling point CF significantly differed from sampling point GW and CHL ($p<0.01$), whereas there was not a significant difference between sampling point GW and CHL. The quantification for each sampling month in Feltre DWTP (Site 1) is illustrated in figure 3.15. Statistical analyses demonstrated the significant difference (ANOVA, Tukey *post hoc test*: $p<0.01$) between the three sampling points, for all the months tested, excepted January, November and December, when CHL samples were not

significantly different from GW samples. In July CHL samples were not significantly different from CF samples and showed a 1.4 fold increase in 16S rDNA gene copies in respect of GW samples. The quantification for each sampling month in Crescenzago DWTP (Site 2) is illustrated in figure 3.16. Statistical analyses demonstrated the significant difference (ANOVA, Tukey *post hoc test*: $p<0.01$) between the three sampling points, for all the months tested. In October 14, Site 2, GW bacteria concentration was higher than CF and CHL bacteria concentration (Fig. 3.16 a). This behaviour was not recorded in November 14, Site 2 (Fig. 3.16 b).

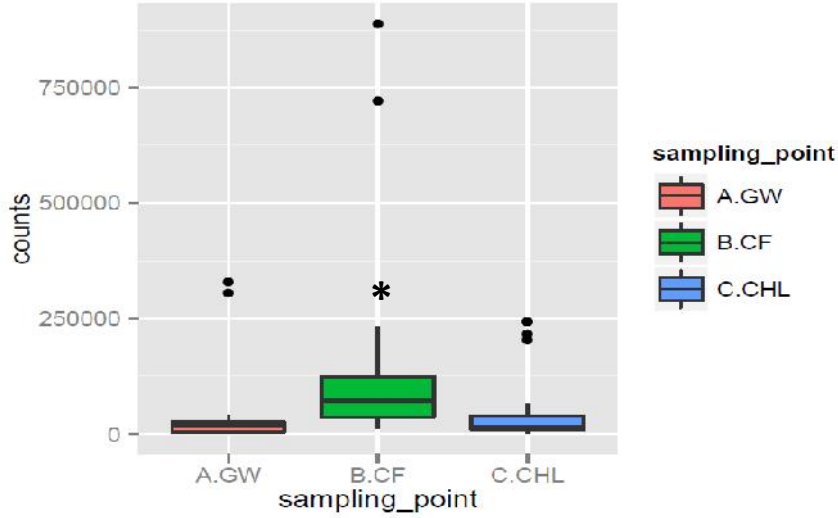


Fig. 3.13. 16S rDNA quantification for sampling points (GW, CF, CHL) of Site 1 (Feltre DWTP). Values are expressed as average of counts.

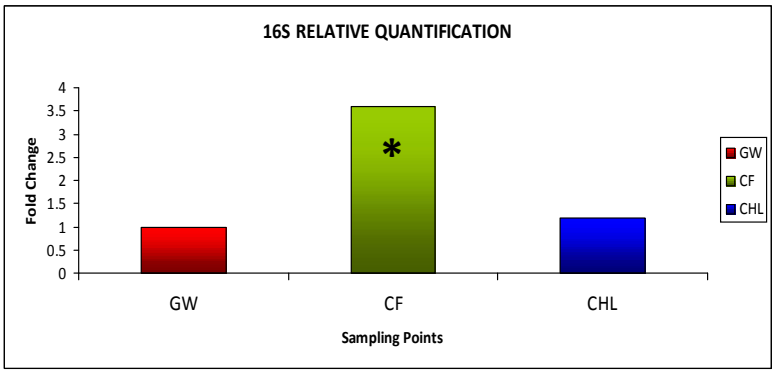
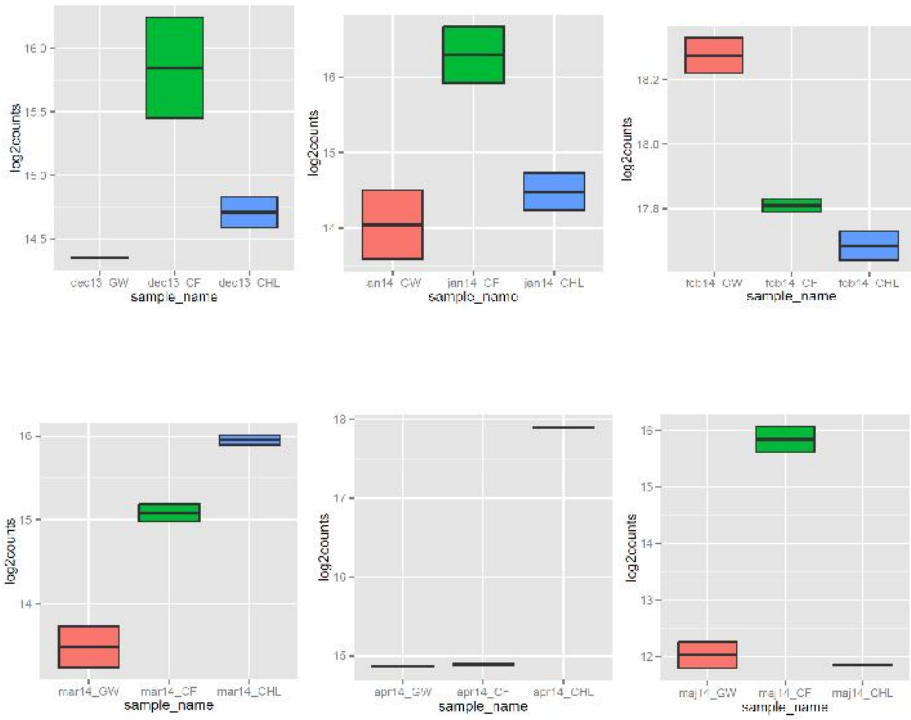


Fig. 3.14. 16S rDNA relative quantification for sampling points (GW, CF, CHL) of site 1 (Feltre DWTP). Values are expressed as average of fold change. In the analysis samples characterized by the carbon filter change are not included.



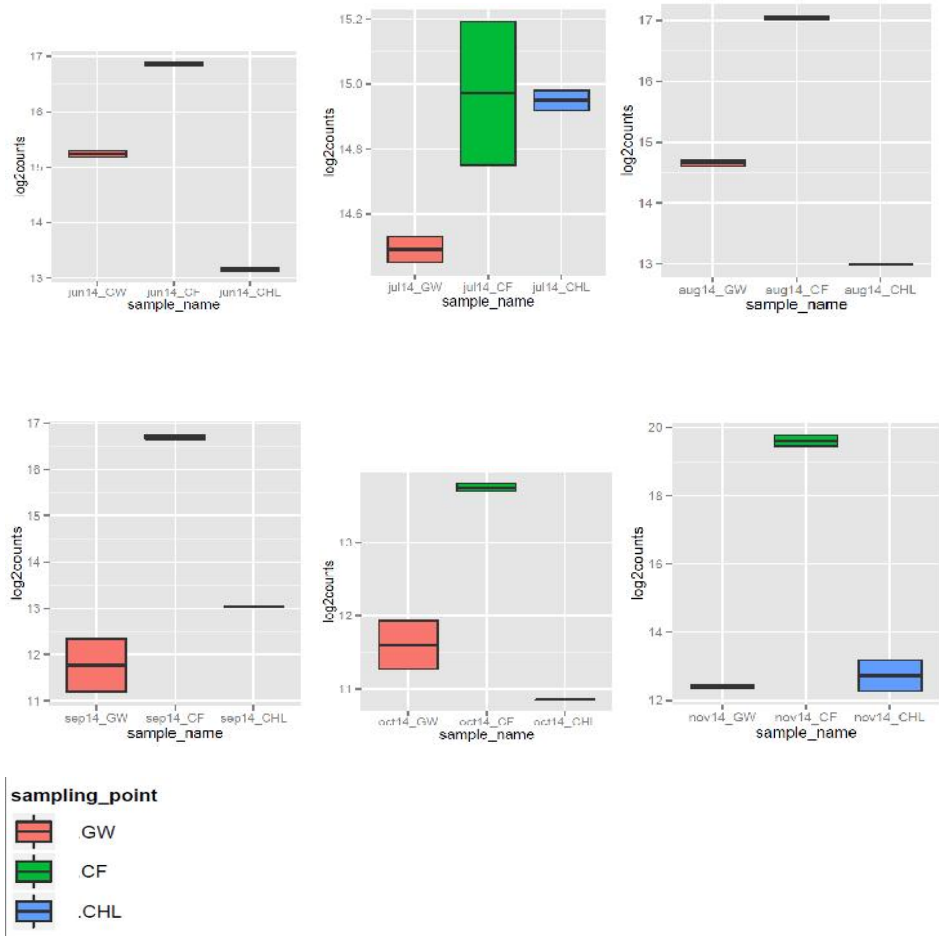


Fig. 3.15. 16S rDNA quantification for sampling points (GW, CF, CHL) of Site 1 (Feltre DWTP) for each month. Values are expressed as $\log_2$(counts).

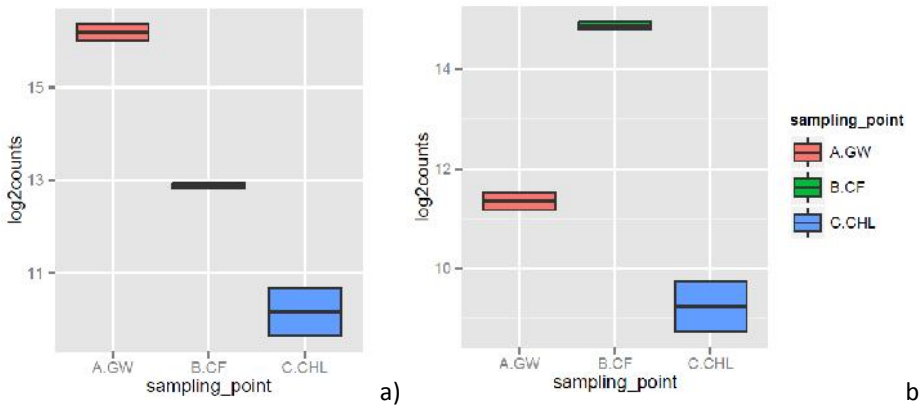
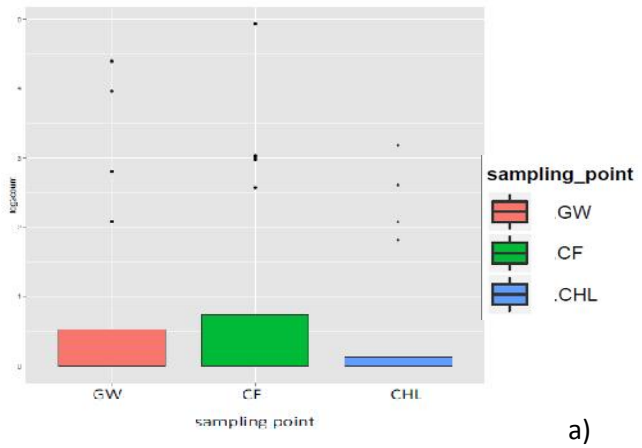


Fig. 3.16. 16S rDNA quantification for sampling points (GW, CF, CHL) of Site 2 (Crescenzago DWTP) in October 14 (a)) and November 14 (b)). Values are expressed as $\log_2$(counts).

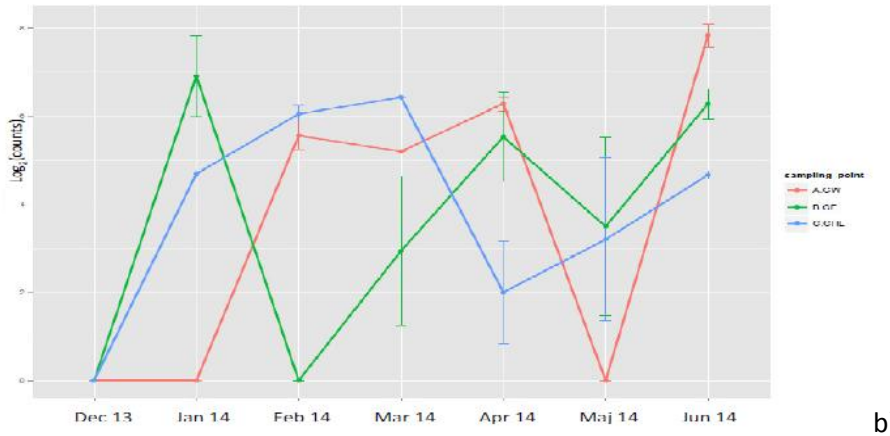
3.7. RESISTANCE GENES

The presence of resistance genes was investigated in a subset of samples, as described in Methods.

It was not possible to correlate the presence of the target **ampicillin** resistance gene with a specific sampling point. December significantly differs ($p<0.05$) from the months January, February, March, April and June, showing low values for all the sampling points. June significantly differs ($p<0.05$) from the months December and May, showing a marked increase in all the sampling points. Anyway there was no evidence of a seasonality in the quantity of target gene detected (ANOVA, Tukey *post hoc* test) (Fig. 3.17).



a)

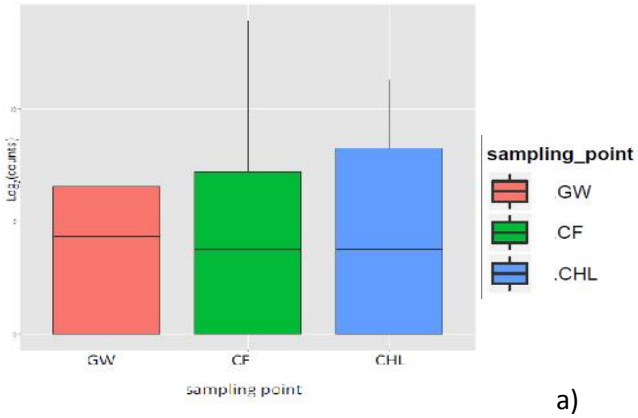


b)

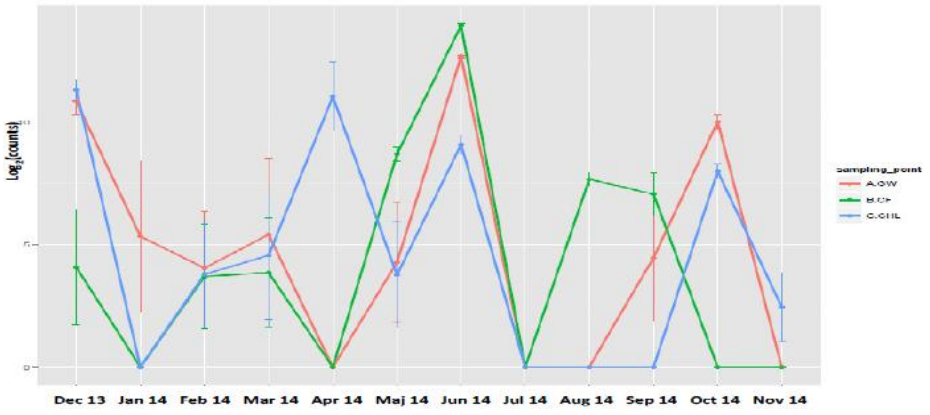
Fig. 3.17. a) **Ampicillin** resistance gene quantification per sampling point, expressed as mean of log_2(counts). Ampicillin resistance gene quantification per month, expressed as log_2(counts).

It was not possible to correlate the presence of the target **β-lactamase** resistance gene with a specific sampling point. June significantly differs ($p<0.05$) from the months February, March, April, May, July, August, September, October and November, showing a marked increase in all the sampling points. July significantly differs ($p<0.05$) from the months January, June and October, showing low values for all the sampling points. December significantly differs ($p<0.05$) from the months

January and November. Anyway there was no evidence of a seasonality in the quantity of target gene detected (ANOVA, Tukey *post hoc* test) (Fig. 3.18).



a)

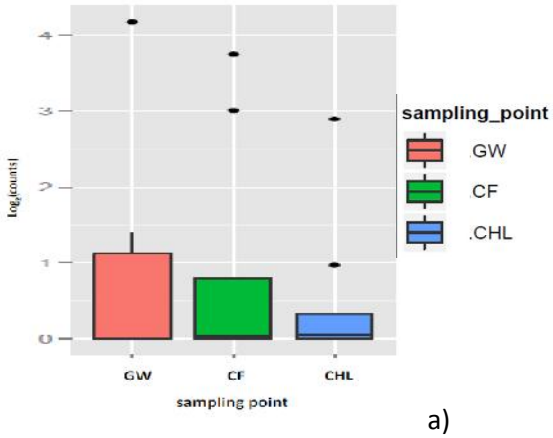


b)

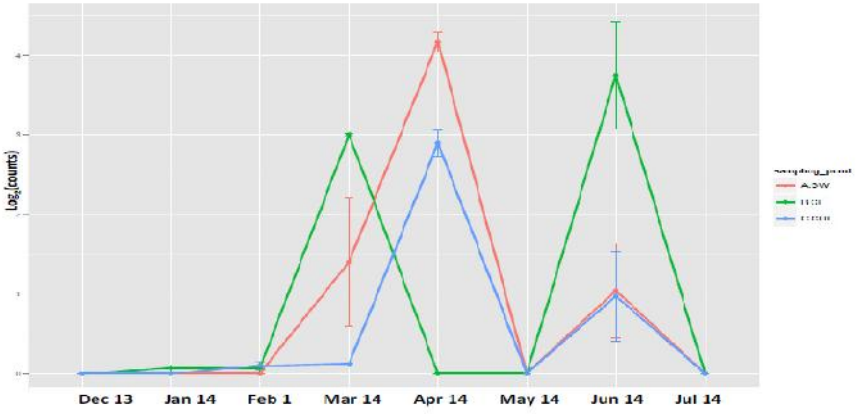
Fig. 3.18. a) **β-lactamase** resistance gene quantification per sampling point, expressed as mean of log_2(counts). β-lactamase resistance gene quantification per month, expressed as log_2(counts).

It was not possible to correlate the presence of the target **Glu-Cys ligase** gene with a specific sampling point. December, January, February and May significantly differ ($p<0.05$) from the months March and June, showing low values for all the sampling

points. July significantly differs ($p<0.05$) from the months March, April and June, showing low values for all the sampling points. June significantly differs ($p<0.05$) from the months December, January, February, May and July, showing an increase in all the sampling points. Anyway there was no evidence of a seasonality in the quantity of target gene detected (ANOVA, Tukey *post hoc* test) (Fig. 3.19).



a)

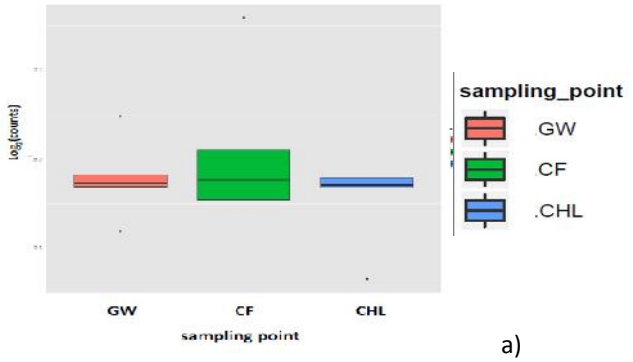


b)

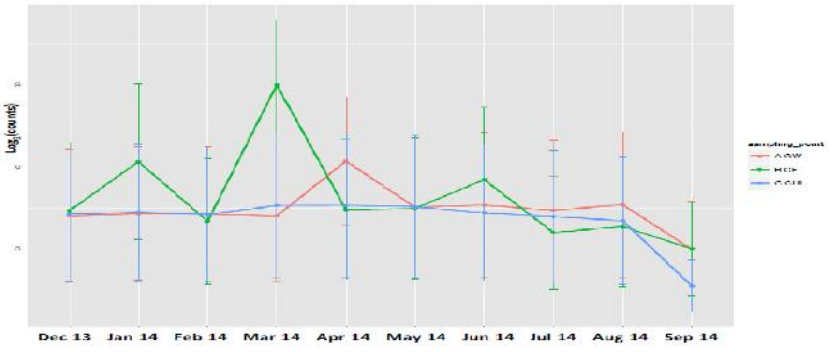
Fig. 3.19. a) **Glu-Cys ligase** gene quantification per sampling point, expressed as mean of log_2(counts). Glu-Cys ligase gene quantification per month, expressed as log_2(counts).

It was not possible to correlate the presence of the target **Glutathione synthase** gene with neither a specific sampling point nor a different distribution during the year (ANOVA, Tukey *post hoc* test) (Fig. 3.20).

Further data are listed in table A.6 in Appendix.



a)



b)

Fig. 3.20. a) **Glutathione synthase** gene quantification per sampling point, expressed as mean of log_2(counts). Glutathione synthase gene quantification per month, expressed as log_2(counts).

3.8. HIGH-THROUGHPUT DNA SEQUENCING

In total about 19 million reads (8474127 + 11287412) were obtained after quality filtering and merging the two Illumina runs. Chimeras filtering and singleton removing (10x) resulted in 9175 OTUs.

First we evaluated the reproducibility of our experimental procedure, testing technical replicates. Bray-Curtis measure of similarity evidenced that replicates are closely related: with a threshold = 0.1, where the *R* statistic value ranges between 0 (complete similarity) and 1 (complete separation), the high similarity between each replica in terms of sequences is evident. This supported the reproducibility of the experimental procedure (Fig. 3.21).

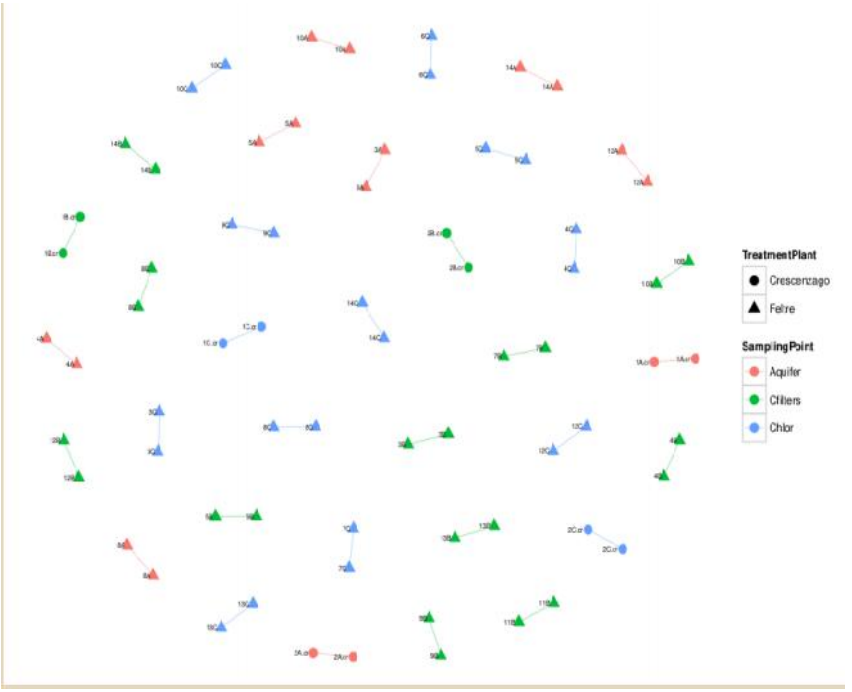


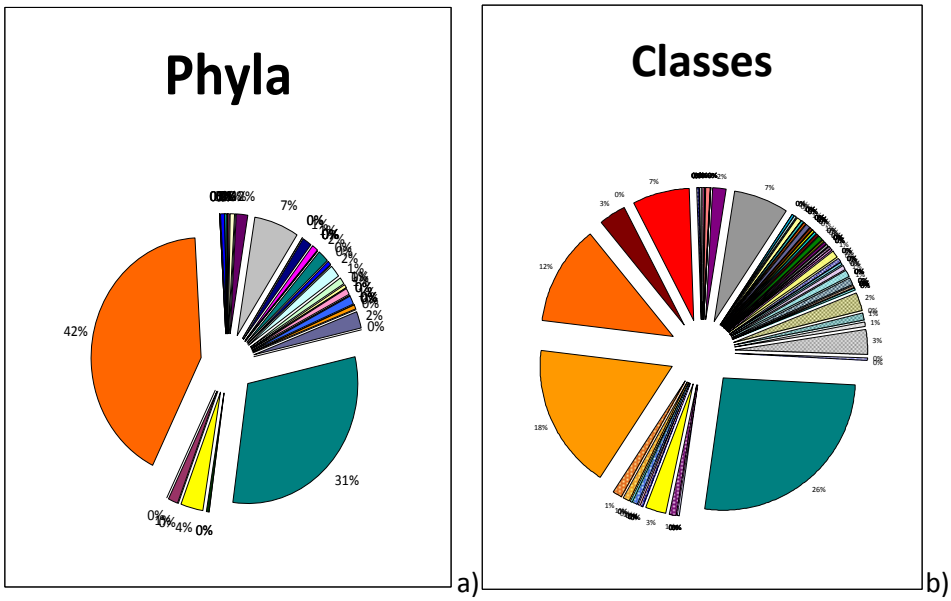
Figure 3.21. Reproducibility. Bray-Curtis network, threshold = 0.1

A total of 44 bacterial phyla and 159 classes were identified (plus 3 Archaea phyla and 10 classes).

Taxonomic analysis revealed that the majority of the sequences in all the samples were associated with the phyla Proteobacteria (42%) and Candidate Division OD1 (31%) with 7% sequences of unclassified bacteria ("Other") (Figure 3.22 a).

Looking inside the taxonomic level of Class, we discovered that Candidate Division OD1-ZB2 were the most abundant (26%), followed by Alpha-Proteobacteria (18%), Beta-Proteobacteria (12%) and Gamma-Proteobacteria (7%) (Figure 3.22 b).

It should be noted that our findings are focused on living organisms or at least intact cells; extracellular DNA released by dead bacteria has been excluded from water samples, due to the filtration procedure used in our experiments.



a) b)

Fig. 3.22. a) Pie chart of phyla detected in water samples. In orange Proteobacteria, in light blue Candidate Division OD1, in grey unclassified bacteria. b) Pie chart of classes detected in water samples. In light orange Alpha-Proteobacteria, in orange Beta-Proteobacteria, in red Gamma-Proteobacteria, in brown Delta-Proteobacteria in light blue Candidate Division OD1, in grey unclassified bacteria.

In Fig 3.23, for each sample relative abundance in phyla (>0.1%) is illustrated. As clearly appears, there is a pattern of distribution of phyla, depending on sampling points.

This peculiarity was evident investigating the taxonomic rank of Class too (3.24).

GW samples are characterized by a high relative abundance of Proteobacteria and unclassified bacteria ("Other"). CF samples showed a switch in composition, with the predominance of Candidate Division OD1. This behaviour was observed even in CHL samples.

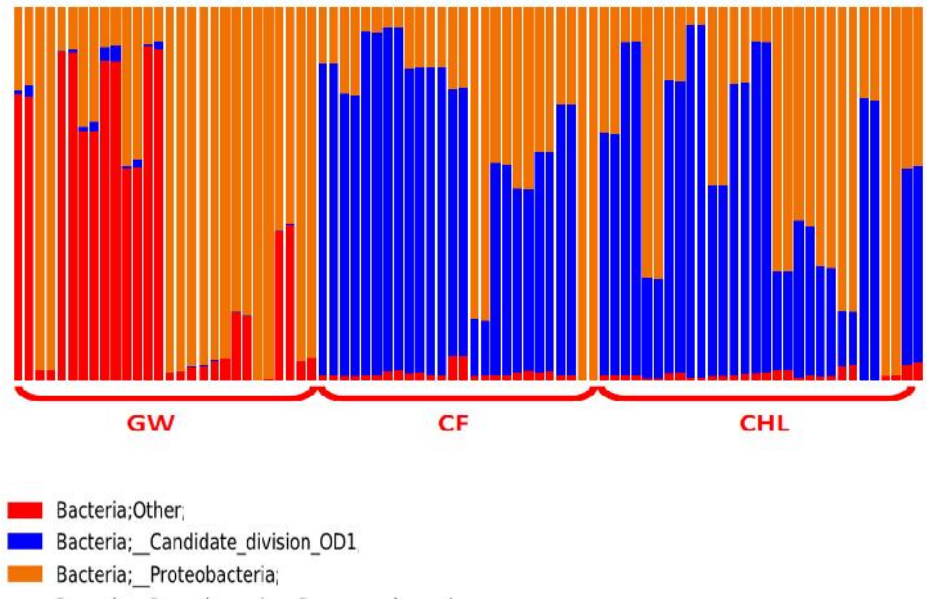


Fig. 3.23. QIIME Bar plot. Taxonomic rank: Phylum. Only the most abundant (>0.1%) are represented. Bar plot describes distribution of bacteria recovered in the different sampling points.

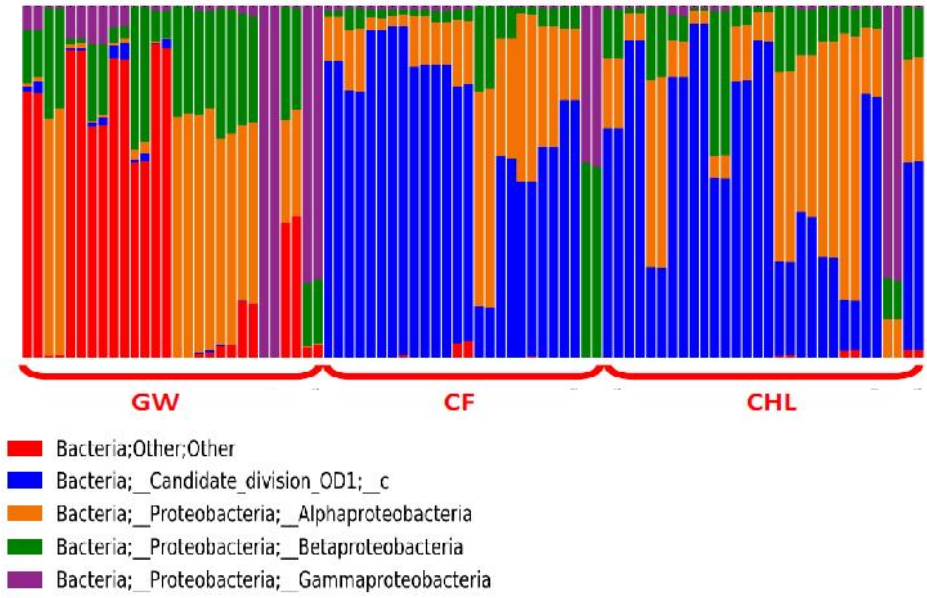


Fig. 3.24. QIIME Bar plot. Taxonomic rank: Class. Only the most abundant (> 0.1%) are represented. Bar plot describes distribution of bacteria recovered in the different sampling points.

When comparing samples derived from different DWTPs, we noticed that there were no great differences in CF and CHL. Nevertheless, the parallel between GW samples from different DWTPs was not possible: in Feltre DWTP (Site 1) the bacteria community is composed by unclassified bacteria, Alpha-Proteobacteria and Beta-Proteobacteria. Only a small proportion is constituted by Gamma-Proteobacteria. By contrast, in Crescenzago DWTP (Site 2) the bacteria community is completely dominated by Gamma-Proteobacteria (Fig. 3.25).

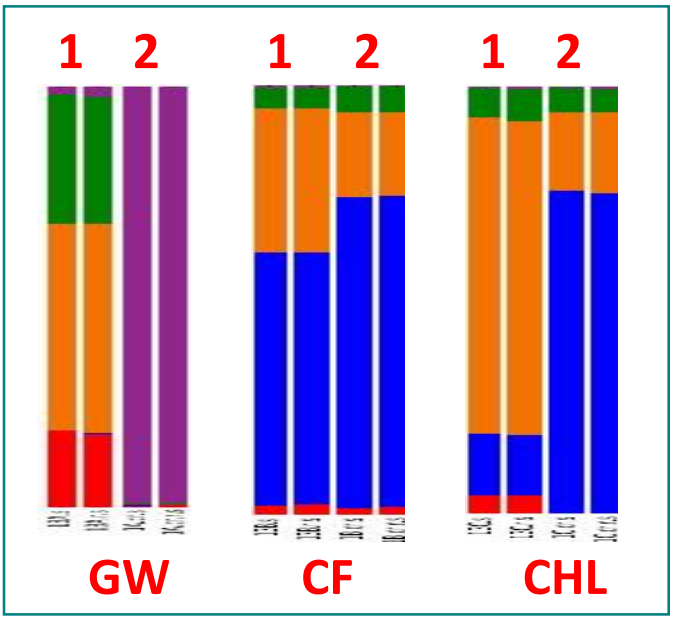


Fig. 3.25. Comparison between two DWTPs. QIIME Barplot. Tax rank: Class. Only the most abundant (> 0.1%) are represented. Barplot describes distribution of bacteria recovered in the different sampling points. 1: Site 1 (Feltre DWTP); 2: Site 2 (Crescenzago DWTP).

In November 14 new (i.e. still not colonized by bacteria) carbon filters were present in Feltre DWTP (Site 1). If no great differences between GW samples collected in October 14 and November 14 were recorded, in CF samples a change in bacteria composition after carbon filters renewing was evident. In particular in CF samples of November 14 the relative abundance of Gamma-Proteobacteria increased and the relative abundance of Candidate Division OD1 decreased, shifting the composition to the typical groundwater bacterial composition. An identical behaviour was measured for CHL samples collected in November 14 (Fig. 3.26).

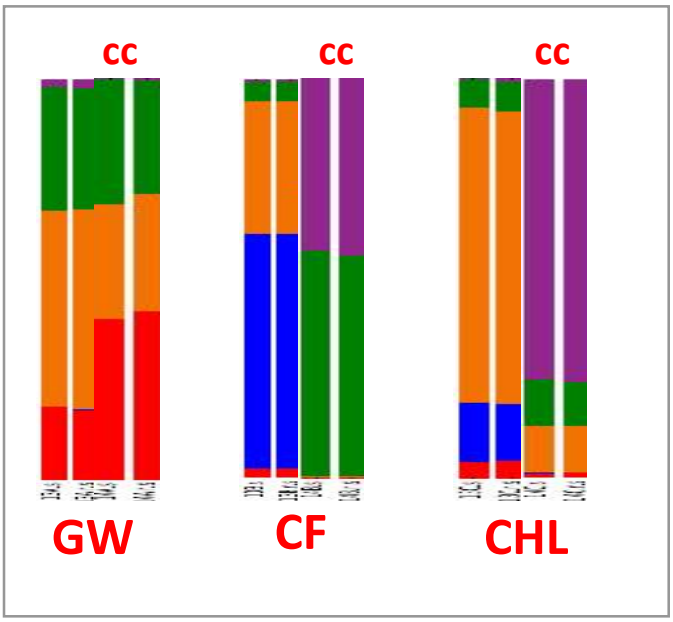


Fig. 3.26. Comparison between old and new carbon filters. QIIME Bar plot. Tax rank: Class. Only the most abundant (> 0.1%) are represented. Bar plot describes distribution of bacteria recovered in the different sampling points. cc: water samples collected in November 14, when carbon filters were renewed.

Statistical analyses were performed to verify these hypotheses. Bray-Curtis metric was used to uncover β-Diversity among sampling points. Statistical significance of differences among samples using the Bray–Curtis measure of similarity (Vegan package within R), ranges between 0 (complete similarity) and 1 (complete separation). As shown in Fig. 3.27, CF and CHL samples clustered together, when the threshold is set to 0.9. An exception was represented by CF and CHL samples collected in November 14 in Site 1 that clustered together, but separately by the other CF and CHL samples. Regarding GW samples, that collected in Site 2 clustered separately in respect of GW samples of Site 1. Comparable results were obtained using a threshold of 0.8 (data not shown).

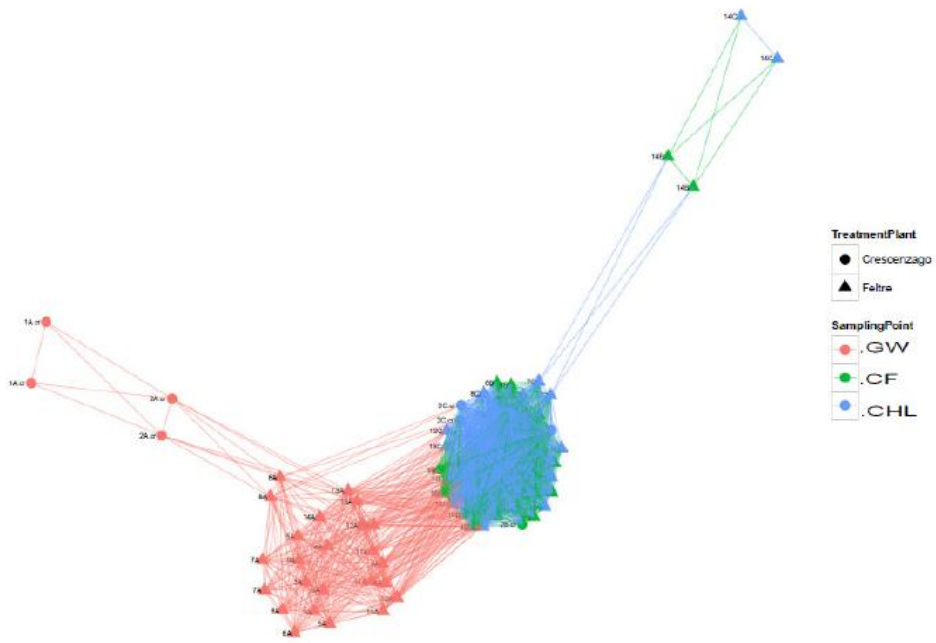


Fig. 3.27. β-Diversity. Bray-Curtis network, threshold = 0.9. Crescenzago: Site 2; Feltre: Site 1.

Non metric multidimensional scaling (nMDS) was performed with the Weighted Unifrac distance matrix using the R statistical package to ordinate the OTU data (samples with similar community structure cluster together, taking into account the relative abundance of each OTU and the taxonomy information).

The resulting nMDS plots (Fig. 3.28, Fig. A.2 in Appendix) highlighted marked bacterial community differences, revealing a noticeable variability of the bacterial communities in the different sampling points. GW is dominated by Candidate Division OP3 and Nitrospirae, whereas Candidate Division OD1 and Proteobacteria are widely distributed.

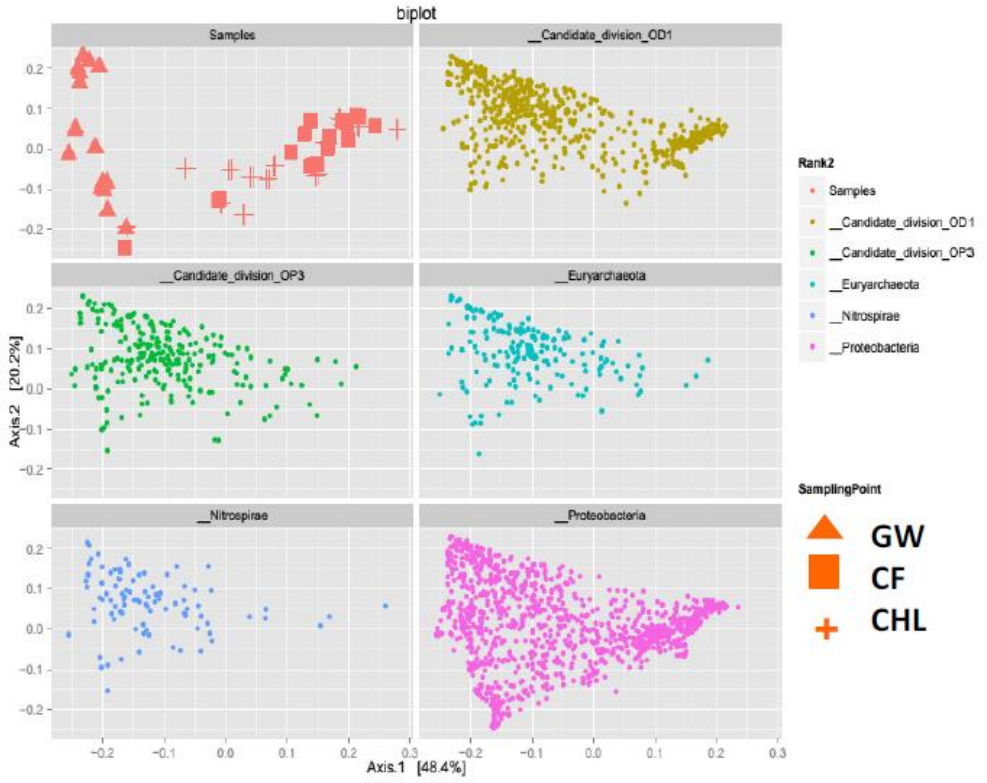


Fig 3.28. Phylum distribution. NMDS (Non Metric Multi Dimensional Scaling) plot.

Weighted UniFrac-based PCoA plots (Fig 3.29) revealed a strong pattern of clustering of community structure by sampling point. Samples from GW clustered together, and within- sampling point UniFrac distances were generally smaller than between-sampling point distances, suggesting the community composition of samples from the same sampling point were more similar to each other. Moreover samples belonging to CF and CHL samples clustered together and separately from GW samples. However CF and CHL samples collected in November 14 (renewed carbon filters) plotted far distant from the other CF and CHL samples and adjacent to GW samples, confirming the evidences collected in the previous analyses.

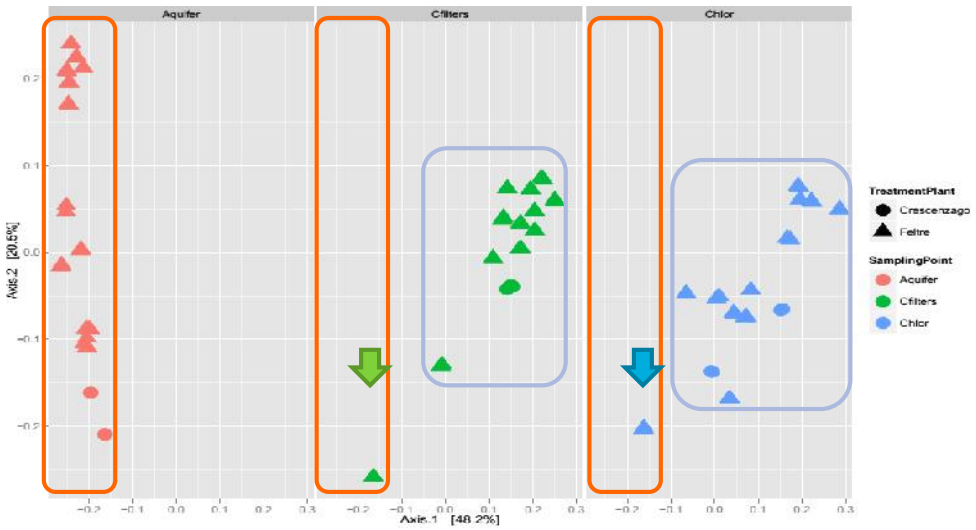


Fig. 3.29. PCoA Weighted Unifrac. p<0.001 for Sampling Point. p<0.001 for Treatment Plant. In light red GW, in green CF and in blue CHL samples are represented.
With arrows are indicated samples characterized by renewed carbon filters (same sampling date, sampling points CF and CHL).

The comparison between untreated water (GW samples) and treated water (CF and CHL samples) at the phylum level is illustrated in Fig. 3.30. Negative binomial difference test (DESeq2, R package) was performed to highlight the significant different OTUs between the two groups. Lentispherae, Candidate Division WS3, Euryarchaeota, Gemmatimonadetes, Candidate Division OP3, Elusimicrobia, OC31, Thaumarchaeota, Nitrospirae are exclusive of GW samples, whereas a high number of different OTUs belonging to Candidate Division OD1 plotted within treated water samples (CF and CHL samples). This evidences showed as untreated water (GW samples) had a higher heterogeneity in bacterial composition compared to treated water (CF and CHL samples) and Candidate Division OD1 OTUs differentiated treated water.

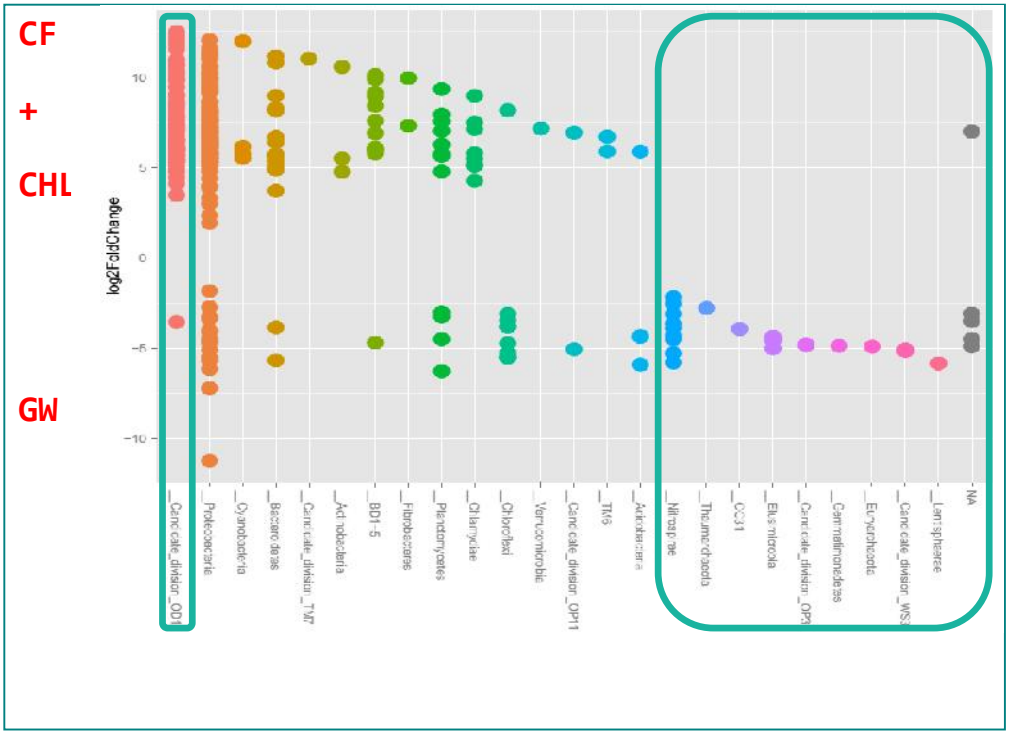


Fig. 3.30. Comparison between treated and untreated water. Significant different OTUs are plotted. Negative binomial difference test (DESeq2), R package.

3.9. NANOBACTERIA ANALYSIS

A total of 3,996,876 reads passed QC step were assigned to nanobacteria phyla. Nanobacteria diversity in our results is a large fraction of the total microbial diversity recovered (36.4%).

Specifically, we found taxa belonging to the nanobacteria candidate phyla/radiations OD1, OP11, TM6, TM7 (as reported as nanobacteria in Luef et al., 2015 and Brown et al., 2015) and to the related OP3, OP1, BRC1 and WS3 phyla/radiations, for a total of 1123 OTUs. The OD1 phylum was even the most represented bacteria group in the DWTP (31% of the entire bacterial community, Fig. 3.22 and Table A.7 in Appendix for

the complete list), whereas all the other nanobacteria and related phyla contributed for the 4.2% of the total.

In order to analyze in depth the microbial diversity at the three potabilization steps, we used phylogenetic entropy as described in Sandionigi et al., 2014. A tree of identified nanobacteria was generated with RAxML (Figure 3.31) and used in an entropy-based approach to estimate the total lineage diversity. Samples and groups (GW, CF, CHL) differentiation was measured with the *phylogenetic turnover*, defined as percentage of the unshared observed lineages. In our data, the nanobacteria community varies among GW, CF and CHL in DWTP, where groundwater is characterized by the highest nanobacteria heterogeneity (alpha diversity values: GW=2.22, CF= 1.54, CHL= 1.57). GW samples share a similar composition during the whole survey and even considering different sampling sites within the city (Fig. 3.32) (*phylogenetic turnover mean* across samples of the same group: 5%).
Water samples deriving from CF and CHL share the same nanobacteria diversity (beta diversity expressed as *phylogenetic turnover*, CF-CHL: 0.57%) that is significantly different from GW (CF-GW: 14% and CHL-GW: 12.6%, $p<0.001$).

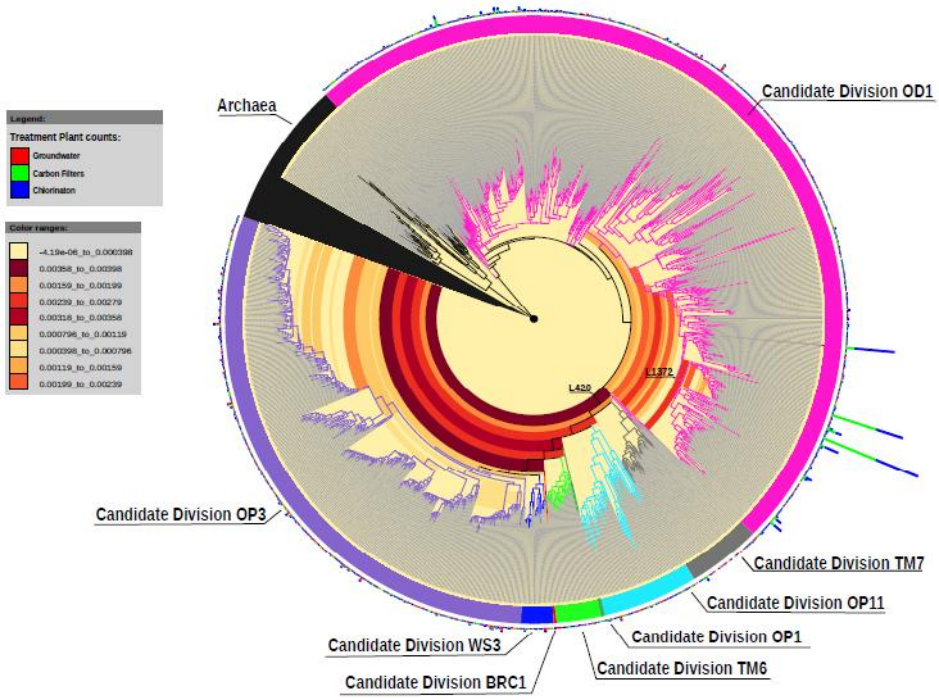


Fig. 3.31. Hairy pacman graphical output from PhyloH analysis. The output couples the phylogenetic information from the RAxML tree of the nanobacteria and the calculated contribution of the mutual information from the lineage involved (expressed as gradient of colours, where yellow is the null contribution and dark red the maximum one). Multiple bars represent the proportion of counts associated to each lineage in respect of the three different sampling points.

The specific lineage L1372 of candidate division OD1 (Fig. 3.31) characterizes treated waters (4%, 68% and 58% of nanobacteria sequences of GW, CF and CHL samples respectively). In carbon filters, the proportion of lineage L1372 increases dramatically reaching about 22% of the total bacterial sequences. On the contrary, lineage L420, that includes all non-OD1 candidate nanobacteria phyla observed, is typical of GW samples (58%, 7% and 10% of GW, CF and CHL samples respectively). The OD1 members not belonging to lineage L1372 are spread across the compartments with low percentage. L1372 and L420 lineages explain 6.4% of the total turnover across the three compartments.

Groundwater (GW) samples are separated by PC1 from Carbon filters (CF) and Chlorination (CHL) samples. Samples belonging to CF and CHL characterised by new (i. e. sterile) carbon filters (CC) are more similar to GW than to CF and CHL samples.

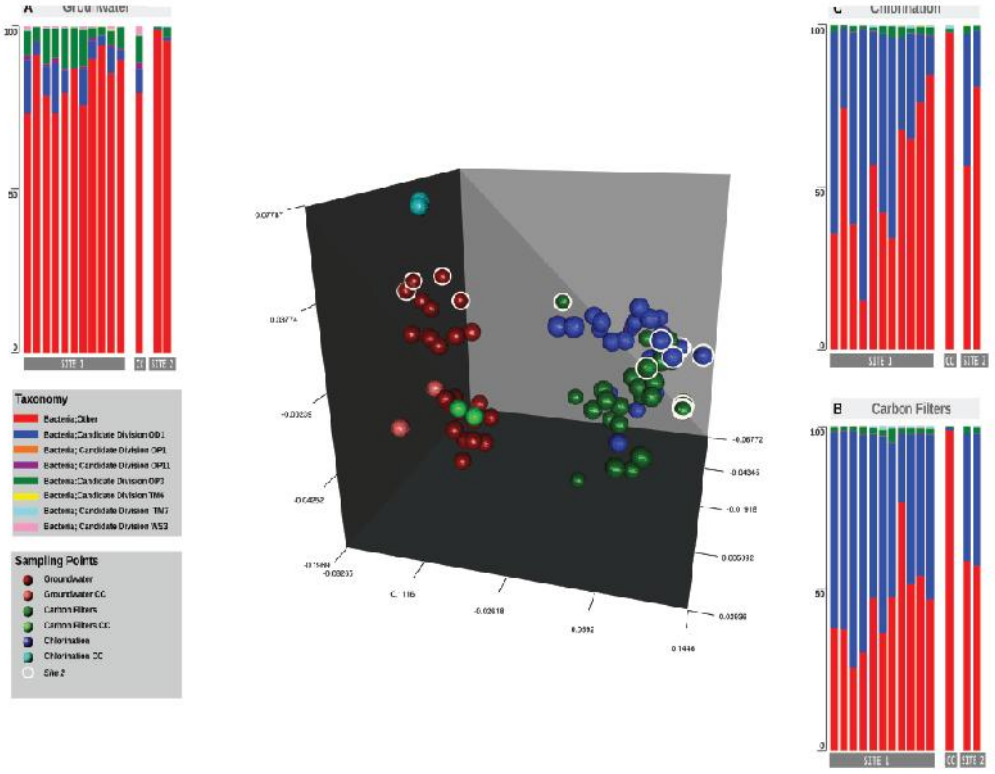


Fig. 3.32. PCoA using phylogenetic turnover as distance metric. Samples deriving from a different DWTP (Site 2) are circled white. A-B-C barplot describes phyla distribution of nanobacteria recovered in the different sampling points.

Samples of November 2014 are different from others in CF and CHL compartments, due to change in carbon filters in the DWTP. Diversity found in the water samples at the CHL step, after flowing through new filters, increased: the mean phylogenetic turnover varied from 5% to 16.5% comparing samples deriving from CHL basin before and after filters replacement.

Notice that excluding Nov-2014, variation across compartment CF and CHL and variation within the three compartments across time are similar and comparable to variation across replicas, indicating a seeding effect of CF on CHL and lack of strong temporal trend. This seeding effect is present in the last sample, given that across compartment difference are smaller than across time in CHL (gray circle), although, in this time, CHL is more different from CF (Fig. 3.33).

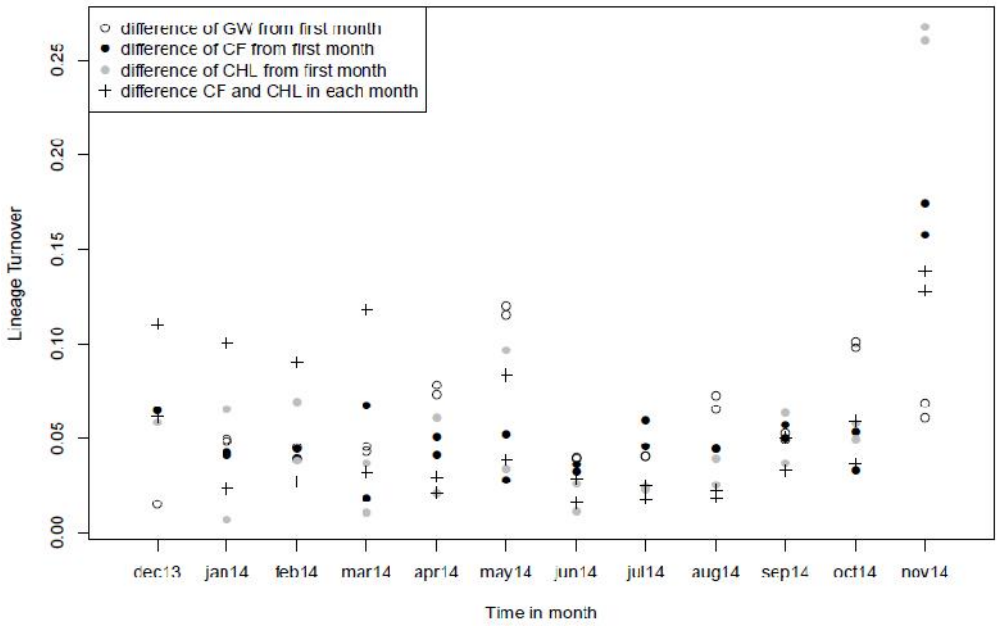


Fig. 3.33. Change over time and compartment within a DWTP. The 3 series of circle show the change respect to the first sample in December 2013 of their respective technical replica and the other sampling times. White, black and gray circles indicate GW, CF, and CHL compartments, respectively. The cross indicates comparison for each replica and each time between compartments CF and CHL.

4. DISCUSSION

Understanding the microbial ecology of drinking water treatment plants is necessary to design innovative and effective control strategies that will ensure safe and high-quality drinking waters. Interactions between bacteria are unaccounted for in current disinfection models.

Drinking water emerging from the tap may contain bacteria (Pinto et al., 2012), archaea (van der Wielen et al., 2009), eukaryotes (Thomas et al., 2010, Pereira et al., 2013), and viruses (Gomez-Alvarez et al., 2012, Lambertini et al., 2012), which together constitute a complex microbial community. Estimations indicate up to millions of microbial cells per litre (Hammes et al., 2008). This is not the case of Milan DWTPs, characterized by a very low abundance in microorganisms. Nevertheless our results highlighted an incredible heterogeneity in microbial composition, depicting groundwater in particular as a high biodiverse habitat.

From a microbiological perspective, the main objectives of drinking water treatments are to ensure the absence of any pathogenic bacteria in drinking water and to limit any uncontrolled regrowth during distribution of the water. Moreover the integrity of the water infrastructure (e.g., corrosion induced by bacteria [Li et al., 2014]), and the aesthetic quality (i.e. colour, taste) of water (Li et al., 2013) are taken into account.

In order to minimize detrimental effects caused by microbes, in DWTPs multiple hygienic barriers are employed, from ozonation, to UV disinfection, from inverse osmosis to chlorination. Nevertheless it is remarkable that the drinking water microbiome can persist under extreme conditions of chronic stress and very low substrate concentrations. We can consider carbon filters and chlorination as sources of chronic stress to water microbial community.

Drinking water microbial community can migrate from the DWTP through the distribution system, to the consumers. Nevertheless our understanding of the microbial ecology of drinking water distribution systems is limited for practical and theoretical reasons: DWTPS are not easily accessible and analyses of specific sections of them is still challenging. Moreover, DWTP environment is still considered too hard for life when compared with other aquatic ecosystems, where microbial entities more easily growth and can be analyzed.

Nowadays there is an increasing awareness regarding recent advances in molecular-based methods to deeply study drinking water ecosystem, since they are an interesting source of unclassified microorganisms and unraveled microbial interactions.

However, manipulating the drinking water microbiome to benefit consumers necessitates the ability to confidently predict its dynamics within DWTPs.

4.1. DEVELOPMENT STANDARDIZATION OF NEW AND IMPROVED MOLECULAR METHODS

Microorganisms detection systems are still strongly based on traditional culture methods. These approaches are reliable, but, at the same time, time-consuming and do not allow to detect uncultivable bacteria (Simon and Daniel, 2011). On the other hand, molecular methods based on microbial DNA detection have been considered unsuitable for assessing water quality, mainly because they do not discriminate between extracellular (dead cells) and intracellular (living cells) DNA. The quest for methods allowing for the rapid, sensitive, selective quantitative detection of living organisms at a reasonable cost is thus considered of great importance in all microbiological analyses.

Given the low densities of target microorganisms in drinking water, concentration of large volumes of water are required before microbial detection. Microbial concentration can be accomplished by several methods including filtration, such as tangential flow filtration and immunocapture; centrifugation, by gradient density separation or continuous flow centrifuge; and flow-cytometry. Tangential flow filtration (TFF) has emerged as a promising technique for the recovery of diverse microbes in water samples. Through parallel fluid flow tangent to the filter surface, there is a reduction of filter clogging, avoiding losses and speeding the concentration procedure. Thus, the concentration of larger volumes is facilitated, leading the way for the preparation of samples for the detection of even viruses.

In particular, tangential flow filtration followed by magnetic beads capture methods, such as Biomerieux automated system, have proved particularly efficient in the recovery of microorganisms. In addition, magnetic capture-beads based methods can be automated and are easy to perform in relatively short times. Most concentration techniques may, however, also lead to the concentration of substances that may hinder nucleic acid extraction, purification and amplification (Jofre and Blanch, 2010). Efficient nucleic acid extraction is therefore also important in order to: i) maximize microbial disruption to make nucleic available for amplification considering the very low number of microorganisms present in drinking water, ii) eliminate potential PCR inhibitors which have been found to be present in drinking water samples and iii) obtain as close as possible the maximum yield of nucleic extraction for a sensitive and accurate quantitative detection of target microorganisms and/or target genes present in water samples.

All the experimental workflow has been tested and validated using artificial contaminated water: a mock community composed by Gram positive and Gram negative bacteria, live or dead, was used to seed sterile water. The filtration process

showed efficiency in the recovery and did not affect microorganisms viability, as demonstrated through qPCR combined with microscopy visualization.

It seems unlikely that overestimation of recovery efficiencies is due to filtration process. More likely the issue resides in the starting DNA quantity: only through the dilution of the mock samples and of the concentrated samples it is possible to obtain reasonable recoveries.

This issue can be crucial when estimating DNA relative quantities of samples with no comparable nucleic acid starting material.

The protocol was optimized for reliable and reproducible recovery from post-filtration.

On samples 00GW, 00CF, 00CHL, several DNA extraction methods were tested and evaluated in order to both reduce risk of contamination and obtain a good quality DNA. The compared DNA extraction methods were very different from each other and use different principles: chemical, enzymatic and/or mechanical lysis as well as an automatic beads-based extraction system. Differences were measured also in performances. The comparison among the extraction methods was difficult since the impossibility to compare the three samples chosen for the analysis. Impaired DNA extraction when the bacterial composition differs can be the explanation. Two different protocols were finally performed: a one-step lysis DNA extraction (Chelex®, Bio-Rad) and the automatic DNA extraction system NucliSENS® easyMAG (BioMerieux). In the case of Chelex®, the great advantages were the rapidity and the reduced risk of introducing contamination tanks to the one-step protocol. On the other hand, NucliSENS® easyMAG, with its automated system, has a high reproducibility and a reduced risk of contamination and, thanks to magnetic beads method, increases significantly the yield of DNA obtained. Good quality DNA was obtained even if, as expected, the yield was not high. A further confirmation of the

effectiveness of the method was obtained with drinking water samples collected at the DWTPs.

An optimization of the protocol of libraries preparation for Illumina sequencing was required, since DNA quantity was near the inferior limit of detection and difficult to quantify. After library preparation, library quantification confirmed the good quality of the DNA obtained and its sufficient quantity. Library measurements strongly replicated qPCR measurements performed with the same primer pairs (described in section 3.6 and discussed later) (Fig. A.1 in Appendix).

Anyway, the main limitation of DNA-based methods is that the presence of microbial DNA in an environment is not a direct measure of viable organisms (Jofre et al., 2010; Nocker et al., 2010). Microscopy visualization after live/dead staining aims to overcome this limitation.

4.2. DRINKING WATER MICROBIOTA

Joshua Lederberg coined the term *microbiota* to define "the ecological community of commensal, symbiotic and pathogenic microorganisms that literally share our body space" (Lederberg and McCray, 2001).

Many scientific articles distinguish *microbiome* to describe the collective genomes of the microorganisms that reside in an environmental niche and *microbiota* to describe the microorganisms themselves (Turnbaugh et al., 2007). However, by the original definitions these terms are largely synonymous.

The Earth Microbiome Project (EMP) was launched in August 2010, with the ambitious aim of constructing a global catalogue of the uncultured microbial diversity of this planet. Drinking water microbiome is still not included in the catalogue, highlighting one more time the knowledge gap in this field.

From a quantitative point of view, as expected, water sample collected during our one-year survey were all characterized by a very low concentration of microorganisms. Quantification through absorbance-based methods (i.e. OD_{600} for intact cells or Nanodrop for DNA extracts) was not reliable, so we decided to quantify bacteria through the amplification of a target region of DNA (16S rDNA V3-V4 regions).

qPCR data showed a significant increase in bacteria load after water passage through granular activated carbon filters. After chlorination the relative quantity of bacteria decreased, reaching values not significantly different from that of groundwater.

Considering Crescenzago DWTP (Site 2), an exception is represented by the samples collected in October 2014, when the concentration of bacteria recorded in groundwater samples was higher than in the other sampling points. Further analyses on Site 2 will clarify the dynamics beside microbial community changes. No significant differences in bacteria concentration trends were measured between October 2014 and November 2014 in Feltre DWTP (Site 1), when carbon filters were renewed. The increase in free chlorine concentration in chlorination basin (0.24 mg/L, over the limit of 0.20 mg/L) must be pointed out in November 2014.

Anyway, molecular techniques, as discussed before, have the limit of distinction between DNA belonging to live and dead bacteria. We used live/dead staining with fluorescent dyes coupled with microscopy visualization to estimate live dead ratio for each sampling point. The majority of bacteria collected both from groundwater and carbon filters are live, as expected. After the disinfection action of chlorination we found bacteria still alive and the percentage reach the 50%. This data strongly demonstrated that chlorination does not exert a full scale effect. As a consequence, the DNA detected in samples belonging to chlorination basin is for the half of bacteria that survived to disinfection, even if they are not necessarily pathogens.

Some European countries (e.g. Switzerland, Germany and The Netherlands) even distributed drinking water without disinfection residuals, but through the control of grow-limiting substrates (Van der Kooij, 2002: Thayanukul et al., 2013). Concentrations in the range of 10^4–10^5 cells mL^{-1} of diverse microbial populations are normal in drinking water, since it is common for bacteria to regrow during treatment and distribution, irrespective of the different disinfection treatments (Hammes et al., 2008).

HTS workflow was used to produce an accurate picture of the biological diversity present in drinking water.

The available technologies enable the evaluation of bacterial diversity and their relative abundance in various environments at a level never reached before. HTS-based metagenomic can detect very low-abundance, uncultivable members of the microbial community that could be otherwise missed or that are too expensive to identify using other methods.

We reported and analyzed, for the first time, the presence of **nanobacteria** along different steps of the DWTP, even after the end of the potabilization process.

We applied entropy-based approach (i.e. PhyloH, Sandionigi et al. 2014) for the analysis of nanobacterial diversity across DWTP. A strong point of our approach was the capacity of identifying a critical taxon with no or few previous taxonomic information. The phylogenetic information allows investigating the contribution of different lineages instead of summarising the results as a simple check-list of predefined taxa.

As described in Results, taxonomic analysis of all the microbial world inside DWTPs revealed that the majority of the sequences in all the samples were associated with the phyla Proteobacteria (42%) and Candidate Division OD1 (31%) with 7% sequences of unclassified bacteria ("Other") (Figure 3.22a). The predominance of Proteobacteria is consistent with previous drinking water studies performed in different geographic locations (Pinto et al., 2012; Gomez-Alvarez, 2015).

This can help to depict the health status of **groundwater** and the outcomes of potabilization processes.

Despite the low concentration in microorganisms, groundwater is characterized by the highest bacterial and nanobacterial diversity. Considering nanobacteria, we found taxa belonging to the candidate phyla/radiations OD1, OP11, TM6, TM7 (as reported as nanobacteria in Luef et al., 2015 and Brown et al., 2015) and to the related OP3, OP1, BRC1 and WS3 phyla/radiations. Even if OD1 candidate phylum was spread across the entire DWTPs, lineage L420, that includes all non-OD1 candidate nanobacteria phyla observed, is typical of groundwater. Thus less abundant phyla strongly contribute to α-diversity in groundwater.

Interestingly, there is no evidence of seasonality affecting microbial composition, showing groundwater as a stable ecosystem, not easily affected by external conditions.

Although all the drinking waters around the world are treated before human consumption to remove chemical and biological contaminants, relatively little is known about the changes in microorganisms composition during potabilization processes.

Only recent studies (e. g. Pinto et al., 2012; Gilbert et al., 2013; Wang et al., 2013) demonstrated that **granular activated carbon filters** harbour stable bacterial communities or biofilm that shape the bacterial composition downstream the water

treatment plant. These bacteria are able to persist even after **chlorine** treatments (Chiao et al., 2014). We also demonstrated that the native and opportunistic microbial communities deriving from groundwater are able to colonize carbon filters and significantly affect drinking water quality (Pinto et al., 2012; Lindgren, 2012). Interestingly, similar evidences arose starting from different DWTPs, located in the same metropolitan area (as in our case study) or in different continents (as in Pinto et al., 2012), and through different DNA sequencing chemistries, supporting the robustness of results obtained.

In the case of carbon filters microbial community, we reported a certain level of stability along a temporal scale, without measuring changes related to seasonality. Other recent studies (Pinto et al., 2014) showed opposite evidences: a seasonal pattern is exhibited by bacterial community and month and season were strong explanatory factors for changes in bacterial community structure.

The analysis of the forces that affect microbial dynamics provides new insights in drinking water treatment process. It is therefore evident that a greater capacity of microbial organism identification is essential to address relevant improvement in prevention strategies.

In conclusion, our data suggest that carbon filters are acting as a substrate enhancing microorganism growth and contribute to seed water downstream, since chlorination do not modify greatly the incoming living bacterial community in terms of global diversity. Pinto et al. (2012) observed a similar pattern, but they did not uncover the nanobacteria diversity that in our results is a large fraction of the total microbial diversity recovered (36.4%). We cannot exclude that this may derive from the use in those experiments of membrane filters with pore size ≥0.22 μm for filtration process and the consequent loss of ultra-small microorganisms.

Our further support to the seeding role of carbon filters came from the bacteria diversity of new (i.e. sterile) filters that is more similar to GW than to CF or CHL samples in operating conditions. Such a finding is highlighted by the diversity found in the water samples at the CHL step, after flowing through new filters.

4.3. PRELIMINARY ANALYSES TOWARD RESISTOME

Historically, concerns about the microbial quality of drinking water have focused on the occurrence of pathogens in drinking water distribution systems. Typical signature of genes encoding resistance to 'old' antibiotics such as tetracyclines, sulfonamides, aminoglycosides and β-lactams can be found in aquatic ecosystem. In the last few years a few scientific researches investigated the antibiotic resistance genes presence in the water treatment plant showing enrichment from the source to the tap of specific classes of resistance genes (Shi et al., 2013).

Most of these genes are located in plasmids and some are part of the variable gene cassettes of integrons and, probably, can easily be mobilized amongst bacteria (Garcillán-Barcia et al., 2011; Partridge, 2011, Vaz-Moreira, 2014). Interestingly, Chao et al. (2013) demonstrated that the microorganisms in drinking water treatment plant contained higher protective genes responding to the selective pressure of chlorination, such as glutathione related genes, indicating a possible co-selection of chlorine/chloride and antibiotic resistance.

In this preliminary study we analysed the presence and the relative abundance of resistance genes at three collection points. In particular, we tested with qPCR ampicillin (Shi et al., 2013), β-lactamase (Shi et al., 2013), Glu-Cys ligase (Chao et al., 2013) and glutathione synthetase (Chao et al., 2013) resistance genes. Considering the antibiotic resistance genes for ampicillin and β-lactam, no significant difference was detected among the three sampling points, showing no evidences of selective pressures acting on these mechanisms of resistance. A similar result was obtained

analysing glutathione related resistance genes: no significant difference was detected among the three sampling points and there was not a significant enrichment after chlorination.

However, these results do not exclude that water treatments could act as a selective system, amplifying the resistance genes signal, and further analyses are needed.

Unexpectedly we measured resistance genes even in groundwater samples, with a non predictable temporal trend across months. For example, samples collected in June 2014, in all the three sampling points, had relative high concentration in ampicillin, β-lactamase and Glu-Cys ligase genes compared to the other months, in particular July 2014, when was recorded a significant decrease in concentration for β-lactamase and Glu-Cys ligase genes (ampicillin gene data not available). It is not clear under which circumstances groundwater bacteria are sources of native resistance genes or whether they simply act as carriers or helper elements that, somehow, facilitate the spread of antibiotic resistance in different conditions.

A large scale analysis of resistance genes is the next step required to expand the knowledge at a microbial community scale and to unveil the *Resistome* (i. e. the collection of all the resistance genes of an ecosystem) of drinking water. Whole genome shotgun sequencing techniques will help revealing the role of resistance genes in survival of bacteria in condition of stress. Moreover, horizontal gene transfer and mobile genetic elements concur in the spread of resistance genes, thus affecting dynamics in microbial communities (Vaz-Moreira et al., 2014).

5. CONCLUSION AND FUTURE PERSPECTIVES

Our results clearly demonstrate that some bacterial taxa survive the passage along DWTP and that a large proportion of that is poorly known (i.e. the so-called “nanobacteria”). According to the parameters provided by international directives (e.g. the European 98/83/CE), drinking water analyzed during this survey is potable according all the existing laws and directives. Nowadays, nanobacteria are not considered pathogenic and are not routinely screened. Nevertheless, the detection of this group of uncultivable bacteria in drinking water opens new scenarios. It is likely that nanobacteria depend on other bacteria to survive (Luef et al., 2015; Brown et al., 2015). What are the bacterial interaction networks in our drinking water? Are their occurrence and concentration linked to the peculiarities of drinking water? Could they affect the water plant management?

Further analyses will consist in the DNA sequencing of carbon filters itself at an initial, intermediate and full level of usage. The discrimination and identification of live bacteria will be carried out coupling DNA sequencing with RNA sequencing, in order to distinguish active bacteria from DNA deriving from dead cells.

Finally, the exploration of nanobacterial world should start studying the co-occurrence of these ultra-small bacteria with bacteria in which they could depend on. Nelson and Stegen suggested that OD1 genomes have typical symbionts signatures, and acquire many fundamental metabolites from a partner organism through close contact. Interestingly, pili genes are well represented in the WWE3, OP11 and OD1 genomes, and may be involved in the uptake of environmental DNA and may aid the

cells in inter-organism, and in general environmental interactions (Chen et al., 2004; Proft et al., 2009).

Interestingly, my work raised even more questions than answers at the original questions, but it is clear that DWTPs should be treated as complex ecosystems rather than inert systems, where a tangled network of microbial interactions take place, from the source (groundwater, river, lake and so on) to the tap in our house. A better knowledge of these networks is crucial to improve the management of drinking water facilities.

APPENDIX

SAMPLE	BLAST HIT	CHARACTERISTICS
00.GW	1. *Uncultured bacterium clone EMIRGE_OTU_s1t2b_4296*	1. Soil bacteria
	2. *Uncultured bacterium clone 3BR-3H*	2. Fresh water environmental bacteria
	3. *Uncultured candidate division OD1 bacterium clone S2-082*	3. Soil bacteria (Chen et al., 2014)
	4. *Uncultured bacterium clone APC-3439-J3C9*	4. Aquatic bacteria
	5. *Uncultured bacterium clone ncd2023b08c1*	5. Environmental bacteria
00.CF	6. *Uncultured bacterium clone A13*	6. Br-reducing bacteria found in carbon filters
	7. *Uncultured bacterium clone LC0153bO5*	7. Environmental bacteria found in carbonatic mines
	8. *Polaromonas sp. BAC311 (Betaproteobacteria)*	8. Bacteria found in carbon filters in DWTPs (Magic-Knezev et al., 2009)
	9. *Uncultured organism clone SBZP_450*	9. Environmental bacteria found in soil (Harris et al., 2013)
	10. *Uncultured bacterium clone 2B*	10. Environmental bacteria (Bastida et al., 2010)
	11. *Uncultured bacterium clone A6B_39*	11. Bacteria found in carbon filters in DWTPs
00.CHL	12. *Polaromonas sp. BAC311*	12. Bacteria found in carbon filters in DWTPs (Magic-Knezev et al., 2009)
	13. *Uncultured delta proteobacterium clone Skagenf2*	13. Bacteria found in waste-water treatment plant (Kong et al., 2007)
	14. *Uncultured planctomycete clone Pln-17*	14. Environmental bacteria
	15. *Uncultured bacterium clone RS-B49*	15. Fresh water bacteria
	16. *Uncultured gamma proteobacterium clone sf-93*	16. Chlorinated water bacteria
	17. *Uncultured bacterium clone SEAA1AE121*	17. Fresh water bacteria (Bouali et al., 2014)
	18. *Uncultured bacterium clone EJ10-Ash11-53*	18. Fresh water bacteria
	19. *Uncultured planctomycete*	19. Fresh water and saline water bacteria
	20. *Uncultured bacterium*	20. Waste water bacteria (Matsumoto et al., 2010)

Table A.1. BLAST hit of cloned fragments (amplified with V3-V4 16S panbacterial primers).

Sample name	Treatment Plant	Sampling Point	Treatment	Month	Year	Season	Weather	Cold - hot	External Temperature (°C)	Humidity*
GW Dec-13	Feltre	Groundwater	no	Dec	2013	winter	sunny	cold	7	B4
CF Dec-13	Feltre	Carbon Filters	yes	Dec	2013	winter	sunny	cold	7	B4
CHL Dec-13	Feltre	Chlorination basin	yes	Dec	2013	winter	sunny	cold	7	B4
GW Jan-14	Feltre	Groundwater	no	Jan	2014	winter	cloudy	cold	1	B4
CF Jan-14	Feltre	Carbon Filters	yes	Jan	2014	winter	cloudy	cold	1	B4
CHL Jan-14	Feltre	Chlorination basin	yes	Jan	2014	winter	cloudy	cold	1	B4
GW Feb-14	Feltre	Groundwater	no	Feb	2014	winter	rain	cold	8	B4
CF Feb-14	Feltre	Carbon Filters	yes	Feb	2014	winter	rain	cold	8	B4
CHL Feb-14	Feltre	Chlorination basin	yes	Feb	2014	winter	rain	cold	8	B4
GW Mar-14	Feltre	Groundwater	no	Mar	2014	spring	sunny	warm	14	B2
CF Mar-14	Feltre	Carbon Filters	yes	Mar	2014	spring	sunny	warm	14	B2
CHL Mar-14	Feltre	Chlorination basin	yes	Mar	2014	spring	sunny	warm	14	B2
GW Apr-14	Feltre	Groundwater	no	Apr	2014	spring	sunny	warm	18	B3
CF Apr-14	Feltre	Carbon Filters	yes	Apr	2014	spring	sunny	warm	18	B3
CHL Apr-	Feltre	Chlorinati	yes	Apr	2014	spri	sunny	warm	18	B3

14		on basin				ng				
GW May-14	Feltre	Groundwater	no	May	2014	spring	sunny	warm	19	C2
CF May-14	Feltre	Carbon Filters	yes	May	2014	spring	sunny	warm	19	C2
CHL May-14	Feltre	Chlorination basin	yes	May	2014	spring	sunny	warm	19	C2
GW Jun-14	Feltre	Groundwater	no	Jun	2014	summer	thunderous	warm	16	B4
CF Jun-14	Feltre	Carbon Filters	yes	Jun	2014	summer	thunderous	warm	16	B4
CHL Jun-14	Feltre	Chlorination basin	yes	Jun	2014	summer	thunderous	warm	16	B4
GW Jul-14	Feltre	Groundwater	no	Jul	2014	summer	sunny	hot	28	B3
CF Jul-14	Feltre	Carbon Filters	yes	Jul	2014	summer	sunny	hot	28	B3
CHL Jul-14	Feltre	Chlorination basin	yes	Jul	2014	summer	sunny	hot	28	B3
GW Aug-14	Feltre	Groundwater	no	Aug	2014	summer	sunny	hot	25	B3
CF Aug-14	Feltre	Carbon Filters	yes	Aug	2014	summer	sunny	hot	25	B3
CHL Aug-14	Feltre	Chlorination basin	yes	Aug	2014	summer	sunny	hot	25	B3
GW Sep-14	Feltre	Groundwater	no	Sep	2014	autumn	sunny	hot	21	B3
CF Sep-14	Feltre	Carbon Filters	yes	Sep	2014	autumn	sunny	hot	21	B3
CHL Sep-14	Feltre	Chlorination basin	yes	Sep	2014	autumn	sunny	hot	21	B3
GW Oct-14	Feltre	Groundwater	no	Oct	2014	autumn	rain	warm	16	B4

CF Oct-14	Feltre	Carbon Filters	yes	Oct	2014	autumn	rain	warm	16	B4
CHL Oct-14	Feltre	Chlorination basin	yes	Oct	2014	autumn	rain	warm	16	B4
GW Nov-14	Feltre	Groundwater	no	Nov	2014	winter	sunny	cold	10	B3
CF Nov-14	Feltre	Carbon Filters	yes	Nov	2014	winter	sunny	cold	10	B3
CHL Nov-14	Feltre	Chlorination basin	yes	Nov	2014	winter	sunny	cold	10	B3
GW Oct-14 Site2	Crescenzago	Groundwater	no	Oct	2014	autumn	rain	warm	16	B4
CF Oct-14 Site2	Crescenzago	Carbon Filters	yes	Oct	2014	autumn	rain	warm	16	B4
CHL Oct-14 Site2	Crescenzago	Chlorination basin	yes	Oct	2014	autumn	rain	warm	16	B4
GW Nov-14 Site2	Crescenzago	Groundwater	no	Nov	2014	winter	sunny	cold	10	B3
CF Nov-14 Site2	Crescenzago	Carbon Filters	yes	Nov	2014	winter	sunny	cold	10	B3
CHL Nov-14 Site2	Crescenzago	Chlorination basin	yes	Nov	2014	winter	sunny	cold	10	B3

Table A.2. Environmental variables recorded during the sampling campaign. *B4: 100%-80% humidity; B3: 79.9%-60% humidity; B2: 59.9%-40% humidity; B1: 39.9%-20% humidity; C2: 19.9%-0% humidity.

Sample name	Alcalinity (mg/L)	Cl2 (mg/L)	Colour (CU)	Conductivity at 20°C (µS/cm)	[H+] (pH)	Total hardness (°F)	Index of aggressivity	Dry residue at 180°C (mg/L)	Temperature (°C)	Turbidity (NTU)
GW Dec-13	200	NA	0.00	716	7.4	30	12.14	465.4	14.6	0.01
CF Dec-13	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CHL Dec-13	203	0.02	0.19	704	7.45	28.00	12.38	457.6	12.90	0.06
GW Jan-14	180.94	NA	<0.01	689.119	7.64	32.31	12.4	447.927	14.3	<0.010
CF Jan-14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CHL Jan-14	171.56	0.03	<0.01	617.309	7.65	30.01	12.36	401.251	12.9	<0.010
GW Feb-14	174.28	NA	<0.01	720.201	7.69	28.69	12.34	468.131	13.8	<0.010
CF Feb-14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CHL Feb-14	172.05	0.01	<0.01	703.086	7.6	28.57	12.25	457.006	14.5	<0.010
GW Mar-14	224.94	NA	<0.01	717.68	8.04	33	12.26	466.492	15.1	<0.010
CF Mar-14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CHL Mar-14	221.74	0.01	<0.01	702.726	7.81	NA	NA	456.772	14.7	<0.010
GW Apr-14	210.58	NA	<0.01	728.947	7.62	29.21	12.41	473.816	15.2	<0.010
CF Apr-14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

CHL Apr-14	209.66	0.02	<0.01	724.798	7.63	29.15	12.42	471.119	15.6	<0.010
GW May-14	207.57	NA	<0.01	702.467	7.55	28.99	12.33	456.604	15.5	NA
CF May-14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CHL May-14	207.59	0.02	<0.01	680.313	7.66	<0.01	<1.00	442.203	16	NA
GW Jun-14	207.34	NA	<0.01	711.002	7.64	26.15	12.37	462.151	17.1	NA
CF Jun-14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CHL Jun-14	207.52	0.04	<0.01	700.165	7.53	29.48	12.32	455.107	16.3	NA
GW Jul-14	193.95	NA	<0.01	732.79	7.62	28.34	12.36	476.314	17	NA
CF Jul-14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CHL Jul-14	191.08	0.01	<0.01	697.342	7.52	31.87	12.31	453.272	17.7	NA
GW Aug-14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CF Aug-14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CHL Aug-14	210.94	0	<0.01	725.479	7.38	33.09	12.23	471.561	18.8	NA
GW Sep-14	206.69	NA	NA	735.464	7.77	27.32	12.53	478.052	NA	NA
CF Sep-14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CHL Sep-14	210.21	0.09	NA	715.535	7.61	28.02	12.38	465.098	16.3	NA
GW Oct-14	212.17	NA	NA	730.142	7.81	26.91	12.57	474.592	NA	<0.010

CF Oct-14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CHL Oct-14	201.82	0.1	NA	703.133	7.71	30.11	12.5	457.036	15.8	<0.010
GW Nov-14	205.37	NA	NA	715.958	7.74	26.55	12.47	465.373	NA	<0.010
CF Nov-14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CHL Nov-14	204.95	0.24	NA	704.303	7.66	32.21	12.48	457.797	14.9	<0.010
GW Oct-14_Site2	203.81	NA	NA	727.981	7.77	27.41	12.51	473.188	NA	<0.010
CF Oct-14_Site2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CHL Oct-14_Site2	202.33	0.1	NA	716.098	7.59	31.36	12.39	465.464	15.5	<0.010
GW Nov-14_Site2	207.82	NA	NA	729.129	7.64	27.75	12.4	473.934	NA	<0.010
CF Nov-14_Site2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CHL Nov-14_Site2	199.06	0.03	NA	698.868	7.57	32.35	12.38	454.264	14.9	<0.010

Table A.3. Chemical analyses on samples used in this study (data from MM).

Sample name	Sampling point	colonies at 37 °C	colonies at 22 °C	Bacteria identified at 37 °C	Bacteria identified at 22 °C
GW Jul-14	GW	0	1	/	Not identified
CF Jul-14	CF	4	5	*Bacillus spp.* *Bacillus idriensis* *Bacillus firmus*	*Bacillus spp.* *Bacillus idriensis* *Caulobacter spp.*
CHL Jul-14	CHL	4	1	*Bacillus spp.*	Not identified

Table A.4. Colonies grown after plating concentrated samples.

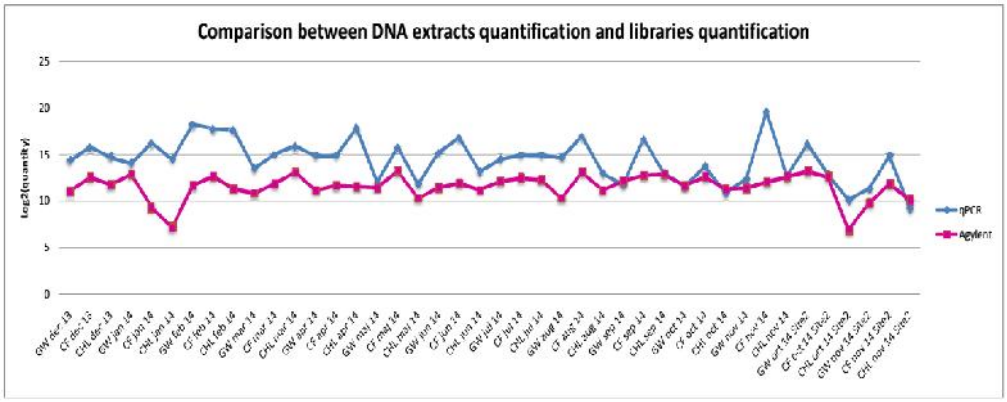


Fig. A.1. In blue is reported the quantification of 16S rDNA of DNA extracts belonging to DWTPs samples; values were obtained through qPCR and are expressed as log2(counts). In fuchsia is reported the quantification of 16S rDNA after library preparation, starting from the same samples; values were measured through Bioanalyzer (Agilent Technologies) and are expressed as log2(pg/microL).

Sample name	log2(counts)	sd
GW dec 13	14.355	0.007
CF dec 13	15.845	0.56
CHL dec 13	14.71	0.17
GW jan 14	14.045	0.64
CF jan 14	16.295	0.53
CHL jan 14	14.485	0.35
GW feb 14	18.275	0.08
CF feb 14	17.81	0.03
CHL feb 14	17.685	0.06
GW mar 14	13.485	0.35
CF mar 14	15.085	0.15
CHL mar 14	15.95	0.09
GW apr 14	14.86	0.01
CF apr 14	14.89	0.001
CHL apr 14	17.895	0.007
GW maj 14	12.025	0.33
CF maj 14	15.835	0.32
CHL maj 14	11.845	0.02
GW jun 14	15.245	0.09
CF jun 14	16.865	0.05
CHL jun 14	13.16	0.04
GW jul 14	14.49	0.06

CF jul 14	14.97	0.31
CHL jul 14	14.95	0.04
GW aug 14	14.655	0.06
CF aug 14	17.04	0.04
CHL aug 14	12.995	0.01
GW sep 14	11.77	0.81
CF sep 14	16.705	0.06
CHL sep 14	13.04	0.03
GW oct 14	11.6	0.47
CF oct 14	13.755	0.08
CHL oct 14	10.85	0.00
GW nov 14	12.395	0.06
CF nov 14	19.61	0.21
CHL nov 14	12.715	0.64
GW Oct 14 Site2	16.185	0.25
CF Oct-14 Site2	12.875	0.08
CHL Oct-14 Site2	10.155	0.73
GW Nov-14 Site2	11.35	0.24
CF Nov-14 Site2	14.86	0.11
CHL Nov-14 Site2	9.24	0.72

Table A.5. Bacteria quantification through qPCR. Mean values are reported, expressed as log2(counts)); sd: standard deviation.

Sample name	log2(counts)	sd
apr14_GW_ampicillin	6.28	0.32
aug14_GW_ampicillin	NA	NA
dec13_GW_ampicillin	0	0
feb14_GW_ampicillin	5.55	0.61
jan14_GW_ampicillin	0	0
jul14_GW_ampicillin	NA	NA
jun14_GW_ampicillin	7.83	
may14_GW_ampicillin	0	0
mar14_GW_ampicillin	5.19	0.00
nov14_GW_ampicillin	NA	NA
oct14_GW_ampicillin	NA	NA
sep14_GW_ampicillin	NA	NA
apr14_GW_betalactamase	0	0
aug14_GW_betalactamase	0	0
dec13_GW_betalactamase	10.84	1.06
feb14_GW_betalactamase	4.04	4.66
jan14_GW_betalactamase	5.33	6.15
jul14_GW_betalactamase	0	0
jun14_GW_betalactamase	12.67	0.08
may14_GW_betalactamase	4.27	4.93
mar14_GW_betalactamase	5.40	6.24

nov14_GW_betalactamase	0	0
oct14_GW_betalactamase	10.01	0.58
sep14_GW_betalactamase	4.45	5.14
apr14_GW_glucys_ligase	4.17	0.24
aug14_GW_glucys_ligase	NA	NA
dec13_GW_glucys_ligase	0	0
feb14_GW_glucys_ligase	0	0
jan14_GW_glucys_ligase	0	0
jul14_GW_glucys_ligase	0	0
jun14_GW_glucys_ligase	1.04.	1.20
may14_GW_glucys_ligase	0	0
mar14_GW_glucys_ligase	1.40	1.61
nov14_GW_glucys_ligase	NA	NA
oct14_GW_glucys_ligase	NA	NA
sep14_GW_glucys_ligase	NA	NA
apr14_GW_glutathione_synthase	6.20	4.68
aug14_GW_glutathione_synthase	4.60	5.3
dec13_GW_glutathione_synthase	4.19	4.84
feb14_GW_glutathione_synthase	4.26	4.92
jan14_GW_glutathione_synthase	4.27	4.93
jul14_GW_glutathione_synthase	4.40	5.08
jun14_GW_glutathione_synthase	4.59	5.30
may14_GW_glutathione_synthase	4.51	5.21

mar14_GW_glutathione_synthase	4.19	4.84
nov14_GW_glutathione_synthase	NA	NA
oct14_GW_glutathione_synthase	NA	NA
sep14_GW_glutathione_synthase	2.97	3.44
apr14_CF_ampicillin	5.53	2.01
aug14_CF_ampicillin	NA	NA
dec13_CF_ampicillin	0	0
feb14_CF_ampicillin	0	0
jan14_CF_ampicillin	6.90	1.82
jul14_CF_ampicillin	NA	NA
jun14_CF_ampicillin	6.28	0.69
may14_CF_ampicillin	3.49	4.03
mar14_CF_ampicillin	2.93	3.39
nov14_CF_ampicillin	NA	NA
oct14_CF_ampicillin	NA	NA
sep14_CF_ampicillin	NA	NA
apr14_CF_betalactamase	0	0
aug14_CF_betalactamase	7.66	0.59
dec13_CF_betalactamase	4.07	4.69
feb14_CF_betalactamase	3.69	4.26
jan14_CF_betalactamase	0	0
jul14_CF_betalactamase	0	0
jun14_CF_betalactamase	13.93	0.21

may14_CF_betalactamase	8.68	0.57
mar14_CF_betalactamase	3.85	4.44
nov14_CF_betalactamase	0	0
oct14_CF_betalactamase	0	0
sep14_CF_betalactamase	7.05	1.78
apr14_CF_glucys_ligase	0	0
aug14_CF_glucys_ligase	NA	NA
dec13_CF_glucys_ligase	0	0
feb14_CF_glucys_ligase	0.062	0.07
jan14_CF_glucys_ligase	0.06	0.07
jul14_CF_glucys_ligase	0	0
jun14_CF_glucys_ligase	3.75	1.36
may14_CF_glucys_ligase	0	0
mar14_CF_glucys_ligase	3.00	0.02
nov14_CF_glucys_ligase	NA	NA
oct14_CF_glucys_ligase	NA	NA
sep14_CF_glucys_ligase	NA	NA
apr14_CF_glutathione_synthase	4.42	5.11
aug14_CF_glutathione_synthase	3.80	4.39
dec13_CF_glutathione_synthase	4.38	5.06
feb14_CF_glutathione_synthase	4.00	4.62
jan14_CF_glutathione_synthase	6.18	5.68
jul14_CF_glutathione_synthase	3.59	4.14

jun14_CF_glutathione_synthase	5.52	5.29
may14_CF_glutathione_synthase	4.47	5.17
mar14_CF_glutathione_synthase	8.99	4.78
nov14_CF_glutathione_synthase	NA	NA
oct14_CF_glutathione_synthase	NA	NA
sep14_CF_glutathione_synthase	2.99	3.45
apr14_CHL_ampicillin	2.00	2.31
aug14_CHL_ampicillin	NA	NA
dec13_CHL_ampicillin	0	0
feb14_CHL_ampicillin	6.04	0.39
jan14_CHL_ampicillin	4.68	0.11
jul14_CHL_ampicillin	NA	NA
jun14_CHL_ampicillin	4.66	0.17
may14_CHL_ampicillin	3.20	3.69
mar14_CHL_ampicillin	6.42.	0.03
nov14_CHL_ampicillin	NA	NA
oct14_CHL_ampicillin	NA	NA
sep14_CHL_ampicillin	NA	NA
apr14_CHL_betalactamase	11.03	2.84
aug14_CHL_betalactamase	0	0
dec13_CHL_betalactamase	11.31	0.83
feb14_CHL_betalactamase	3.78	4.37
jan14_CHL_betalactamase	0	0

jul14_CHL_betalactamase	0	0
jun14_CHL_betalactamase	9.08	0.78
may14_CHL_betalactamase	3.75	4.33
mar14_CHL_betalactamase	4.55	5.25
nov14_CHL_betalactamase	2.44	2.82
oct14_CHL_betalactamase	7.99	0.57
sep14_CHL_betalactamase	0	0
apr14_CHL_glucys_ligase	2.89	0.32
aug14_CHL_glucys_ligase	NA	NA
dec13_CHL_glucys_ligase	0	0
feb14_CHL_glucys_ligase	0.09	0.11
jan14_CHL_glucys_ligase	0	0
jul14_CHL_glucys_ligase	0	0
jun14_CHL_glucys_ligase	0.97	1.12
may14_CHL_glucys_ligase	0	0
mar14_CHL_glucys_ligase	0.11	0.00
nov14_CHL_glucys_ligase	NA	NA
oct14_CHL_glucys_ligase	NA	NA
sep14_CHL_glucys_ligase	NA	NA
apr14_CHL_glutathione_synthase	4.58	5.29
aug14_CHL_glutathione_synthase	4.02	4.64
dec13_CHL_glutathione_synthase	4.26	4.92
feb14_CHL_glutathione_synthase	4.25	4.91

jan14_CHL_glutathione_synthase	4.33	5.00
jul14_CHL_glutathione_synthase	4.17	4.81
jun14_CHL_glutathione_synthase	4.30	4.97
may14_CHL_glutathione_synthase	4.53	5.23
mar14_CHL_glutathione_synthase	4.57	5.28
nov14_CHL_glutathione_synthase	NA	NA
oct14_CHL_glutathione_synthase	NA	NA
sep14_CHL_glutathione_synthase	1.63.	1.88

Table A.6. Rsistance genes quantification through qPCR. Mean values are reported, expressed as log2(counts); sd: standard deviation.

Taxonomy	Total	GW Dec-13	CF Dec-13	CHL Dec-13	GW Jan-14	CF Jan-14	CHL Jan-14
Unclassified;Other	0.1%	0.4%	0.1%	0.0%	0.1%	0.0%	0.1%
Archaea;Other	0.1%	0.3%	0.1%	0.1%	0.1%	0.1%	0.0%
Archaea; Crenarchaeota	0.6%	1.3%	0.3%	0.1%	0.5%	0.3%	0.0%
Archaea; Euryarchaeota	0.0%	0.2%	0.0%	0.0%	0.1%	0.0%	0.0%
Archaea; [Parvarchaeota]	1.7%	5.9%	1.4%	1.0%	2.4%	0.6%	0.4%
Bacteria;Other	6.6%	16.4%	5.6%	2.9%	6.5%	4.2%	1.4%
Bacteria;	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; AC1	0.1%	0.0%	0.0%	0.0%	0.1%	0.1%	0.0%
Bacteria; Acidobacteria	1.3%	2.4%	0.3%	0.2%	1.6%	0.7%	0.2%
Bacteria;	1.0%	1.6%	0.6%	1.0%	0.3%	0.2%	0.2%

Actinobacteria							
Bacteria; AncK6	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Armatimonadetes	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%
Bacteria; BHI80-139	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; BRC1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Bacteroidetes	1.6%	1.6%	2.6%	2.5%	0.9%	1.5%	0.5%
Bacteria; Chlamydiae	0.4%	0.4%	0.2%	0.3%	0.0%	0.2%	0.2%
Bacteria; Chlorobi	0.2%	0.4%	0.1%	0.0%	0.0%	0.2%	0.0%
Bacteria; Chloroflexi	1.5%	3.4%	0.5%	0.3%	3.3%	0.5%	0.1%
Bacteria; Cyanobacteria	0.9%	0.6%	0.7%	1.3%	0.1%	0.7%	0.9%
Bacteria; Elusimicrobia	0.5%	1.8%	0.2%	0.3%	0.5%	0.2%	0.1%
Bacteria; FBP	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Firmicutes	0.9%	1.2%	0.3%	0.7%	0.2%	0.1%	0.3%
Bacteria; Fusobacteria	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
Bacteria; GAL15	0.1%	0.1%	0.0%	0.0%	0.1%	0.1%	0.0%
Bacteria; GN02	1.0%	2.2%	4.1%	1.5%	0.3%	2.2%	0.3%
Bacteria; GN04	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Gemmatimonadetes	0.1%	0.3%	0.0%	0.1%	0.1%	0.0%	0.0%
Bacteria; Lentisphaerae	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; NC10	0.4%	0.2%	0.1%	0.1%	1.0%	0.1%	0.0%
Bacteria; NKB19	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Bacteria; Nitrospirae	2.2%	6.9%	0.8%	0.5%	2.3%	0.7%	0.4%
Bacteria; OC31	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; OD1	30.8%	16.5%	60.0%	62.2%	3.9%	61.0%	24.8%
Bacteria; OP1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; OP11	0.3%	1.5%	0.3%	0.0%	0.3%	0.3%	0.0%
Bacteria; OP3	3.5%	7.6%	1.6%	2.1%	4.5%	1.3%	0.8%
Bacteria; PAUC34f	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Planctomycetes	1.3%	1.8%	1.0%	1.8%	0.7%	0.2%	0.1%
Bacteria; Poribacteria	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Proteobacteria	41.8%	22.7%	18.7%	20.7%	69.6%	24.0%	68.9%
Bacteria; SBR1093	0.1%	0.1%	0.0%	0.0%	0.1%	0.1%	0.0%
Bacteria; SR1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Spirochaetes	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; TM6	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; TM7	0.3%	0.1%	0.1%	0.0%	0.0%	0.1%	0.0%
Bacteria; Verrucomicrobia	0.2%	0.4%	0.2%	0.1%	0.1%	0.1%	0.0%
Bacteria; WPS-2	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%
Bacteria; WS3	0.2%	1.3%	0.1%	0.0%	0.2%	0.1%	0.0%
Bacteria; ZB3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; [Caldithrix]	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; [Thermi]	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%

Taxonomy	GW Feb-14	CF Feb-14	CHL Feb-14	GW Mar-14	CF Mar-14	CHL Mar-14	GW Apr-14	CF Apr-14	CHL Apr-14
Unclassified;Other	0.1%	0.0%	0.1%	0.3%	0.1%	0.0%	0.0%	0.0%	0.0%
Archaea;Other	0.3%	0.1%	0.1%	1.0%	0.2%	0.0%	0.4%	0.0%	0.1%
Archaea; Crenarchaeota	1.8%	0.2%	0.2%	2.5%	0.4%	0.1%	3.2%	0.5%	0.3%
Archaea; Euryarchaeota	0.1%	0.0%	0.0%	0.4%	0.0%	0.0%	0.2%	0.0%	0.0%
Archaea; [Parvarchaeota]	4.2%	0.8%	1.3%	7.9%	1.5%	0.7%	4.8%	1.2%	1.8%
Bacteria;Other	18.4%	3.9%	6.8%	17.2%	4.6%	1.8%	20.6%	5.9%	5.2%
Bacteria;	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; AC1	0.2%	0.0%	0.1%	0.3%	0.1%	0.0%	0.0%	0.0%	0.0%
Bacteria; Acidobacteria	3.7%	0.5%	0.4%	4.0%	0.5%	0.2%	3.4%	0.8%	0.3%
Bacteria; Actinobacteria	0.5%	0.1%	0.3%	0.5%	0.2%	0.2%	0.4%	0.4%	0.3%
Bacteria; AncK6	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.1%	0.0%	0.0%
Bacteria; Armatimonadetes	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; BHI80-139	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; BRC1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Bacteroidetes	1.1%	1.9%	0.8%	1.6%	6.0%	0.5%	1.4%	3.1%	1.3%
Bacteria; Chlamydiae	0.2%	0.1%	0.3%	0.3%	0.3%	0.1%	0.2%	0.4%	0.3%
Bacteria; Chlorobi	0.3%	0.1%	0.1%	0.8%	0.2%	0.0%	1.0%	0.1%	0.1%
Bacteria; Chloroflexi	5.9%	0.6%	0.7%	3.5%	0.3%	0.2%	5.5%	1.2%	0.5%

Bacteria; Cyanobacteria	0.5%	0.6%	0.8%	0.3%	0.1%	0.3%	0.8%	0.6%	0.2%
Bacteria; Elusimicrobia	1.5%	0.2%	0.2%	1.4%	0.1%	0.3%	1.0%	0.4%	0.1%
Bacteria; FBP	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Firmicutes	0.1%	0.0%	0.1%	1.1%	0.2%	0.1%	0.6%	0.7%	0.1%
Bacteria; Fusobacteria	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; GAL15	0.4%	0.0%	0.0%	0.2%	0.0%	0.0%	0.6%	0.0%	0.0%
Bacteria; GN02	1.1%	1.6%	2.0%	0.3%	1.5%	0.5%	1.3%	1.3%	1.9%
Bacteria; GN04	0.1%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Gemmatimonadetes	0.3%	0.0%	0.0%	0.2%	0.0%	0.0%	0.3%	0.1%	0.0%
Bacteria; Lentisphaerae	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; NC10	1.0%	0.2%	0.2%	1.6%	0.4%	0.2%	2.3%	0.2%	0.0%
Bacteria; NKB19	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%
Bacteria; Nitrospirae	6.8%	0.8%	1.0%	4.9%	1.0%	0.6%	4.7%	1.3%	0.8%
Bacteria; OC31	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; OD1	8.9%	73.1%	59.2%	15.2%	67.0%	83.4%	6.8%	50.6%	41.2%
Bacteria; OP1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; OP11	0.6%	0.0%	0.3%	1.9%	0.3%	0.2%	0.4%	0.1%	0.3%
Bacteria; OP3	10.9%	1.3%	1.7%	8.8%	2.5%	1.1%	12.6%	1.8%	1.1%
Bacteria; PAUC34f	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Planctomycetes	2.0%	0.3%	0.7%	2.4%	0.5%	0.5%	1.6%	0.8%	0.9%
Bacteria;	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Poribacteria									
Bacteria; Proteobacteria	27.3%	13.0%	21.8%	19.9%	11.7%	8.7%	24.2%	27.6%	42.2%
Bacteria; SBR1093	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.1%	0.0%
Bacteria; SR1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Spirochaetes	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
Bacteria; TM6	0.1%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; TM7	0.2%	0.0%	0.1%	0.2%	0.0%	0.1%	0.2%	0.7%	0.5%
Bacteria; Verrucomicrobia	0.5%	0.2%	0.2%	0.3%	0.0%	0.2%	0.3%	0.0%	0.2%
Bacteria; WPS-2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%
Bacteria; WS3	0.5%	0.1%	0.1%	0.6%	0.1%	0.0%	0.3%	0.0%	0.1%
Bacteria; ZB3	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.1%	0.0%	0.0%
Bacteria; [Caldithrix]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; [Thermi]	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%

Taxonomy	GW May-14	CF May-14	CHL May-14	GW Jun-14	CF Jun-14	CHL Jun-14	GW Jul-14	CF jul14	CHL Jul-14
Unclassified;Other	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.1%	0.0%	0.0%
Archaea;Other	0.2%	0.1%	0.0%	0.6%	0.1%	0.0%	0.2%	0.1%	0.1%
Archaea; Crenarchaeota	1.1%	0.4%	0.2%	1.7%	0.6%	0.4%	0.8%	0.5%	0.3%
Archaea; Euryarchaeota	0.0%	0.0%	0.0%	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%
Archaea; [Parvarchaeota]	3.0%	1.3%	1.1%	7.2%	1.7%	2.4%	2.8%	0.7%	1.2%
Bacteria;Other	11.5%	4.1%	2.4%	21.4%	8.8%	3.9%	6.5%	3.8%	5.0%

Bacteria;	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; AC1	0.0%	0.0%	0.0%	0.3%	0.2%	0.0%	0.4%	0.2%	0.1%
Bacteria; Acidobacteria	3.9%	0.2%	0.3%	2.4%	0.6%	0.3%	0.9%	0.2%	0.5%
Bacteria; Actinobacteria	1.4%	0.4%	1.1%	0.9%	0.4%	0.9%	0.3%	0.4%	0.4%
Bacteria; AncK6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Armatimonadetes	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; BHI80-139	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; BRC1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
Bacteria; Bacteroidetes	2.0%	1.3%	1.1%	1.6%	1.3%	1.0%	0.8%	0.5%	0.8%
Bacteria; Chlamydiae	1.7%	0.3%	0.4%	0.3%	0.5%	0.5%	0.0%	0.1%	0.2%
Bacteria; Chlorobi	0.0%	0.0%	0.0%	0.3%	0.0%	0.0%	0.1%	0.0%	0.1%
Bacteria; Chloroflexi	0.6%	0.5%	0.2%	3.6%	0.9%	0.2%	1.8%	0.4%	0.5%
Bacteria; Cyanobacteria	1.5%	0.4%	0.9%	0.7%	0.6%	0.1%	0.3%	0.2%	0.3%
Bacteria; Elusimicrobia	0.6%	0.2%	0.2%	1.3%	0.3%	0.3%	0.2%	0.1%	0.3%
Bacteria; FBP	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Firmicutes	4.6%	0.2%	2.1%	1.4%	0.1%	1.0%	0.4%	0.4%	0.2%
Bacteria; Fusobacteria	0.0%	0.0%	0.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; GAL15	0.5%	0.0%	0.0%	0.3%	0.1%	0.0%	0.0%	0.0%	0.1%
Bacteria; GN02	0.2%	1.5%	1.4%	0.9%	1.3%	0.5%	0.5%	0.8%	0.2%
Bacteria; GN04	0.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Bacteria; Gemmatimonadetes	0.3%	0.0%	0.0%	0.3%	0.1%	0.0%	0.1%	0.0%	0.0%
Bacteria; Lentisphaerae	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; NC10	1.6%	0.3%	0.1%	1.1%	0.1%	0.1%	0.6%	0.0%	0.4%
Bacteria; NKB19	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Nitrospirae	3.0%	0.6%	0.9%	6.0%	1.3%	1.2%	2.5%	0.7%	1.1%
Bacteria; OC31	0.2%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; OD1	0.3%	60.5%	55.2%	11.0%	47.4%	61.9%	5.7%	21.0%	28.4%
Bacteria; OP1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; OP11	0.0%	0.1%	0.0%	0.9%	0.3%	0.0%	0.8%	0.3%	0.1%
Bacteria; OP3	12.1%	1.9%	2.1%	11.2%	4.1%	3.5%	3.3%	1.5%	3.3%
Bacteria; PAUC34f	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Planctomycetes	5.3%	0.4%	0.6%	1.6%	0.5%	0.9%	0.5%	0.2%	0.7%
Bacteria; Poribacteria	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Proteobacteria	42.9%	23.8%	28.6%	20.0%	27.7%	20.4%	69.7%	66.9%	55.0%
Bacteria; SBR1093	0.2%	0.0%	0.0%	0.1%	0.0%	0.1%	0.0%	0.0%	0.1%
Bacteria; SR1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Spirochaetes	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%
Bacteria; TM6	0.2%	0.1%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.1%
Bacteria; TM7	0.3%	1.1%	0.3%	0.3%	0.7%	0.1%	0.1%	0.6%	0.2%
Bacteria; Verrucomicrobia	0.4%	0.1%	0.3%	0.5%	0.1%	0.1%	0.0%	0.0%	0.1%
Bacteria; WPS-2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Bacteria; WS3	0.0%	0.0%	0.0%	0.9%	0.1%	0.2%	0.4%	0.2%	0.1%
Bacteria; ZB3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; [Caldithrix]	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; [Thermi]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Taxonomy	GW Aug-14	CF Aug-14	CHL Aug -14	GW Sep-14	CF Sep-14	CHL Sep-14	GW Oct-14	CF Oct-14	CHL Oct-14
Unclassified;Other	0.1%	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%
Archaea;Other	0.1%	0.0%	0.0%	0.3%	0.1%	0.0%	0.1%	0.0%	0.1%
Archaea; Crenarchaeota	1.1%	0.8%	0.1%	0.8%	0.3%	0.1%	1.7%	0.2%	0.6%
Archaea; Euryarchaeota	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%
Archaea; [Parvarchaeota]	1.5%	0.9%	0.4%	3.1%	1.4%	1.1%	1.8%	0.5%	0.3%
Bacteria;Other	5.7%	3.9%	3.9%	8.5%	5.5%	2.2%	10.4%	4.2%	6.2%
Bacteria;	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; AC1	0.0%	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%
Bacteria; Acidobacteria	2.4%	0.5%	1.1%	1.9%	0.4%	0.5%	6.8%	0.5%	0.4%
Bacteria; Actinobacteria	0.5%	0.6%	0.5%	2.0%	0.5%	0.9%	4.7%	0.4%	1.1%
Bacteria; AncK6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%
Bacteria; Armatimonadetes	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; BHI80-139	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; BRC1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.0%	0.0%
Bacteria;	1.6%	1.4%	0.5%	2.9%	1.6%	1.6%	1.3%	1.2%	1.3%

Bacteroidetes									
Bacteria; Chlamydiae	0.0%	0.2%	0.4%	0.0%	0.3%	0.7%	0.3%	0.3%	1.3%
Bacteria; Chlorobi	0.3%	0.1%	0.0%	0.5%	0.1%	0.1%	1.1%	0.4%	0.2%
Bacteria; Chloroflexi	3.2%	1.0%	0.9%	2.4%	0.7%	1.2%	6.2%	0.5%	0.8%
Bacteria; Cyanobacteria	0.2%	0.6%	0.5%	0.7%	0.3%	1.3%	0.7%	0.4%	1.0%
Bacteria; Elusimicrobia	0.8%	0.1%	0.3%	1.0%	0.3%	0.1%	1.1%	0.1%	0.0%
Bacteria; FBP	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Firmicutes	0.5%	0.0%	0.6%	0.4%	0.1%	0.6%	3.9%	0.1%	0.7%
Bacteria; Fusobacteria	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; GAL15	0.2%	0.0%	0.1%	0.1%	0.0%	0.1%	0.2%	0.0%	0.0%
Bacteria; GN02	0.3%	1.4%	1.0%	1.1%	1.8%	0.4%	0.5%	2.1%	0.5%
Bacteria; GN04	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.1%
Bacteria; Gemmatimonadetes	0.1%	0.0%	0.0%	0.1%	0.0%	0.0%	1.0%	0.0%	0.0%
Bacteria; Lentisphaerae	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; NC10	0.6%	0.1%	0.1%	0.8%	0.2%	0.5%	0.8%	0.5%	0.0%
Bacteria; NKB19	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Nitrospirae	4.7%	0.6%	0.9%	6.2%	2.0%	0.5%	6.5%	0.8%	3.2%
Bacteria; OC31	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; OD1	3.0%	46.2%	32.7%	7.6%	43.7%	20.7%	3.0%	50.7%	11.8%
Bacteria; OP1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; OP11	0.1%	0.1%	0.2%	0.8%	0.5%	0.4%	0.4%	0.2%	0.4%
Bacteria; OP3	2.6%	1.8%	1.4%	4.4%	1.3%	2.2%	6.7%	1.7%	2.6%

Bacteria; PAUC34f	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Planctomycetes	2.5%	0.6%	1.3%	2.0%	0.5%	3.2%	3.2%	0.2%	0.5%
Bacteria; Poribacteria	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Proteobacteria	66.0%	37.8%	51.8%	49.9%	37.4%	60.7%	36.5%	33.8%	66.2%
Bacteria; SBR1093	0.0%	0.0%	0.1%	0.0%	0.0%	0.5%	0.2%	0.1%	0.2%
Bacteria; SR1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; Spirochaetes	0.0%	0.1%	0.0%	0.1%	0.0%	0.0%	0.0%	0.2%	0.0%
Bacteria; TM6	0.1%	0.0%	0.0%	0.3%	0.1%	0.3%	0.0%	0.1%	0.1%
Bacteria; TM7	0.1%	0.8%	0.7%	0.1%	0.5%	0.2%	0.1%	0.9%	0.4%
Bacteria; Verrucomicrobia	0.2%	0.1%	0.2%	0.6%	0.1%	0.1%	0.2%	0.1%	0.1%
Bacteria; WPS-2	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; WS3	0.0%	0.0%	0.1%	1.0%	0.2%	0.0%	0.2%	0.0%	0.0%
Bacteria; ZB3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; [Caldithrix]	1.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bacteria; [Thermi]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Taxonomy	GW Nov-14	CF Nov-14	CHL Nov-14	GW Oct-14 Site2	CF Oct-14 Site2	CHL Oct-14 Site2	GW Nov-14 Site2	CF Nov-14 Site2	CHL Nov-14 Site2
Unclassified;Other	0.1%	0.1%	0.0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%
Archaea;Other	0.3%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%
Archaea; Crenarchaeota	1.8%	0.1%	0.0%	0.10%	0.20%	0.20%	0.10%	0.20%	0.00%

Archaea; Euryarchaeota	0.0%	0.0%	0.0%	0.00%	0.00%	0.00%	0.10%	0.00%	0.00%
Archaea; [Parvarchaeota]	3.7%	0.4%	0.2%	0.10%	0.30%	0.00%	0.10%	0.50%	0.00%
Bacteria;Other	14.4%	0.8%	0.9%	1.10%	4.30%	2.20%	4.20%	7.60%	1.30%
Bacteria;	0.0%	0.0%	0.0%	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria; AC1	0.5%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria; Acidobacteria	3.8%	0.1%	0.1%	0.50%	0.30%	0.00%	2.80%	0.30%	2.50%
Bacteria; Actinobacteria	0.7%	0.0%	7.8%	0.90%	0.80%	1.30%	0.90%	0.20%	3.80%
Bacteria; AncK6	0.0%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria; Armatimonadetes	0.0%	0.0%	0.3%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria; BHI80-139	0.0%	0.0%	0.0%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%
Bacteria; BRC1	0.0%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria; Bacteroidetes	1.1%	0.3%	4.7%	3.80%	0.80%	0.60%	2.10%	0.30%	0.70%
Bacteria; Chlamydiae	0.3%	0.0%	0.2%	1.20%	0.20%	0.40%	0.30%	0.30%	1.10%
Bacteria; Chlorobi	0.8%	0.0%	0.1%	0.20%	0.00%	0.00%	0.60%	0.10%	0.00%
Bacteria; Chloroflexi	7.5%	0.1%	0.3%	0.60%	0.40%	0.90%	1.60%	0.30%	0.20%
Bacteria; Cyanobacteria	0.3%	0.4%	1.2%	0.10%	2.40%	3.70%	0.00%	1.60%	7.20%
Bacteria; Elusimicrobia	1.7%	0.0%	0.1%	0.10%	0.30%	0.60%	1.30%	0.20%	0.00%
Bacteria; FBP	0.0%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria; Firmicutes	0.3%	0.0%	1.3%	0.20%	1.60%	2.90%	2.70%	0.20%	5.10%

Bacteria; Fusobacteria	0.0%	0.0%	0.0%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%
Bacteria; GAL15	0.6%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%
Bacteria; GN02	0.1%	0.1%	0.3%	0.10%	0.50%	0.00%	0.10%	0.70%	0.00%
Bacteria; GN04	0.0%	0.0%	0.0%	0.00%	0.00%	0.00%	0.40%	0.00%	0.00%
Bacteria; Gemmatimonadetes	0.3%	0.0%	0.0%	0.10%	0.10%	0.00%	0.20%	0.00%	0.00%
Bacteria; Lentisphaerae	0.0%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria; NC10	0.7%	0.0%	0.1%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria; NKB19	0.0%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria; Nitrospirae	9.4%	0.3%	0.7%	1.50%	0.80%	0.00%	2.70%	0.80%	0.00%
Bacteria; OC31	0.0%	0.0%	0.0%	0.00%	0.00%	0.00%	0.20%	0.00%	0.00%
Bacteria; OD1	7.1%	0.8%	0.0%	0.30%	39.40%	40.70%	1.10%	40.60%	17.30%
Bacteria; OP1	0.0%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria; OP11	2.3%	0.1%	0.1%	0.00%	0.10%	0.00%	0.10%	0.10%	0.00%
Bacteria; OP3	8.1%	0.5%	0.9%	0.70%	1.90%	2.30%	3.00%	2.10%	1.50%
Bacteria; PAUC34f	0.0%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria; Planctomycetes	3.3%	0.1%	1.6%	0.70%	1.60%	2.00%	2.10%	0.70%	4.30%
Bacteria; Poribacteria	0.0%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria; Proteobacteria	26.4%	95.5%	77.4%	86.90%	43.60%	41.60%	72.40%	42.40%	54.70%
Bacteria; SBR1093	0.3%	0.0%	0.1%	0.00%	0.00%	0.30%	0.00%	0.00%	0.00%
Bacteria; SR1	0.0%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria;	0.2%	0.0%	0.0%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%

Spirochaetes									
Bacteria; TM6	0.0%	0.0%	0.0%	0.00%	0.00%	0.30%	0.00%	0.00%	0.00%
Bacteria; TM7	0.2%	0.0%	1.2%	0.20%	0.10%	0.00%	0.30%	0.30%	0.20%
Bacteria; Verrucomicrobia	0.9%	0.0%	0.2%	0.30%	0.10%	0.00%	0.60%	0.10%	0.00%
Bacteria; WPS-2	0.0%	0.0%	0.0%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%
Bacteria; WS3	2.7%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria; ZB3	0.0%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria; [Caldithrix]	0.0%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bacteria; [Thermi]	0.0%	0.0%	0.0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table A.7. Percentage of bacteria assigned to each sample. Taxonomic rank: phylum.

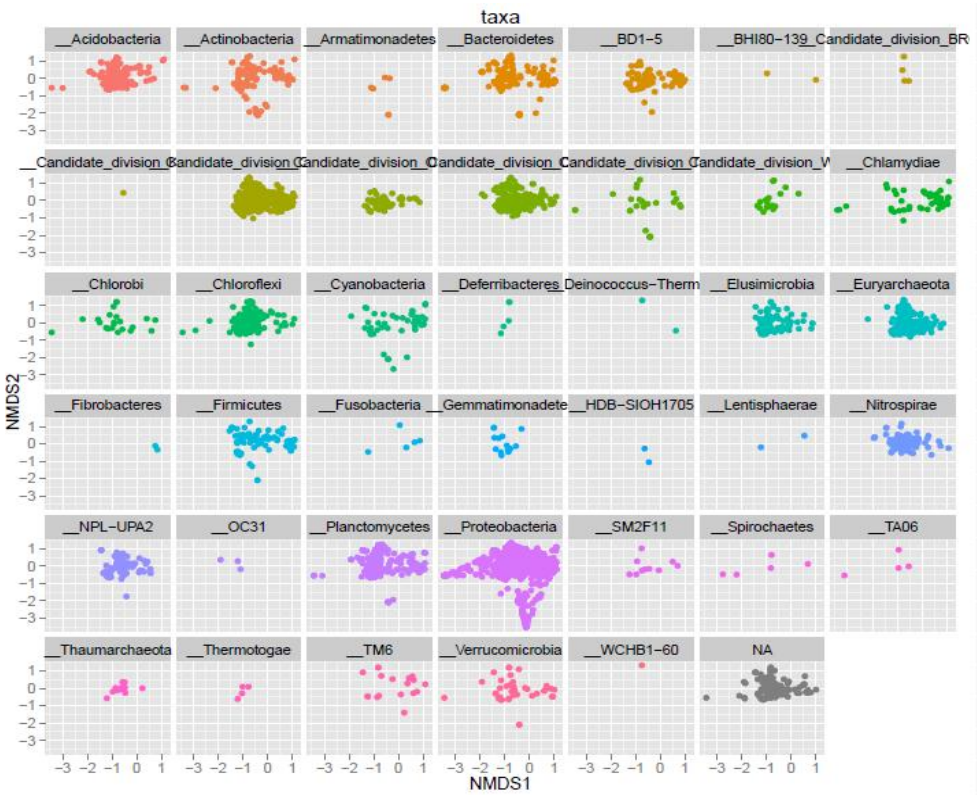


Figure A.2 Phylum distribution. NMDS (Non Metric Multi Dimensional Scaling) plot, Bray-Curtis.

REFERENCES

Ashbolt, N. J., Grabow, W. O. K., & Snozzi, M. (2001). Indicators of microbial water quality. *IWA Publishing,* 289-316.

Auguet, J. C., Barberan, A., & Casamayor, E. O. (2010). Global ecological patterns in uncultured Archaea. *The ISME journal*, *4*(2), 182-190.

Bali, L., Diman, A., Bernard, A., Roosens, N. H., & De Keersmaecker, S. C. (2014). Comparative Study of Seven Commercial Kits for Human DNA Extraction from Urine Samples Suitable for DNA Biomarker-Based Public Health Studies. *Journal of biomolecular techniques: JBT*, *25*(4), 96.

Baquero, F., Tedim, A. P., & Coque, T. M. (2013). Antibiotic resistance shaping multi-level population biology of bacteria. *Frontiers in microbiology*, *4*.

Beretta, G. P., Pagotto, A., Vandini, R., & Zanni, S. (1992). Aquifer overexploitation in the Po Plain: hydrogeological, geotechnical and hydrochemical aspects. *Simmers I., Villaroya F., Rebollo LF Editors: Selected Papers on Aquifer Overexplotation, IAH*, *3*.

Beretta, G. P., Avanzini, M., & Pagotto, A. (2004). Managing groundwater rise: experimental results and modelling of water pumping from a quarry lake in Milan urban area (Italy). *Environmental geology*, *45*(5), 600-608.

Berry, D., Xi, C., & Raskin, L. (2006). Microbial ecology of drinking water distribution systems. *Current opinion in biotechnology*, *17*(3), 297-302.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., ... & Shao, Y. (1997). The complete genome sequence of Escherichia coli K-12. *Science*, *277*(5331), 1453-1462.

Block, J. C., & Rolland, D. (1979). Method for Salmonella concentration from water at pH 3.5, using micro-fiber glass filters. *Applied and environmental microbiology*, *38*(1), 1-6.

Board, S. S. (1999). Size Limits of Very Small Microorganisms: Proceedings of a Workshop. National Academies Press.

Brown, C. T., Hug, L. A., Thomas, B. C., Sharon, I., Castelle, C. J., Singh, A., ... & Banfield, J. F. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature*, *523*(7559), 208-211.

Cai, L., Yang, Y., Jiao, N., & Zhang, R. (2015). Evaluation of Tangential Flow Filtration for the Concentration and Separation of Bacteria and Viruses in Contrasting Marine Environments. *PloS one*, *10*(8), e0136741.

Calderon, R. L. (2000). "The Epidemiology of Chemical Contaminants of Drinking Water". *Food and Chemical Toxicology* ***38*** (1 Suppl), S13–S20.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K.,... & Huttley, G. A. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature methods*, *7*(5), 335-336.

Cardenas, E., Cole, J. R., Tiedje, J. M., & Park, J. (2009). Microbial community analysis using RDP II (Ribosomal Database Project II): methods, tools and new advances. *Environmental Engineering Research*, *14*(1), 3-12.

Chao, Y., Ma, L., Yang, Y., Ju, F., Zhang, X. X., Wu, W. M., & Zhang, T. (2013). Metagenomic analysis reveals significant changes of microbial compositions and protective functions during drinking water treatment. *Scientific reports*, *3*.

Chesney, J. A., Eaton, J. W., & Mahoney, J. R. (1996). Bacterial glutathione: a sacrificial defense against chlorine compounds. *Journal of bacteriology*, *178*(7), 2131-2135.

Chiao, T. H., Clancy, T. M., Pinto, A., Xi, C., & Raskin, L. (2014). Differential resistance of drinking water bacterial populations to monochloramine disinfection. *Environmental science & technology, 48*(7), 4038-4047.

Colombo, L., Francani, V., & Alberti, L. (2015). SPEED CONTROL OF GROUNDWATER FLOW BY MEANS OF INJECTION EXTRACTION WELL SYSTEMS IN A URBAN AQUIFER (MILAN, ITALY). *International Multidisciplinary Scientific GeoConference: SGEM: Surveying Geology & mining Ecology Management*, *2*, 737.

Connon, S. A., & Giovannoni, S. J. (2002). High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolate. *Applied and Environmental Microbiology*, *68*(8), 3878-3885.

Davies, J., Spiegelman, G.B. & Yim, G. (2006). The world of subinhibitory antibiotic concentrations. *Curr Opin Microbiol 9*: 445–453.

Davies, J. & Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev 74*: 417–433.

D'Costa VM, King CE, Kalan L et al. (2011) Antibiotic resistance is ancient. *Nature 477*: 457–461.

Douterelo, I., Boxall, J. B., Deines, P., Sekar, R., Fish, K. E., & Biggs, C. A. (2014). Methodological approaches for studying the microbial ecology of drinking water distribution systems. *Water research*, *65*, 134-156.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460-2461.

European Drinking Water Directive http://eur-lex.europa.eu/legal-content/EN/LSU/?uri=CELEX:31998L0083

Fu, T.-J., K. Reineke, T. Krupinski, O. VanPelt, O., Maks-Warren, N. Parisi, B., et al. (2005). Integrating tangential flow filtration with rapid tests for detection of Escherichia coli O157:H7, Salmonella and Listeria monocytogenes in large volumes of spent sprout irrigation water. *IAFP Annual Meeting*. August 13-17, Baltimore, MD.

Furtak, V. A., Dabrazhynetskaya, A., Volokhov, D. V., & Chizhikov, V. (2015). Use of tangential flow filtration for improving detection of viral adventitious agents in cell substrates. *Biologicals*, *43*(1), 23-30.

Galimberti, A., Bruno, A., Mezzasalma, V., De Mattia, F., Bruni, I., & Labra, M. (2015). Emerging DNA-based technologies to characterize food ecosystems. *Food Research International*, *69*, 424-433.

Garcillán-Barcia, M. P., Alvarado, A., & de la Cruz, F. (2011). Identification of bacterial plasmids based on mobility and plasmid population biology. *FEMS microbiology reviews*, *35*(5), 936-956.

Gibert, O., Lefèvre, B., Fernández, M., Bernat, X., Paraira, M., Calderer, M., & Martínez-Lladó, X. (2013). Characterising biofilm development on granular activated carbon used for drinking water production. Water research, 47(3), 1101-1110.

Gillings, M. R. (2013). Evolutionary consequences of antibiotic use for the resistome, mobilome and microbial pangenome. *Frontiers in microbiology*, *4*.

Griebler, C., & Lueders, T. (2009). Microbial biodiversity in groundwater ecosystems. *Freshwater Biology*, *54*(4), 649-677.

Griebler, C., Kellermann, C., Jürgen Hahn, H., Stein, H., Brielmann, H., Berkhoff, S., & Fuchs, A. (2014a). Attempts for an integrative (ecological) assessment of groundwater ecosystems status. In *EGU General Assembly Conference Abstracts* (Vol. 16, p. 11614).

Griebler, C., Malard, F., &Lefébure, T. (2014b). Current developments in groundwater ecology—from biodiversity to ecosystem function and services. *Current opinion in biotechnology*, *27*, 159-167.

Griebler, C., & Avramov, M. (2015). Groundwater ecosystem services: a review. *Freshwater Science*, *34*(1), 355-367.

Guidelines for Drinking-water Quality, Fourth Edition; World Health Organisation; 2011.

Hammes, F., Berney, M., Wang, Y., Vital, M., Köster, O., & Egli, T. (2008). Flow-cytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes. *Water Research*, *42*(1), 269-277.

Hill, V. R., Polaczyk, A. L., Hahn, D., Narayanan, J., Cromeans, T. L., Roberts, J. M., & Amburgey, J. E. (2005). Development of a rapid method for simultaneous recovery of diverse microbes in drinking water by ultrafiltration with sodium polyphosphate and surfactants. *Applied and Environmental Microbiology*, *71*(11), 6878-6884.

Hill, V. R., Kahler, A. M., Jothikumar, N., Johnson, T. B., Hahn, D., & Cromeans, T. L. (2007). Multistate evaluation of an ultrafiltration-based procedure for simultaneous recovery of enteric microbes in 100-liter tap water samples. *Applied and environmental microbiology*, *73*(13), 4218-4225.

Hugenholtz, P. (2002). Exploring prokaryotic diversity in the genomic era. *Genome Biol*, *3*(2), 1-0003.

Jofre, J., & Blanch, A. R. (2010). Feasibility of methods based on nucleic acid amplification techniques to fulfil the requirements for microbiological analysis of water quality. Journal of applied microbiology, 109(6), 1853-1867.

Kahler, A. M., Johnson, T. B., Hahn, D., Narayanan, J., Derado, G., & Hill, V. R. (2015). Evaluation of an Ultrafiltration-Based Procedure for Simultaneous Recovery of Diverse Microbes in Source Waters. *Water*, *7*(3), 1202-1216.

Kantor, R. S., Wrighton, K. C., Handley, K. M., Sharon, I., Hug, L. A., Castelle, C. J., ... & Banfield, J. F. (2013). Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. *MBio*, *4*(5), e00708-13.

Kleijnen, R.G. (2011). The Chlorine Dilemma. *Eindhoven University of Technology.*

Kogure, K., Simidu, U., & Taga, N. (1979). A tentative direct microscopic method for counting living marine bacteria. *Canadian Journal of Microbiology*, *25*(3), 415-420.

Lambertini, E., Borchardt, M. A., Kieke Jr, B. A., Spencer, S. K., & Loge, F. J. (2012). Risk of viral acute gastrointestinal illness from nondisinfected drinking water distribution systems. *Environmental science & technology*, *46*(17), 9299-9307.

Larentis, M., Psenner, R., & Alfreider, A. (2015). Prokaryotic community structure in deep bedrock aquifers of the Austrian Central Alps. *Antonie van Leeuwenhoek*, *107*(3), 687-701.

Lederberg, J., & Mccray, A. (2001). The Scientist:\'Ome Sweet\'Omics--A Genealogical Treasury of Words. *The Scientist*, *17*(7).

Lindgren, E., Andersson, Y., Suk, J. E., Sudre, B., & Semenza, J. C. (2012). Monitoring EU emerging infectious disease risk due to climate change. *Science, 336*(6080), 418-419.

Li, X. K., Chu, Z. R., Liu, Y. J., Zhu, M. T., Yang, L., & Zhang, J. (2013). Molecular characterization of microbial populations in full-scale biofilters treating iron, manganese and ammonia containing groundwater in Harbin, China. *Bioresource technology*, *147*, 234-239.

Li, X., Wang, H., Zhang, Y., Hu, C., & Yang, M. (2014). Characterization of the bacterial communities and iron corrosion scales in drinking groundwater distribution systems with chlorine/chloramine. *International Biodeterioration & Biodegradation*, *96*, 71-79.

Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., & Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. The ISME journal, 5(2), 169.

Luef, B., Frischkorn, K. R., Wrighton, K. C., Holman, H. Y. N., Birarda, G., Thomas, B. C., ... & Downing, K. H. (2015). Diverse uncultivated ultra-small bacterial cells in groundwater. *Nature communications*, *6*.

Masetti, M., Poli, S., & Sterlacchini, S. (2007). The use of the weights-of-evidence modeling technique to estimate the vulnerability of groundwater to nitrate contamination. *Natural Resources Research*, *16*(2), 109-119.

McMahon, S., & Parnell, J. (2014). Weighing the deep continental biosphere. *FEMS microbiology ecology*, *87*(1), 113-120.

McMURDIE, P. J., & Holmes, S. (2012). Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data. In *Pac Symp Biocomput* (Vol. 17, pp. 235-246).

McOrist, A. L., Jackson, M., & Bird, A. R. (2002). A comparison of five methods for extraction of bacterial DNA from human faecal samples. *Journal of microbiological methods*, *50*(2), 131-139.

Onorati, G., Di Meo, T., Bussettini, M., Fabiani, C., Farrace, M. G., Fava, A., ... & Mazzoni, M. (2006). Groundwater quality monitoring in Italy for the implementation of the EU water framework directive. *Physics and Chemistry of the Earth, Parts A/B/C*, *31*(17), 1004-1014.

Partridge, S. R. (2011). Analysis of antibiotic resistance regions in Gram-negative bacteria. *FEMS microbiology reviews*, *35*(5), 820-855.

Payment, P., Bérubé, A., Perreault, D., Armon, R., & Trudel, M. (1989). Concentration of Giardia lamblia cysts, Legionella pneumophila, Clostridium perfringens, human enteric viruses, and coliphages from large volumes of drinking water, using a single filtration. *Canadian journal of microbiology*, *35*(10), 932-935.

Pereira, V. J., Marques, R., Marques, M., Benoliel, M. J., & Crespo, M. B. (2013). Free chlorine inactivation of fungi in drinking water sources. *Water research*, *47*(2), 517-523.

Pierucci, O. (1978). Dimensions of Escherichia coli at various growth rates: model for envelope growth. *Journal of bacteriology*, *135*(2), 559-574.

Pinto, A. J., Xi, C., & Raskin, L. (2012). Bacterial community structure in the drinking water microbiome is governed by filtration processes. *Environmental science & technology*, *46*(16), 8851-8859.

Pinto, A. J., Schroeder, J., Lunn, M., Sloan, W., & Raskin, L. (2014). Spatial-temporal survey and occupancy-abundance modeling to predict bacterial community dynamics in the drinking water microbiome. *MBio*, *5*(3), e01135-14.

Polaczyk, A. L., Roberts, J. M., & Hill, V. R. (2007). Evaluation of 1MDS electropositive microfilters for simultaneous recovery of multiple microbe classes from tap water. *Journal of microbiological methods*, *68*(2), 260-266.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*, gks1219.

Ramirez-Farias C, Slezak K, Fuller Z, Duncan A, Holtrop G, Louis P. Effect of inulin on the human gut microbiota: stimulation of Bifidobacterium adolescentis and Faecalibacterium prausnitzii. Br. J. Nutr. 2009; 101: 541–550.

Saby, S., Leroy, P., & Block, J. C. (1999). Escherichia coli resistance to chlorine and glutathione synthesis in response to oxygenation and starvation. *Applied and environmental microbiology*, *65*(12), 5600-5603.

Sandionigi, A., Vicario, S., Prosdocimi, E. M., Galimberti, A., Ferri, E., Bruno, A., ... & Casiraghi, M. (2014). Towards a better understanding of *Apis mellifera* and Varroa destructor microbiomes: introducing 'phyloh' as a novel phylogenetic diversity analysis tool. Molecular ecology resources.

Simon, C., & Daniel, R. (2011). Metagenomic analyses: past and future trends. *Applied and environmental microbiology, 77*(4), 1153-1161.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics, 22*(21), 2688-2690.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., ... & Sahl, J. W. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology, 75*(23), 7537-7541.

Staley, J. T., & Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Reviews in Microbiology*, *39*(1), 321-346.

Thayanukul, P., Kurisu, F., Kasuga, I., & Furumai, H. (2013). Evaluation of microbial regrowth potential by assimilable organic carbon in various reclaimed water and distribution systems. *Water research*, *47*(1), 225-232.

Thomas, J. M., & Ashbolt, N. J. (2010). Do free-living amoebae in treated drinking water systems present an emerging health risk?. *Environmental science & technology*, *45*(3), 860-869.

Torsvik, V., Salte, K., Sørheim, R., & Goksøyr, J. (1990). Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. *Applied and Environmental Microbiology*, *56*(3), 776-781.

Torsvik, V., & Øvreås, L. (2002). Microbial diversity and function in soil: from genes to ecosystems. *Current opinion in microbiology*, *5*(3), 240-245.

Tufenkji, N., Ryan, J. N., & Elimelech, M. (2002). Peer reviewed: The promise of bank filtration. *Environmental science & technology*, *36*(21), 422A-428A.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C., Knight, R., & Gordon, J. I. (2007). The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature*, *449*(7164), 804.

van der Kooij, D. (2002). Assimilable organic carbon (AOC) in treated water: determination and significance. G. Bitton (Ed.), *Encyclopedia of Environmental Microbiology, Wiley, Hoboken, NJ, USA (2002)*, pp. 312–327.

van der Wielen, P. W., Voost, S., & van der Kooij, D. (2009). Ammonia-oxidizing bacteria and archaea in groundwater treatment and drinking water distribution systems. *Applied and environmental microbiology*, *75*(14), 4687-4695.

Vaz-Moreira, I., Nunes, O. C., & Manaia, C. M. (2014). Bacterial diversity and antibiotic resistance in water habitats: searching the links with the human microbiome. *FEMS microbiology reviews*, *38*(4), 761-778.

U.S. Geological Survey. (2015). "The World's Water".

Wang, H., Pryor, M. A., Edwards, M. A., Falkinham III, J. O., & Pruden, A. (2013). Effect of GAC pre-treatment and disinfectant on microbial community structure and opportunistic pathogen occurrence. *Water research, 47*(15), 5760-5772.

Yu, K., & Zhang, T. (2012). Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge. *PloS one, 7*(5), e38183.

Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z., & Forney, L. J. (2012). Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One*, *7*(3), e33865.

MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2015) **15**, 697–710 doi: 10.1111/1755-0998.12341

Towards a better understanding of *Apis mellifera* and *Varroa destructor* microbiomes: introducing 'PHYLOH' as a novel phylogenetic diversity analysis tool

A. SANDIONIGI,*[1] S. VICARIO,†[1] E. M. PROSDOCIMI,‡ A. GALIMBERTI,* E. FERRI,* A. BRUNO,* B. BALECH,§ V. MEZZASALMA* and M. CASIRAGHI*

**ZooPlantLab, Department of Biotechnology and Biosciences, University of Milan-Bicocca, Piazza della Scienza 2, 20126 Milan, Italy, †Institute of Biomedical and Technologies (ITB), National Research Council (CNR), Via Giovanni Amendola, 122/D, 70126 Bari, Italy, ‡DEFENS, University of Milan, Via Mangiagalli, 25, 20133 Milan, Italy, §Institute of Biomembrane and Bioenergetics (IBBE), National Research Council (CNR), Via Giovanni Amendola, 165/A, 70126 Bari, Italy*

Abstract

The study of diversity in biological communities is an intriguing field. Huge amount of data are nowadays available (provided by the innovative DNA sequencing techniques), and management, analysis and display of results are not trivial. Here, we propose for the first time the use of phylogenetic entropy as a measure of bacterial diversity in studies of microbial community structure. We then compared our new method (i.e. the web tool PHYLOH) for partitioning phylogenetic diversity with the traditional approach in diversity analyses of bacteria communities. We tested PHYLOH to characterize microbiome in the honeybee (*Apis mellifera*, Insecta: Hymenoptera) and its parasitic mite varroa (*Varroa destructor*, Arachnida: Parasitiformes). The rationale is that the comparative analysis of honeybee and varroa microbiomes could open new perspectives concerning the role of the parasites on honeybee colonies health. Our results showed a dramatic change of the honeybee microbiome when varroa occurs, suggesting that this parasite is able to influence host microbiome. Among the different approaches used, only the entropy method, in conjunction with phylogenetic constraint as implemented in PHYLOH, was able to discriminate varroa microbiome from that of parasitized honeybees. In conclusion, we foresee that the use of phylogenetic entropy could become a new standard in the analyses of community structure, in particular to prove the contribution of each biological entity to the overall diversity.

Keywords: bioinformatics, high-throughput DNA sequencing, microbial community structure, phylogenetic entropy, symbioses

Received 10 July 2014; revision received 23 October 2014; accepted 24 October 2014

Introduction

In ecology, the analysis and interpretation of community diversity is a hot topic. In particular, with the advent of high-throughput DNA sequencing (HTS), the attention of the researchers on this topic dramatically increased. Among communities, symbioses, and particularly host–parasite interactions, are intriguing themes. In a symbiotic relationship, the partners reciprocally influence their physiology and, in general, their evolution. Nowadays, the characterization of the microbiome (intended as the sum of microscopic living beings found in a symbiotic relationship in different host body compounds, ranging from the gut to the skin) is considered pivotal to understand physiological changes occurring in a symbiosis (Mazmanian *et al.* 2005). In recent years, the scientific researchers focused on microbiome composition and variation in different hosts or physiological/environmental conditions using HTS (Sanchez *et al.* 2012; Dimitriu *et al.* 2013; Meriweather *et al.* 2013). The link between microbiome diversity and host health condition was discernible since the first published manuscripts. For instance, there is evidence that humans and mice subjected to different kind of stresses (such as diseases, parasites or ecological factors) show intense modifications in their own microbiomes in terms of initial colonization, final composition and overall stabilization (Candela *et al.* 2012; Lozupone *et al.* 2012).

However, the comprehension of mechanisms and dynamics influencing microbial diversity in hosts and

Correspondence: Maurizio Casiraghi, Fax: +39-02-64483568; E-mail: maurizio.casiraghi@unimib.it

[1]These authors contributed equally to this work.

© 2014 John Wiley & Sons Ltd

symbionts is much more complicated due to (i) the occurrence of several interacting variables (both abiotic and biotic) and (ii) the neglected contribution of the evolutionary history of single biological entities on the overall diversity when conventional analytical methods are applied (Fig. S1, Supporting information). Currently, researchers have just started incorporating historical constraints (represented as phylogenies) into their analyses. This innovation is motivated by the aim of filling the gap between evolutionary and ecological analyses of microbial communities (Lozupone *et al.* 2007, 2011).

Here we tested two classes of approaches for microbial community analysis: distance method approaches (DMAs) and partitioning phylogenetic diversity (PPD). DMAs became a standard in microbiome analyses, whereas PPD, which uses the phylogenetic entropy as a measure of microbiome diversity (Jost 2007; Chao *et al.* 2010), is here applied for the first time. The phylogenetic entropy is a generalization of Shannon entropy based on the fact that different observed categories are not all equally different from each others, having a similar structure that could be modelled using a phylogenetic tree. We implemented PPD in the user-friendly web application, PHYLOH.

We applied DMAs and PPD to characterize microbiomes in the model honeybee (*Apis mellifera*) and its parasitic mite varroa (*Varroa destructor*, Arachnida: Varroidae). The rationale is that the comparative analysis of both honeybee and the parasitic varroa microbiomes could open new perspectives about the role of the parasite on health of honeybee colonies. Indeed, *Varroa destructor* is considered responsible of the increasing incidence of deformed wing virus (Möckel *et al.* 2011), and it was reported as a vector of bacterial pathogens causing for example the European foulbrood (e.g. *Melissococcus plutonius*) (Forsgren 2010; Evans & Schwarz 2011).

Varroa destructor parasites honeybee larvae in their brood cells, where female mites feed on honeybee haemolymph and lay eggs. Mites have a large dispersal capability and, in absence of reiterate chemical and/or antibiotic treatments, infested honeybee colonies typically collapse in few years. For these reasons, the occurrence of varroa has serious consequences on ecological, social and economic contexts (Rinderer *et al.* 2010; Rosenkranz *et al.* 2010; Annoscia *et al.* 2012; Guzman-Novoa *et al.* 2012).

In spite of the interest, the ecological dynamics of the honeybee–varroa parasitic symbiosis are still largely unknown. Studies conducted on adult honeybees showed a characteristic microbiome (Jeyaprakash *et al.* 2003; Dillon & Dillon 2004; Mohr & Tebbe 2006; Martinson *et al.* 2011; Sabree *et al.* 2012). However, the microbiome of the larval stages, as well as that of the parasitic mite, remains largely unexplored, excluding few researches on the transmission of specific pathogens (Mouches *et al.* 1984; Cornman *et al.* 2010; Forsgren 2010; Martinson *et al.* 2012). At the light of present knowledge, alterations of honeybee microbiome due to the symbiosis with varroa are expected, even if several aspects influence the final outcome. The honeybee microbiome undergoes a peculiar dynamic over the life cycle of the insect: the pupa is almost sterile, as a consequence of the physiologic characteristics of the gut tract and the diet of mature larvae during the 6 days before capping (i.e. the closure of the brood cell) (Martinson *et al.* 2012). It is reasonable to assume that the bacterial load within the brood cells partially reflects the total bacterial count of the hives and that microbial communities characterizing the hives are partially present in the cells even after capping (Martinson *et al.* 2012). But, what happens when varroa alters this equilibrium? The perturbation caused by varroa in the developmental phase of honeybee larvae, and the consequent formation of the nutrition hole caused by the parasite, could lead to the intrusion of external bacteria (both from a potential vector or from the environment) into the larva, with a substantial modification of the original microbial community. The hypothesis is that varroa mites play a fundamental role in the alteration of bacterial composition of honeybee larvae, acting not only as a vector, but also as a sort of 'swing door' through which exogenous bacteria can enter into the larva and alter the mechanisms of primary succession of honeybee microbiome.

To validate our hypothesis, we studied varroa and honeybee bacterial communities through DNA-barcoded amplicon pyrosequencing, taking advantage of the HTS methods (Blow 2008; Metzker 2009), which also allow the detection of uncultured bacteria. We compared the results of the phylogenetic entropy-based approach (implemented in PHYLOH) with a classical method based on pairwise distances. We critically evaluated the strength and weakness of both approaches and the importance of phylogenetic constraint. At the same time, we aimed to reach a more complete vision of the relationships between honeybee and varroa in the microbiome interchange.

Materials and methods

A schematic overview of the experimental pipeline is shown in Fig. 1. Our work is divided into three sections as follows: (i) laboratory procedures including sampling; (ii) DNA extraction, amplification and pyrosequencing; (iii) sequence analysis and microbial community differential analysis.

Laboratory procedures

Sampling—Honeybee larvae and varroa mites were sampled directly from capped brood cells in eight

© 2014 John Wiley & Sons Ltd

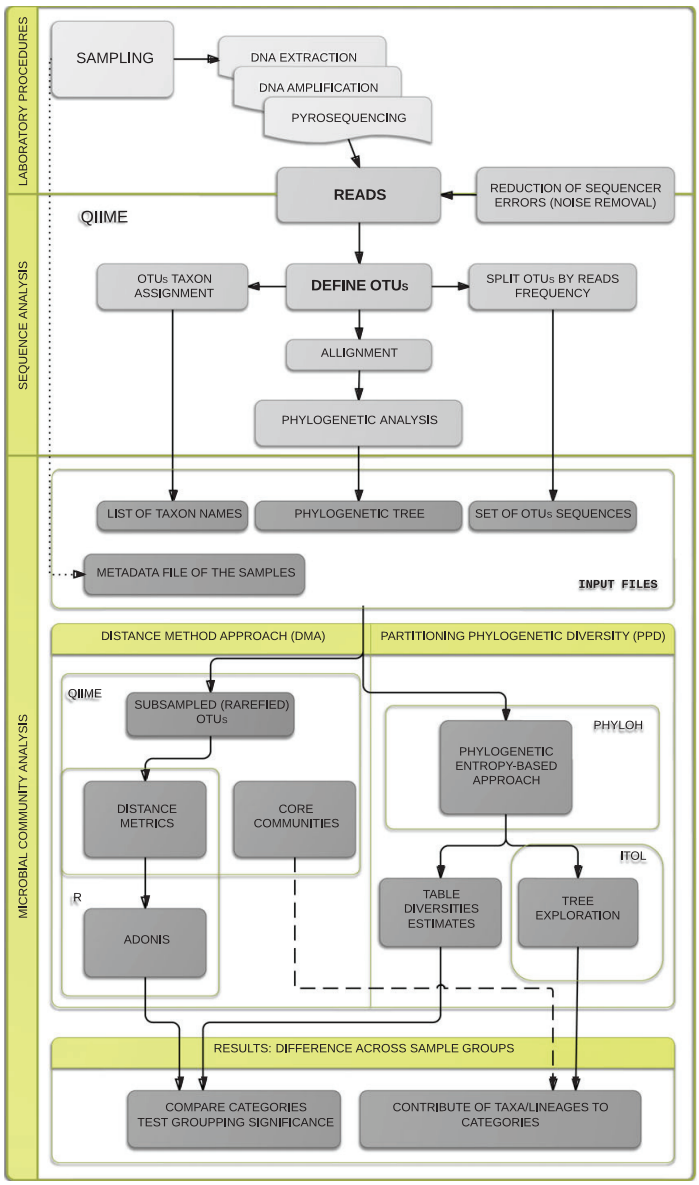


Fig. 1 Workflow of the procedures used in our project. Our work is divided into three sections: (i) laboratory procedures; (ii) sequence analysis; and (iii) microbial community analysis. In particular, the microbial community analysis was conducted following two approaches: microbial diversity analyses and partitioning phylogenetic distances.

apiaries in Northern Italy. We performed our analyses on a total of 21 individuals of honeybee larvae from seven different apiaries, and 21 varroa mites found in the same brood cells. As a negative control, a pool of five healthy honeybee larvae from a noninfected site was analysed.

Opercula of cells were opened with sterile instruments. Honeybee larvae and varroa were immediately removed and put in 2-mL tubes filled with absolute ethanol. The samples were stored at −20 °C until DNA extraction. In the study area, *V. destructor* is abundant and widespread, and consequently, we found only one noninfested apiary. This apiary was determined to be healthy after a careful inspection of all the hives by expert bee-keepers. The same experts determined that the presence of varroa in the other seven apiaries was

© 2014 John Wiley & Sons Ltd

high. In absence of preventive acaricide treatments, these honeybee colonies would have certainly collapsed before winter.

DNA extraction—All the extraction steps were performed in a sterile laminar flow cabinet. After the removal of the head, only the first segments of *A. mellifera* specimens (after cuticular removal) were used for the DNA extraction, while for *V. destructor*, DNA was extracted from the whole organism. The dissections were made in sterile conditions with a scalpel in a Petri dish. Each sample was then rehydrated for 4 h in sterile water at room temperature, and mechanically grinded with the scalpel. Total DNA was then extracted using a commercial kit (DNeasy Blood and Tissue Kit; Qiagen, Milan, Italy) and eluted in 50 μL sterile water. A pretreatment of Qiagen columns was performed to wash away any trace of contaminating bacterial DNA (Evans *et al.* 2003; Mohammadi *et al.* 2005). DNA extracts of the five larvae from the noninfested apiary were pooled.

16S rRNA amplification and pyrosequencing—The 16S rDNA gene fragment corresponding to the V3 hypervariable region was PCR-amplified with Roche 454 FLX (Titanium reagents) using the primer pair 341F (5′-CC TACGGGAGGCAGCAG-3′) and 518R (5′-ATTACCGC GGCTGCTGG-3′) (Watanabe *et al.* 2001). The reaction was performed in a 20 μL volume with the following reagents: 1X Taq-buffer with $MgCl_2$ 1,5 mM, dNTPs 2 mM, forward and reverse primers 1 μM each, Taq polimerase 0.5 U, DNA 50 ng, milliQ H_2O to the volume. The thermal cycle was: 94 °C for 90 s, 29 cycles at 94 °C for 20 s, 58 °C for 30 s, 72 °C for 20 s followed by a final extension at 72 °C for 10 min and 60 °C for 5 min.

A subsequent nested PCR step using the products of the first one as template was performed with standard 52 bp primers, comprising pyrosequencing primers A or B, multiplex identifiers (MID) and 518R or 341F primers. PCR mix and reaction volumes were the same as described above, except for the primers (10 μM). The thermal cycle was 94 °C for 90 s, 40 cycles at 94 °C for 20 s, 58 °C for 30 s, 72 °C for 20 s and a final extension at 72 °C for 10 min and 60 °C for 5 min.

For each sample, we used a unique combination of MIDs on the forward and reverse primers. PCR products were quantified using Bioanalyzer 2100 (Agilent) and normalized for quantity. Pyrosequencing was then performed on Roche 454 GS-FLX titanium by BMR Genomics Service at the Interdepartmental Biotechnology Centre of the University of Padua (CRIBI).

Sequences analysis—Sequence analysis was performed according to the following steps: (i) reads denoising; (ii) operational taxonomic units (OTUs) definition; (iii) OTUs taxon assignment; (iv) phylogenetic analysis; (v) OTUs splitting based on reads frequencies.

(i)All reads were trimmed, filtered and assigned to the corresponding sample according to their tag. Sequences shorter than 100 bp with quality average <30 or containing unresolved nucleotides were removed from the data set. ACACIA software version 1.52 (Bragg *et al.* 2012) was used for pyrosequencing noise removal considering Balzer error model and a maximum k-mer distance between reads of 13 (default parameter for error correction). The detection of chimera reads was performed using a pipeline based on USEARCH (Edgar 2010) and UCHIME (Edgar *et al.* 2011) included in Quantitative Insights Into Microbial Ecology (QIIME) software suite (version 1.7.0) (Caporaso *et al.* 2010).

(ii)UCLUST wrapper was used to cluster sequences into OTUs, based on 97% sequence similarity. For each OTU, a cluster centroid (i.e. a representative sequence) was chosen. To estimate diversity and reduce noise in patterns of beta diversity, singleton OTUs (i.e. OTUs represented by a single sequence) were removed before community analysis (Zhou *et al.* 2011).

(iii)Using a PYTHON2.7 script, we merged the Greengenes 16S rRNA database prefiltered at 97% identity (McDonald *et al.* 2011) with a bacterial OTUs data set constituted by symbionts previously described in studies conducted on *Apis mellifera* (Mohr & Tebbe 2006; Martinson *et al.* 2011, 2012; Mattila *et al.* 2012; Moran *et al.* 2012; Sabree *et al.* 2012). To create a reference database for taxonomic assignment, bacterial 16S rRNA sequences were retrieved from GenBank and clustered using UCLUST at 97% sequence similarity. The taxonomic attribution of cluster centroid sequences was carried out using RDP Bayesian classifier (Wang *et al.* 2007) with the new merged data set obtained using a 0.8 confidence level. OTUs were assigned by the RDP classifier, considering the fifth and sixth taxonomic levels wherever possible, which, in most cases, corresponded to family and genus ranks. If RDP assignment was uncertain (probability between 0.8 and 0.9), the QIIME-selected representative sequence was used to query with *blastn* algorithm on NCBI nucleotide database. Only in case of perfect match (i.e. max identity 100%, query coverage 100%), the NCBI taxonomy was used (results are shown in Table S1 in Supporting information and/or DRYAD entry doi:10.5061/dryad.j4d15).

(iv)A maximum-likelihood (ML) phylogenetic tree was built according to default parameters using FASTTREE software (Price *et al.* 2010) integrated in QIIME.

(v)The community abundance profile, produced by UCLUST and labelled by RDP Bayesian classifier, was split into two groups, as their global frequency was lower or higher than 1%. This threshold is the advised value in QIIME manual; it is often used in the literature and

© 2014 John Wiley & Sons Ltd

matches a gap in the frequency distribution of this data set (see Supporting information and/or DRYAD entry doi:10.5061/dryad.j4d15).

After this step, three data sets were produced: 'All Frequency Cluster' (i.e. AFC, including all OTUs), 'Low Frequency Cluster' (i.e. LFC, including OTUs with frequency lower than 1%) and 'High Frequency Clusters' (i.e. HFC, including OTUs with frequency higher than 1%). This partitioning allowed us to explore the effect of dominant and rare taxa among the microbiomes of honeybee and varroa.

Microbial community analyses

Microbial communities were examined using two approaches: the conventional distance matrices analysis (DMA) and the partitioning phylogenetic diversity (PPD). The last method was here applied for the very first time to the analysis of microbiomes.

Distance matrices analysis, coupled with ANOVA family statistics, is a well-known statistical framework in microbial community analysis that allows comparing the effect of different explanatory variables. This method is sensitive to unbalanced sampling, and it requires data rarefaction. In a HTS framework, the term 'unbalanced sampling' refers to the amount of sequences generated for each biological sample (e.g. if you get 100 000 sequences from the organism 1 and 50 000 sequences from the organism 2, the randomization leads to a loss of information from the larger sample). The principal pitfall of this approach is that it does not take into account the influence of rare OTUs on the global diversity of a sample (see also Fig. S1, Supporting information). To overcome this limit, we here propose the use of a PPD approach (Chao *et al*. 2010). PPD is being framed within information theory and can deal directly with discrete values, without producing distance matrices. It can also incorporate information deriving from unbalanced sampling, therefore avoiding a preliminary step of data rarefaction. Finally, being based on the phylogenetic structure of the data, PPD takes into account the influence of rare lineages on the microbial composition. In fact, similar sequences with low counts can build up consistent contribution to beta diversity on the branch leading to their most recent common ancestor.

The distribution of variability among bacterial communities was described for three environmental variables: (i) 'Cells' to show differences between single honeybee and the corresponding parasite found in the same brood cell, highlighting possible relationships between host and parasite microbiomes; (ii) 'Localities' to draw attention to differences existing between the microbiome of the seven apiaries; (iii) 'Status' to show microbiome differences among the pools of healthy honeybees, parasitized honeybees and mites.

Distance method approach

All the analyses were performed on the rarefied OTU tables to permit comparisons of diversity patterns within and between communities. The number of OTUs (based on the 97% sequences similarity) was determined for each sample. As depicted in Fig. 1, community analyses were performed with qualitative *(jaccard* and *unweighted UniFrac* (Lozupone *et al*. 2011) and quantitative distance metrics [*squared chord* (Cavalli Sforza & Edwards 1967; Orloci 1967) and *weighted UniFrac* (Lozupone *et al*. 2011)] using QIIME and R for statistical computing (R Development Core Team 2012).

Jaccard and *squared chord* were chosen as complementary metrics to *unweighted* and *weighted UniFrac* to test how the community pattern changes with or without phylogenetic information. We chose the *squared-chord* distance because it was identified in previous works as a metric fitting well at an exploratory analysis of communities where sampling was conducted blindly [see for example (Legendre & Gallagher 2001)].

To interpret the distance matrix, we used UPGMA hierarchical clustering method and we tested the robustness of results with jackknife analysis (1000 permutations). Further, to determine whether the grouping of samples by a given category was statistically significant, we used *adonis* (Oksanen *et al*. 2007) a permutational MANOVA allowing the use of distance matrix as dependent variable. This procedure is included within the QIIME suite and was implemented in *vegan* R package. The model used in *adonis* was the following:

$$\text{Dist}\,(i, j) \sim \textit{Status} + \textit{Localities} + \textit{Cells}$$

where i and j are all possible pairs of samples without redundancy and identity.

In this model, distances were considered as the response variable, whereas each of the environmental variables was considered as predictor.

We identified the microbial core community in honeybee and in the mite using *compute_core_microbiome.py* script from QIIME. OTUs were grouped according to their occurrence in a specific percentage of the total samples. The grouping steps were defined as elevens threshold between 0.5 and 1, corresponding to the 50% and 100% of the samples, respectively. This allowed defining the core community of each host species and more specifically to recognize the OTUs present in the majority of the samples of a given host.

Figures S2 and S3 (Supporting information and/or DRYAD entry doi:10.5061/dryad.j4d15), showing the

© 2014 John Wiley & Sons Ltd

taxonomic assignment and the abundance distribution, were generated with *phyloseq* (McMurdie & Holmes 2013) and *ggplot2* (Wickham 2009) R packages.

Partitioning phylogenetic diversity

Following the framework proposed by Jost (2007), it is possible to parse the total phylogenetic entropy of a data set (γ component), in intragroups entropy (α component, and intergroups entropy (β component). Jost (2007) distinguishes between entropy measures, having bit has unit (or nats, or bans, depending on the logarithm base), and diversity measure having as unit the number of equally abundant categories that would produce the same amount of entropy.

Partitioning operations are performed using entropy components, while the final result is transformed into diversity by elevating to the base of the used logarithm. Assuming that cluster label of observation is collected in vector $\boldsymbol{X}$ and that group label is collected in vector $\boldsymbol{Y}$, this framework allows to define H_γ as entropy of $\boldsymbol{X}$, H_α as entropy of $\boldsymbol{X}$ conditional to $\boldsymbol{Y}$, and H_β entropy as $H_\gamma - H_\alpha$, defined also as the mutual information between $\boldsymbol{X}$ and $\boldsymbol{Y}$ (MacKay 2003).

It is important to notice that *beta* diversity (D_β), the exponential of H_β, has as unit the number of equally abundant and different samples (the categories of $\boldsymbol{Y}$), while D_γ and D_α are measured in number of equally abundant clusters (the categories of $\boldsymbol{X}$). Within microbial community analysis, the interest lays generally in estimating D_β. To assess whether this measure was significantly different from 1 (i.e. the diversity value under the hypothesis of no difference among groups), we compared the realized statistics with a null distribution obtained by a permutation of $\boldsymbol{X}$ values onto $\boldsymbol{Y}$ ones. This procedure keeps a constant number of observations per group, allowing accounting for different sampling efforts. As described here, this procedure does not consider the phylogenetic structure that links the categories of the vector $\boldsymbol{X}$. This limitation is critical for biological data. Indeed, radically different findings from a biological perspective would produce the same Shannon-based beta diversity (see Fig. S1, Supporting information). This becomes possible using the phylogenetic entropy. The phylogenetic entropy is a generalization of Shannon entropy where the different observed categories are not all equally different from each other, but have a similarity structure that could be modelled using a phylogenetic tree. Following Chao *et al.* (2010), we assume that variable $\boldsymbol{X}$ is the abundance distribution of the clusters defined by UCLUST, and its similarity structure can be modelled with a phylogenetic tree t. The phylogenetic entropy measure could be defined as follows:

$$H_p(\boldsymbol{X}) = -\sum_{i \in B_t} \frac{L_i}{T} p_i \log p_i$$

where L_i is the value of the branch length for the ith branch while T is the average distance from tip observation to root in the tree, as defined by the formula

$$T = \sum L_i p_i$$

where $i \in B_t$ is the set of branches of the tree t, and p_i is the frequency of the descendant of branch i. Once this point is set, to generalize the partitioning of diversity to include phylogenetic information, it is sufficient to apply the previous definition of γ, α, β diversity using the phylogenetic entropy instead of the Shannon entropy. Phylogenetic entropy γ (H_γ) is equal to H_p ($\boldsymbol{X}$), while phylogenetic entropy α is equal to the weighted mean of the phylogenetic entropies per group, where weights are proportional to the number observation carried out in each group. More formally, this can be written as follows:

$$H_{p\alpha} = \sum_{y \in \boldsymbol{Y}} p_y \sum_{x \in \boldsymbol{X}} H_p(\boldsymbol{X} \mid \boldsymbol{Y} = y) \tag{1}$$

where p_y is the relative frequency of observation in each group while

$$\sum_{x \in \boldsymbol{X}} H_p(\boldsymbol{X} \mid \boldsymbol{Y} = y) \tag{2}$$

is the phylogenetic entropy measured in the different group y defined in vector $\boldsymbol{Y}$. Phylogenetic entropy beta ($H_{p\beta}$) remains defined as the $H_{p\gamma} - H_{p\alpha}$. This phylogenetic entropy beta, or phylogenetic mutual information, behaves in analogous way as Shannon-based mutual information given the Kullback–Leiber divergence. This matches the different way to estimate the classical mutual information (Marcon *et al.* 2012).

Given that the $H_{p\beta}$ is a difference of two summations in which each term is relative to a branch in the tree, it is possible to reorder the terms and obtain the contribution of each branch in the final $H_{p\beta}$.

The mutual information $H_{p\beta}$, or its exponential D_β, cannot be directly compared across different grouping variable $\boldsymbol{Y}$, given that the cardinality of each variable $\boldsymbol{Y}$ (the number of possible states) defines a different upper boundary to the value of each β diversity. To normalize diversity measures across different partitioning variables, it is necessary to transform diversity in overlap or effective average proportion of shared lineages in an

© 2014 John Wiley & Sons Ltd

individual environment. The value ranges from 1 (all lineages are shared) to 0 according to the following formula:

$$\text{Overlap} = \frac{(H_\beta - \max H_\beta)}{\max H_\beta}$$

where the maximum value of H_β is the logarithm of the number of groups or the entropy of $\boldsymbol{Y}$ depending on whether differential sampling needs to be included in the measure or not.

In this work, this approach was applied using the tree obtained from FASTTREE (Price *et al.* 2010) and the tree with the same topology, but internal branches with length 0 and terminal branches with length 1. The latter modified tree is used to perform the Shannon entropy analysis without taking into account the phylogenetic information, using the same software implementation. In fact, setting all internal branch lengths to zero leaves in the summation only the terms present in Shannon entropy formula. These two alternative settings allow to better evaluating the importance of phylogeny when interpreting the results.

Note that fractions of overlap are always higher in the phylogenetic entropy case, given that there are some phylogenetic similarities between different OTUs, while the classic Shannon-based approach assumes that each OTU is totally different from the others.

The mutual information (i.e. the beta entropy) deals with the covariation between two variables, so we explored the relationship between sequences and a single explanatory variable at the time. Within the framework of information theory, it could be possible to take into account the network of multiple interactions, but its application to phylogenetic entropy requires further investigations.

Partitioning phylogenetic diversity approach was implemented as a stand-alone Python script (PHYLOH available at https://github.com/svicario/phyloH) and includes a visualization routine based on ITOL (Letunic & Bork 2007) that allows to visualize the distribution of diversity and the contribution of the different OTUs to the partition as an html file. Input format follows PHYLOCOM input standard (Webb *et al.* 2008). To facilitate the use for the microbiologist community, we wrapped in a Web Service (https://www.biodiversitycatalogue.org/rest_methods/143) the script coupling it within a workflow (http://www.myexperiment.org/workflows/3570.html) with some parsing script enabling the use of output files from QIIME suite as input. The workflow could be run locally using a TAVERNA desktop engine (Wolstencroft *et al.* 2013) or as web application in the BioVeL portal. Access to the portal could be obtained from the BioVel website (www.biovel.eu) or contacting directly the authors.

Results

Results are organized according to the pipeline showed in Fig. 1.

Sequence analysis

Define OTUs—After sorting sequence reads for quality scores, sequencing errors and chimeras, we obtained 34 816 sequences. UCLUST returned 295 OTUs (data set AFC).

Taxonomic assignment—The complete list of OTUs, with corresponding taxa names and acronyms, is provided as Supporting information and/or DRYAD entry doi:10.5061/dryad.j4d15 (Table S1, Supporting information). Taxon assignment at family level of the three types of organisms involved (healthy honeybee, parasitized honeybee and varroa) is shown in Fig. 2.

Split OTUs by sequence frequency—Of 295 OTUs, 21 exceeded the threshold of 1% of minimum total observations (data set 'HFC', total sequences: 24 005), and 274 were defined as a rare OTUs not reaching the 1% threshold (data set 'LFC', tot sequences: 10 811).

Phylogenetic analysis—The maximum-likelihood tree generated by FASTTREE is shown as internal tree in PHYLOH output (see Supporting information and/or DRYAD entry doi:10.5061/dryad.j4d15).

Microbial community analysis

Distance method approach—The UPGMA analysis, considering *jaccard* distances for all the three sets of observations (AFC, HFC, LFC), shows a single cluster including all samples belonging to parasitized honeybees and mites and a separate cluster including the pool of healthy honeybees. On the contrary, with *squared-chord* metrics, which consider abundances information of OTUs in HFC and AFC, the analysis shows two different groups between parasitized honeybees and mites. These results are partially replicated with *UniFrac* (both *weighted* and *unweighted*), although separation is less sharp (see Fig. S4, Supporting information).

The *adonis* test was performed separately for each variable (*Cells*, *Localities* and *Status*) and using all variables together without interactions. Results were highly similar; for this reason only the coefficients of the model with all predictors combined are reported (Supporting

© 2014 John Wiley & Sons Ltd

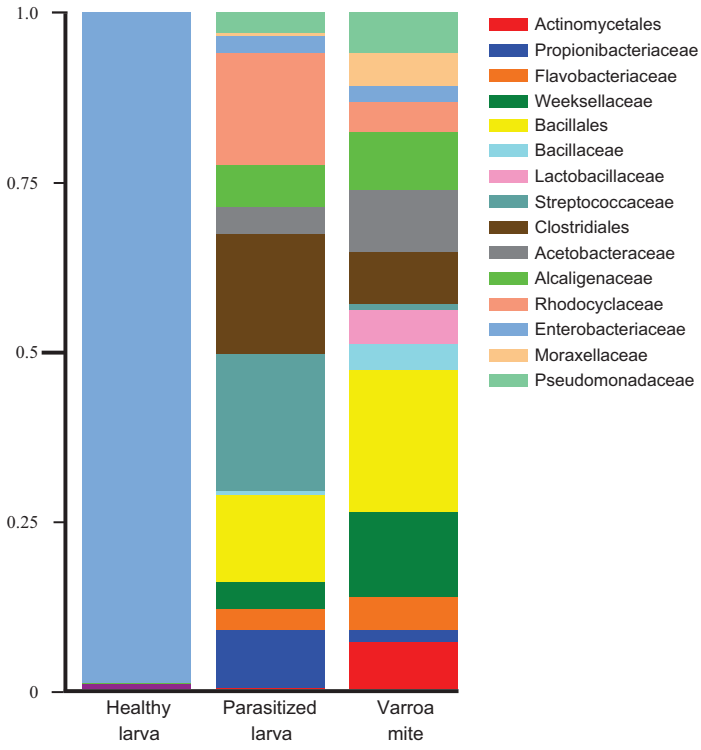


Fig. 2 Structure of microbiomes from healthy and parasitized honeybees and varroas. The histogram shows the 15 families detected and relative abundances. In two cases (Bacillales and Clostridiales), the family rank was not assigned, and consequently, the order rank is shown.

information Table S2 and/or DRYAD entry doi:10.5061/dryad.j4d15).

The *adonis* test grouping the samples for brood cells (*Cells*) has limited significance (*P*-values between 0.05 and 0.01) and small effect compared to degrees of freedom used up (34–35% explained variance, 1.6% the mean value for single cell). In addition, *adonis* test is significant only using the *jaccard* distance and *unweighted UniFrac* in the LFC data set. We found the same variance in the AFC data set, but only using the *unweighted UniFrac* (35% variance explained and *P*-value of 0.04).

A slightly stronger signal is detected in *Localities*, where *jaccard*, *unweighted UniFrac* and *squared chord* show a significant grouping in the AFC data set (*P*-value <0.01 and effect 17–20% with per-*Localities* mean variance explained between 3% and 2%). Similar signal is also visible with HFC and LFC, but only *jaccard* is significant (Supporting information Table S2 and/or DRYAD entry doi:10.5061/dryad.j4d15).

The *Status* grouping shows a very different pattern. The mean variance explained per state of *Status* builds up to 4–5% in the AFC data set and 9% for the *weighted UniFrac* in HFC. Only the LFC subset is not significant for the majority of distances used, and in any case the mean variance explained is quite low (1%).

Microbial core communities. In healthy honeybees, we found only three OTUs, one of them (Proteo-7, a member of the genus *Serratia*) accounting for the 99% of sequences. The remaining 1% is shared between two OTUs: Proteo-2 and Firmi-7, respectively, identified as *Achromobacter* sp. and *Lactobacillus* sp.

On the curves returned by QIIME script, we defined the host and parasite core microbiomes using a 0.8 threshold (Fig. S5, Supporting information). The threshold was chosen according to the steepest point in the curve, namely the point in which the least addition of OTU in the core ensures the largest fraction of samples to be compliant with the core representation. This means that a single OTU has to be present in the 80% of samples of a certain species to be considered 'core' for that species.

The OTUs considered as microbial core for honeybee and varroa are shown in Table 1.

Partitioning phylogenetic diversity—We used PHYLOH to perform partitioning tests on three environmental variables (*Cells*, *Localities* and *Status*) for the three frequency data sets (AFC or LFC or HFC) considering the

© 2014 John Wiley & Sons Ltd

Table 1 OTUs associated with healthy and parasitized honeybees and varroa by *compute_core_microbiome.py* script

Parasitized honeybee	Varroa mite	Shared
Propionibacterium sp. (Actino-1)	*Chryseobacterium* sp. (Bacte-1)	*Chryseobacterium* sp. (Bacte-1)
Chryseobacterium sp. (Bacte-1)	Flavobacteriaceae (Bacte-2)	*Streptococcus* sp. (Firmi-1)
Streptococcus sp. (Firmi-1)	*Streptococcus* sp. (Firmi-1)	*Lactobacillus* sp. (Firmi-7)
Clostridiales (Firmi-3)	*Bacillus* sp. (Firmi-2)	*Hydrogenophilus* sp. (Proteo-1)
Lactobacillus sp. (Firmi-7)	*Geobacillus* (Firmi-5)	*Achromobacter* sp. (Proteo-2)
Hydrogenophilus sp. (Proteo-1)	*Anoxybacillus* (Firmi-6)	
Achromobacter sp. (Proteo-2)	*Lactobacillus* sp. (Firmi-7)	
Escherichia sp. (Proteo-6)	*Hydrogenophilus* sp. (Proteo-1)	
	Achromobacter sp. (Proteo-2)	
	Pseudomonas sp. (Proteo-4)	
	Escherichia sp. (Proteo-6)	

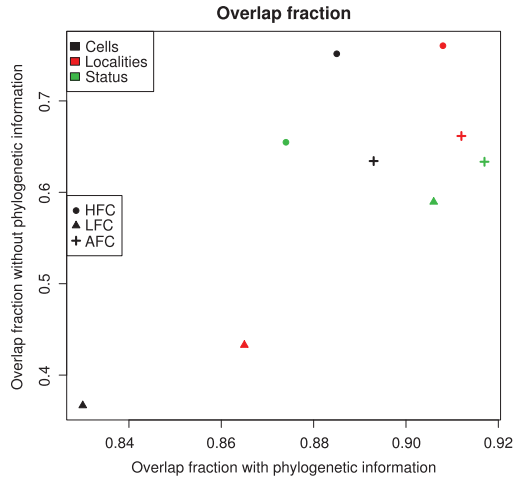


Fig. 3 Comparison of the percentage of overlap of the six data sets with or without taking into account the phylogenetic information. The percentage of overlap was calculated using estimate of beta diversity and phylogenetic beta diversity for AFC, HFC, LFC cluster frequencies and *Cells*, *Localities*, *Status* environmental variables.

phylogenetic information available or not. We can observe several qualitative differences among the 18 analyses transforming the mutual information into percentage of overlapping to allow a comparison among variables. As shown in Fig. 3, *Localities* and *Cells* variables are not heavily influenced by phylogenetic information. The relative position of LFC, HFC and AFC, measured in fraction of overlap counted in OTUs or lineages, is similar. In fact, LFC has, in both cases, low overlap, whereas AFC and HFC data sets show the maximum overlapping.

The variable *Status* describes a very different pattern from the previous two. Taking into account phylogeny, HFC differentiates more among groups (overlap 0.87), with the other two data sets showing about 0.91 overlapping. On the contrary, if phylogenetic information is not considered, the pattern is similar to the other two variables. The biggest difference is observable in LFC, while the smallest difference is observable in HFC.

The contribution of the branches to the mutual information between sequences and the Status *variable*. The experimental design output and the numerical partitioning of phylogenetic diversity for *Status* variable considering all OTUs found (AFC) are shown in Table 2. A general overview of the branches contribution to beta diversity across groups is shown in Fig. 4. The 295 AFC OTUs are well distributed on the phylogenetic tree, and consequently, their impact on the phylogenetic index is higher. It is observable that Bacte-1, -2, -3 are more typical of the varroa and are all grouped in the same lineage (L208). On the contrary, in Proteobacteria and Firmicutes, it is possible to find OTUs preferentially found in honeybee or varroa. Proteo-12, Proteo-14, Proteo-46, Proteo-24 and Proteo-17 OTUs belong to lineage L268 and are preferentially present in honeybee. There are three isolated OTUs preferentially found in varroa (Proteo-3, Proteo-4 and Proteo-5). Firmi-1, Firmi-9 and Firmi-4, typical of honeybee, are mixed with varroa's OTUs Firmi-6, Firmi-14, Firmi-2, Firmi-8 (both descending from lineage L387).

This lack of strong phylogenetic signal could be caused by recent specialization or random community assembly, given that Firmi-6 has a lower frequency sister taxa Firmi-14 also present mainly in varroa we prefer the first hypothesis.

Discussion

The analysis of complex communities (such as microbiomes) is today standard in different fields of biology and medicine. The number of published works is increasing daily, but there are some concerns on the real quality of the results showed. We are here comparing a 'traditional' approach to the community analysis [i.e. distance method approach (DMA)], vs. the innovative partitioning phylogenetic diversity (PPD, here implemented in

© 2014 John Wiley & Sons Ltd

Table 2 Variability in reads count for each sample. Variability in reads count and sample number in groups. Overall gamma in each group. Overall alpha and contribution to alpha in each group. Beta across samples within groups and between groups. E is always within S, given that each sample belongs to only one environment type or sample group

Experimental design diversity: entropy and diversity of observation in the different groups							
H_Environment							
H(E)			MaxDiversity				Diversity
0.691			2				2
Experimental design diversity: entropy and diversity of observation in the sample within the groups							
H_Sample							
H(E)			MaxDiversity				Diversity
3.68			42				39.5
Gamma diversity: diversity using all data and in each group							
H_gamma		H_gamma_parasitized_honeybee			H_gamma_mite		
H(T)	alpha_Diversity	H(T | E = host_p)	gamma_Diversity		H(T | E = pars)	gamma_Diversity	
1.76	5.83	1.75	5.73		1.69	5.42	
Alpha diversity: mean within group diversity							
H_alphaByEnvironment			H_alphaBySamples				
H(T | E)	alpha_Diversity		H(T | S)	alpha_Diversity			
1.72	5.56		1.35	3.85			
Beta diversity or mutual information between the phylogeny and a given grouping: diversity across group and across sample within same group							
MI_treeAndEnvironment				MI_treeAndSampleGivenEnvironment			
I(T,E)	beta_Diversity	Percentage_Overlap	*P* value	I(T,S | E)	beta_Diversity	Percentage_Overlap	*P* value
0.0465	1.05	0.933	0	0.369	1.45	0.466	0
Difference of each group from total: phylogenetic Kullback–Leiber divergence between each group and the overall sample							
KL_of_host_p Observed							KL_of_pars Observed
0.0162							0.0303

H_gamma = H_alphabyEnvironment + H_beta = H(T) = H(T | E) + I(T,E) while taking into account sample info: H_gamma = H_alphaBySample + H_betabySamplegivenEnvironment+H_beta = H(T) = H(T | S) + I(T,S | E) + I(T,E).

our newly proposed software PHYLOH). The final aim is to propose our approach as a new standard in community diversity analysis. We tested the performance and utility of DMAs and PPD in the case of honeybee and varroa microbiomes, a biological scenario intriguing and complex at the same time. Our results showed that healthy honeybees have a simplified microbiome, constituted of few bacterial OTUs, while varroa is characterized by a more complex microbiome, qualitatively not different from that of parasitized honeybee. A simple conclusion could be that varroa microbes infected honeybees. However, the most abundant OTUs of parasitized honeybee do not derive directly from the mite, but are generalist or environmental bacteria. This is undoubtedly a peculiar result suggesting that these microbes could play a role of pioneer species, with a potential pathogenic activity (i.e. Firmi-1, -9, L387, *Streptococcus*; Firmi-4, -14, L387, Clostridiales) (Lozupone *et al*. 2012).

Our results show that both DMAs and PPD approach give comparable results, but only PPD shows explicit support and allows, within the same statistical framework, to observe our data both from the general pattern to the contribution of single or group of OTUs.

We partitioned data according to three environmental variables, which represent three different forces shaping microbiome diversity: *Localities*, *Cells* and *Status*. When *Localities* variable is considered, in DMAs approach, the UPGMA method and the *adonis* test analyses show a weak, but significant signal for most of the distances and data set used. When all OTUs (AFC) combined with *unweighted UniFrac* are considered, the signal is stronger.

© 2014 John Wiley & Sons Ltd

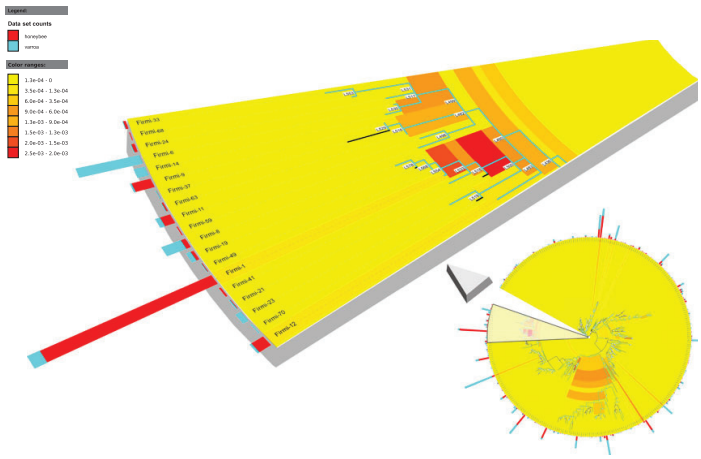


Fig. 4 In the figure are shown the main tips and relative branches of one of the most discriminating lineages across *Status* groups [L387, see tree.html (Appendix S1) in Supporting information and/or DRYAD entry doi:10.5061/dryad.j4d15 for further details] as example of the 'hairy pacman graphical output' derived from a PHYLOH analysis. Three types of data are shown on the tree: (i) the cyan colour of the branches indicates a significant contribution to I(T | E) (beta entropy or mutual information); (ii) the background of each branch is a gradient (colour range) from yellow to red for increased contribution to I(T | E); (iii) bar plot on each tip indicates the number of reads count in each group. For more details, see README file (Appendix S1) in Supporting information and/or DRYAD entry doi:10.5061/dryad.j4d15.

The results are similar using PPD, but with a stronger support. In fact, the variable *Localities* produces a significant effect, although this is better appreciable for rare OTUs (LFC) rather than considering the most abundant OTUs (HFC). In other words, low-frequency OTUs (better defined using PPD) represent the fraction of bacteria characterizing the single apiary. This finding is in agreement with previous results, showing that geographical distance is not crucial in shaping the core microbiome of beehives (Sabree *et al*. 2012).

In our hypothesis, the bacteria are transmitted from varroas to the honeybee larvae. This is corroborated by the analysis of the variable *Cells*, where differences between each larva and the corresponding mite were estimated. This grouping has limited effects in terms of mean explained variance when the DMA approach is used, but it is still significant for some distances in AFC and LFC data sets. In PPD, the overlap across categories is obtained by *Cells* with the LFC data set, showing a strong effect. This situation is difficult to explain: only the more frequent bacteria are shared among cells, whereas rare bacteria are not. However, the rare bacteria are more interesting because they act as a clear signature to discriminate different apiaries.

Considering the variable *Status*, both DMAs and PPD approaches can discriminate the three different categories: healthy honeybees, parasitized honeybees and varroas. Healthy honeybees have a simply and distinctive community with only one dominant OTU and two other low-frequency OTUs. The most representative phylotype (Proteo-7) belongs to the genus *Serratia* that was isolated from the intestinal contents of healthy foraging worker honeybees (Jeyaprakash *et al*. 2003), a well-known symbiont (generally harmless) in many insect taxa (Dillon & Dillon 2004).

The scarcity of bacteria in the healthy larvae could be attributed to their particular gut morphology, physiology and nutrition (Martinson *et al*. 2012). Indeed, the larva retains its faeces from the early days of development, due to the temporary absence of a connection between the large mid-gut and the hindgut. The mature larva defaecates just before spinning a cocoon, when the capping has already happened. As the cocooned pupa does not eat, we can assume that there is no further colonization by bacteria present in the brood cell. Through these mechanisms, the early microbiome characterizing honeybee larvae is maintained constant in composition and ubiquitous in space (Jeyaprakash *et al*. 2003; Mohr & Tebbe 2006).

The higher level of bacterial communities diversity showed in varroa and parasitized larvae suggests a transmission from parasite to the host. These two microbiomes are identical from a qualitative point of view (taking into account the HFC), but the relative abundance of the different OTUs clearly differentiates the two bacterial communities. When phylogenetic information is included, the level of discrimination between parasitized larvae and varroa is even higher. In fact, only using

© 2014 John Wiley & Sons Ltd

PPD and *weighted UniFrac* in DMA (i.e. the approaches that take into account the phylogenetic signal), the two clusters are highly discriminated. However, only in PPD, it is possible to show the contribution of the single OTU and groups of OTUs to the differences observed among samples.

The enhanced performance of PPD is exemplified in PHYLOH output tree [Supporting information tree.html (Appendix S1) and/or DRYAD entry doi:10.5061/dryad.j4d15 and as an example Fig. 4] where the lineages L208, L268 and L387 are those discriminating the microbiome of varroa and infected honeybee. In particular, L208 lineage encompasses OTUs present in varroa, while L268 and L387 contain OTUs present in both varroa and honeybees. The most discriminating OTUs of lineage L208 belong to the genus *Chryseobacterium* (Bacte-1, -2, -3). In contrast to the majority of bacteria belonging to Flavobacteria, typically found in soil and water environments, two (Bacte-1, -2) of these three OTUs were found as pathogens of soft ticks (Burešová *et al.* 2006). Given the phylogenetic closeness between mites and ticks, it would be interesting to investigate the role of this genus in mites. The OTUs belonging to the lineage L268 were assigned to genus *Haemophilus*, in which bacteria recognized as pathogenic to bees but not associated with the presence of varroa were found. Indeed, there are numerous studies related to the antimicrobial properties of honey with references to *Haemophilus* (i.e. Jeffrey & Echazarreta 1996; Antúnez *et al.* 2008; Al-Waili *et al.* 2011).

Proteo-4 (genus *Pseudomonas*) is strongly present in varroa, probably because it is common on the mite cuticle (Tang *et al.* 2012), which had not been removed in our study.

Lineage L387 (Bacillaceae: Firmi-6, -8) includes OTUs associated with honeybee (Mohr & Tebbe 2006; Evans & Schwarz 2011; Moran *et al.* 2012), but here we found associated mainly with varroa. It is possible that these bacteria may be generalist present in the hive.

In conclusion, we would like to underline that the partitioning of phylogenetic diversity is a powerful method to analyse community diversity. Using PPD, it is possible to evaluate the different analyses using the percentage of overlap across groups as a comparable statistic. The use of relative read frequencies as distance among samples, instead of the summary statistics typical of DMAs, allows to identify which lineages, or groups of lineages, generate the significant differences. Furthermore, the permutation procedure on all sequences prevents any subsampling procedure (i.e. rarefaction), a practice that has been recently criticized because it reduces the resolution power hiding the signal coming from rare OTUs (McMurdie & Holmes 2014).

Recently, Chiu *et al.* (2014) proposed a different formulation for alpha diversity than the one proposed in Jost (2006, 2007). Here, the alpha entropy becomes the joint entropy of observation and environment minus the logarithm of the number of environments. The new formulation allows beta diversity to reach its theoretical maximum (number of environments) whatever unbalanced design is used, but it causes the lower bound of the beta diversity to be higher than 1 depending on sampling design. Furthermore, the alpha diversity does not match anymore the concept of mean diversity within each environment. Consequently, we are not following this new formulation because it is distant from the canonical information theory and further evaluations are needed.

A brief example could illustrate the reason of our preference. Let us assume two communities with the same six equal abundant species, but one locality has 100 observations and the other 1000. According to our definitions, alpha and beta diversities values are, respectively, 6 and 1, while according to Chiu *et al.* (2014) definition corresponding values are 4.07 and 1.475. We think much more logic, and close to the original information theory realm, the values of 6 and 1 that reflect the fact that in each environment we expect to find six species and that the two samples behave as 1. Chiu *et al.* (2014) could oppose that using a very similar example, but with the six species being different in the two environments, the value of beta diversity would be 1.35 for the formulation of this article and 2.00 for theirs. This apparent mismatch could be corrected by realizing that the real maximum beta diversity is given by the exponential of the entropy of the sampling vector that is exactly 1.35.

Partitioning phylogenetic diversity method was here implemented in our newly proposed tool PHYLOH. We tested our analytic software using large trees (i.e. constituted by several thousand tips) with no significant delay. The only problem we observed is related to the number of permutation that is proportional to the number of observations following the rule of $N\log(N)$ permutation for N observations. There are no possibilities to avoid this, and several thousands of observations require hours of computations. However, the service, given by the National Institute for Nuclear Physics (INFN), allows 24 h as maximal running time, sufficient to cope with medium–large data sets. Larger data sets could be handled downloading the software and using it in local computation facilities.

The software is not parallelizable, and replicates cannot be distributed on several CPUs. We do not think that parallelization is needed, but we will work on distributing permutations on several CPUs. A complete tutorial of PHYLOH is available on Biovel web page. Input data used in the present article are available on DRYAD entry doi:10.5061/dryad.j4d15.

© 2014 John Wiley & Sons Ltd

On the whole, we think that PHYLOH will be a relevant innovation to study community diversity, a field where data analysis is complex, but representation of results is even more difficult. Indeed, PHYLOH produces clear vectorial graphical outputs, in html format, that are easy to explore using functions such as text finder and zoom without reducing image quality. PHYLOH shows its performances when large amounts of interlinked data are available. This is the case of the researches involving high-throughput DNA sequencing, such as our pilot study on microbiomes. Furthermore, it can be used whenever a depth analysis of diversity distribution, from microorganisms to macroorganisms, is required.

Acknowledgements

This work was supported by Fondazione Cariplo Grant C91H09000010003 with the project entitled 'Insetti Pronubi: mezzi di connessione e diffusione di specie vegetali rare ed endemiche del parco regionale della Grigna Settentrionale'. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We are thankful to BIOVEL (FP7 project no. 283359) for funding and Stefano Pagnotta (University of Sannio) for early discussion on the use of phylogenetic entropy. A special thank is for all the bee-keepers who gave us the possibility to obtain biological samples from their hives. Finally, we would like to thank the anonymous reviewers for their helpful and stimulating comments.

References

Al-Waili NS, Salom K, Butler G, Al Ghamdi AA (2011) Honey and microbial infections: a review supporting the use of honey for microbial control. *Journal of Medicinal Food*, **14**, 1079–1096.

Annoscia D, Del Piccolo F, Nazzi F (2012) How does the mite *Varroa destructor* kill the honeybee *Apis mellifera*? Alteration of cuticular hydrocarbons and water loss in infested honeybees. *Journal of Insect Physiology*, **58**, 1548–1555.

Antúnez K, Harriet J, Gende L, Maggi M, Eguaras M, Zunino P (2008) Efficacy of natural propolis extract in the control of American Foulbrood. *Veterinary Microbiology*, **131**, 324–331.

Blow N (2008) Metagenomics: exploring unseen communities. *Nature*, **453**, 687–690.

Bragg L, Stone G, Imelfort M, Hugenholtz P, Tyson GW (2012) Fast, accurate error correction of amplicon pyrosequences using Acacia. *Nature Methods*, **9**, 425–426.

Burešová V, Franta Z, Kopáček P (2006) A comparison of *Chryseobacterium indologenes* pathogenicity to the soft tick *Ornithodoros moubata* and hard tick *Ixodes ricinus*. *Journal of Invertebrate Pathology*, **93**, 96–104.

Candela M, Biagi E, Maccaferri S, Turroni S, Brigidi P (2012) Intestinal microbiota is a plastic factor responding to environmental changes. *Trends in Microbiology*, **20**, 385–391.

Caporaso JG, Kuczynski J, Stombaugh J *et al*. (2010) QIIME allows analysis of high throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Cavalli Sforza LL, Edwards AW (1967) Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics*, **19**(3 Pt 1), 233.

Chao A, Chiu CH, Jost L (2010) Phylogenetic diversity measures based on Hill numbers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 3599–3609.

Chiu CH, Jost L, Chao A (2014) Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. *Ecological Monographs*, **84**, 21–44.

Cornman RS, Schatz MC, Johnston JS *et al*. (2010) Genomic survey of the ectoparasitic mite *Varroa destructor*, a major pest of the honey bee *Apis mellifera*. *Bmc Genomics*, **11**, 602.

Dillon RJ, Dillon VM (2004) The gut bacteria of insects: nonpathogenic interactions. *Annual Reviews in Entomology*, **49**, 71–92.

Dimitriu PA, Boyce G, Samarakoon A, Hartmann M, Johnson P, Mohn WW (2013) Temporal stability of the mouse gut microbiota in relation to innate and adaptive immunity. *Environmental Microbiology Reports*, **5**, 200–210.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Evans JD, Schwarz RS (2011) Bees brought to their knees: microbes affecting honey bee health. *Trends in Microbiology*, **19**, 614–620.

Evans GE, Murdoch DR, Anderson TP, Potter HC, George PM, Chambers ST (2003) Contamination of Qiagen DNA extraction kits with Legionella DNA. *Journal of Clinical Microbiology*, **41**, 3452–3453.

Forsgren E (2010) European foulbrood in honey bees. *Journal of Invertebrate Pathology*, **103**, S5S9.

Guzman-Novoa E, Emsen B, Unger P, EspinosaMontaño LG, Petukhova T (2012) Genotypic variability and relationships between mite infestation levels, mite damage, grooming intensity, and removal of *Varroa destructor* mites in selected strains of worker honey bees (*Apis mellifera* L.). *Journal of Invertebrate Pathology*, **110**, 314–320.

Jeffrey AE, Echazarreta CM (1996) Medical uses of honey. *Revista Biomedica*, **7**, 4349.

Jeyaprakash A, Hoy MA, Allsopp MH (2003) Bacterial diversity in worker adults of *Apis mellifera capensis*, *Apis mellifera scutellata* (Insecta: Hymenoptera) assessed using 16S rRNA sequences. *Journal of Invertebrate Pathology*, **84**, 96–103.

Jost L (2006) Entropy and diversity. *Oikos*, **113**, 363–375.

Jost L (2007) Partitioning diversity into independent alpha and beta components. *Ecology*, **88**, 2427–2439.

Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia*, **129**, 271–280.

Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, **23**, 127–128.

Lozupone CA, Hamady M, Kelley ST, Knight R (2007) Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. *Applied and Environmental Microbiology*, **73**, 1576–1585.

Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R (2011) UniFrac: an effective distance metric for microbial community comparison. *The ISME Journal*, **5**, 169.

Lozupone C, Faust K, Raes J *et al*. (2012) Identifying genomic and metabolic features that can underlie early successional and opportunistic lifestyles of human gut symbionts. *Genome Research*, **22**, 1974–1984.

MacKay DJ (2003) *Information Theory, Inference, and Learning Algorithms*, vol. 7. Cambridge University Press, Cambridge.

Marcon E, Hérault B, Baraloto C, Lang G (2012) The decomposition of Shannon's entropy and a confidence interval for beta diversity. *Oikos*, **121**, 516–522.

Martinson VG, Danforth BN, Minckley RL, Rueppell O, Tingek S, Moran NA (2011) A simple and distinctive microbiota associated with honey bees and bumble bees. *Molecular Ecology*, **20**, 619–628.

Martinson VG, Moy J, Moran NA (2012) Establishment of characteristic gut bacteria during development of the honeybee worker. *Applied and Environmental Microbiology*, **78**, 2830–2840.

Mattila HR, Rios D, WalkerSperling VE, Roeselers G, Newton IL (2012) Characterization of the active microbiotas associated with honey bees

© 2014 John Wiley & Sons Ltd

reveals healthier and broader communities when colonies are genetically diverse. *PLoS One*, **7**, e32962.
Mazmanian SK, Liu CH, Tzianabos AO, Kasper DL (2005) An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell*, **122**, 107–118.
McDonald D, Price MN, Goodrich J *et al.* (2011) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal*, **6**, 610–618.
McMurdie PJ, Holmes S (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, **8**, e61217.
McMurdie PJ, Holmes S (2014) Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Computational Biology*, **10**, e1003531.
Meriweather M, Matthews S, Rio R, Baucom RS (2013) A 454 Survey reveals the community composition and core microbiome of the common bed bug (*Cimex lectularius*) across an urban landscape. *PLoS One*, **8**, e61465.
Metzker ML (2009) Sequencing technologies—the next generation. *Nature Reviews Genetics*, **11**, 31–46.
Möckel N, Gisder S, Genersch E (2011) Horizontal transmission of deformed wing virus: pathological consequences in adult bees (*Apis mellifera*) depend on the transmission route. *Journal of General Virology*, **92**, 370–377.
Mohammadi T, Reesink HW, Vandenbroucke-Grauls CM, Savelkoul PH (2005) Removal of contaminating DNA from commercial nucleic acid extraction kit reagents. *Journal of Microbiological Methods*, **61**, 285–288.
Mohr KI, Tebbe CC (2006) Diversity and phylotype consistency of bacteria in the guts of three bee species (Apoidea) at an oilseed rape field. *Environmental Microbiology*, **8**, 258–272.
Moran NA, Hansen AK, Powell JE, Sabree ZL (2012) Distinctive gut microbiota of honey bees assessed using deep sampling from individual worker bees. *PLoS One*, **7**, e36393.
Mouches C, Bové JM, Albisetti J (1984) Pathogenicity of *Spiroplasma apis* and other spiroplasmas for honey-bees in Southwestern France. *Annales de l'Institut Pasteur/Microbiologie*, **135**, 151–155.
Oksanen J, Kindt R, Legendre P, O'Hara B (2007) *Vegan: community ecology package. R package version* 1.8-5. Available from http://cran.r-project.org/ (accessed 07 November 2014).
Orloci L (1967) An agglomerative method for classification of plant communities. *The Journal of Ecology*, **55**, 193–206.
Price MN, Dehal PS, Arkin AP (2010) FASTTREE 2–approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**, e9490.
R Development Core Team (2012) *R: A language and environment for statistical computing*. R Development Core Team, Vienna, Austria. Available from http://www.r-project.org/ (accessed 07 November 2014).
Rinderer TE, Harris JW, Hunt GJ, de Guzman LI (2010) Breeding for resistance to *Varroa destructor* in North America. *Apidologie*, **41**, 409–424.
Rosenkranz P, Aumeier P, Ziegelmann B (2010) Biology and control of *Varroa destructor*. *Journal of Invertebrate Pathology*, **103**, S96–S119.
Sabree ZL, Hansen AK, Moran NA (2012) Independent studies using deep sequencing resolve the same set of core bacterial species dominating gut communities of honey bees. *PLoS One*, **7**, e41250.
Sanchez LM, Wong WR, Riener RM, Schulze CJ, Linington RG (2012) Examining the fish microbiome: vertebrate-derived bacteria as an environmental niche for the discovery of unique marine natural products. *PLoS One*, **7**, e35398.
Tang X, Freitak D, Vogel H *et al.* (2012) Complexity and variability of gut commensal microbiota in polyphagous lepidopteran larvae. *PLoS One*, **7**, e36978.
Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, **73**, 5261–5267.
Watanabe K, Kodama Y, Harayama S (2001) Design and evaluation of PCR primers to amplify bacterial 16S ribosomal DNA fragments used for community fingerprinting. *Journal of Microbiological Methods*, **44**, 253–262.
Webb CO, Ackerly DD, Kembel SW (2008) PHYLOCOM: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, **24**, 2098–2100.
Wickham H (2009) *ggplot2: Elegant graphics for data analysis*. Available from http://cran.r-project.org/ (accessed 07 November 2014).
Wolstencroft K, Haines R, Fellows D *et al.* (2013) The TAVERNA workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud. *Nucleic Acids Research*, **41**, W557.
Zhou J, Wu L, Deng Y et al. (2011) Reproducibility and quantitation of amplicon sequencing-based detection. *The ISME Journal*, **5**, 1303–1313.

A.S., S.V. and M.C. wrote the manuscript. E.M.P., A.G., E.F., A.B. and B.B. implemented the manuscript. E.M.P., E.F. and M.C. designed the project. S.V. designed and implemented the software PHYLOH. A.S., S.V. and B.B. evaluated the codes. A.S., A.G., E.F., A.B., V.M. and M.C. provided biological background.

Data Accessibility

DNA sequences: NCBI SRA: SRP046312.

Source code of PHYLOH, user manual and example data set are available on github: https://github.com/svicario/phyloH.

Access to the web page of PHYLOH: https://www.biodiversitycatalogue.org/rest_methods/143.

Input and output data of PHYLOH, Supporting information Figures and Tables: DRYAD entry doi:10.5061/dryad.j4d15.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 The figure depicts how Shannon entropy is totally blind to some meaningful biological pattern.

Fig. S2 Distribution of OTUs of High Frequency Cluster (HFC) between parasitised honey bee and varroa considering relative abundance.

Fig. S3 Distribution of OTUs of High Frequency Cluster (HFC) between healthy and parasitised honey bee and varroa.

Fig. S4 UPMGA trees showing the different relationships among samples for all the distance metrics considered in DMA (*jaccard*, *squared-chord*, *unweighted UniFrac* and *weighted UniFrac*), considering Localities and Status variables.

Fig. S5 Output of the script *compute_core_microbiome.py*.

Table S1 Summary of all OTUs taxon assignment.

Table S2 Summary of results of adonis statistic.

Appendix S1 PHYLOH output in html format for 'Status' variable.

Appendix S2 Phylogenetic tree used in PHYLOH analysis in Newick tree format.

© 2014 John Wiley & Sons Ltd

Food Research International 69 (2015) 424–433

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

Review

Emerging DNA-based technologies to characterize food ecosystems

Andrea Galimberti [a,*], Antonia Bruno [a], Valerio Mezzasalma [a], Fabrizio De Mattia [b], Ilaria Bruni [a], Massimo Labra [a]

[a] ZooPlantLab®, Department of Biotechnology and Biosciences, University of Milano-Bicocca, P.za Della Scienza 2, 20126 Milan, Italy
[b] FEM2 Ambiente s.r.l., P.za Della Scienza 2, 20126 Milan, Italy

ARTICLE INFO

Article history:
Received 15 December 2014
Accepted 24 January 2015
Available online 31 January 2015

Keywords:
Biotransformation
DNA barcoding
Food traceability
Food microbiota
High Throughput Sequencing
HTS

ABSTRACT

Food safety and quality depend on raw material characteristics and on the chemical, physical and biotechnological processes adopted during food transformation. Since a huge number of microorganisms are involved in food production, foodstuffs should be considered as complex matrices where any microbial component has a precise role and evolves in response to physical and chemical composition of food. Moreover, knowing the dynamics of microbial community involved in a food supply chain it is useful to reduce food spoilage, enhance the industrial processes and extend product shelf-life. In a more comprehensive vision, a precise understanding of the metabolic activity of microorganisms can be used to drive biotransformation steps towards the improvement of quality and nutritional value of food. High Throughput Sequencing (HTS) technologies are nowadays an emerging and widely adopted tool for microbial characterization of food matrices. Differently from traditional culture-dependent approaches, HTS allows the analysis of genomic regions of the whole biotic panel inhabiting and constituting food ecosystems. Our intent is to provide an up-to-date review of the principal fields of application of HTS in food studies. In particular, we devoted major attention to the analysis of food microbiota and to the applied implications deriving from its characterization in the principal food categories to improve biotransformation processes.

© 2015 Elsevier Ltd. All rights reserved.

Contents

1. DNA barcoding to characterize food raw material and derived products

Along the food supply chain, characteristics of raw materials strongly influence the quality of the final food products. This is a postulate of traditional and modern food-related disciplines. In this perspective, the selection of high-quality vegetables, meat or fish and the availability of suitable tools for their traceability represented so far the main goals of food producers (Aung & Chang, 2014; Imazio et al., 2002; Opara & Mazaud, 2001). The demand for reliable traceability systems is indeed essential to authenticate the geographical provenance of food (also in the case of protected designation of origin products, PDO), and to prevent commercial frauds and adulteration cases. Such emerging topics addressed the scientific research, hence producing different analytical approaches to the problem (El Sheikha et al., 2009; Mafra, Ferreira, & Oliveira, 2008; Myers, 2011).

The validation of food authenticity relies mostly on the analysis of chemical compounds, proteins and/or DNA sequences. While being

* Corresponding author. Tel.: +39 02 6448 3472; fax: +39 02 6448 3450.
E-mail addresses: andrea.galimberti@unimib.it (A. Galimberti), a.bruno2@campus.unimib.it (A. Bruno), valerio.mezzasalma@unimib.it (V. Mezzasalma), fabrizio.demattia@fem2ambiente.com (F. De Mattia), ilaria.bruni@unimib.it (I. Bruni), massimo.labra@unimib.it (M. Labra).

http://dx.doi.org/10.1016/j.foodres.2015.01.017
0963-9969/© 2015 Elsevier Ltd. All rights reserved.

Table 1
List of references concerning the DNA barcoding characterization of raw materials or processed food products.

Foodstuff category	Raw material/food product	References
Fruit	Mango	Hidayat, Kusumawaty, and Pancoro (2013)
	Citrus species	Yu, Yan, Lu, and Zhou (2011)
	Goji	Xin et al. (2013)
	Berries	Jaakola et al. (2010)
	Pineapple	Hidayat, Abdullah, Kuppusamy, Samad, and Wagiran (2012)
	Olives and olive oil	Agrimonti, Vietina, Pafundo, and Marmiroli, 2011 and Ganopoulos et al. (2013)
	Cocoa	Kane et al. (2012)
	Dates	Enan and Ahamed (2014)
Vegetables	*Capsicum* cultivars	Jarret (2008)
	Legume seeds	Ganopoulos et al. (2012) and Madesis, Ganopoulos, Anagnostis, and Tsaftaris (2012)
	Soybean and other crops	Kim et al. (2014)
Aromatic plants	Fresh and processed spices	De Mattia et al. (2011), Federici et al. (2013), Gismondi, Fanali, Labarga, Caiola, and Canini (2013), Kojoma et al. (2002), Parvathy et al. (2014), Theodoridis et al. (2012) and Wang et al. (2013)
Herbal infusions	Tea	Stoeckle et al. (2011)
	Plant-based beverages	Li et al. (2012)
Mushrooms	Wild and cultivated mushrooms	Dentinger, Didukh, and Moncalvo (2011), Khaund and Joshi (2014) and Raja, Baker, Little, and Oberlies (2014)
Honey	Honey	Bruni et al. (2015) and Valentini, Miquel, and Taberlet (2010)
Jams	Fruit jams	Arleo et al. (2012)
Medicinal plants	Medicinal plants	Pansa et al. (2011) and Zuo et al. (2011)
Seafood	Various fishes	Ardura, Linde, Moreira, and Garcia-Vazquez (2010), Ardura, Planes, and Garcia-Vazquez (2013), Carvalho et al. (2015), Galal-Khallaf, Ardura, Mohammed-Geba, Borrell, and Garcia-Vazquez (2014) and Lamendin, Miller, and Ward (2015).
	Tuna and other scombrid species	Abdullah and Rehbein (2014) and Botti and Giuffra (2010)
	Smoked fish products	Smith, McVeagh, and Steinke (2008)
	Crab meat products	Haye, Segovia, Vera, Gallardo, and Gallardo-Escárate (2012)
	Philippine fish products	Maralit et al. (2013)
Meat	Bovidae species	Cai et al. (2011)
	Bovine, ovine, caprine meat	Saderi, Saderi, and Rahimi (2013)
	Game meat	D'Amato, Alechine, Cloete, Davison, and Corach (2013)
Dairy products	Milk source	Gonçalves, Pereira, Amorim, and van Asch (2012) and Guerreiro, Fernandes, and Bardsley (2012)
	Plant traces in milk	Ponzoni et al. (2009)

effective in testing fresh products, chemical and protein-based approaches can be biased by the strong food manufacturing processes, the limited number of detectable isozymes, or the high tissue and developmental stage specificity of the markers. DNA markers are more informative than protein or chemical based methods because DNA better resists physical and chemical industrial processes (Madesis, Ganopoulos, Sakaridis, Argiriou, & Tsaftaris, 2014). DNA is also detectable in the presence of small traces of organic material therefore permitting the detection of low-concentration biological adulterants. As a consequence, DNA markers and in particular PCR-based methods have rapidly become the most used tools in the field of food control. Among these, discontinuous molecular markers such as RAPDs, AFLPs, and their variants (e.g., ISSR, SSAP) as well as sequencing-based systems such as SNPs and SSRs have been successfully adopted for the characterization of food raw materials. However, being highly species specific, these approaches require access to the correct DNA sequence of the organisms and their application is often limited to a single species. In the last decade, DNA barcoding, a standardized method providing species identification through the analysis of the variability in a short DNA gene region — the "barcode", was proposed as a universal DNA-based tool for species identification (Hebert, Ratnasingham, & deWaard, 2003).

Recently, Galimberti et al. (2013) reviewed the usefulness of DNA barcoding to certify food identity by tracking origin and provenance of raw materials at different levels of their transformation. DNA barcoding permits to discriminate biological entities analyzing the variability in a single or in a few standard molecular marker(s) (Casiraghi, Labra, Ferri, Galimberti, & De Mattia, 2010). In this context, DNA sequence(s) identify different food products in the same way that a supermarket scanner uses the black stripes of the UPC barcode to identify any purchase. The application of this tool opened new opportunities to track not only common crops and breeds, but also those minor crops and local products still lacking a reference genetic fingerprint (Galimberti et al., 2014). As an example, DNA barcoding was extensively applied in the last decade to verify the origin of seafood (Becker, Hanner, & Steinke, 2011) and to exclude commercial frauds occurring in its production and distribution (Barbuto et al., 2010; Carvalho, Palhares, Drummond, & Frigo, 2015; Cutarelli et al., 2014). The success of seafood molecular identification allowed the US Food and Drug Administration to propose DNA barcoding as a routine approach for the authentication of fish-based commercial products (Yancy et al., 2008). Both consumers and foodstuff producers may take advantage of a DNA barcoding screening, especially concerning items distributed as shredded or powdered material, which otherwise results as unidentifiable by a simple morphological analysis (Cornara et al., 2013). Among these, promising results were obtained in studies on commercial spices (De Mattia et al., 2011), herbal teas (Li et al., 2012) and fruit juices (Faria, Magalhães, Nunes, & Oliveira, 2013). Table 1 provides an updated list of references on identification and traceability of raw materials/processed foodstuffs by using DNA barcoding. Analysis of the case studies provided in Table 1 suggests that DNA barcoding is a sensitive, fast and cheap approach, able to identify and track a wide panel of raw materials and deriving food commodities. The cost and time-effectiveness of DNA barcoding and the recent development of innovative sequencing technologies allow a certain degree of automation in species identification, which is particularly useful in simultaneous monitoring activities of multiple foodstuffs and batches.

Moreover, works listed in Table 1 highlight the principal advantages of using DNA barcoding for both producers and consumers. The firsts can value their products by certifying composition and provenance of raw materials and can have access to a sort of a universal certification system (a pivotal requisite as we are in the era of globalization). On the other hand, consumers can defend themselves against frauds and species substitution cases, as well as knowing the full composition of foodstuffs. This growing awareness is useful in mitigating the health impact of allergenic reactions, intolerances and other outbreaks, as also outlined by international regulations (e.g., the recently adopted Reg. (EU) No 1169/2011; EC No, 1169/2011, 2011).

International agencies or institutions, which are responsible for quality control of raw materials or food commodities, can cooperate

by exchanging their data, hence creating reference databases, the lack of which is the main limit of the method. In fact, whereas some groups of organisms (e.g., fish) are well represented, a lot of work is required to provide reference DNA barcoding data for poorly investigated taxonomic groups (e.g., minor crops).

As a diagnostic tool, the DNA barcoding approach can be more or less fallacious, and it should be taken into account that failures are mainly in the essence of biological species rather than in the method (Casiraghi et al., 2010). DNA barcoding performance is strongly influenced by the molecular variability of the organisms. As an example, the method cannot to date be easily applied to the differentiation of GM (Genetically Modified) food raw material, breeds and cultivars, basing on the standard molecular markers. The modified genomic tracts usually do not involve the plastidial or nuclear regions analyzed in a classical DNA barcoding approach. However, given the increasing demand for a fast and reliable traceability system for these kinds of products, a panel of additional markers (i.e., promoters, reporter genes) could be applied in combination with classical DNA barcodes. As an example, very recent studies showed the potential of High-Resolution-Melting (HRM) analyses when coupled to the investigation of DNA barcoding markers (barHRM) to differentiate cultivars and closely related species and to authenticate Protected Designation of Origin (PDO) of some food products (Druml & Cichna-Markl, 2014; Ganopoulos, Bazakos, Madesis, Kalaitzis, & Tsaftaris, 2013; Ganopoulos, Madesis, Darzentas, Argiriou, & Tsaftaris, 2012; Jaakola, Suokas, & Häggman, 2010).

2. The complex ecosystem of food biotransformation processes

Food quality does not rely on raw material characteristics only, but also on manufacturing and biotransformation processes involved during their conversion into final food products. Since time immemorial, biotechnological procedures are involved in food production. These take advantage of environmental microorganisms such as bacteria and yeasts and of their metabolisms, transforming raw materials into enriched foodstuffs. Well-known examples refer to the production of wine, beer and other alcoholics, where biotransformation increases their organoleptic properties and extends their shelf-life; yogurt and dairy products, where microorganisms transform milk into products exhibiting peculiar sensory and functional (e.g., probiotics) characteristics; bread and other bakery products obtained by the fermentation activity of selected yeasts. Pools of microorganisms can modify chemical and physical features of raw materials to get new metabolites and materials and therefore influencing sensory, safety and nutritional properties of the final transformed food products (Bull, Plummer, Marchesi, & Mahenthiralingam, 2013; Caplice & Fitzgerald, 1999).

Generally, in food industries physical and chemical modifications of raw materials are well calibrated at any step of the production chain to preserve the organoleptic properties of the final product (De Filippis, La Storia, Stellato, Gatti, & Ercolini, 2014; Doyle & Buchanan, 2013). However, the calibration of biotransformation procedures is even more difficult.

Before discussing the complexity of microbial ecosystems it is necessary to describe the three main categories of food biotransformation processes: fermentation, biopreservation, and functionalisation.

The fermentation process consists in the oxidation of carbohydrates to obtain the major end products such as alcohol and carbon dioxide, as well as vitamins and secondary metabolites, thanks to the metabolic pathways of microorganisms (Ray & Daeschel, 1992). In the last 20 years, due to the continuous discoveries in biotechnology and genetic engineering, fermentation has definitively moved to industrialized and life-science driven technology Waites, Morgan, Rockey, and Higton (2009). Nowadays, there is an astonishing variety of fermented foods covering a broad range of food substrates (e.g., plants, milk, and many others). Considering that fermented foods constitute 1/3 of the human diet Campbell-Platt (1994) and due to the importance of this process in many industrial compartments, the next chapters of this review will focus on case studies and novel techniques to explore microbial ecosystems involved in this biotransformation process.

Concerning biopreservation, most food and beverages require treatments that prolong their shelf-life, in order to maintain an acceptable level of quality and safety from manufacturing to consumption. Modern food preservation approaches are based on the use of microorganisms producing antimicrobial compounds (i.e., organic acids, ethanol, hydrogen peroxide and bacteriocins) that are able to inhibit or contrast food spoilage (Ross, Morgan, & Hill, 2002). For example, a considerable number of starter strains used mainly in fermented foods derives from the activity of lactic acid bacteria (LAB). LAB are able to produce antimicrobial metabolites such as lactic acid, acetic acid and other organic acids therefore determining a low pH environment that prevents the growth of several pathogenic and spoilage microorganisms (Cizeikiene, Juodeikiene, Paskevicius, & Bartkiene, 2013; Crowley, Mahony, & van Sinderen, 2013). Nowadays, more than 170 bacteriocins have been described and are used for food preservation purposes (Hammami, Zouhir, Le Lay, Hamida, & Fliss, 2010). The last frontier of biopreservation is the use of microbial antagonistic molecules to functionalize food packages (Appendini & Hotchkiss, 2002). Active packaging systems include natural antimicrobials as additives, among which is nisin, one of the most studied and commercialized bacteriocins. As an example, bacteriocins applied to food packaging materials were found to inhibit *Listeria monocytogenes* on meat products Gálvez, Abriouel, López, & Omar, 2007). The exploitation of such natural biopreservation strategies holds great potentials, especially in the last years, as the awareness of the consumer towards the so-called “green technologies” (i.e., minimally processed foods, free from chemical and harmful preservatives) is growing and growing.

Functionalization is the production of new metabolites or functions mediated by microorganisms which can be delivered to the consumer through diet. These kinds of food, known as functional foods or nutraceuticals (Shah, 2007), share three basic characteristics: they derive from naturally occurring ingredients; they have to be consumed as a part of daily diet and they have significant benefits to human health. The most common functional foods can be grouped into three categories: probiotics, prebiotics and synbiotics (Pfeiler & Klaenhammer, 2013). A probiotic is a live microorganism that confers a health benefit on the host when administered in adequate amounts. Prebiotics are non-digestible food ingredients that stimulate growth and/or activity of other bacteria, with positive effects on the host's health. When both prebiotics and probiotics are present in the same food product, those functional foods are referred to as synbiotics. As a direct consequence of this new nutritional trend, a wide panel of functional foods became suitable for large-scale industrial production (Stanton, Ross, Fitzgerald, & Sinderen, 2005). A great number of genera of bacteria are used as probiotics, but the main species showing probiotic characteristics are *Lactobacillus acidophilus*, *Bifidobacterium* spp., and *Lactobacillus casei* (Bull et al., 2013). Yeasts also play an important role as probiotics, with *Saccharomyces boulardii* as the most known probiotic fungus which has been successfully used for curing intestinal diseases (Czerucka, Piche, & Rampal, 2007). Several applications of probiotics and/or prebiotics have been studied: from the enhancement of immune response to positive effects in contrasting allergies and even AIDS or other pathologies.

Fermentation, biopreservation, and functionalization processes involve microorganism communities, sensitive to different environmental parameters (Bokulich, Thorngate, Richardson, & Mills, 2014; Minervini, De Angelis, Di Cagno, & Gobbetti, 2014). Moreover, community structure and relationships among different bacteria, yeasts and other microorganisms undergo substantial changes during biotransformation. Thus, only an exhaustive evaluation of microbial community structure and of its dynamics during food production could help optimize industrial transformation steps in order to get high-quality products.

Except for traditionally biotransformed foods and beverages, an astounding number of edible products, including the emerging ‘functional

foods', involves the activity of microorganisms during at least one step of their industrial production. Thus, several microorganisms gained an important role in human food production and this trend rapidly increased with the advances and industrialization of modern food manufacturing procedures (Betoret, Betoret, Vidal, & Fito, 2011; Roberfroid, 2000). For this reason, at the industrial level, biotransformation steps could be partially controlled by using selected microorganisms as reaction starters. For example, in the case of winemaking, selected *Saccharomyces cerevisiae* strains are used for activating the alcoholic fermentation of must. However, other microorganisms naturally inhabiting raw materials or the surrounding environment, could also be involved during food transformation. Again, in the case of wine, the wine cellar yeasts and bacteria could actively contribute to the chemical modification of grape juice to obtain wine with specific organoleptic properties (Bokulich, Ohta, Richardson, & Mills, 2013; David et al., 2014). Environmental microorganisms represent an important source of biodiversity to differentiate a certain food product from the others, even at a reduced spatial production scale (Quigley et al., 2012; Riquelme et al., 2015). For these reasons, modern food companies should not underestimate the importance of knowing the composition of a microbial community accompanying food from farms to consumer's fork or glass.

Moreover, during food production, undesirable microorganisms could also enter into the food supply chain (Bondi, Messi, Halami, Papadopoulou, and de Niederhausern (2014); Newell et al., 2010). External microbial components can reduce the quality of food products (spoilage microorganisms) or even negatively affect their safety (foodborne pathogens) (Doyle & Buchanan, 2013). In these cases, an in-depth analysis of food microbial community is essential to assess the safety of raw materials and related final products (Fusco & Quero, 2014; Solieri, Dakal, & Giudici, 2013).

Given the complex dynamics occurring in food ecosystems, one of the emerging topics of food science is the development of revolutionary analytical systems that are able to characterize the microbial community as well as the DNA barcoding approach is able to characterize raw materials.

Nowadays, the occurrence and abundance of microbes in a given food ecosystem can be evaluated by studying its microbiota (Ercolini, 2013), which refers to the sum of microscopic living beings and their genomes (i.e., the microbiome) in the environment under investigation. In this review, we discuss the potential of modern technological advances in the molecular characterization of food-related microorganisms. Only the combination of high quality raw materials with fine regulated biotransformation processes will lead to the improvement of food nutritional quality.

3. Novel molecular approaches to investigate food ecosystems

Since the advent of disciplines devoted to the study of food, the investigation of microbial ecology has dramatically changed and this process is in constant evolution (Solieri et al., 2013). For a long period, food-associated microorganisms and their dynamics have been studied through culture based-methods (Doyle & Buchanan, 2013). However, these revealed to be often weak to accomplish a complete microbial characterization of many food ecosystems (Ceuppens et al., 2014). Problems and shortcomings of culturing methods basically involve the underestimation of microbial diversity, and even the failure of a precise detection of some species or genera.

Following the advent of molecular biology, a plethora of laboratory techniques have been developed and most of these are now extensively adopted in food control activities (see for example, Ercolini, 2013; Solieri et al., 2013). Molecular approaches permit to identify food-related microorganisms and estimate their relative abundance, providing a fast, accurate and economic detection tool. Most techniques rely on the analysis of genetic DNA markers and become increasingly important in food microbiology. They identify microorganisms rapidly and accurately, complementing or substituting classical methods (Ceuppens et al., 2014; Chakraborty, Doss, Patra, & Bandyopadhyay, 2014).

Denaturing gradient gel electrophoresis (DGGE) is one of the most used fingerprinting techniques in food microbiology. It is based on the separation of polymerase chain reaction (PCR) amplicons of the same size but different sequences. Fragments are separated in a denaturing gradient gel based on their differential denaturation (melting) profile (Ercolini, 2004). In recent years, PCR-DGGE has been largely used to characterize bacteria and yeasts in fermented products (Muyzer, De Waal, & Uitterlinden, 1993; Peres, Barlet, Loiseau, & Montet, 2007) and to define the origin of raw material starting from the characteristics of its yeast or bacterial communities as in the case of fruit (El Sheikha, Bouvet, & Montet, 2011; El Sheikha, Durand, Sarter, Okullo, & Montet, 2012; El Sheikha, Métayer, & Montet, 2011) and fish (El Sheikha & Montet, 2014; Le Nguyen, Ngoc, Dijoux, Loiseau, & Montet, 2008). However, it is not always possible to resolve DGGE fragments when the difference in sequence is not wide enough or when different DNA fragments have identical melting behavior (Ercolini, 2004).

Since advances in technology have always driven discoveries and changes in microorganism taxonomy, taxonomic identification is an issue of primary importance when approaching the study of food microbiota. In this scenario, genomics now underlies a renaissance in food microbiology therefore accelerating food safety monitoring and food production processes (Ceuppens et al., 2014). Considering bacteria, the present taxonomy is still a complex topic for biologists as well as an area of growing interest, because the definition of microbial species as a taxonomic unit lacks a commonly accepted theoretical basis (Felis & Dellaglio, 2007). Microbial taxonomy directly influences a number of basic scientific and applied fields where microorganisms are involved (Tautz, Arctander, Minelli, Thomas, & Vogler, 2003) including food production, conservation and probiotic activity. Depending on the level of investigation required, the taxonomic resolution of microorganisms can vary. For example, the genus rank could be sufficient when monitoring changes in microbial community during a biotransformation or treatment process of food raw material (e.g., fermentation, pasteurization) (Quigley et al., 2012). In contrast, species or strains have to be precisely identified in the case of pathogen detection analyses, or to assess the efficacy of a certain probiotic.

Aiming to differentiate microorganisms at the species level, methods based on DNA sequencing are currently the most adopted. In many cases, when a fast and accurate response is needed, a 'DNA barcoding-like' approach is the most reliable (Chakraborty et al., 2014). Many scientists used 16S rRNA gene as a universal marker for species-level typing of microorganisms (Bokulich, 2012; Claesson et al., 2010; Janda & Abbott, 2007). This genomic region is considered a 'bacterial barcode' due to its peculiar properties (Patel, 2001): it is present in all the bacterial species, it contains sufficient information (1500 bp long) to differentiate species and, in some cases, strains (Muñoz-Quezada et al., 2013) and finally, the 16S rRNA relies upon an impressive archive of reference sequences such as Greengenes (De Santis et al., 2006) and SILVA (Pruesse et al., 2007). Amplicons belonging to whole genomic extraction conducted on the matrices under investigation (e.g., food products) are sequenced and reads are compared to reference databases to identify the Operational Taxonomic Units — OTUs (Sandionigi et al., in press).

Several studies test analytical approaches for the DNA-based detection of emergent food microbial contaminants in a wide panel of food products (see for example Fusco & Quero, 2014; Velusamy, Arshak, Korostynska, Oliwa, & Adley, 2010 and related references). Such techniques allow to detect specific bacteria and strains in different steps of the food supply chain as reported for example in the cases of seafood and meat manufacturing (Amagliani, Brandi, & Schiavano, 2012; Norhana, Poole, Deeth, & Dykes, 2010; Zbrun et al., 2013). In international trade, major food categories are commonly shipped over very long distances and are therefore exposed to various contaminants such as *Salmonella*, *Listeria* and *Campylobacter*. PCR and Real-time PCR

based methods are nowadays routinely used for the detection of these pathogens. Primer combinations also permit the simultaneous identification of a panel of foodborne pathogens in a single reaction (see for example Jofré et al., 2005).

Progresses in sequencing technologies and bioinformatic analysis of data, led nowadays to a more complex scenario of food control activities. Detection approaches targeting one or few microorganisms are not sufficient to have a reliable characterization of quality and safety of foodstuff. Recent technological advances offer a panel of analytical tools able to screen the whole microbial community of food matrices. The use of universal markers produces several DNA barcode fragments, corresponding to the each bacterial species present in a food sample. With the ultimate goal of characterizing the complete spectrum of microorganisms, the traditional Sanger sequencing approach is inadequate to uncover this huge diversity. To date, several novel approaches, referred to as 'Next Generation Sequencing' (NGS) and, more recently, 'High Throughput Sequencing' (HTS), have been developed (Ercolini, 2013; Mayo et al., 2014; Solieri et al., 2013).

HTS techniques are able to provide sequence data around a hundred times faster and cheaper than the conventional Sanger approach. Sequencers from 454 Life Sciences/Roche (producing about a million sequences of 800 to 1000 base length), Solexa/Illumina and Applied Biosystems SOLiD technology (producing over a billion sequences of 50 to 500 base length) were produced as second generation technologies and other competitive instruments appeared on the market such as the Ion Torrent and PacBio. HTS technologies also permit to prepare several DNA samples from different extracts and to mark them with different DNA tags, mixed and processed at the same time. Thanks to these practical advantages, it is possible to analyze in parallel a very high number of samples, and hence lower the analysis cost. The reduction in cost and time for generating DNA sequence data has resulted in a range of new successful applications, including food traceability and especially food microbiology (Madesis et al., 2014).

Table 2 encompasses recent and emblematic case studies concerning the adoption of HTS approaches to study the microbial ecosystem (in terms of diversity and dynamics) of different food categories. In most cases, the obtained results could be of great impact in the food supply chain to improve industrial biotransformation processes, enhance quality of the final products, extend the shelf-life and valuating local productions.

In the following sections, we selected two of the most representative food categories to highlight the role of novel molecular approaches in characterizing food microbial ecosystems. The first category refers to foodstuffs having plant organisms as starting raw material and where HTS analysis was used to characterize the microbiota of some food products from field to table. Similarly, the second section describes emblematic case studies involving dairy products, which are characterized by complex and sometimes unconventional biotransformation processes.

4. Microbiota composition and dynamics in plant fermentation processes

Fermentation is considered one of the oldest biotechnological methods to convert sugars, starches, or other carbohydrates, into alcohol, and organic acids, by microorganisms. Archeologists have found molecular evidence for the production of fermented beverages dated back to 7000 and 5400 BC. In the Neolithic, fermentation ensured vegetable preservation (McGovern, Glusker, Exner, & Voigt, 1996; Ross et al., 2002) and was based on spontaneous microorganisms inhabiting fruits and seeds. Nowadays, many selected strains of microorganisms are used to transform raw materials in foodstuffs having additional nutritional properties. HTS analyses also clarified the key role of spontaneous microorganisms in biotransformation processes (Table 2). The equilibrium among spontaneous and commercial microorganisms during fermentation depends on many factors, including the microbial biodiversity present in the food and the environmental conditions occurring during

Table 2
Case studies concerning the use of emerging DNA-based technologies to characterize food microbiota. Potential implications for the food supply chain are reported for each food category.

Raw material/food category	Aims	Implications for the food supply chain	References
Grapevine	Study of bacterial consortia inhabiting grapevine surfaces	Valuing cultivars and wine production at the regional scale	Bokulich et al. (2014)
Must and Wine	Study of microbial community of must and its dynamics during alcoholic fermentation	Improvement of wine fermentation processes	Bokulich, Bamforth, and Mills (2012), Bokulich et al. (2013) and David et al. (2014)
Beer	Study of microbial community involved during beer production	Improvement of brewery at both artisanal and industrial scale	Bokulich, Bamforth, and Mills (2012) and Jung, Nam, Roh, and Bae (2012)
Soybean, rice and vegetables	Study of microbial community of fermented products	Quality improvement of final foodstuffs. Valorization of production by enhancing sensorial characteristics of local and commercial products	Jung et al. (2011), Kim et al. (2011), Nam, Lee, & Lim (2012), Park et al. (2012) and Sakamoto, Tanaka, Sonomoto, and Nakayama (2011)
Olives	Study of olive fermentation dynamics and bacterial biodiversity	Improvement of the sensory quality of table olives	Cocolin et al. (2013)
Raw milk	Assessing the effects of cattle's diet on milk quality	Enhance and preserve organoleptic quality and shelf-life of raw milk and dairy products by calibrating cattle diet	Kuehn et al. (2013), Masoud et al. (2012) and Zhang et al. (in press)
Processed Milk	Influence of milk origin and treatments on microbiota	Selection of new strains or strains with novel properties for their use as dairy starters	Delgado et al. (2013), Dobson, O'Sullivan, Cotter, Ross, and Hill (2011) and Leite et al. (2012)
PDO cheeses	Characterization of the microbiota involved in cheese production	Improvement of fermenation processes to obtain high quality cheese	Aldrete-Tapia et al. (2014), Alegría, Szczesny, Mayo, Bardowski, and Kowalczyk (2012), De Filippis et al. (2014), De Pasquale et al. (2014), Lusk et al. (2012), Quigley et al. (2012) and Riquelme et al. (2015)
Seafood	Study of microbial community of fermented seafood	Improvement of fermentation and conservation processes	Koyanagi et al. (2011) and Roh et al. (2009)
Seafood	Investigating relationships between seafood microbiota and products' shelf-life	Shelf-life extension of seafood products	Broekaert, Heyndrickx, Herman, Devlieghere, and Vlaemynck (2013), Chaillou et al. (2014), Kim et al. (2014) and Koyanagi et al. (2011)
Meat	Characterization of microbial communities and dynamics associated to meat products	Improvement of organoleptic characteristics and quality of typical products.	Chaillou et al. (2014), Nieminen et al. (2012), Połka, Rebecchi, Pisacane, Morelli, and Puglisi (2015)

biotransformation. An HTS approach allows studying the evolution of food microbiota in time and in response to different parameters such as temperature, pH, substrate chemical composition and others. For example, David et al. (2014) mapped microbial population dynamics in wine musts (organic and conventional) and showed substantial changes during each biotransformation phase in response to must characteristics. These data could be used by winemakers to drive fermentation processes and to set up the most suitable environmental conditions to enhance wine characteristics (Bokulich, Joseph, Allen, Benson, & Mills, 2012). Similar analyses were conducted for brewing. Data suggested that beer is characterized by consistent modification in microbial activity at every stage, from raw material production and malting to stability in the package. Again, the HTS approach allowed to evaluate this diversity and to exclude the presence of undesirable bacteria (Vriesekoop, Krahl, Hucker, & Menz, 2012).

In table olive fermentation, HTS techniques were used to evaluate the impact of NaOH treatment (Cocolin et al., 2013). No treated olives were characterized by the presence of halophilic bacteria, which were substituted by *Lactobacillus* at the later stages of the fermentation, whereas Enterobacteria were dominant when the olives were treated with sodium hydroxide. Higher biodiversity was found for *Lactobacillus plantarum* isolated during untreated fermentation: different biotypes were found on the olive surface and in the brines. When the debittering process was carried out, a decrease in the number of *L. plantarum* biotypes was observed and those originating from the surface of the olive did not differ from those occurring in the brines. These changes in microbiota structure could lead to a modification of the sensory quality of olives.

In plant products, the microbial community of the cultivation area could also influence the quality and nutritional value of the final food products. Using HTS analyses, Bokulich et al. (2013), identified the "wine microbial *terroir*" and elucidated the relationship between production region, climate, and microbial patterns. This information may help to enhance biological control of vineyard, improving the wine supply and to enhance the economic value of important agricultural commodities, as also suggested by Baldan et al. (2014).

Microbiome analysis could also be used to evaluate and enhance the nutritional value of food products. For example, analysis performed on different commercial brands and local productions of *doenjang*, a traditional fermented soybean product, revealed consistent differences in microbial community structure (see Table 2 for references). Such differences largely influence the flavor and nutritional properties of *doenjang* (Nam, Park, & Lim, 2012). Commercial brands contain simple microbial communities dominated by *Tetragenococcus* and *Staphylococcus* that homogenize the taste and composition of the product. In contrast, local products showed conspicuous variability in microbial populations, providing products of completely different fermentations.

The analysis of spontaneous microbiota associated with original raw materials and the evaluation of antimicrobial components is another important element to drive biotransformation processes. For example, the consistent demands of new flours from cereals and other crops lead to the test of different mixtures. Chestnut flour was considered one of the most interesting raw materials due to its content of proteins with essential amino acids (4–7%), mineral salts and vitamins; however, the occurrence of phenolic compounds with antimicrobial activity prevents the use of this raw material for the fermented products De Vasconcelos, Bennett, Rosa, and Ferreira-Cardoso (2010). The combination of chestnut flour with wheat (Dall'Asta et al., 2013), rice (Demirkesen, Mert, Sumnu, & Sahin, 2010) and rye flours could reduce the chestnut antibacterial components. A mix of raw materials resulted in a mix of microbiota that can contribute to improve the efficacy of biotransformation (Aponte et al., 2014).

Finally, the modern molecular approaches to study microbial ecosystems of plant-derived foods could also reduce food spoilage occurrence due to undesirable microorganisms. In general, food alteration derives from contamination mediated by specific microorganisms, but sometimes several pathogens can simultaneously contaminate a food matrix (Fusco & Quero, 2014). For example, brewing could be negatively affected by different classes of bacteria such as lactic acid bacteria, acetic acid bacteria, Enterobacteriaceae and *Zymomonas* (Vriesekoop et al., 2012) that can coexist. In these cases, HTS analysis is reliable for identifying any undesirable microorganism and could be used to enhance food sanitation and preservation measures.

5. The evolution of microbial community in artisanal and industrial dairy production

Dairy products are the result of a long history and local traditions (Cordain et al., 2005) that led nowadays to the recognition of hundreds of Protected Designation of Origin (PDO) products. Such brand refers to peculiarities in their flavor, consistency and methods of production that are characteristic of a certain geographical site and increase their market value. Due to the economic relevance, health and social issues related to this category of foodstuff, many DNA-based techniques are currently available to assess authenticity and adulteration of milk-derived food (Mafra et al., 2008). Among the applications of these molecular tools, there is the possibility of detecting the adulteration of higher value milk by nondeclared cow's milk (Galimberti et al., 2013) and even to detect traces of feed-derived plant DNA fragments in raw milk and in its fractions (Ponzoni, Mastromauro, Gianì, & Breviario, 2009). In contrast, the characterization of their microbial component is much more difficult. Microbial dynamics occurring within major ingredients involved in the manufacturing of typical cheeses (i.e., milk, rennet, salt) shape the production of the different varieties and can contribute to aroma and taste defects. As a result, the microbiota of different cheeses varies considerably depending on the type of fermentation adopted (Quigley et al., 2012). Due to the complexity of biotransformation processes, diversity, not only at the species level but also at the strain one is pivotal for industrial purposes. This aspect requires the availability of reliable methods for strain discrimination and monitoring (De Filippis et al., 2014). Indeed, a deep knowledge of raw material indigenous microbiota could permit proper selection and dosage of a starter culture to enhance the transformation steps and increase sensorial properties of the final product (see Table 2 for examples).

Microbial populations in cheese can be split into two distinct groups i.e., starter and non-starter microorganisms. Homofermentative lactic acid bacteria (LAB) are the dominant and most important component of the microbiota of fermented milk products as they act as starter cultures, causing rapid acidification via the production of lactic acid. In some fermented dairy products, additional yeasts, molds, as well as bacteria such as non-starter lactic acid bacteria (NSLAB), are involved in the production of flavor compounds or carbon dioxide (De Pasquale, Di Cagno, Buchin, De Angelis, & Gobbetti, 2014; Fox, Guinee, Cogan, & McSweeney, 2000; Quigley et al., 2012). However, they can also be associated with the occurrence of defects. The relative importance of the starter culture and other added microorganisms varies from product to product (Johnson & Steele, 2013), as well as the microbial composition in different parts of a ripened product (e.g., internal part, rind). A precise control of microbial strains and their proportions is fundamental to minimize cheese defects and enhance its quality (O'Sullivan, Giblin, McSweeney, Sheehan, & Cotter, 2013).

The basic goal of characterizing microbial diversity and community dynamics in relation to dairy microbiology is to understand the relationships between microorganisms and their impact on food sensorial properties and safety (Solieri et al., 2013). The modern molecular approach to study microbiota composition can contribute to clarify the role of raw milk quality and added ingredients in dairy transformation processes. Many studies showed how cheese microbiota structure can vary according to the animal origin of the milk (Coppola, Blaiotta, Ercolini, & Moschetti, 2001; Quigley et al., 2012), its preliminary treatments (e.g., pasteurization, Delgado et al., 2013) and additional ingredients used during production (Ercolini, De Filippis, La Storia, & Iacono,

2012). In a survey based on HTS analyses conducted on the microbiota of 62 artisanal Irish cheeses, Quigley et al. (2012) provided evidence for a different microbial richness (in terms of genera of bacteria) in milk of different sources, with a maximum (i.e., 21 genera) for cow milk cheeses and a minimum (i.e., 2 genera) for sheep milk cheeses. They also highlighted, in some cheeses, a negative effect of salt content on the presence of certain genera (e.g., *Leuconostoc* and *Pseudomonas*) as well as a different microbial community structure when herbs and species were involved during cheese manufacturing.

In 2012, Ercolini et al., demonstrated the importance of the microbiota of natural whey culture (NWC) added to raw milk to drive fermentation processes and shaping the final bacterial community of water buffalo mozzarella, a highly appreciated Italian nonripened cheese. Although completely different production technologies are employed, some products such as Grana Padano, Parmigiano Reggiano and other PDO cheeses share the use of the NWC as starter for curd acidification. Studies on their microbial communities and dynamics revealed by HTS approach (e.g., De Filippis et al., 2014), showed how, starting from similar NWCs, temperature and pH drive selection of a characteristic core microbiota, responsible in the achieving the typical sensory characteristics of each cheese type.

Animal diet was thought to be of primary importance for determining milk composition, microbial structure and quality. Using a 454 pyrosequencing approach, Zhang, Huo, Zhu, and Mao (in press) found that high-concentrate feeding had significant effects on shaping the milk microbial community of dairy cows. This kind of diet resulted in a greater proportion of psychrotrophic bacteria in milk, such as *Pseudomonas*, *Brevundimonas*, *Sphingobacterium*, *Alcaligenes*, *Enterobacter* and *Lactobacillus*. A possible conclusion was that inappropriate cattle feeding may lower the organoleptic quality of raw milk and dairy products, also limiting the shelf-life of processed fluid milk.

HTS analysis of microbiota composition can also give information about the dairy production methods. Generally, traditional manufacturing processes (i.e., artisanal production) are characterized by a complex microbial community. In contrast, industrially obtained foods are characterized by more-simple microbial consortia (De Filippis et al., 2014; Ercolini, 2013).

Several researches also revealed that different cheese-making units within the same broad geographic area share a common core microbiota (see for example De Filippis et al., 2014; Quigley et al., 2012). A precise knowledge of such bacterial consortia may help in transferring certain productions from the artisanal to the industrial level with consequent economical benefits.

However, in dairy production, one of the possible risks occurring in the passage from artisanal to industrial manufacturing could be the loss of flavors and aromas which are characteristic of the product. This goal requires the standardization of the cheese production process, using for example pasteurized milk instead of the raw one. The standardization of fermented dairy manufacturing is not trivial because different products which are similar in appearance can exhibit unique bacterial profiles and unique sensorial properties (Lusk et al., 2012). In a recent study, Aldrete-Tapia, Escobar-Ramírez, Tamplin, and Hernández-Iturriaga (2014) used HTS techniques to establish the denomination of origin for the Mexican artisanal Poro cheese: they provided an insight into the composition and dynamics of bacterial communities present during its production and ripening. Since molecular data determined the relative composition and bacterial species in the artisanal production process of Poro cheese, it could be possible to identify not only the microbial communities but also those bacteria that could be potentially used in starter cultures.

Another emblematic case is that of Pico Cheese, an artisanal dairy cattle product manufactured by few Azorean (Portugal) producers without the addition of starter cultures (Riquelme et al., 2015). Given the ongoing loss of local producers and the necessity to preserve its peculiarity and enhance its marketability even at a semi-industrial scale production, Riquelme et al. (2015) examined in depth the microbiota diversity and dynamics during ripening of Pico Cheese. Researchers characterized the core bacterial components (*Lactococcus*, *Streptococcus* and some unclassified Enterobacteriaceae) of artisanal Pico cheese microbiota, a first step to recreate certain conditions for a potential industrial production.

The microbiota of the processing environment also influences the microbial community and its succession of fermented dairy products. During manufacturing, raw milk and its fermented intermediates, encounter many different surfaces, all acting as potential vectors for microbes. HTS analyses conducted by Bokulich and Mills (2013) on two artisanal cheesemaking plants revealed that similar communities of microbes occupied the same surface types, reflecting the selection for distinct communities on the basis of the production stage. Such a situation may play an important role in populating cheese microbial communities, beneficially directing the course of sequential fermentation and the quality of the final products (see for example the cases of water buffalo mozzarella and other artisanal cheeses: Aldrete-Tapia et al., 2014; Mauriello, Moio, Genovese, & Ercolini, 2003; Randazzo, Pitino, Ribbera, & Caggia, 2010). Interestingly, De Filippis et al. (2014), in a study on three highly-appreciated PDO Italian cheeses, found many subdominant OTUs of environmental provenance, probably arising from soil and agricultural environment and established into the final product.

The spatial distribution of microbes in foods is also a very interesting issue. It was demonstrated that structurally complex foodstuffs can host a different microbiota within their parts, such as the crust, veins, and core in a blue cheese (Ercolini, 2013). The use of HTS technologies is successful in assessing the location of different microbes across food matrices (Gkatzionis, Yunita, Linforth, Dickinson, & Dodd, 2014) and this information can have important consequences in understanding and enhancing the ripening and flavoring processes of high-value products.

6. Conclusions

High Throughput Sequencing technologies are nowadays an emerging and widely adopted tool for microbial characterization of a huge number of matrices and ecosystems, among which foodstuffs. In the field of food quality and safety assessment, the vast majority of published studies focus on fermented beverages and dairy products, in spite of their relevance and economic value in the global market. Other food categories such as meat and seafood are widely distributed worldwide but many aspects of their microbial ecology are largely unknown. In recent years, thanks to the growing accessibility of modern analytical technologies (i.e., HTS), the first studies on these apparently less complex food matrices are emerging.

In contrast to environmental microbiology, few studies have been conducted to identify the metabolic pathways and active compounds involved during the main food transformation processes. A more detailed knowledge on the role of different microorganisms in food would help in enhancing production processes, reducing wastes and extending product shelf-life. In this context, recent advances in 'omic' can have great relevance in food science. In the very next-future an effective integration among different sources of biological information is desirable in order to better understand and manipulate flavor formation, taste and the nutritional quality of foodstuff.

References

Abdullah, A., & Rehbein, H. (2014). Authentication of raw and processed tuna from Indonesian markets using DNA barcoding, nuclear gene and character-based approach. *European Food Research and Technology*, *239*, 695–706.

Agrimonti, C., Vietina, M., Pafundo, S., & Marmiroli, N. (2011). The use of food genomics to ensure the traceability of olive oil. *Trends in Food Science & Technology*, *22*, 237–244.

Aldrete-Tapia, A., Escobar-Ramírez, M.C., Tamplin, M.L., & Hernández-Iturriaga, M. (2014). High-throughput sequencing of microbial communities in Poro cheese, an artisanal Mexican cheese. *Food Microbiology*, *44*, 136–141.

Alegría, Á., Szczesny, P., Mayo, B., Bardowski, J., & Kowalczyk, M. (2012). Biodiversity in Oscypek, a traditional Polish cheese, determined by culture-dependent and -independent approaches. *Applied and Environmental Microbiology*, *78*, 1890–1898.

Amagliani, G., Brandi, G., & Schiavano, G.F. (2012). Incidence and role of *Salmonella* in seafood safety. *Food Research International*, *45*, 780–788.

Aponte, M., Boscaino, F., Sorrentino, A., Coppola, R., Masi, P., & Romano, A. (2014). Effects of fermentation and rye flour on microstructure and volatile compounds of chestnut flour based sourdoughs. *LWT—Food Science and Technology*, *58*, 387–395.

Appendini, P., & Hotchkiss, J.H. (2002). Review of antimicrobial food packaging. *Innovative Food Science & Emerging Technologies*, *3*, 113–126.

Ardura, A., Linde, A.R., Moreira, J.C., & Garcia-Vazquez, E. (2010). DNA barcoding for conservation and management of Amazonian commercial fish. *Biological Conservation*, *143*, 1438–1443.

Ardura, A., Planes, S., & Garcia-Vazquez, E. (2013). Applications of DNA barcoding to fish landings: Authentication and diversity assessment. *ZooKeys*, *365*, 49–65.

Arleo, M., Ruibal, F., Pereyra, J., Miquel, E., Fernàndez, M., & Martínez, C. (2012). A DNA-based approach to discriminate between quince and apple in quince jams. *International Food Research Journal*, *19*, 1471–1477.

Aung, M.M., & Chang, Y.S. (2014). Traceability in a food supply chain: Safety and quality perspectives. *Food Control*, *39*, 172–184.

Baldan, E., Nigris, S., Populin, F., Zottini, M., Squartini, A., & Baldan, B. (2014). Identification of culturable bacterial endophyte community isolated from tissues of *Vitis vinifera* 'Glera'. *Plant Biosystems*, *148*, 508–516.

Barbuto, M., Galimberti, A., Ferri, E., Labra, M., Malandra, R., Galli, P., et al. (2010). DNA barcoding reveals fraudulent substitutions in shark seafood products: The Italian case of "palombo"(*Mustelus* spp.). *Food Research International*, *43*, 376–381.

Becker, S., Hanner, R., & Steinke, D. (2011). Five years of FISH-BOL: Brief status report. *Mitochondrial DNA*, *22*, 3–9.

Betoret, E., Betoret, N., Vidal, D., & Fito, P. (2011). Functional foods development: Trends and technologies. *Trends in Food Science & Technology*, *22*, 498–508.

Bokulich, N.A. (2012). Next-generation approaches to the microbial ecology of food fermentations. *Biochemistry and Molecurar Biology Reports*, *45*, 377–389.

Bokulich, N.A., Bamforth, C.W., & Mills, D.A. (2012). Brewhouse-resident microbiota are responsible for multi-stage fermentation of American coolship ale. *PLoS ONE*, *7*, e35507.

Bokulich, N.A., Joseph, C.L., Allen, G., Benson, A.K., & Mills, D.A. (2012). Next-generation sequencing reveals significant bacterial diversity of botrytized wine. *PLoS ONE*, *7*, e36357.

Bokulich, N.A., & Mills, D.A. (2013). Facility-specific "house" microbiome drives microbial landscapes of artisan cheesemaking plants. *Applied and Environmental Microbiology*, *79*, 5214–5223.

Bokulich, N.A., Ohta, M., Richardson, P.M., & Mills, D.A. (2013). Monitoring seasonal changes in winery-resident microbiota. *PLoS ONE*, *8*, e66437.

Bokulich, N.A., Thorngate, J.H., Richardson, P.M., & Mills, D.A. (2014). Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. *Proceedings of the National Academy of Sciences*, *111*, E139–E148.

Bondi, M., Messi, P., Halami, P.M., Papadopoulou, C., & de Niederhausern, S. (2014). Emerging microbial concerns in food safety and new control measures. *BioMedical Research International* (Article ID 251512).

Botti, S., & Giuffra, E. (2010). Oligonucleotide indexing of DNA barcodes: Identification of tuna and other Scombrid species in food products. *BMC Biotechnology*, *10*, 1–7.

Broekaert, K., Heyndrickx, M., Herman, L., Devlieghere, F., & Vlaemynck, G. (2013). Molecular identification of the microbiota of peeled and unpeeled brown shrimp (*Crangon crangon*) during storage on ice and at 7.5 °C. *Food Microbiology*, *36*, 123–134.

Bruni, I., Galimberti, A., Caridi, L., Scaccabarozzi, D., De Mattia, F., Casiraghi, M., et al. (2015). A DNA barcoding approach to identify plant species in multiflower honey. *Food Chemistry*, *170*, 308–315.

Bull, M., Plummer, S., Marchesi, J., & Mahenthiralingam, E. (2013). The life history of *Lactobacillus acidophilus* as a probiotic: A tale of revisionary taxonomy, misidentification and commercial success. *FEMS Microbiology Letters*, *349*, 77–87.

Cai, Y., Zhang, L., Shen, F., Zhang, W., Hou, R., Yue, B., et al. (2011). DNA barcoding of 18 species of Bovidae. *Chinese Science Bulletin*, *56*, 164–168.

Campbell-Platt, G. (1994). Fermented foods — A world perspective. *Food Research International*, *27*, 253–257.

Caplice, E., & Fitzgerald, G.F. (1999). Food fermentations: Role of microorganisms in food production and preservation. *International Journal of Food Microbiology*, *50*, 131–149.

Carvalho, D.C., Palhares, R.M., Drummond, M.G., & Frigo, T.B. (2015). DNA barcoding identification of commercialized seafood in South Brazil: A governmental regulatory forensic program. *Food Control*, *50*, 784–788.

Casiraghi, M., Labra, M., Ferri, E., Galimberti, A., & De Mattia, F. (2010). DNA barcoding: a six-question tour to improve users' awareness about the method. *Briefings in Bioinformatics*, *11*, 440–453.

Ceuppens, S., Li, D., Uyttendaele, M., Renault, P., Ross, P., Ranst, M.V., et al. (2014). Molecular methods in food safety microbiology: Interpretation and implications of nucleic acid detection. *Comprehensive Reviews in Food Science and Food Safety*, *13*, 551–577.

Chaillou, S., Chaulot-Talmon, A., Caekebeke, H., Cardinal, M., Christieans, S., Denis, C., et al. (2014). Origin and ecological selection of core and food-specific bacterial communities associated with meat and seafood spoilage. *The ISME Journal*, *2014*, 1–14.

Chakraborty, C., Doss, C.G.P., Patra, B.C., & Bandyopadhyay, S. (2014). DNA barcoding to map the microbial communities: Current advances and future directions. *Applied Microbiology and Biotechnology*, *98*, 3425–3436.

Cizeikiene, D., Juodeikiene, G., Paskevicius, A., & Bartkiene, E. (2013). Antimicrobial activity of lactic acid bacteria against pathogenic and spoilage microorganism isolated from food and their control in wheat bread. *Food Control*, *31*, 539–545.

Claesson, M.J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P., et al. (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Research*, *38*, e200.

Cocolin, L., Alessandria, V., Botta, C., Gorra, R., De Filippis, F., Ercolini, D., et al. (2013). NaOH-debittering induces changes in bacterial ecology during table olives fermentation. *PLoS ONE*, *8*, e69074.

Coppola, S., Blaiotta, G., Ercolini, D., & Moschetti, G. (2001). Molecular evaluation of microbial diversity occurring in different types of Mozzarella cheese. *Journal of Applied Microbiology*, *90*, 414–420.

Cordain, L., Eaton, S.B., Sebastian, A., Mann, N., Lindeberg, S., Watkins, B.A., et al. (2005). Origins and evolution of the Western diet: Health implications for the 21st century. *The American Journal of Clinical Nutrition*, *81*, 341–354.

Cornara, L., Borghesi, B., Canali, C., Andrenacci, M., Basso, M., Federici, S., et al. (2013). Smart drugs: Green shuttle or real drug? *International Journal of Legal Medicine*, *127*, 1109–1123.

Crowley, S., Mahony, J., & van Sinderen, D. (2013). Current perspectives on antifungal lactic acid bacteria as natural bio-preservatives. *Trends in Food Science & Technology*, *33*, 93–109.

Cutarelli, A., Amoroso, M.G., De Roma, A., Girardi, S., Galiero, G., Guarino, A., et al. (2014). Italian market fish species identification and commercial frauds revealing by DNA sequencing. *Food Control*, *37*, 46–50.

Czerucka, D., Piche, T., & Rampal, P. (2007). Review article: Yeast as probiotics — *Saccharomyces boulardii*. *Alimentary Pharmacology & Therapeutics*, *26*, 767–778.

D'Amato, M.E., Alechine, E., Cloete, K.W., Davison, S., & Corach, D. (2013). Where is the game? Wild meat products authentication in South Africa: A case study. *Investigative Genetics*, *4*, 1–13.

Dall'Asta, C., Cirlini, M., Morini, E., Rinaldi, M., Ganino, T., & Chiavaro, E. (2013). Effect of chestnut flour supplementation on physico-chemical properties and volatiles in bread making. *LWT—Food Science and Technology*, *53*, 233–239.

David, V., Terrat, S., Herzine, K., Claisse, O., Rousseaux, S., Tourdot-Maréchal, R., et al. (2014). High-throughput sequencing of amplicons for monitoring yeast biodiversity in must and during alcoholic fermentation. *Journal of Industrial Microbiology & Biotechnology*, *41*, 811–821.

De Filippis, F., La Storia, A., Stellato, G., Gatti, M., & Ercolini, D. (2014). A selected core microbiome drives the early stages of three popular Italian cheese manufactures. *PLoS ONE*, *9*, e89680.

De Mattia, F., Bruni, I., Galimberti, A., Cattaneo, F., Casiraghi, M., & Labra, M. (2011). A comparative study of different DNA barcoding markers for the identification of some members of Lamiacaea. *Food Research International*, *44*, 693–702.

De Pasquale, I., Di Cagno, R., Buchin, S., De Angelis, M., & Gobbetti, M. (2014). Microbial ecology dynamics reveal a succession in the core microbiota involved in the ripening of pasta filata caciocavallo pugliese cheese. *Applied and Environmental Microbiology*, *80*, 6243–6255.

De Santis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, *72*, 5069–5072.

De Vasconcelos, M.C., Bennett, R.N., Rosa, E.A., & Ferreira-Cardoso, J.V. (2010). Composition of European chestnut (*Castanea sativa* Mill.) and association with health effects: Fresh and processed products. *Journal of the Science of Food and Agriculture*, *90*, 1578–1589.

Delgado, S., Rachid, C.T., Fernández, E., Rychlik, T., Alegría, Á., Peixoto, R.S., et al. (2013). Diversity of thermophilic bacteria in raw, pasteurized and selectively-cultured milk, as assessed by culturing, PCR-DGGE and pyrosequencing. *Food Microbiology*, *36*, 103–111.

Demirkesen, I., Mert, B., Sumnu, G., & Sahin, S. (2010). Utilization of chestnut flour in gluten-free bread formulations. *Journal of Food Engineering*, *101*, 329–336.

Dentinger, B.T., Didukh, M.Y., & Moncalvo, J.M. (2011). Comparing COI and ITS as DNA barcode markers for mushrooms and allies (Agaricomycotina). *PLoS ONE*, *6*, e25081.

Dobson, A., O'Sullivan, O., Cotter, P.D., Ross, P., & Hill, C. (2011). High-throughput sequence-based analysis of the bacterial composition of kefir and an associated kefir grain. *FEMS Microbiology Letters*, *320*, 56–62.

Doyle, M.P., & Buchanan, R.L. (2013). *Food microbiology: Fundamentals and frontiers* (4th ed.). Washington, DC: ASM Press.

Druml, B., & Cichna-Markl, M. (2014). High resolution melting (HRM) analysis of DNA — Its role and potential in food analysis. *Food Chemistry*, *158*, 245–254.

EC No 1169/2011 (2011). Commission Regulation (EC) No 1169/2011 of 25 October 2011 on the provision of food information to consumers. *Official Journal of the European Union*, *L 304*(18).

El Sheikha, A.F., Bouvet, J.M., & Montet, D. (2011). Biological bar-code for the determination of geographical origin of fruits by using 28S rDNA fingerprinting of fungal communities by PCR-DGGE: An application to Shea tree fruits. *Quality Assurance & Safety of Crops & Foods*, *3*, 40–47.

El Sheikha, A.F., Condur, A., Métayer, I., Le Nguyen, D.D., Loiseau, G., & Montet, D. (2009). Determination of fruit origin by using 26S rDNA fingerprinting of yeast communities by PCR-DGGE: An application to *Physalis* fruits from Egypt. *Yeast*, *26*, 567–573.

El Sheikha, A.F., Durand, N., Sarter, S., Okullo, J.B.L., & Montet, D. (2012). Study of the microbial discrimination of fruits by PCR-DGGE: Application to the determination of the geographical origin of *Physalis* fruits from Colombia, Egypt, Uganda and Madagascar. *Food Control*, *24*, 57–63.

El Sheikha, A.F., Métayer, I., & Montet, D. (2011). A biological bar-code for determining the geographical origin of fruit by using 28S rDNA fingerprinting of fungi communities by PCR-DGGE: An application to *Physalis* fruits from Egypt. *Food Biotechnology*, *25*, 115–129.

El Sheikha, A.F., & Montet, D. (2014). Fermented fish and fish products: Snapshots on culture and health. In R.C. Ray, & D. Montet (Eds.), *Microorganisms and food fermentation* (pp. 188–222). Florida, USA: Science Publishers, Inc., New Hampshire; CRC Press.

Enan, M.R., & Ahamed, A. (2014). DNA barcoding based on plastid matK and RNA polymerase for assessing the genetic identity of date (*Phoenix dactylifera* L.) cultivars. *Genetics and Molecular Research*, *13*, 3527–3536.

Ercolini, D. (2004). PCR-DGGE fingerprinting: Novel strategies for detection of microbes in food. *Journal of Microbiological Methods*, *56*, 297–314.

Ercolini, D. (2013). High-throughput sequencing and metagenomics: Moving forward in the culture-independent analysis of food microbial ecology. *Applied and Environmental Microbiology*, *79*, 3148–3155.

Ercolini, D., De Filippis, F., La Storia, A., & Iacono, M. (2012). "Remake" by high-throughput sequencing of the microbiota involved in the production of water buffalo mozzarella cheese. *Applied and Environmental Microbiology*, *78*, 8142–8145.

Faria, M.A., Magalhães, A., Nunes, M.E., & Oliveira, M.B.P.P. (2013). High resolution melting of *trnL* amplicons in fruit juices authentication. *Food Control*, *33*, 136–141.

Federici, S., Galimberti, A., Bartolucci, F., Bruni, I., Cortis, P., & Labra, M. (2013). DNA barcoding to analyse taxonomically complex groups in plants: The case of *Thymus* (Lamiaceae). *Botanical Journal of the Linnean Society*, *171*, 687–699.

Felis, G.E., & Dellaglio, F. (2007). Taxonomy of lactobacilli and bifidobacteria. *Current Issues in Intestinal Microbiology*, *8*, 44–61.

Fox, P.F., Guinee, T.P., Cogan, T.M., & McSweeney, P.L.H. (2000). *Fundamentals of cheese science*. Gaithersburg, MD, USA: Aspen Publication.

Fusco, V., & Quero, G.M. (2014). Culture-dependent and culture-independent nucleic acid-based methods used in the microbial safety assessment of milk and dairy products. *Comprehensive Reviews in Food Science and Food Safety*, *13*, 493–537.

Galal-Khallaf, A., Ardura, A., Mohammed-Geba, K., Borrell, Y.J., & Garcia-Vazquez, E. (2014). DNA barcoding reveals a high level of mislabeling in Egyptian fish fillets. *Food Control*, *46*, 441–445.

Galimberti, A., De Mattia, F., Losa, A., Bruni, I., Federici, S., Casiraghi, M., et al. (2013). DNA barcoding as a new tool for food traceability. *Food Research International*, *50*, 55–63.

Galimberti, A., Labra, M., Sandionigi, A., Bruno, A., Mezzasalma, V., & De Mattia, F. (2014). DNA barcoding for minor crops and food traceability. *Advances in Agriculture* (Article ID 831875).

Gálvez, A., Abriouel, H., López, R.L., & Omar, N.B. (2007). Bacteriocin-based strategies for food biopreservation. *International Journal of Food Microbiology*, *120*, 51–70.

Ganopoulos, I., Bazakos, C., Madesis, P., Kalaitzis, P., & Tsaftaris, A. (2013). Barcode DNA high-resolution melting (Bar-HRM) analysis as a novel close-tubed and accurate tool for olive oil forensic use. *Journal of the Science of Food and Agriculture*, *93*, 2281–2286.

Ganopoulos, I., Madesis, P., Darzentas, N., Argiriou, A., & Tsaftaris, A. (2012). Barcode high resolution melting (Bar-HRM) analysis for detection and quantification of PDO "Fava Santorinis"(*Lathyrus clymenum*) adulterants. *Food Chemistry*, *133*, 505–512.

Gismondi, A., Fanali, F., Labarga, J.M.M., Caiola, M.G., & Canini, A. (2013). *Crocus sativus* L. genomics and different DNA barcode applications. *Plant Systematics and Evolution*, *299*, 1859–1863.

Gkatzionis, K., Yunita, D., Linforth, R.S., Dickinson, M., & Dodd, C.E. (2014). Diversity and activities of yeasts from different parts of a Stilton cheese. *International Journal of Food Microbiology*, *177*, 109–116.

Gonçalves, J., Pereira, F., Amorim, A., & van Asch, B. (2012). New method for the simultaneous identification of cow, sheep, goat, and water buffalo in dairy products by analysis of short species-specific mitochondrial DNA targets. *Journal of Agricultural and Food Chemistry*, *60*, 10480–10485.

Guerreiro, J.S., Fernandes, P., & Bardsley, R.G. (2012). Identification of the species of origin of milk in cheeses by multivariate statistical analysis of polymerase chain reaction electrophoretic patterns. *International Dairy Journal*, *25*, 42–45.

Hammami, R., Zouhir, A., Le Lay, C., Hamida, J.B., & Fliss, I. (2010). BACTIBASE second release: A database and tool platform for bacteriocin characterization. *BMC Microbiology*, *10*, 1–5.

Haye, P.A., Segovia, N.I., Vera, R., Gallardo, M.D.L.Á., & Gallardo-Escárate, C. (2012). Authentication of commercialized crab-meat in Chile using DNA barcoding. *Food Control*, *25*, 239–244.

Hebert, P.D.N., Ratnasingham, S., & deWaard, J.R. (2003). Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*, *270*, S96–S99.

Hidayat, T., Abdullah, F.I., Kuppusamy, C., Samad, A.A., & Wagiran, A. (2012). Molecular identification of Malaysian pineapple cultivar based on internal transcribed spacer region. *APCBEE Procedia*, *4*, 146–151.

Hidayat, T., Kusumawaty, D., & Pancoro, A. (2013). Utility of matK gene as DNA barcode to assess evolutionary relationship of important tropical forest tree genus *Mangifera* (Anacardiaceae) in Indonesia and Thailand. *Biotropia*, *18*, 74–80.

Imazio, S., Labra, M., Grassi, F., Winfield, M., Bardini, M., & Scienza, A. (2002). Molecular tools for clone identification: The case of the grapevine cultivar 'Traminer'. *Plant Breeding*, *121*, 531–535.

Jaakola, L., Suokas, M., & Häggman, H. (2010). Novel approaches based on DNA barcoding and high-resolution melting of amplicons for authenticity analyses of berry species. *Food Chemistry*, *123*, 494–500.

Janda, J.M., & Abbott, S.L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *Journal of Clinical Microbiology*, *45*, 2761–2764.

Jarret, R.L. (2008). DNA barcoding in a crop genebank: The *Capsicum annuum* species complex. *The Open Biology Journal*, *1*, 35–42.

Jofré, A., Martin, B., Garriga, M., Hugas, M., Pla, M., Rodríguez-Lázaro, D., et al. (2005). Simultaneous detection of *Listeria monocytogenes* and *Salmonella* by multiplex PCR in cooked ham. *Food Microbiology*, *22*, 109–115.

Johnson, M.E., & Steele, J.L. (2013). Fermented dairy products. In M.P. Doyle, & R.L. Buchanan (Eds.), *Food microbiology: Fundamentals and frontiers* (pp. 825–839). Washington, DC: ASM Press.

Jung, J.Y., Lee, S.H., Kim, J.M., Park, M.S., Bae, J.W., Hahn, Y., et al. (2011). Metagenomic analysis of kimchi, a traditional Korean fermented food. *Applied and Environmental Microbiology*, *77*, 2264–2274.

Jung, M.J., Nam, Y.D., Roh, S.W., & Bae, J.W. (2012). Unexpected convergence of fungal and bacterial communities during fermentation of traditional Korean alcoholic beverages inoculated with various natural starters. *Food Microbiology*, *30*, 112–123.

Kane, N., Sveinsson, S., Dempewolf, H., Yang, J.Y., Zhang, D., Engels, J.M., et al. (2012). Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *American Journal of Botany*, *99*, 320–329.

Khaund, P., & Joshi, S.R. (2014). DNA barcoding of wild edible mushrooms consumed by the ethnic tribes of India. *Gene*, *550*, 123–130.

Kim, Y.S., Kim, M.C., Kwon, S.W., Kim, S.J., Park, I.C., Ka, J.O., et al. (2011). Analyses of bacterial communities in meju, a Korean traditional fermented soybean bricks, by cultivation-based and pyrosequencing methods. *The Journal of Microbiology*, *49*, 340–348.

Kim, H.J., Kim, M.J., Turner, T.L., Kim, B.S., Song, K.M., Yi, S.H., et al. (2014). Pyrosequencing analysis of microbiota reveals that lactic acid bacteria are dominant in Korean flat fish fermented food, gajami-sikhae. *Bioscience Biotechnology and Biochemistry*, *78*, 1611–1618.

Kim, Y.H., Park, H.M., Hwang, T.Y., Lee, S.K., Choi, M.S., Jho, S., et al. (2014). Variation block-based genomics method for crop plants. *BMC Genomics*, *15*, 1–14.

Kojoma, M., Kurihara, K., Yamada, K., Sekita, S., Satake, M., & Iida, O. (2002). Genetic identification of cinnamon. *Planta Medica*, *68*, 94–96.

Koyanagi, T., Kiyohara, M., Matsui, H., Yamamoto, K., Kondo, T., Katayama, T., et al. (2011). Pyrosequencing survey of the microbial diversity of 'narezushi', an archetype of modern Japanese sushi. *Letters in Applied Microbiology*, *53*, 635–640.

Kuehn, J.S., Gorden, P.J., Munro, D., Rong, R., Dong, Q., Plummer, P.J., et al. (2013). Bacterial community profiling of milk samples as a means to understand culture-negative bovine clinical mastitis. *PLoS ONE*, *8*, e61959.

Lamendin, R., Miller, K., & Ward, R.D. (2015). Labelling accuracy in Tasmanian seafood: An investigation using DNA barcoding. *Food Control*, *47*, 436–443.

Le Nguyen, D.D., Ngoc, H.H., Dijoux, D., Loiseau, G., & Montet, D. (2008). Determination of fish origin by using 16S rDNA fingerprinting of bacterial communities by PCR-DGGE: An application on Pangasius fish from Viet Nam. *Food Control*, *19*, 454–460.

Leite, A.M.O., Mayo, B., Rachid, C.T.C.C., Peixoto, R.S., Silva, J.T., Paschoalin, V.M.F., et al. (2012). Assessment of the microbial diversity of Brazilian kefir grains by PCR-DGGE and pyrosequencing analysis. *Food Microbiology*, *31*, 215–221.

Li, M., Wong, K.L., Chan, W.H., Li, J.X., But, P.P.H., Cao, H., et al. (2012). Establishment of DNA barcodes for the identification of the botanical sources of the Chinese 'cooling' beverage. *Food Control*, *25*, 758–766.

Lusk, T.S., Ottesen, A.R., White, J.R., Allard, M.W., Brown, E.W., & Kase, J.A. (2012). Characterization of microflora in Latin-style cheeses by next-generation sequencing technology. *BMC Microbiology*, *12*, 254.

Madesis, P., Ganopoulos, I., Anagnostis, A., & Tsaftaris, A. (2012). The application of Bar-HRM (barcode DNA-high resolution melting) analysis for authenticity testing and quantitative detection of bean crops (Leguminosae) without prior DNA purification. *Food Control*, *25*, 576–582.

Madesis, P., Ganopoulos, I., Sakaridis, I., Argiriou, A., & Tsaftaris, A. (2014). Advances of DNA-based methods for tracing the botanical origin of food products. *Food Research International*, *60*, 163–172.

Mafra, I., Ferreira, I.M., & Oliveira, M.B.P. (2008). Food authentication by PCR-based methods. *European Food Research and Technology*, *227*, 649–665.

Maralit, B.A., Aguila, R.D., Ventolero, M.F.H., Perez, S.K.L., Willette, D.A., & Santos, M.D. (2013). Detection of mislabeled commercial fishery by-products in the Philippines using DNA barcodes and its implications to food traceability and safety. *Food Control*, *33*, 119–125.

Masoud, W., Vogensen, F.K., Lillevang, S., Abu Al-Soud, W., Sørensen, S.J., & Jakobsen, M. (2012). The fate of indigenous microbiota, starter cultures, *Escherichia coli*, *Listeria innocua* and *Staphylococcus aureus* in Danish raw milk and cheeses determined by pyrosequencing and quantitative real time (qRT)-PCR. *International Journal of Food Microbiology*, *153*, 192–202.

Mauriello, G., Moio, L., Genovese, A., & Ercolini, D. (2003). Relationships between flavoring capabilities, bacterial composition, and geographical origin of natural whey cultures used for traditional water-buffalo Mozzarella cheese manufacture. *Journal of Dairy Science*, *86*, 486–497.

Mayo, B., Rachid, C., Alegría, Á., Leite, A., S Peixoto, R., & Delgado, S. (2014). Impact of next generation sequencing techniques in food microbiology. *Current Genomics*, *15*, 293–309.

McGovern, P.E., Glusker, D.L., Exner, L.J., & Voigt, M.M. (1996). Neolithic resinated wine. *Nature*, *381*, 480–481.

Minervini, F., De Angelis, M., Di Cagno, R., & Gobbetti, M. (2014). Ecological parameters influencing microbial diversity and stability of traditional sourdough. *International Journal of Food Microbiology*, *171*, 136–146.

Muñoz-Quezada, S., Chenoll, E., María Vieites, J., Genovés, S., Maldonado, J., Bermúdez-Brito, M., et al. (2013). Isolation, identification and characterisation of three novel probiotic strains (*Lactobacillus paracasei* CNCM I-4034, *Bifidobacterium breve* CNCM I-4035 and *Lactobacillus rhamnosus* CNCM I-4036) from the faeces of exclusively breast-fed infants. *British Journal of Nutrition*, *109*, S51–S62.

Muyzer, G., De Waal, E.C., & Uitterlinden, A.G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology*, *59*, 695–700.

Myers, M.J. (2011). Molecular identification of animal species in food: Transition from research laboratories to the regulatory laboratories. *The Veterinary Journal*, *190*, 7–8.
Nam, Y.D., Lee, S.Y., & Lim, S.I. (2012). Microbial community analysis of Korean soybean pastes by next-generation sequencing. *International Journal of Food Microbiology*, *155*, 36–42.
Nam, Y.D., Park, S.L., & Lim, S.I. (2012). Microbial composition of the Korean traditional food "kochujang" analyzed by a massive sequencing technique. *Journal of Food Science*, *77*, M250–M256.
Newell, D.G., Koopmans, M., Verhoef, L., Duizer, E., Aidara-Kane, A., Sprong, H., et al. (2010). Food-borne diseases—The challenges of 20 years ago still persist while new ones continue to emerge. *International Journal of Food Microbiology*, *139*, S3–S15.
Nieminen, T.T., Koskinen, K., Laine, P., Hultman, J., Säde, E., Paulin, L., et al. (2012). Comparison of microbial communities in marinated and unmarinated broiler meat by metagenomics. *International Journal of Food Microbiology*, *157*, 142–149.
Norhana, W., Poole, S.E., Deeth, H.C., & Dykes, G.A. (2010). Prevalence, persistence and control of *Salmonella* and *Listeria* in shrimp and shrimp products: A review. *Food Control*, *21*, 343–361.
Opara, L.U., & Mazaud, F. (2001). Food traceability from field to plate. *Outlook on Agriculture*, *30*, 239–247.
O'Sullivan, D.J., Giblin, L., McSweeney, P.L., Sheehan, J.J., & Cotter, P.D. (2013). Nucleic acid-based approaches to investigate microbial-related cheese quality defects. *Frontiers in Microbiology*, *4*, 1–15.
Pansa, M., Runglawan, S., Tawatchai, T., Kowit, N., Nat, B., & Arunrat, C. (2011). Species diversity, usages, molecular markers and barcode of medicinal Senna species (Fabaceae, Caesalpinioideae) in Thailand. *Journal of Medicinal Plants Research*, *5*, 6173–6181.
Park, E.J., Chun, J., Cha, C.J., Park, W.S., Jeon, C.O., & Bae, J.W. (2012). Bacterial community analysis during fermentation of ten representative kinds of kimchi with barcoded pyrosequencing. *Food Microbiology*, *30*, 197–204.
Parvathy, V.A., Swetha, V.P., Sheeja, T.E., Leela, N.K., Chempakam, B., & Sasikumar, B. (2014). DNA barcoding to detect chilli adulteration in traded black pepper powder. *Food Biotechnology*, *28*, 25–40.
Patel, J.B. (2001). 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Molecular Diagnosis*, *6*, 313–321.
Peres, B., Barlet, N., Loiseau, G., & Montet, D. (2007). Review of the current methods of analytical traceability allowing determination of the origin of foodstuffs. *Food Control*, *18*, 228–235.
Pfeiler, E.A., & Klaenhammer, T.R. (2013). Probiotics and prebiotics. In M.P. Doyle, & R.L. Buchanan (Eds.), *Food microbiology: Fundamentals and frontiers* (pp. 949–971) (4th ed.). Washington, DC: ASM Press.
Połka, J., Rebecchi, A., Pisacane, V., Morelli, L., & Puglisi, E. (2015). Bacterial diversity in typical Italian salami at different ripening stages as revealed by high-throughput sequencing of 16S rRNA amplicons. *Food Microbiology*, *46*, 342–356.
Ponzoni, E., Mastromauro, F., Gianì, S., & Breviario, D. (2009). Traceability of plant diet contents in raw cow milk samples. *Nutrients*, *1*, 251–262.
Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., et al. (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, *35*, 7188–7196.
Quigley, L., O'Sullivan, O., Beresford, T.P., Ross, R.P., Fitzgerald, G.F., & Cotter, P.D. (2012). High-throughput sequencing for detection of subpopulations of bacteria not previously associated with artisanal cheeses. *Applied and Environmental Microbiology*, *78*, 5717–5723.
Raja, H.A., Baker, T.R., Little, J.G., & Oberlies, N.H. (2014). DNA barcoding for identification of species in mushrooms: A component of product certification. *Planta Medica*, *80*, 846–847.
Randazzo, C.L., Pitino, I., Ribbera, A., & Caggia, C. (2010). Pecorino Crotonese cheese: Study of bacterial population and flavour compounds. *Food Microbiology*, *27*, 363–374.
Ray, B., & Daeschel, M. (1992). *Food biopreservatives of microbial origin*. USA: CRC Press.
Riquelme, C., Câmara, S., Maria de Lurdes, N., Vinuesa, P., da Silva, C.C.G., Malcata, F.X., et al. (2015). Characterization of the bacterial biodiversity in Pico cheese (an artisanal Azorean food). *International Journal of Food Microbiology*, *192*, 86–94.
Roberfroid, M.B. (2000). A European consensus of scientific concepts of functional foods. *Nutrition*, *16*, 689–691.
Roh, S.W., Kim, K.H., Nam, Y.D., Chang, H.W., Park, E.J., & Bae, J.W. (2009). Investigation of archaeal and bacterial diversity in fermented seafood using barcoded pyrosequencing. *The ISME Journal*, *4*, 1–16.
Ross, P.R., Morgan, S., & Hill, C. (2002). Preservation and fermentation: Past, present and future. *International Journal of Food Microbiology*, *79*, 3–16.
Saderi, M., Saderi, A.H., & Rahimi, G. (2013). Identification of bovine, ovine and caprine pure and binary mixtures of raw and heat processed meats using species specific size markers targeting mitochondrial genome. *Iranian Journal of Veterinary Research*, *14*, 29–34.
Sakamoto, N., Tanaka, S., Sonomoto, K., & Nakayama, J. (2011). 16S rRNA pyrosequencing-based investigation of the bacterial community in nukadoko, a pickling bed of fermented rice bran. *International Journal of Food Microbiology*, *144*, 352–359.
Sandionigi, A., Vicario, S., Prosdocimi, E.M., Galimberti, A., Ferri, E., Bruno, A., et al. (2015s). Towards a better understanding of *Apis mellifera* and Varroa destructor microbiomes: Introducing 'phyloh' as a novel phylogenetic diversity analysis tool. *Molecular Ecology Resources* (in press).
Shah, N.P. (2007). Functional cultures and health benefits. *International Dairy Journal*, *17*, 1262–1277.
Smith, P.J., McVeagh, S.M., & Steinke, D. (2008). DNA barcoding for the identification of smoked fish products. *Journal of Fish Biology*, *72*, 464–471.
Solieri, L., Dakal, T.C., & Giudici, P. (2013). Next-generation sequencing and its potential impact on food microbial genomics. *Annals of Microbiology*, *63*, 21–37.
Stanton, C., Ross, R.P., Fitzgerald, G.F., & Sinderen, D.V. (2005). Fermented functional foods based on probiotics and their biogenic metabolites. *Current Opinion in Biotechnology*, *16*, 198–203.
Stoeckle, M.Y., Gamble, C.C., Kirpekar, R., Young, G., Ahmed, S., & Little, D.P. (2011). Commercial teas highlight plant DNA barcode identification successes and obstacles. *Scientific Reports*, *1*, 1–7.
Tautz, D., Arctander, P., Minelli, A., Thomas, R.H., & Vogler, A.P. (2003). A plea for DNA taxonomy. *Trends in Ecology & Evolution*, *18*, 70–74.
Theodoridis, S., Stefanaki, A., Tezcan, M., Aki, C., Kokkini, S., & Vlachonasios, K.E. (2012). DNA barcoding in native plants of the Labiatae (Lamiaceae) family from Chios Island (Greece) and the adjacent Çeşme-Karaburun Peninsula (Turkey). *Molecular Ecology Resources*, *12*, 620–633.
Valentini, A., Miquel, C., & Taberlet, P. (2010). DNA barcoding for honey biodiversity. *Diversity*, *2*, 610–617.
Velusamy, V., Arshak, K., Korostynska, O., Oliwa, K., & Adley, C. (2010). An overview of foodborne pathogen detection: In the perspective of biosensors. *Biotechnology Advances*, *28*, 232–254.
Vriesekoop, F., Krahl, M., Hucker, B., & Menz, G. (2012). 125th anniversary review: Bacteria in brewing: The good, the bad and the ugly. *Journal of the Institute of Brewing*, *118*, 335–345.
Waites, M.J., Morgan, N.L., Rockey, J.S., & Higton, G. (2009). *Industrial microbiology: An introduction*. Oxford, UK: Wiley-Blackwell.
Wang, M., Zhao, H.X., Wang, L., Wang, T., Yang, R.W., Wang, X.L., et al. (2013). Potential use of DNA barcoding for the identification of *Salvia* based on cpDNA and nrDNA sequences. *Gene*, *528*, 206–215.
Xin, T., Yao, H., Gao, H., Zhou, X., Ma, X., Xu, C., et al. (2013). Super food *Lycium barbarum* (Solanaceae) traceability via an internal transcribed spacer 2 barcode. *Food Research International*, *54*, 1699–1704.
Yancy, H.F., Zemlak, T.S., Mason, J.A., Washington, J.D., Tenge, B.J., Nguyen, N.L.T., et al. (2008). Potential use of DNA barcodes in regulatory science: Applications of the Regulatory Fish Encyclopedia. *Journal of Food Protection*, *71*, 210–217.
Yu, J., Yan, H.X., Lu, Z.H., & Zhou, Z.Q. (2011). Screening potential DNA barcode regions of chloroplast coding genome for citrus and its related genera. *Scientia Agricultura Sinica*, *2*, 015.
Zbrun, M.V., Romero-Scharpen, A., Olivero, C., Rossler, E., Soto, L.P., Rosmini, M.R., et al. (2013). Occurrence of thermotolerant *Campylobacter* spp. at different stages of the poultry meat supply chain in Argentina. *New Zealand Veterinary Journal*, *61*, 337–343.
Zhang, R., Huo, W., Zhu, W., & Mao, S. (2015s). Characterization of bacterial community of raw milk from dairy cows during subacute ruminal acidosis challenge by high-throughput sequencing. *Journal of the Science of Food and Agriculture* (in press).
Zuo, Y., Chen, Z., Kondo, K., Funamoto, T., Wen, J., & Zhou, S. (2011). DNA barcoding of *Panax* species. *Planta Medica*, *77*, 182–187.

Hindawi Publishing Corporation
Advances in Agriculture
Volume 2014, Article ID 831875, 8 pages
http://dx.doi.org/10.1155/2014/831875

Review Article

DNA Barcoding for Minor Crops and Food Traceability

Andrea Galimberti, Massimo Labra, Anna Sandionigi, Antonia Bruno, Valerio Mezzasalma, and Fabrizio De Mattia

ZooPlantLab, Dipartimento di Biotecnologie e Bioscienze, Università degli Studi di Milano-Bicocca, Piazza della Scienza 2, 20126 Milano, Italy

Correspondence should be addressed to Fabrizio De Mattia; fabrizio.demattia@unimib.it

Received 27 March 2014; Accepted 11 June 2014; Published 23 June 2014

Academic Editor: Pawan L. Kulwal

Copyright © 2014 Andrea Galimberti et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This outlook paper addresses the problem of the traceability of minor crops. These kinds of cultivations consist in a large number of plants locally distributed with a modest production in terms of cultivated acreage and quantity of final product. Because of globalization, the diffusion of minor crops is increasing due to their benefit for human health or their use as food supplements. Such a phenomenon implies a major risk for species substitution or uncontrolled admixture of manufactured plant products with severe consequences for the health of consumers. The need for a reliable identification system is therefore essential to evaluate the quality and provenance of minor agricultural products. DNA-based techniques can help in achieving this mission. In particular, the DNA barcoding approach has gained a role of primary importance thanks to its universality and versatility. Here, we present the advantages in the use of DNA barcoding for the characterization and traceability of minor crops based on our previous or ongoing studies at the ZooPlantLab (Milan, Italy). We also discuss how DNA barcoding may potentially be transferred from the laboratory to the food supply chain, from field to table.

1. DNA Barcoding for Plant Identification

Plants as primary producers are the basis of human nutrition from time immemorial. It is estimated that about 7,000 species of plants have been cultivated for consumption in human history (FAO data) and a large number of cultivars and varieties are also recognized. The Commission on Genetic Resources for Food and Agriculture (http://www.fao.org/nr/cgrfa/cthemes/plants/en/) estimated that 30 crops are usually referred currently as major agricultural products since they provide 95% of human food energy needs (e.g., rice, wheat, maize, and potato). These resources are widely monitored and well characterized with the analysis of DNA markers specifically developed for each cultivar (see, e.g., [1–3]). On the contrary, reliable characterization tools for the minor varieties are far from being defined. Minor crops include plants for food, pharmaceutical, cosmetic, and ornamental purposes with a modest production in terms of cultivated acreage and quantity of final product [4]. There are no fixed standard values to define a minor crop; however, conventionally, all the local varieties could be placed in this category. Most of these species or varieties show peculiar traits from the alimentary, pharmaceutical, or ornamental points of view. Some examples of minor crops that are now widely cultivated and worldwide distributed are Goji (*Lycium barbarum* L. [5]), Chokeberry (*Aronia melanocarpa* (Michx.), [6]), Peach Palm (*Bactris gasipaes* Kunth [7]), Teff (*Eragrostis tef* (Zucc.) [8]), and Okra (*Abelmoschus esculentus* (L.) Moench [9]). A large number of minor crops were usually produced and consumed locally [10] but, nowadays, the continuous demand by developed countries for identifying new active metabolites for human health and nutrition has increased their diffusion at global level [11–14]. This phenomenon implies a major risk for species substitution or uncontrolled admixture of manufactured plant products. Substitution or adulteration can be deliberate (e.g., to maximize financial gains) or inadvertent (e.g., due to an insufficient knowledge by farmers) but they can have serious consequences for consumers at any rate [14–19].

Given these premises, it is clear that the definition of a reliable traceability system is an aspect of major concern when plants, parts of plants, or plant extracts are used in food

industry. The need for an unequivocal identification is also essential to start quality assurance procedures for agricultural products, to authenticate their geographical provenance (in the case of protected designation of origin), and to prevent commercial frauds and adulteration cases.

Agricultural products are subjected to strong processing and manufacturing before they are released as final products to the consumer. These processes alter the plant structure, thereby impeding the use of morphological characters to identify most of the agricultural products. To overcome this limit, the analysis of proteins and/or DNA is nowadays used as the main tool for plant traceability. However, although chemical or protein-based approaches are useful in characterizing the composition of fresh products, these methods can be biased by several factors such as the strong food manufacturing processes, the limited number of detectable isozymes, or the high tissue and developmental stage specificity of the markers [20]. DNA markers are more informative than protein or chemical based methods because DNA better resists industrial processes such as shredding, boiling, pressure cooking, or transformations mediated by chemical agents (see, e.g., [18, 21, 22]). This property allows a successful identification of plant material, even when it is present in small traces [23, 24]. Moreover, the availability of advanced technologies and efficient commercial kits for DNA extraction permits obtaining an acceptable yield of genetic material from processed or degraded plant material [25].

As a consequence, DNA markers have rapidly become the most used tools in the genetic analyses of crops and cultivars, as well as in the tracking and certification of the raw materials in food industry processes [26–32]. PCR-based methods are more sensitive and faster than other technologies in characterizing agricultural products [1–3]. Among these, discontinuous molecular markers such as RAPDs, AFLPs, and their variants (e.g., ISSR, SSAP) have been successfully adopted for the characterization of crop species [24]. Moreover, sequencing-based systems such as single nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs) are also used because of their high level of polymorphism and high reproducibility [30]. However, being highly species specific, these approaches require access to the correct DNA sequence of the organisms and their application is often limited to a single species.

In the last decade, DNA barcoding was proposed as a universal DNA-based tool for species identification [33]. The name "DNA barcoding" figuratively refers to the way an infrared scanner univocally identifies a product by using the stripes of the universal product code (UPC). At the same time, this approach is based on the analysis of the variability within one or a few standard regions of the genome called "DNA barcode/s" [33]. The rationale of the method is that the DNA barcoding sequence/s univocally corresponds to each species (i.e., low intraspecific variability) but largely differs between taxa (i.e., high interspecific variability) [33, 34]. DNA barcoding has the advantage of combining three important innovations: molecularization of the identification approach (i.e., the investigation of DNA variability to differentiate taxa), standardization of the process (from sample collection to the analysis of molecular results), and computerization (i.e., the not redundant transposition of the data using informatics) [34].

Several plastidial and nuclear regions have been proposed as barcode regions for plants [35–37] and some of them are now used for the identification of crop species, as recently reviewed by [38]. In 2009, the Plant Working Group of CBOL (consortium for the barcode of life) defined a standard core-barcode panel of markers based on the combination of portions of two coding plastidial regions: *matK* and *rbcL* [39, 40]. Despite their high universality in terms of amplification and sequencing success, the analysis of these coding regions fails in some cases due to the interspecific sharing of sequences [41]. Internal transcribed spacer regions of nuclear ribosomal DNA (ITS) were recommended as additional marker being highly variable in angiosperms [40]. ITS works well in many plant groups but, in some cases, incomplete concerted evolution and intraindividual variation make it unsuitable as universal plant barcode [40]. However, the combination of *matK* and *rbcL* with the plastidial intergenic noncoding region *trnH-psbA* increases the identification performance of DNA barcoding. As a consequence, the use of *trnH-psbA* is growing due to its easy amplification, and its high genetic variability among closely related taxa [15, 35, 42].

At the University of Milano-Bicocca (Milan, Italy), the ZooPlantLab group (http://www.zooplantlab.btbs.unimib.it/) is one of the most active centers where DNA barcoding is used as a universal traceability system. The ZooPlantLab research team investigates concrete problems dealing with agricultural production of minor crops by transferring the analytical pipeline from the laboratory to food supply chain. This approach aims to overcome technical traceability problems in order to offer solid solutions to the market.

In the following sections, we present some of the potential applications and advantages of DNA barcoding for the identification and traceability along the food supply chain of minor crops. We also examine the most innovative approaches dealing with DNA barcoding that have been recently adopted to characterize these kinds of agricultural products.

2. Traceability of Minor Crops in the Supply Chain: The Case of Spices

Spices represent a clear example of minor crops. Most of these belong to Lamiaceae, a large family of 264 genera and almost 7,000 described species [78] characterized by aromatic oils and secondary metabolites. Thanks to their peculiar chemical profiles, these plants are commonly used as flavor for cooking, essences for cosmetics, and active components in medicines. Given their economical importance, many members of Lamiaceae have been investigated widely with different approaches ranging from morphology to chemistry and genetics in order to characterize their variability and improve the quality of cultivated varieties [25, 26, 79, 80].

Although some species showed distinctive morphological traits, this family encompasses many critical genera such as *Thymus* [43], where differences among closely related taxa are limited to few minor morphological characters. However, morphology could be ineffective for tracing spices along

the supply chain (i.e., from the crop cultivation sites to the final products) which usually encompasses strong manufacturing processes such as crushing, powdering, or aqueous/alcoholic extraction of plant material.

International agencies such as the American Spice Trade Association (ASTA, http://www.astaspice.org) and the European Spice Association (ESA, http://www.esa-spices.org/) support the characterization of the phytochemical profile to assess the quality of herbs and spices. The evaluation of chemical characteristics is essential to standardize the industrial production of spices-derived products; however, in most cases, the analysis of chemical compounds is not able to univocally identify the original plants at the species level [26]. For this reason, we proposed the DNA barcoding approach as a universal and suitable tool to characterize and trace aromatic species. DNA analyses were conducted starting from different plant portions [22] or their derived products (e.g., oils, extracts) stored at different conditions (i.e., dried, frozen). In our study [22], we investigated 6 major groups of cooking spices (i.e., mint, basil, oregano, sage, thyme, and rosemary) also including their most relevant cultivars and hybrids. We collected samples at different stages of the industrial supply chain starting from seeds and plants cultivated by private farmers or in garden centers to commercial dried spices or other manufactured products. We also tested the performances of DNA barcoding starting from plant extracts. A good yield of high quality DNA was obtained through extraction protocols from all of the considered samples and then used for the next steps of the analysis (i.e., PCR and sequencing). A sufficient amount of DNA was also extracted from several of the plant extracts (Labra M., unpublished data) by using commercial kits. This first result confirmed that the industrial processes to transform the raw plant material such as drying, crushing, and aqueous or alcoholic extractions do not excessively degrade DNA. Among the four tested DNA barcoding regions (i.e., *rbcL*, *matK*, *trnH-psbA,* and *rpoB*), the *trnH-psbA* ranked the first in genetic divergence values among species, followed by *matK* and *rbcL*. On the contrary, *rpoB* showed the lowest sequence divergence among the tested taxa (see [22] for further details).

Our results partially supported the guidelines provided by the CBOL [40]. Indeed, the two core-barcode markers (i.e., *matK* + *rbcL*) properly assigned the tested spices to the expected genus and, in most cases, they also reached the species level. However, the highest identification performances were achieved by using the additional *trnH-psbA* barcode region. A clear example is that of basil (genus *Ocimum*), a group consisting of 30–160 species with many recognized cultivars [81]. In our study, exclusive *trnH-psbA* haplotypes, were found for almost all the tested cultivars, providing a reliable system for their identification. This result deserves to be highlighted because it is one of the first pieces of evidence supporting the usefulness of DNA barcoding in discriminating organisms at a taxonomic level lower than the species one.

Other important data revealed by our analyses concerned the capability of DNA barcoding to identify parental and hybrid species in some members of Lamiaceae. An example is represented by the case of peppermint (*M. piperita* L.), a sterile hybrid between *M. aquatica* L. × *M. spicata* L. [82, 83]. The plastidial markers used in this study confirmed that *M. spicata* L. is the maternal parental of *M. piperita* L. because both taxa showed the same DNA profile. However, to confirm definitively the hybrid origin of *M. piperita* L. and to identify the exact parental inheritance, the ITS2 codominant marker was sequenced (Labra M., unpublished data).

On the whole, the most relevant result of our work consisted in the assessment of the universality of DNA barcoding in a context of minor crops traceability. Using a single primer combination for each one of the few DNA barcoding markers and following standard laboratory protocols, it is possible to recognize the original species starting from different plant portions or derived processed materials. The same approach is also useful for validating several other herbal products commonly distributed on the market such as tea [50], saffron [44, 84], ginseng [69], black pepper [59], and many others (see also Table 1). These cases clearly emphasize the high versatility of DNA barcoding. It is an authentic functional tool for molecular traceability of agricultural products, as most of the minor crops have not yet been characterized with private markers such as SSR or SNP in order to allow a reliable DNA fingerprinting system. Moreover, DNA barcoding does not require any previous knowledge of the plant genome for the investigated species and the analytical procedures can be easily adopted by any laboratory equipped for molecular biology.

3. Commercial Frauds and Dangerous Substitutions

Nowadays, the global diffusion of several minor crops in the absence of suitable traceability protocols is leading to frequent cases of plant substitution and inadvertent or deliberate adulteration. There are several documented examples of commercial frauds where minor crops were substituted with related taxa showing a higher productivity or biomass but without the agronomical and nutritional characteristics of the original species/cultivars [27, 85, 86] (see also Table 1). Astounding cases of this phenomenon were observed for some of the most common spices such as the Mediterranean oregano adulterated with *Cistus incanus* L., *Rubus caesius* L. [87–89] and saffron substituted with *Crocus vernus* (L.) Hill, *Carthamus,* and *Curcuma* [19, 44, 84]. In this context, the use of DNA barcoding can be decisive because it can not only verify the presence/absence of the original species, but also identify the nature of the replaced species. One of the most striking substitution cases ever revealed by our investigations refers to fish meat (e.g., sold as slices, fillets, blocks, surimi, fish sticks, and fins). In this product category, the manufacturing processes often lead to the loss of any morphological diagnostic feature that may correctly identify the original species. In our molecular investigation [90], we documented the frequent substitutions of Palombo (i.e., the Italian vernacular name for *Mustelus mustelus* and *Mustelus asterias*) with other less valuable shark species. Our test showed that about 80% of the screened fish products did not correspond to these two species but to other species or genera,

Table 1: List of studies dealing with DNA barcoding identification of minor crops.

DNA barcoding application	Minor crop/food product	Notes	References
Traceability of minor crops in the food supply chain	Aromatic plants	Identification of spices from fresh samples to manufactured or processed products	[22, 43–47]
	Legumes	Legume seeds traceability	[48, 49]
	Herbal infusions	Traceability of tea products	[50]
	Fruit	Identification and traceability of mango	[51]
		Identification of *Citrus* species	[52]
		Identification of Goji	[53]
		Identification of berries	[54]
	Vegetables	Identification of *Capsicum* cultivars	[55]
	Medical plant and food supplements	Traceability of medicinal plants	[56–58]
Commercial frauds and dangerous substitutions	Aromatic products	Identification of spices adulterants	[59, 60]
	Vegetal flour	Identification of buckwheat in commercial foodstuffs	[61]
	Legumes	Seed admixture and adulteration	[62, 63]
	Fruit	Identification and adulteration of fruit-based products	[60, 64]
	Oil	Oil adulteration	[65]
	Medicinal plants/food supplements	Dangerous substitution of *Solanum lyratum* with *Aristolochia mollissima*	[66]
		Adulteration of herbal products	[67]
	Tea	Contamination of tea products	[68]
Molecular identification of minor crops in complex matrices	Natural health products	Identification of pharmaceutical plants in commercial products	[69]
	Juice and vegetal beverages	Juice authentication	[70–72]
	Honey	Identification of pollen and plant residuals	[73]
	Jams or yogurt	Identification of fruit in commercial products	[74, 75]
	Food supplements	Identification of allergenic plants	[76, 77]

some of which are fished or marketed illegally. Starting from this experience, we tested the usefulness of DNA barcoding to evaluate the contamination of plant-based products. For example, in a pilot study on spices conducted by our group, we detected contaminant DNA in commercial samples of sage (i.e., *Salvia*) produced by local farmers. This DNA corresponded to species belonging to the family Poaceae (i.e., *Festuca* sp.). We hypothesized that these contaminant plants were accidentally grown together with the sage and fragments of them were erroneously collected, shredded, and consequently admixed to the final commercial products (Labra M., unpublished data). These conditions are dangerous if the contaminant taxon is toxic or allergenic for humans. A typical example is that of nuts and almonds which cause allergies in many people [91]. Several commercial foodstuffs (e.g., bakery, pastry, and snacks) showed contamination by these plants (see, e.g., [76, 92]). Also in this case, DNA barcoding acts as a very versatile tool, allowing the detection of both species (and many other allergenic taxa) also when they were present in traces [76].

Similarly, DNA barcoding can be efficient in identifying those plant species causing intoxication or poisoning in consumers. In recent years, plant exposures are among the most frequent poisoning cases reported by poison control centers [15, 93, 94]. Many of these are due to inadvertent misidentification as reported in [95] where the authors documented the exchange of spontaneous salad (*Lactuca alpine* (L.) Wallr.) with *Aconitum* spp. and wild garlic (*Allium ursinum* L.) with *Colchicum* sp. Both *Aconitum* and *Colchicum* contain toxic metabolites with severe consequences for human health after ingestion [96, 97]. Our analysis showed that DNA barcoding allowed us to detect the presence of poisonous plants and identify specific sequence-characterized amplified regions (SCARs) useful in a real-time PCR approach for rapid diagnosis in poison centers [60].

4. Plant Molecular Identification in Complex Matrices

Most food and cosmetic products are made up of a pool of plant species, major and minor crops, and spontaneous species. These are considered complex matrices [31] and, to establish traceability, the availability of universal tools able to univocally identify each plant species is needed. We underline that the assumptions for which DNA barcoding

region(s) and the primers used are universal [33] imply that when the method is applied to complex matrices, PCR amplifications will produce several DNA barcoding amplicons, corresponding to different species. For this reason we tested this diagnostic method to identify the plant composition on different mixed products such as the commercial potpourris [14] and multiflower honeys (Bruni et al., submitted). For most of these herbal products, a detailed list of ingredients is not reported on the label; as a consequence, it is difficult to understand which species are used for their preparation and especially how safe these are for human health. In the case of potpourris, our results showed that the principal ingredients are simple aromatic plants (e.g., species of Lamiaceae) which are sometimes edible (e.g., *Salvia officinalis* L.; *Ocimum basilicum* L.) or ornamental (e.g., *Salvia splendens* Sellow ex J.A. Schultes, *Lavandula angustifolia* Miller) without negative effects on human health. In other cases these products revealed the presence of plants which produce natural toxic metabolites, such as alkaloids that are dangerous for human health [14, 98–100]. However, the main critical element for the identification of plant-based complex matrices is the availability of DNA barcoding reference databases [101, 102]. To date, the Barcode of Life Data System (i.e., BOLD, http://www.boldsystems.org/ [103]) contains 52,767 plant DNA sequences although several minor crops and local varieties are missing. Recent works, edited by our laboratory and other groups, highlighted the need for dedicated reference archives of DNA barcoding data for these kinds of plants [31, 67, 101, 102, 104, 105]. In another study, we demonstrated that, starting from a robust local database, it is possible to characterize the pollen composition of multiflower honey, one of the most complex food matrices. Our tests, conducted on honey samples produced in the Italian Alps, showed the conspicuous presence of endemic taxa. This result allowed us to assess not only the composition of honeys, but also their geographical origin (Bruni et al., submitted). See also Table 1 for further examples.

In comparison to agricultural products made by a single plant, the molecular characterization of complex matrices requires some technical advances, especially concerning the sequencing step. The traditional DNA-sequencing method [106] can only be adopted for direct sequencing of amplicons deriving from a single taxon. Complex matrices often contain mixtures of DNA from many individuals belonging to a certain taxonomic group (e.g., angiosperms) and DNA amplification may generate amplicons of the same size for a certain locus (e.g., a DNA barcode region for plant identification), therefore impeding direct sequencing with the Sanger approach. A possible solution could be the adoption of a preliminary cloning step to separate single DNA templates but this strategy has its own limitations (e.g., high costs) and can introduce biases (e.g., low representation of the sequenced colonies in the case of highly complex matrices [107, 108]). Recovering DNA sequences from the tens to thousands of specimens present in a complex food matrix requires the ability to read DNA from multiple templates in parallel. Since 2005, advances in the field of next-generation sequencing (NGS) technologies [109] have been helping in addressing this issue with ever-lowering costs. To date, several models of high-throughput sequencing devices have been commercially introduced based on different chemistries and detection techniques [108]. NGS technologies can generate up to tens of millions of sequencing reads in parallel and these approaches are being used in a variety of applications, including the traceability of food matrices containing agricultural products [73, 74, 110].

In conclusion, given the rapid evolution and standardization of NGS advances, we think that a universal approach such as DNA barcoding combined with them can offer a new opportunity for the traceability of minor crops from field to table.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] J. S. C. Smith, E. C. L. Chin, H. Shu et al., "An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPS and pedigree," *Theoretical and Applied Genetics*, vol. 95, no. 1-2, pp. 163–173, 1997.

[2] F. De Mattia, G. Lovicu, J. Tardaguila et al., "Genetic relationships between Sardinian and Spanish viticulture: the case of "Cannonau" and 'Garnacha'," *Journal of Horticultural Science and Biotechnology*, vol. 84, no. 1, pp. 65–71, 2009.

[3] S. R. McCouch, K. Zhao, M. Wright et al., "Development of genome-wide SNP assays for rice," *Breeding Science*, vol. 60, no. 5, pp. 524–535, 2010.

[4] J. Womach, *Agriculture: A Glossary of Terms, Programs, and Laws*, Congressional Research Service, Library of Congress, Washington, DC, USA, 2005.

[5] H. Amagase and N. R. Farnsworth, "A review of botanical characteristics, phytochemistry, clinical relevance in efficacy and safety of *Lycium barbarum* fruit (Goji)," *Food Research International*, vol. 44, no. 7, pp. 1702–1717, 2011.

[6] S. E. Kulling and H. M. Rawel, "Chokeberry (*Aronia melanocarpa*)-a review on the characteristic components and potential health effects," *Planta Medica*, vol. 74, no. 13, pp. 1625–1634, 2008.

[7] J. M. Urpí, J. C. Weber, and C. R. Clement, *Peach Palm, Bactris Gasipaes Kunth*, vol. 20, Bioversity international, Rome, Italy, 1997.

[8] S. Ketema, *Tef-Eragrostis tef (Zucc.)*, vol. 12, Bioversity international, Rome, Italy, 1997.

[9] M. Camciuc, M. Deplagne, G. Vilarem, and A. Gaset, "Okra—*Abelmoschus esculentus* L. (Moench.) a crop with economic potential for set aside acreage in France," *Industrial Crops and Products*, vol. 7, no. 2-3, pp. 257–264, 1998.

[10] K. T. Moe, S. Kwon, and Y. Park, "Trends in genomics and molecular marker systems for the development of some underutilized crops," *Genes and Genomics*, vol. 34, no. 5, pp. 451–466, 2012.

[11] E. Ernst, "The efficacy of herbal medicine-an overview," *Fundamental and Clinical Pharmacology*, vol. 19, no. 4, pp. 405–409, 2005.

[12] H. A. Tindle, R. B. Davis, R. S. Phillips, and D. M. Eisenberg, "Trends in use of complementary and alternative medicine

by us adults: 1997–2002," *Alternative Therapies in Health and Medicine*, vol. 11, no. 1, pp. 42–49, 2005.

[13] G. Heubl, "New aspects of DNA-based authentication of Chinese medicinal plants by molecular biological techniques," *Planta Medica*, vol. 76, no. 17, pp. 1963–1974, 2010.

[14] L. Cornara, B. Borghesi, C. Canali et al., "Smart drugs: green shuttle or real drug?" *International Journal of Legal Medicine*, vol. 127, no. 6, pp. 1109–1123, 2013.

[15] I. Bruni, F. De Mattia, A. Galimberti et al., "Identification of poisonous plants by DNA barcoding approach," *International Journal of Legal Medicine*, vol. 124, no. 6, pp. 595–603, 2010.

[16] S. L. Taylor and J. L. Baumert, "Cross-contamination of foods and implications for food allergic patients," *Current Allergy and Asthma Reports*, vol. 10, no. 4, pp. 265–270, 2010.

[17] Z. P. Zeng and J. G. Jiang, "Analysis of the adverse reactions induced by natural product-derived drugs," *British Journal of Pharmacology*, vol. 159, no. 7, pp. 1374–1391, 2010.

[18] J. Costa, I. Mafra, J. S. Amaral, and M. B. P. P. Oliveira, "Detection of genetically modified soybean DNA in refined vegetable oils," *European Food Research and Technology*, vol. 230, no. 6, pp. 915–923, 2010.

[19] S. Babaei, M. Talebi, and M. Bahar, "Developing an SCAR and ITS reliable multiplex PCR-based assay forsafflower adulterant detection in saffron samples," *Food Control*, vol. 35, no. 1, pp. 323–328, 2014.

[20] Y. J. Park, J. K. Lee, and N. S. Kim, "Simple sequence repeat polymorphisms (SSRPs) for evaluation of molecular diversity and germplasm classification of minor crops," *Molecules*, vol. 14, no. 11, pp. 4546–4569, 2009.

[21] S. Soares, I. Mafra, J. S. Amaral, and M. B. P. P. Oliveira, "A PCR assay to detect trace amounts of soybean in meat sausages," *International Journal of Food Science and Technology*, vol. 45, no. 12, pp. 2581–2588, 2010.

[22] F. De Mattia, I. Bruni, A. Galimberti, F. Cattaneo, M. Casiraghi, and M. Labra, "A comparative study of different DNA barcoding markers for the identification of some members of Lamiacaea," *Food Research International*, vol. 44, no. 3, pp. 693–702, 2011.

[23] A. K. Lockley and R. G. Bardsley, "DNA-based methods for food authentication," *Trends in Food Science and Technology*, vol. 11, no. 2, pp. 67–77, 2000.

[24] I. Mafra, I. M. Ferreira, and M. B. P. Oliveira, "Food authentication by PCR-based methods," *European Food Research and Technology*, vol. 227, no. 3, pp. 649–665, 2008.

[25] J. Novak, S. Grausgruber-Gröger, and B. Lukas, "DNA-based authentication of plant extracts," *Food Research International*, vol. 40, no. 3, pp. 388–392, 2007.

[26] M. Labra, M. Miele, B. Ledda, F. Grassi, M. Mazzei, and F. Sala, "Morphological characterization, essential oil composition and DNA genotyping of Ocimum basilicum L. cultivars," *Plant Science*, vol. 167, no. 4, pp. 725–731, 2004.

[27] M. Woolfe and S. Primrose, "Food forensics: using DNA technology to combat misdescription and fraud," *Trends in Biotechnology*, vol. 22, no. 5, pp. 222–226, 2004.

[28] S. Imazio, M. Labra, F. Grassi, A. Scienza, and O. Failla, "Chloroplast microsatellites to investigate the origin of grapevine," *Genetic Resources and Crop Evolution*, vol. 53, no. 5, pp. 1003–1011, 2006.

[29] F. De Mattia, F. Grassi, S. Imazio, and M. Labra, "Chloroplast and nuclear DNA markers to characterize cultivated and spontaneous *Ribes*," *Plant Biosystems*, vol. 142, no. 2, pp. 204–212, 2008.

[30] P. Kumar, V. K. Gupta, A. K. Misra, D. R. Modi, and B. K. Pandey, "Potential of molecular markers in plant biotechnology," *Plant Omics: Journal of Plant Molecular Biology & Omics*, vol. 2, no. 4, pp. 141–162, 2009.

[31] A. Galimberti, F. De Mattia, A. Losa et al., "DNA barcoding as a new tool for food traceability," *Food Research International*, vol. 50, no. 1, pp. 55–63, 2013.

[32] H. Chuang, H. Lur, K. Hwu, and M. Chang, "Authentication of domestic Taiwan rice varieties based on fingerprinting analysis of microsatellite DNA markers," *Botanical Studies*, vol. 52, no. 4, pp. 393–405, 2011.

[33] P. D. Hebert, S. Ratnasingham, and J. R. de Waard, "Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species," *Proceedings of the Royal Society B: Biological Sciences*, vol. 270, supplement 1, pp. S96–S99, 2003.

[34] M. Casiraghi, M. Labra, E. Ferri, A. Galimberti, and F. de Mattia, "DNA barcoding: a six-question tour to improve users'awareness about the method," *Briefings in Bioinformatics*, vol. 11, no. 4, Article ID bbq003, pp. 440–453, 2010.

[35] J. Shaw, E. B. Lickey, E. E. Schilling, and R. L. Small, "Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the Tortoise and the hare III," *American Journal of Botany*, vol. 94, no. 3, pp. 275–288, 2007.

[36] A. J. Fazekas, K. S. Burgess, P. R. Kesanakurti et al., "Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well," *PLoS ONE*, vol. 3, no. 7, Article ID e2802, 2008.

[37] A. J. Fazekas, P. R. Kesanakurti, K. S. Burgess et al., "Are plant species inherently harder to discriminate than animal species using DNA barcoding markers?" *Molecular Ecology Resources*, vol. 9, no. 1, pp. 130–139, 2009.

[38] D. Mathew, "Biotechnology," in *Horticulture: Methods and Applications*, K. V. Peter, Ed., chapter 2, pp. 25–50, New India Publishing Agency, New delhi, India, 1st edition, 2014.

[39] M. L. Hollingsworth, A. Andra Clark, L. L. Forrest et al., "Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants," *Molecular Ecology Resources*, vol. 9, no. 2, pp. 439–457, 2009.

[40] P. M. Hollingsworth, S. W. Graham, and D. P. Little, "Choosing and using a plant DNA barcode," *PLoS ONE*, vol. 6, no. 5, Article ID e19254, 2011.

[41] I. Bruni, F. De Mattia, S. Martellos et al., "DNA barcoding as an effective tool in improving a digital plant identification system: a case study for the area of Mt. Valerio, Trieste (NE Italy)," *PloS one*, vol. 7, no. 9, Article ID e43256, 2012.

[42] W. J. Kress, D. L. Erickson, N. G. Swenson, J. Thompson, M. Uriarte, and J. K. Zimmerman, "Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a puerto rican forest dynamics plot," *PLoS ONE*, vol. 5, no. 11, Article ID e15409, 2010.

[43] S. Federici, A. Galimberti, F. Bartolucci et al., "DNA barcoding to analyse taxonomically complex groups in plants: the case of Thymus (Lamiaceae)," *Botanical Journal of the Linnean Society*, vol. 171, no. 4, pp. 687–699, 2013.

[44] A. Gismondi, F. Fanali, J. M. M. Labarga, M. G. Caiola, and A. Canini, "*Crocus sativus* L. genomics and different DNA barcode applications," *Plant Systematics and Evolution*, vol. 299, no. 10, pp. 1859–1863, 2013.

[45] S. Theodoridis, A. Stefanaki, M. Tezcan, C. Aki, S. Kokkini, and K. E. Vlachonasios, "DNA barcoding in native plants of the

Labiatae (Lamiaceae) family from Chios Island (Greece) and the adjacent Çeşme-Karaburun Peninsula (Turkey)," *Molecular Ecology Resources*, vol. 12, no. 4, pp. 620–633, 2012.

[46] M. Kojoma, K. Kurihara, K. Yamada, S. Sekita, M. Satake, and O. Iida, "Genetic identification of cinnamon (*Cinnamomum* spp.) based on the trnL-trnF chloroplast DNA," *Planta Medica*, vol. 68, no. 1, pp. 94–96, 2002.

[47] M. Wang, H. Zhao, L. Wang et al., "Potential use of DNA barcoding for the identification of *Salvia* based on cpDNA and nrDNA sequences," *Gene*, vol. 528, no. 2, pp. 206–215, 2013.

[48] I. Ganopoulos, P. Madesis, N. Darzentas, A. Argiriou, and A. Tsaftaris, "Barcode High Resolution Melting (Bar-HRM) analysis for detection and quantification of PDO "fava Santorinis" (*Lathyrus clymenum*) adulterants," *Food Chemistry*, vol. 133, no. 2, pp. 505–512, 2012.

[49] P. Madesis, I. Ganopoulos, A. Anagnostis, and A. Tsaftaris, "The application of Bar-HRM (Barcode DNA-High Resolution Melting) analysis for authenticity testing and quantitative detection of bean crops (Leguminosae) without prior DNA purification," *Food Control*, vol. 25, no. 2, pp. 576–582, 2012.

[50] M. Y. Stoeckle, C. C. Gamble, R. Kirpekar, G. Young, S. Ahmed, and D. P. Little, "Commercial teas highlight plant DNA barcode identification successes and obstacles," *Scientific Reports*, vol. 1, p. 42, 2011.

[51] T. Hidayat, A. Pancoro, and D. Kusumawaty, "Utility of matK gene to assess evolutionary relationship of genus *Mangifera* (anacardiaceae) in Indonesia and Thailand," *Biotropia*, vol. 18, no. 2, pp. 74–80, 2011.

[52] J. Yu, H. X. Yan, Z. H. Lu, and Z. Q. Zhou, "Screening potential DNA barcode regions of chloroplast coding genome for *citrus* and its related genera," *Scientia Agricultura Sinica*, vol. 44, no. 2, pp. 341–348, 2011.

[53] T. Xin, H. Yao, H. Gao et al., "Super food *Lycium barbarum* (Solanaceae) traceability via an internal transcribed spacer 2 barcode," *Food Research International*, vol. 54, no. 2, pp. 1699–1704, 2013.

[54] L. Jaakola, M. Suokas, and H. Häggman, "Novel approaches based on DNA barcoding and high-resolution melting of amplicons for authenticity analyses of berry species," *Food Chemistry*, vol. 123, no. 2, pp. 494–500, 2010.

[55] R. L. Jarret, "DNA Barcoding in a crop genebank: the Capsicum annuum species complex," *Open Biology Journal*, vol. 1, pp. 35–42, 2008.

[56] S. Chen, H. Yao, J. Han et al., "Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species," *PLoS ONE*, vol. 5, no. 1, Article ID e8613, 2010.

[57] T. Gao, H. Yao, J. Song et al., "Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2," *Journal of Ethnopharmacology*, vol. 130, no. 1, pp. 116–121, 2010.

[58] Y. Zuo, Z. Chen, K. Kondo, T. Funamoto, J. Wen, and S. Zhou, "DNA barcoding of panax species," *Planta Medica*, vol. 77, no. 2, pp. 182–187, 2011.

[59] V. A. Parvathy, V. P. Swetha, T. E. Sheeja, N. K. Leela, B. Chempakam, and B. Sasikumar, "DNA barcoding to detect chilli adulteration in traded black pepper powder," *Food Biotechnology*, vol. 28, no. 1, pp. 25–40, 2014.

[60] S. Federici, D. Fontana, A. Galimberti et al., "A rapid diagnostic approach to identify poisonous plants using DNA barcoding data," *Plant Biosystems*. In press.

[61] T. Hirao, S. Imai, H. Sawada, N. Shiomi, S. Hachimura, and H. Kato, "PCR method for detecting trace amounts of buckwheat (Fagopyrum spp.) in food," *Bioscience, Biotechnology and Biochemistry*, vol. 69, no. 4, pp. 724–731, 2005.

[62] I. Ganopoulos, P. Madesis, and A. Tsaftaris, "Universal ITS2 Barcoding DNA Region Coupled with High-Resolution Melting (HRM) Analysis for Seed Authentication and Adulteration Testing in Leguminous Forage and Pasture Species," *Plant Molecular Biology Reporter*, vol. 30, no. 6, pp. 1322–1328, 2012.

[63] I. Bosmali, I. Ganopoulos, P. Madesis, and A. Tsaftaris, "Microsatellite and DNA-barcode regions typing combined with High Resolution Melting (HRM) analysis for food forensic uses: a case study on lentils (*Lens culinaris*)," *Food Research International*, vol. 46, no. 1, pp. 141–147, 2012.

[64] C. C. Ng, C. Y. Lin, W. S. Tzeng, C. C. Chang, and Y. T. Shyu, "Establishment of an internal transcribed spacer (ITS) sequence-based differentiation identification procedure for mei (*Prunus mume*) and plum (*Prunus salicina*) and its use to detect adulteration in preserved fruits," *Food Research International*, vol. 38, no. 1, pp. 95–101, 2005.

[65] I. Ganopoulos, C. Bazakos, P. Madesis, P. Kalaitzis, and A. Tsaftaris, "Barcode DNA high-resolution melting (Bar-HRM) analysis as a novel close-tubed and accurate tool for olive oil forensic use," *Journal of the Science of Food and Agriculture*, vol. 93, no. 9, pp. 2281–2286, 2013.

[66] M. Li, K. Au, H. Lam et al., "Identification of Baiying (Herba Solani Lyrati) commodity and its toxic substitute Xungufeng (Herba Aristolochiae Mollissimae) using DNA barcoding and chemical profiling techniques," *Food Chemistry*, vol. 135, no. 3, pp. 1653–1658, 2012.

[67] S. G. Newmaster, M. Grguric, D. Shanmughanandhan, S. Ramalingam, and S. Ragupathy, "DNA barcoding detects contamination and substitution in North American herbal products," *BMC Medicine*, vol. 11, no. 1, p. 222, 2013.

[68] B. Dhiman and M. Singh, "Molecular detection of Cashew Husk (*Anacardium occidentale*) adulteration in market samples of dry tea (*Camellia sinensis*)," *Planta Medica*, vol. 69, no. 9, pp. 882–884, 2003.

[69] L. J. Wallace, S. M. A. L. Boilard, S. H. C. Eagle, J. L. Spall, S. Shokralla, and M. Hajibabaei, "DNA barcodes for everyday life: routine authentication of Natural Health Products," *Food Research International*, vol. 49, no. 1, pp. 446–452, 2012.

[70] M. A. Faria, A. Magalhães, M. E. Nunes, and M. B. P. P. Oliveira, "High resolution melting of *trnL* amplicons in fruit juices authentication," *Food Control*, vol. 33, no. 1, pp. 136–141, 2013.

[71] M. Li, K. Wong, W. Chan et al., "Establishment of DNA barcodes for the identification of the botanical sources of the Chinese "cooling" beverage," *Food Control*, vol. 25, no. 2, pp. 758–766, 2012.

[72] J. Han, Y. Wu, W. Huang et al., "PCR and DHPLC methods used to detect juice ingredient from 7 fruits," *Food Control*, vol. 25, no. 2, pp. 696–703, 2012.

[73] A. Valentini, C. Miquel, and P. Taberlet, "DNA barcoding for honey biodiversity," *Diversity*, vol. 2, no. 4, pp. 610–617, 2010.

[74] A. Ortola-Vidal, H. Schnerr, M. Rojmyr, F. Lysholm, and A. Knight, "Quantitative identification of plant genera in food products using PCR and Pyrosequencing technology," *Food Control*, vol. 18, no. 8, pp. 921–927, 2007.

[75] M. Arleo, F. Ruibal, J. Pereyra, E. Miquel, M. Fernández, and C. Martínez, "A DNA-based approach to discriminate between quince and apple in quince jams," *International Food Research Journal*, vol. 19, no. 4, pp. 1471–1477, 2012.

[76] T. Yano, Y. Sakai, K. Uchida et al., "Detection of walnut residues in processed foods by polymerase chain reaction," *Bioscience,*

Biotechnology and Biochemistry, vol. 71, no. 7, pp. 1793–1796, 2007.

[77] P. Madesis, I. Ganopoulos, I. Bosmali, and A. Tsaftaris, "Barcode High Resolution Melting analysis for forensic uses in nuts: a case study on allergenic hazelnuts (*Corylus avellana*)," *Food Research International*, vol. 50, no. 1, pp. 351–360, 2013.

[78] G. Singh, *Plant Systematics: an Integrated Approach*, Science Publishers, New York, NY, USA, 2004.

[79] H. Trindade, "Molecular biology of aromatic plants and spices. A review," *Flavour and Fragrance Journal*, vol. 25, no. 5, pp. 272–281, 2010.

[80] M. Viuda-Martos, Y. Ruiz-Navajas, J. Fernández-López, and J. A. Pérez-Álvarez, "Spices as functional foods," *Critical Reviews in Food Science and Nutrition*, vol. 51, no. 1, pp. 13–28, 2011.

[81] A. Paton, M. R. Harley, and M. M. Harley, "*Ocimum*: an overview of classification and relationships," in *Basil: The Genus Ocimum*, pp. 1–38, 1999.

[82] A. O. Tucker, "The truth about mints," *Herb Companion*, vol. 4, pp. 51–52, 1992.

[83] V. Gobert, S. Moja, M. Colson, and P. Taberlet, "Hybridization in the section *Mentha* (Lamiaceae) inferred from AFLP markers," *American Journal of Botany*, vol. 89, no. 12, pp. 2017–2023, 2002.

[84] A. Torelli, M. Marieschi, and R. Bruni, "Authentication of saffron (*Crocus sativus* L.) in different processed, retail products by means of SCAR markers," *Food Control*, vol. 36, no. 1, pp. 126–131, 2014.

[85] K. Dhanya and B. Sasikumar, "Molecular marker based adulteration detection in traded food and agricultural commodities of plant origin with special reference to spices," *Current Trends in Biotechnology and Pharmacy*, vol. 4, no. 1, pp. 454–489, 2010.

[86] P. Posadzki, L. Watson, and E. Ernst, "Contamination and adulteration of herbal medicinal products (HMPs): an overview of systematic reviews," *European Journal of Clinical Pharmacology*, vol. 69, no. 3, pp. 295–307, 2013.

[87] M. Marieschi, A. Torelli, F. Poli, A. Bianchi, and R. Bruni, "Quality control of commercial Mediterranean oregano: development of SCAR markers for the detection of the adulterants *Cistus incanus* L., *Rubus caesius* L. and *Rhus coriaria* L.," *Food Control*, vol. 21, no. 7, pp. 998–1003, 2010.

[88] M. Marieschi, A. Torelli, A. Bianchi, and R. Bruni, "Detecting *Satureja montana* L. and *Origanum majorana* L. by means of SCAR-PCR in commercial samples of Mediterranean oregano," *Food Control*, vol. 22, no. 3-4, pp. 542–548, 2011.

[89] M. Marieschi, A. Torelli, A. Bianchi, and R. Bruni, "Development of a SCAR marker for the identification of *Olea europaea* L.: a newly detected adulterant in commercial Mediterranean oregano," *Food Chemistry*, vol. 126, no. 2, pp. 705–709, 2011.

[90] M. Barbuto, A. Galimberti, E. Ferri et al., "DNA barcoding reveals fraudulent substitutions in shark seafood products: the Italian case of "palombo" (*Mustelus* spp.)," *Food Research International*, vol. 43, no. 1, pp. 376–381, 2010.

[91] Z. Hubalkova and E. Rencova, "One-step multiplex PCR method for the determination of pecan and Brazil nut allergens in food products," *Journal of the Science of Food and Agriculture*, vol. 91, no. 13, pp. 2407–2411, 2011.

[92] J. Costa, I. Mafra, I. Carrapatoso, and M. B. P. P. Oliveira, "Almond allergens: molecular characterization, detection, and clinical relevance," *Journal of Agricultural and Food Chemistry*, vol. 60, no. 6, pp. 1337–1349, 2012.

[93] F. M. Hammouda, A. M. Rizk, M. M. El-Missiry et al., "Poisonous plants contaminating edible ones and toxic substances in plant foods. IV. Phytochemistry and toxicity of *Lolium temulentum*," *International Journal of Crude Drug Research*, vol. 26, no. 4, pp. 240–245, 1988.

[94] R. Walker, "Criteria for risk assessment of botanical food supplements," *Toxicology Letters*, vol. 149, no. 1–3, pp. 187–195, 2004.

[95] M. L. Colombo, F. Assisi, T. D. Puppa et al., "Most commonly plant exposures and intoxications from outdoor toxic plants," *Journal of Pharmaceutical Sciences and Research*, vol. 2, no. 7, pp. 417–425, 2010.

[96] Y. Finkelstein, S. E. Aks, J. R. Hutson et al., "Colchicine poisoning: the dark side of an ancient drug," *Clinical Toxicology*, vol. 48, no. 5, pp. 407–414, 2010.

[97] M. A. Berdai, S. Labib, K. Chetouani, and M. Harandou, "*Atropa Belladonna* intoxication: a case report," *Pan African Medical Journal*, vol. 11, p. 72, 2012.

[98] E. Röder, "Medicinal plants in Europe containing pyrrolizidine alkaloids," *Pharmazie*, vol. 50, no. 2, pp. 83–98, 1995.

[99] C. Franz, R. Chizzola, J. Novak, and S. Sponza, "Botanical species being used for manufacturing plant food supplements (PFS) and related products in the EU member states and selected third countries," *Food and Function*, vol. 2, no. 12, pp. 720–730, 2011.

[100] H. Wiedenfeld and J. Edgar, "Toxicity of pyrrolizidine alkaloids to humans and ruminants," *Phytochemistry Reviews*, vol. 10, no. 1, pp. 137–151, 2011.

[101] K. S. Burgess, A. J. Fazekas, P. R. Kesanakurti et al., "Discriminating plant species in a local temperate flora using the rbcL+matK DNA barcode," *Methods in Ecology and Evolution*, vol. 2, no. 4, pp. 333–340, 2011.

[102] A. Sandionigi, A. Galimberti, M. Labra et al., "Analytical approaches for DNA barcoding data-how to find a way for plants?" *Plant Biosystems*, vol. 146, no. 4, pp. 805–813, 2012.

[103] S. Ratnasingham and P. D. N. Hebert, "BOLD: the barcode of life data system: barcoding," *Molecular Ecology Notes*, vol. 7, no. 3, pp. 355–364, 2007.

[104] F. De Mattia, R. Gentili, I. Bruni et al., "A multi-marker DNA barcoding approach to save time and resources in vegetation surveys," *Botanical Journal of the Linnean Society*, vol. 169, no. 3, pp. 518–529, 2012.

[105] M. L. Kuzmina, K. L. Johnson, H. R. Barron, and P. D. N. Hebert, "Identification of the vascular plants of Churchill, Manitoba, using a DNA barcode library," *BMC Ecology*, vol. 12, p. 25, 2012.

[106] F. Sanger, S. Nicklen, and A. R. Coulson, "DNA sequencing with chain-terminating inhibitors," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, no. 12, pp. 5463–5467, 1977.

[107] M. Hajibabaei, S. Shokralla, X. Zhou, G. A. C. Singer, and D. J. Baird, "Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos," *PLoS ONE*, vol. 6, no. 4, Article ID e17497, 2011.

[108] S. Shokralla, J. L. Spall, J. F. Gibson, and M. Hajibabaei, "Next-generation sequencing technologies for environmental DNA research," *Molecular Ecology*, vol. 21, no. 8, pp. 1794–1805, 2012.

[109] M. L. Metzker, "Sequencing technologies—the next generation," *Nature Reviews Genetics*, vol. 11, no. 1, pp. 31–46, 2010.

[110] M. L. Coghlan, J. Haile, J. Houston et al., "Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns," *PLoS Genetics*, vol. 8, no. 4, Article ID e1002657, 2012.

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 26 (1): 13–24, 2015
OPEN ACCESS
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/11347/pdf
doi:10.4404/hystrix-26.1-11347

Commentary

DNA barcoding in mammals: what's new and where next?

Andrea Galimberti[a], Anna Sandionigi[a], Antonia Bruno[a], Adriana Bellati[b], Maurizio Casiraghi[a,*]

[a] *ZooPlantLab, Dipartimento di Biotecnologie e Bioscienze, Università degli Studi di Milano-Bicocca,Piazza della Scienza 2, 20126 Milano, Italy.*
[b] *Dipartimento di Scienze della Terra e dell'Ambiente, Università degli Studi di Pavia, Via Ferrata 1, 27100 Pavia, Italy*

Keywords:
Biological databases
integrative taxonomy
molecular identification
theriology
wildlife forensics

Article history:
Received: 4 June 2015
Accepted: 25 June 2015

Acknowledgements
The authors are indebted with Monica Pozzi for the linguistic revision of the manuscript. The authors are also grateful to the anonymous referees for their useful comments that substantially improved the final version of this work.

Abstract

DNA barcoding is a universal molecular identification system of living beings for which efficacy and universality have been largely demonstrated in the last decade in many contexts. It is common to link DNA barcoding to phylogenetic reconstruction, and there is indeed an overlap, but identification and phylogenetic positioning/classification are two different processes. In mammals, a better phylogenetic reconstruction, able to dig in fine details the relationships among biological entities, is really welcomed, but do we need DNA barcoding too? In our opinion, the answer is positive, but not only for the identification power, nor for the supposed ability of DNA barcoding to discover new species. We do need DNA barcoding because it is a modern tool, able to create an integrated system, in which it is possible to link the many aspects of the biology of living beings starting from their identification. With 7000 species estimated and a growing interest in knowledge, exploitation and conservation, mammals are one of the best animal groups to achieve this goal.

We organised our review to show how an integrative approach to taxonomy, leaded by DNA barcoding, can be effective in the twenty-first century identification and/or description of species.

Introduction

Mammals represent a relatively small animal group, with 5564 species listed in the Catalogue of Life (ITIS database, http://www.catalogueoflife.org). Being our own class, it is thought that these species are among the most known animals, especially regarding taxonomic aspects (Wilson and Reeder, 2006).

Generally speaking this is correct, but there are relevant exceptions, even on (presumably) well-established species. The case of African bush and forest elephants is emblematic. In 2001 the populations of bush and forest elephants were split in two distinct species, *Loxodonta africana* (Blumenbach, 1797) and *L. cyclotis* (Matschie, 1900), using molecular data to support this separation (Roca et al., 2001). It is clear that there is a hidden biodiversity within the mammal record, the extent of which is still under discussion, but surely in some groups like chiroptera, it has a deep impact on the taxonomy (see for example Galimberti et al., 2012b and Bogdanowicz et al., 2015). On the whole, the estimation of the unknown biodiversity in mammals is not so trivial, but there is an agreement on the number of about 7000 species (Reeder et al., 2007). The question is now simple: how to discover them?

Since 2003, DNA barcoding has been claimed to be an innovative and revolutionary approach to identify living beings, and a way to speed up the writing of "the encyclopedia of life" (Savolainen et al., 2005). In other words, the technique would be a system to increase the efficiency in species discovery. DNA barcoding has many advantages, but criticisms raised against the ability to discover new species (see for a review Casiraghi et al., 2010). The signature of the success of DNA barcoding is evident from the many group-specific or environment-specific campaigns launched in the past years (see an updated list of them at the international Barcode of Life initiative, www.ibol.org). Figure 1 shows a simplistic analysis of the publications on DNA barcoding in vertebrates since the seminal paper by Paul Hebert was issued in 2003 (Hebert et al., 2003). The figure has to be carefully taken into consideration because it does not represent a full bibliometric analysis as many articles do not include barcoding keywords in their title or abstracts (see Fig. 1 caption for more details), making this schematization certainly incomplete. However, Fig. 1 clearly shows that DNA barcoding in vertebrates is still largely diffused among fishes (probably for their importance in the global food market and for the frequent occurrence of frodes, mislabelling, species substitution to which they are subjected, see for instance Barbuto et al., 2010), whereas this tendency is not found in other vertebrates.

The DNA barcoding of mammals is ongoing under the auspices of the iBOL. According to the BOLD System (http://www.boldsystems.org) at the end of May 2015 about 2850 mammal species have been barcoded, and at least 300 unnamed clusters (i.e. not assigned taxonomic rank) are recognised on MammaliaBoL. In Fig. 2, the DNA barcoding coverage in mammal known species is plotted. As a consequence, given the 7000 presumed mammal species, there are DNA barcodes for about 45% of them. This also means that even if we believe in the species discovery power of DNA barcoding, it is difficult to think that this would be the main support for the mammal initiative. It could be a relevant drive in other animals, but not in mammals. In the modern taxonomy, identification and classification are two different processes (Casiraghi et al., 2010) and in mammals the main problem is related to the phylogenetic reconstruction, that is not, in a strict sense, DNA barcoding (Rodrigues et al., 2011; Huang et al., 2012).

DNA barcoding is more than a simple identification system and its major strength is beyond the discrimination power. In this context, DNA barcoding in mammals moved forward from the identification, becoming a "service system" useful for several aspects originating from taxonomy, but being relevant in other areas of the biology of mammals, ranging from distribution to behaviour and conservation.

So, the time is ripe to ask a fundamental question: do we still need DNA barcoding in mammals? We wrote our essay to solve this ques-

*Corresponding author
Email address: maurizio.casiraghi@unimib.it (Maurizio Casiraghi)

©©©©2015 Associazione Teriologica Italiana
doi:10.4404/hystrix-26.1-11347

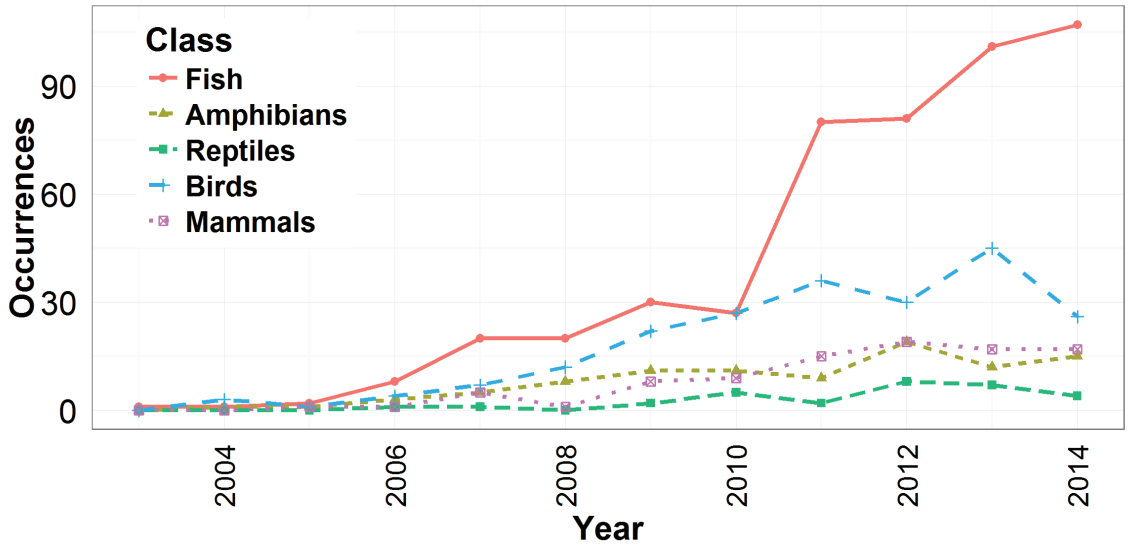


Figure 1 – A schematic overview of the tendencies in published papers on DNA barcoding in vertebrates from the beginning of the initiative (2003) to the end of 2014. Please note that the graphic is not exhaustive and it has been generated interlinking different keywords searches on ISI WEB of Science. Mammalia: barcode mammals; barcoding mammals; barcode mammal; barcoding mammal; barcoding mammalia; barcode mammalia. Aves: barcode bird; barcoding bird; barcode birds; barcoding birds; barcoding aves; barcode aves. Amphibia: barcode amphibian; barcoding amphibian; barcode amphibians; barcoding amphibians; barcoding amphibia; barcode Amphibia. Reptilia: barcode reptiles; barcoding reptile; barcode reptiles; barcoding reptiles; barcoding reptilia; barcode reptilia. Fish: barcoding fishes; barcode fishes; barcoding cartilaginous fish; barcode cartilaginous fish; barcoding fish; barcode fish; barcoding Agnatha; barcode Agnatha; barcoding Osteichthyes; barcode Osteichthyes; barcoding bony fishes; barcode bony fishes.

tion, and the different sections listed below are the different answers we can give.

The importance of reference databases

In DNA barcoding, the identification procedure involves the assignment of taxonomic names to unknown specimens using a DNA reference library of vouchers, previously identified trough different criteria. Such reference accessions and the international platforms in which they are organized, constitute the scaffold of the DNA barcoding initiative. Reference DNA barcodes often derive from natural history museums or private collections (Puillandre et al., 2012) as the role of these institutions has always been that of storing, univocally labelling and sharing the reference biological material for taxonomists. In the not-molecularized biology, most of the work of taxonomists was entirely based on the comparison between newly collected or already archived material and the one of other collections. In the case of mammals, one of the main challenges for a taxonomist relies on the fact that the largest reference collections are scattered among museums. This generated some paradoxes with researchers working in tropical biodiversity hotspots that have to move to North America and Europe to examine the largest collections of mammals inhabiting their own species-rich areas (Francis et al., 2010).

The advent of DNA barcoding moved forward allowing contemporary taxonomists to make comparisons with other taxonomic material, even at a distance with consequent benefits in terms of time and resources saved. In addition, ongoing improvements in molecular technology permit to cheaply obtain high quality sequences from very small and long-time preserved tissue samples like those stored in museums (Mitchell, 2015). These advances boosted the researches in mammalogy for several reasons. First, the possibility of confirming the identification of specimens through DNA barcodes allows museums to establish reference collections that can serve as a basis for future research including the description of new biological entities (Puillandre et al., 2012). Second, the standardized molecular reexamination of museum-deposited voucher specimens and the comparison with other reference data permits to rapidly "flag" the identification mistakes typically occurring during field surveys. As pointed out by Francis and co-workers 2010, field determinations for many mammal species are difficult, because they require the analysis of internal morphology (e.g., skull or dentition) and are often biased by age/sex variations, undescribed/extralimital species and lack of comparative material. Finally, the digital nature of genetic information (the so-called "computerization" *sensu* Casiraghi et al., 2010) makes DNA barcoding data readily comparable through publicly accessible online databases thus providing a wide panel of potential applications ranging from progresses in taxonomy to the fields of forensics and food traceability (see dedicated paragraphs of this review).

Concerning this last point, in the framework of the International Barcode of Life (iBOL) initiative, the building of a comprehensive public library of DNA barcodes, the Barcode of Life Data System (BOLD), was launched to provide a global identification system freely accessible (Ratnasingham and Hebert, 2007, 2013). This platform consists of several components, among which the Identification Engine tool (BOLD-IDS) is one of the most useful. BOLD-IDS provides a species identification tool that accepts DNA barcode sequences and returns a taxonomic assignment at the species level whenever possible.

Unlike other international sequence databases (such as EMBL and GenBank), BOLD has a quality control system built in, and specific information is required to store and publish a specimen or barcode. To be included in BOLD, specimens have to be properly vouchered following the protocol specified by the Global Registry of Biodiversity Repositories (http://grbio.org), and the data standards for BARCODE Records (Hanner, 2009). Moreover, required details on the sample include the collection date and location with GPS coordinates, and the PCR primers used to generate the sequences. Finally, submission of the original trace files is also needed. Noteworthy, barcode sequences in BOLD are associated with specimen records linked to institutional (e.g., museum) material making them the most valuable among putative reference accessions.

The accuracy of DNA barcoding species assignment relies upon the level of taxonomic representation for each group of metazoans and the amount of intraspecific genetic diversity represented in the databases (Gaubert et al., 2014).

In the case of mammals, assembling a reference database of DNA barcode sequences is fundamental for the goals of the iBOL initiative, also considering that the rate of species discovery within this class has recently accelerated due to the growing use of molecular techniques (Reeder et al., 2007).

Differently from larger DNA barcoding campaigns focusing on fishes (i.e., FISH-BOL, Becker et al., 2011), birds (i.e., ABBI, Hebert et al., 2004), insects (Jinbo et al., 2011) and others, there have only been a few references on mammals, generally focusing on a limited number of taxa or geographic areas. As of 2015, more than 69000 barcode mammalian sequences from over 2800 species have been archived in BOLD with more than 50% assembled at the Biodiversity Institute of Ontario in collaboration with the Royal Ontario Museum (ROM) and other institutions. The most part of these data belong to bats, rodents and primates from the Neotropical Region and other tropical biodiversity hotspots (Lim, 2012 and Fig. 2).

To date, the largest published studies on mammals DNA barcoding are those by Francis et al. (2010) and Clare et al. (2011), where the authors examined 1896 specimens belonging to 157 species from the South East Asia and 9076 specimens belonging to 163 species from the Neotropics respectively. Table 1 provides an updated list of the major studies that contributed to populate the current reference DNA barcoding database for mammals. Although most of these are limited to a reduced number of species or geographical extent, they are important in

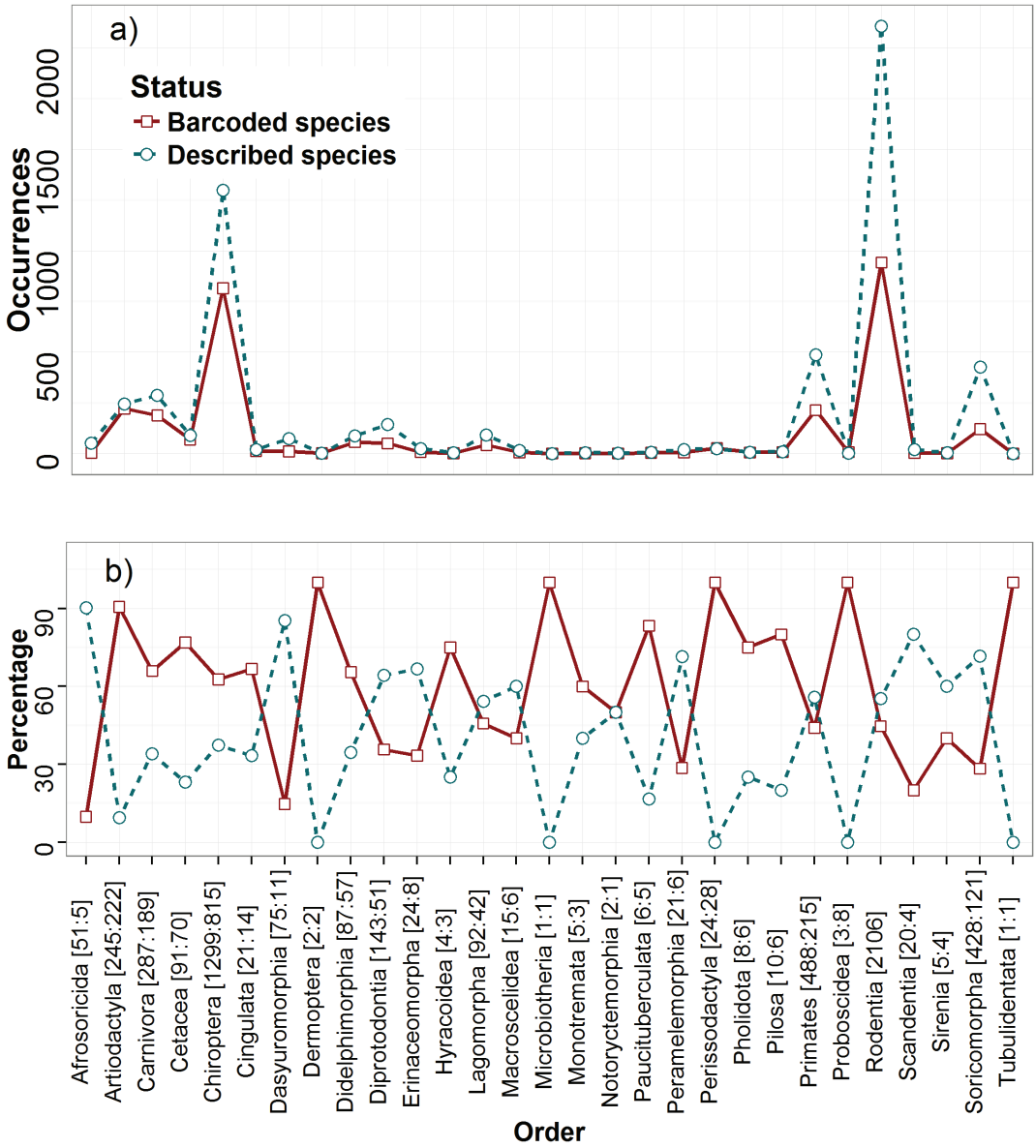


Figure 2 – Overview of the Mammalian DNA barcoding initiative showing the distribution of barcoded species in the different orders. Data on described species is derived from Integrated Taxonomic Information System (ITIS, http://www.itis.gov). Data on barcoded species is derived from the Barcode of Life Data Systems (BOLD System, http://www.boldsystems.org). In a) the number of species described and barcoded is plotted in the various mammal orders. In b) the percentage of species described and barcoded is plotted in the various mammal orders. Dotted line: described species (number or percentage). Continuous line: species with a DNA barcode.

filling the gaps of knowledge for many taxonomic groups, discovering new species or lineages and enabling potential effective conservation planning. The availability of a public database of reference specimens and related genetic data of mammal species is also at the base of wildlife forensics as for example recommended by the International Society for Forensic Genetics Commission (Linacre et al., 2011; Johnson et al., 2014).

Increasing knowledge on biology, distribution and conservation

As a matter of fact, the primary role of DNA barcoding in mammals has been so far, and will long remain, the identification of known species and one of the most rapid approaches to detect new ones, the so-called "DNA barcoding *sensu stricto*". Table 1 provides a list of case studies where DNA barcoding was successfully used in many application contexts to identify mammal species.

However, the "*sensu lato*" face of the approach (see Casiraghi et al., 2010), is even more interesting as it provides new information on the biology, distribution and conservation of mammals.

First of all, DNA-based techniques and consequently DNA barcoding are valid data generators to increase the existing knowledge on rare or poorly investigated taxa. In most cases, the analysis of barcode sequences allowed to confirm the occurrence of certain species in areas out of their known distributional range such as bats (e.g., De Pasquale and Galimberti, 2014) and Artiodactyla (e.g., Wilsonet al., 2014). The implications in a context of conservation are numerous and many studies supported the use of DNA barcoding in recognizing rare or elusive mammal species traditionally monitored with expensive field techniques (i.e., direct observations, captures and camera traps). DNA barcoding proved to be more effective in discriminating morphologically similar species, such as small ungulates and carnivores, which were difficult to recognize using camera traps (Inoue and Akomo-Okoue, 2015). In these cases, great advantage was provided by the possibility of identifying species from a part of the animal (i.e., hair/fur, claws, or skin) or its droppings as well described in recent case studies conducted in Amazonian and other unexplored areas of the planet (Michalski et al., 2011; De Matteo et al., 2014; Stanton et al., 2014; Inoue and Akomo-Okoue, 2015).

In other situations, the DNA barcoding approach could flag the occurrence of newly undescribed lineages that are confined to a certain geographic area or could represent a new taxa. Apart from the light and shadows of the method in a pure taxonomic context, an aspect of primary importance is the possibility of rapidly detecting putative units deserving further investigations to characterize their ecology, distribution and conservation status. Such kind of approach is fundamental to plan early and effective conservation strategies. Several studies proved the role of DNA barcoding in this framework such as in the case of Italian echolocating bats (Galimberti et al., 2012b) where the authors found, starting from DNA barcoding, a new well diverged lineage of *Myotis nattereri* in Southern Italy and several less divergent lineages within *M. bechsteinii* and *Plecotus auritus* from different areas of the Peninsula. A greater diversity was also found within neotropical bats in which Clare and colleagues 2011 found supported evidence of the existence of previously undescribed lineages for at least 44 species out of the 163 examined by DNA barcoding.

Invaluable data on mammal ecology and their conservation also derive from the characterization of their diets which has been conducted in many cases with a DNA barcoding approach. Understanding trophic interactions is fundamental also to assess the importance of certain species for ecosystems functioning and how they respond to variation (Clare et al., 2014a). The recent exploitation of High Throughput DNA Sequencing techniques (see below) allowed to characterize mixed DNA samples (e.g., stomach contents or faecal samples) and to identify the preys consumed by a given predator (Boyer et al., 2015). Such analyses revealed for example temporal and spatial variation patterns in the use of arthropod resources by different bat species (Clare et al., 2014b; Rasgour et al., 2011; Alberdi et al., 2012; Vesterinen et al., 2013; Hope et al., 2014) or diet differentiation between species and/or during different phenological periods (Bohmann et al., 2011; Burgar et al., 2014; Krüger et al., 2014a,b; Sedlock et al., 2014).

In conclusion, we are now aware that in mammals, even more than in other animals, we need to collect complementary data to better understand their biology. The system generated by DNA barcoding has the possibility to rapidly increase these knowledge.

Forensic applications

Given its peculiarities as a universal identification tool, DNA barcoding naturally acquired a role of primary importance in forensic (Dawnay et al., 2007; Iyengar, 2014), including case studies on animal derived foodstuff (e.g., Barbuto et al., 2010; Galimberti et al., 2013). In particular, wildlife forensic is a wide-ranging discipline covering more forms of crimes compared to human forensic. Concerning mammals, typical investigations include: trafficking in live specimens or parts of them, poaching or hunting out of season, cruelty to animals, habitat destruction and species substitution of food products (e.g., the bushmeat). These phenomena are of major concern also considering their economic impact at the global scale. For instance, recent estimates highlighted that a significant portion of the international trade of wildlife and wildlife products is illegal (i.e., 5–8 billion US $ of the total 6-20 billion US $, Baker, 2008) and includes species that are protected by national laws and international conventions (Eaton et al., 2010). Given the illicit nature of these activities, it is almost impossible to monitor and quantify the exact volumes and species involved as well as the real impact on wildlife populations (Gavin et al., 2010; Conteh et al., 2015). However, in the last century, the tremendous global collapse of some species that are object of illegal trade confirms the emerging problem of wildlife crimes (see for example the cases of *Panthera tigris* and *Diceros bicornis* which populations have decreased of 90% and 96% respectively in few decades; Linacre and Tobe, 2011). The biological material that is traded and analyzed in wildlife forensic is vast, ranging from whole animals (live, hunted or inadvertently killed) to skins, skeletons or animal body parts (e.g., meat, horns and teeth) (Huffman and Wallace, 2012; Johnson et al., 2014). In other cases, the only available material is blood, hairs and trace DNA or mixtures of genetic material (Johnson et al., 2014). Apart from clearly unmistakable species (e.g., an elephant tusk or a skin of a big carnivore), the morphological approach used for identification has usually to be undertaken by an expert mammalogist (Huffman and Wallace, 2012). Also microscopy of hairs or the analysis of bones require high-skilled experience to achieve a reliable identification, and even so, in some cases they failed to go further from a general group of putative species (see examples in Moore, 1988). Indeed, the strong processing of the wildlife raw material that can be finally traded as fillets, powders, potions or oils, often impedes unequivocal identification with morphology. In addition, both general operators and specialists are sometimes required to investigate on species that have not previously been studied in a forensic context and therefore lacks of accurate morphological reference data.

Given these premises, it is clear that universal, fast and accurate methods of species identification are necessary to improve the ability of detecting, monitoring and controlling the trade in mammals and other groups of animals (and their processed products).

In the last decades, the advent of DNA-based technologies revolutionized the field of wildlife forensic as DNA tools offered the possibility of overcoming the limits described above. Concerning species identification, several approaches and loci were selected, but in the last 10 years, DNA barcoding and the use of the mitochondrial cytochrome-c-oxidase subunit (i.e. mt-*coxI*) rapidly affirmed their utility in those cases involving crimes against mammals. Literature and examples are numerous, and three main categories of wildlife forensic investigations where DNA barcoding is successfully adopted can be identified:

Illegal hunting and traceability of wild game

The unregulated hunting of wildlife is an emerging issue as it involves the harvesting of millions of tons of wild game -– mostly mammals -– per year (Eaton et al., 2010; Gaubert et al., 2014). Conservation problems are typically referred to the "bushmeat" hunting which includes

Table 1 – Updated list of case studies dealing with mammals DNA barcoding. For each study, the context of application, the taxonomic order of target mammals, the aim of the work and the number of species involved are reported .

Context	Order	Aim	N°species	References
DNA taxonomy	Chiroptera	Characterization of Guyana bat species	87	Clare et al., 2007
DNA taxonomy	Chiroptera	Identification of a new species of Malaysian bat	1	Francis et al., 2007
DNA taxonomy	various	Characterization of small mammal communities of Suriname	74	Borisenko et al., 2008
DNA taxonomy	Didelphimorphia	Identification of cryptic species of opossum	2	Cervantes et al., 2010
DNA taxonomy	Primates	Characterization of primates species	50	Nijman and Aliabadian, 2010
DNA taxonomy	Chiroptera	Characterization of Malaysian wolly bats	6	Khan et al., 2010
DNA taxonomy	Chiroptera	Characterization of Southeast Asian bats	165	Francis et al., 2010
DNA taxonomy	Soricomorpha	Characterization of white-toothed shrews from Vietnam	6	Bannikova et al., 2011
DNA taxonomy	Artiodactyla	Characterization of Tanzanian antelopes	20	Bitanyi et al., 2011
DNA taxonomy	Artiodactyla	Characterization of Chinese bovidae	18	Cai et al., 2011
DNA taxonomy	Chiroptera	Characterization of Neotropical bats	163	Clare, 2011
DNA taxonomy	Chiroptera	Characterization of Ecuadorian bats	45	McDonough et al., 2011
DNA taxonomy	Soricomorpha	Characterization of shrews from Guinea	10	Jacquet et al., 2012
DNA taxonomy	Didelphimorphia	Characterization of opossum species in Brazilian Atlantic Rainforest	2	Sousa et al., 2012
DNA taxonomy	Cetacea	Characterization of Cetacean species	61	Viricel and Rosel, 2012
DNA taxonomy	Rodentia	Characterization of Chinese small mammals	11	Lu et al., 2012
DNA taxonomy	Rodentia	Characterization of species in the Praomyini tribe	40	Nicolas et al., 2012
DNA taxonomy	Chiroptera	Characterization of Neotropical *Myotis* bats	18	Larsen et al., 2012
DNA taxonomy	Chiroptera	Characterization of Italian echolocating bats	31	Galimberti et al., 2012b
DNA taxonomy	Chiroptera	Charachterization of the Mexican funnel-eared bats	2	López-Wilchis et al., 2012
DNA taxonomy	Chiroptera	Characterization of Yucatan phyllostomid bats	20	Hernández-Dávila et al., 2012
DNA taxonomy	Didelphimorphia	Characterization of atlantic forest didelphid marsupials	11	Agrizzi et al., 2012
DNA taxonomy	Rodentia	Characterization of minibarcode regions for rodents identification	103	Galan et al., 2012
DNA taxonomy	Chiroptera	Characterization of genetic diversity of northeastern Palearctic bats	38	Kruskop et al., 2012
DNA taxonomy	Rodentia	Characterization of Brazilian Sigmodontine Rodents	45	Müller et al., 2013
DNA taxonomy	various	Identification of marine mammals along the French Atlantic coast	15	Alfonsi et al., 2013
DNA taxonomy	Chiroptera	Identification of cryptic species in the New World bat *Pteronotus parnellii*	1	Clare et al., 2013
DNA taxonomy	Chiroptera	Identification of a new bat species in Vietnam	1	Kruskop and Borisenko, 2013
DNA taxonomy	various	Identification of Brazilian forest mammals	7	Cerboncini et al., 2014
DNA taxonomy	Chiroptera	Identification of cryptic bat species in French Guiana and Brazil	2	Thoisy et al., 2014
DNA taxonomy	Primates	Characterization of Peruvian primate species	2	Ruiz-García et al., 2014
DNA taxonomy	Chiroptera	Characterization of Kerivoula bats in Thailand	7	Douangboubpha et al., 2015
DNA taxonomy	Chiroptera	Identification of Malaysian bat species	9	Wilsonet al., 2014
DNA taxonomy	Rodentia	Identification of alien *Callosciurus* squirrels in Argentina	5	Gabrielli et al., 2014
DNA taxonomy	Rodentia	Characterization of Chinese species of Murinae and Arvicolinae	54	Li et al., 2015b
DNA taxonomy	Artiodactyla	Characterization of Chinese Cervidae	21	Cai et al., 2015
DNA taxonomy	Chiroptera	Characterization of two Southeast Asian *Miniopterus* species	2	Li et al., 2015a
DNA taxonomy	Rodentia	Characterization of Eurasian Ground Squirrels	16	Ermakov et al., 2015
Forensic	various	Traceability of bushmeat origin from Central African and South American countries	12	Eaton et al., 2010
Forensic	Artiodactyla	Identification of wildlife crime cases in South Africa	2	Dalton and Kotze, 2011
Forensic	various	Investigation of illegal hunting cases of Brazilian wildlife	2	Sanches et al., 2012

Table 1 – Updated list of case studies dealing with mammals DNA barcoding. For each study, the context of application, the taxonomic order of target mammals, the aim of the work and the number of species involved are reported *(continued)*.

Context	Order	Aim	N°species	References
Forensic	Artiodactyla	Identification of African bushmeat items	15	Bitanyi et al., 2013
Forensic	various	Identification of organs of threatened species	10	Luo et al., 2013
Forensic	Primates	Identification of primate bushmeat in Guinea-Bissau markets	6	Minhós et al., 2013
Forensic	Artiodactyla	Traceability of animal horn products in China	10	Yan et al., 2013
Forensic	various	Authentication of South African wild meat products	10	D'Amato et al., 2013
Forensic	Artiodactyla	Identification of ungulates used in traditional chinese medicine	8	Chen et al., 2015
Forensic	various	Development of a traceability system for African forest bushmeat	59	Gaubert et al., 2014
Non-invasive sampling	Artiodactyla	etection of Kenyan mountain bongo from faecal samples	1	Faria et al., 2011
Non-invasive sampling	Carnivora	Identification of Carnivore species from faecal samples	33	Chaves et al., 2012
Non-invasive sampling	Carnivora	Identification of felid species from scat samples	4	De Matteo et al., 2014
Non-invasive sampling	various	Species identification from faeces	14	Inoue and Akomo-Okoue, 2015
Non-invasive sampling	various	Species identification from blowfly guts content	40	Lee et al., 2015
Parasitology investigation	various	Identification of bloodmeal hosts of ectoparasite species	16	Alcaide et al., 2009
Parasitology investigation	various	Identification of bloodmeal African hosts of tsetse flies	7	Muturi et al., 2011
Parasitology investigation	various	Identification of bloodmeal hosts of biting midges	3	Lassen et al., 2011
Parasitology investigation	various	Development of a rapid diagnostic approach to identify bloodmeal hosts of mosquitoes	5	Thiemann et al., 2012
Parasitology investigation	various	Identification of bloodmeal hosts of ticks	10	Gariepy et al., 2012

most mammals. Although considered illegal, the bushmeat hunting is an increasing economic activity in many countries among which Western and central Africa and other tropical regions (Nasi et al., 2008). In these countries the practice has historically been conducted for subsistence consumption or for local trade and now has reached unsustainable levels (Jenkins et al., 2011; Harrison et al., 2013; Borgerson, 2015).

Several studies, recently examined the utility of DNA barcoding as a standard tool to monitor the traffic of mammals (i.e., whole animals, meat, and other products), with particular emphasis on species commonly traded in bushmeat markets or to determine the species of unknown samples deriving from local cases of poaching or species substitution (see for example Eaton et al., 2010; Dalton and Kotze, 2011; Gaubert et al., 2014). These studies encompassed different groups of mammals such as: bovids (Bitanyi et al., 2011; Cai et al., 2011), suids (Eaton et al., 2010) and primates (Minhós et al., 2013) or covered a wider panel of taxa in an attempt to generate reference datasets for future applications. Concerning this last category, a clear example is given by the DNABUSHMEAT dataset developed by Gaubert and colleagues (2014). Four mitochondrial gene fragments (including the barcode *coxI*), were sequenced in more than 300 African bushmeat samples belonging to nine orders and 59 species. Sequences were then included as references in a query database, called DNABUSHMEAT, which provides an efficient DNA typing decision pipeline to trace the origin of bushmeat items. The DNABUSHMEAT project also contributed in filling the existing gap of African mammals representations in the international archives (i.e., NCBI and BOLD). The availability of a well populated reference dataset is a necessary condition for a successful application of DNA-based identification techniques. The relevance of reference databases has been underlined in recent studies, where a DNA barcoding survey on bushmeat food items traded in Tanzania (Bitanyi et al., 2013) and South Africa (D'Amato et al., 2013) revealed a low correctness of species identification by consumers (i.e., 59% of 124 analysed samples, Bitanyi et al., 2013) and a high rate of species substitution in local markets (i.e., 76.5% of 146 samples, D'Amato et al., 2013). Such problem is not uncommon in the context of the global food market and many published works highlighted the suitability of DNA barcoding in monitoring and hopefully reducing the overexploitation of wildlife species (see for example, Barbuto et al., 2010; Ardura et al., 2013).

Use of animal parts in traditional medicine

The use of animal organs or parts in traditional medicine involves many mammalian species that are currently known for their threatened or endangered conservation status. Among the most frequent cases there is the illegal hunting and trading of rhinoceros horn, saiga antelope horn, bear bile crystals and many others which are commonly used as ingredients in traditional Asian medicine (Luo et al., 2013; Yan et al., 2013). Despite the existing international legislation for the safeguard of these species (i.e., the CITES and the IUCN Red List), the trade of organs still remains an issue of major concern for wildlife conservation and is accelerating the extinction of many species.

As reported in several studies, animal organs/parts are usually processed to obtain powder, tablets, capsules and oils (Coghlan et al., 2012; Cao et al., 2014). Such processes impede any kind of morphological identification and therefore it is almost impossible to set up a suitable traceability pipeline along the supply chain. A method to characterize the biological origin of processed materials is thus mandatory to overcome the limits of morphological-based approaches. In recent

years, some studies highlighted the efficacy of DNA barcoding in authenticating mammal traded organs/parts or their occurrence in traditional medicine products (Luo et al., 2013; Yan et al., 2013). Most of these studies focused on the identification of horns and horn powder, mainly belonging to Cervidae and Bovidae such as the Saiga antelope (*Saiga tatarica*), a protected migratory ungulate living in central Asia and south-eastern Europe, whose horns are one of the main ingredients of the "Lingyangjiao", a traditional Chinese remedy (Chen et al., 2015).

Also in this case, DNA barcoding shows great potentials and should be considered as a valid tool for enforcing local and international legislation and to prosecute cases of illegal trade of mammal organs/parts.

Pet trade and monitoring of alien species

Another issue of major concern involving wildlife conservation and in particular mammals is the trade of organisms as pets. Nowadays, the pet trade is a common pathway of species introduction at the global scale (Bertolino, 2009; Bomford et al., 2009; Genovesi et al., 2012). Frequently, traded individuals are able to establish wild populations as a consequence of either accidental escapes or deliberate releases thus provoking severe problems to the indigenous communities. As a matter of fact, the introduction of alien species is one of the most important causes of biodiversity loss and represents a long-term threat to ecosystem functioning (Mack et al., 2000; Ehrenfeld, 2010; Strayer, 2012). When monitoring or preventive actions are required to control the spread of invasive species, as well as tracking their potential pathways of introduction, the first step is the correct identification of the invasive taxon (Boykin et al., 2012; Pisanu et al., 2013).

In this context, DNA barcoding showed great potential, for instance in the case of squirrels. Many squirrel species belonging to different continents have been introduced through the international pet trade for aesthetic reasons, or to increase hunting opportunities (Long, 2003), and in most cases they established as successful invaders (Bertolino, 2009; Martinoli et al., 2010). Some studies also suggested a lack of taxonomic knowledge within this well studied groups of mammals (Gabrielli et al., 2014; Ermakov et al., 2015). *coxI* barcode sequences were used to investigate the taxonomic status of a group of invasive tree squirrels belonging to the genus *Callosciurus* introduced in Argentina. Unexpectedly, the captured animals were found to be grouped in a previously uncharacterized molecular lineage closer to *C. finlaysonii* rather than to *C. erythraeus* as initially expected from morphological comparisons (Gabrielli et al., 2014). Ermakov and co-workers (2015) used DNA barcoding to characterize the whole diversity of Eurasian ground squirrels. They found unexpected levels of *coxI* divergence in four species out of the 16 investigated, suggesting the occurrence of undescribed cryptic species.

In conclusion, the system generated from DNA barcoding is really useful in the forensic field, and mammals indeed represent a group of organisms in which this application is really welcomed.

Parasitological analyses

Mammals are the natural hosts for a wide panel of parasites. In a broader vision, the parasites typically harbored by mammals could be grouped in macroparasites (e.g., helminths and arthropods) and microbial pathogens (e.g., viruses and bacteria) (Price, 1980; Pedersen et al., 2007; Hatcher and Dunn, 2011). The attack by one or more group of parasites can negatively affect the fitness of the host and even cause significant population declines or boost the extinction risk in already threatened species (Pedersen et al., 2007). In addition, it has been estimated that since the end of 20th century, at least 75% of the emerging infection diseases for humans were zoonotic (Taylor et al., 2001). For this reason, the monitoring and control of zoonotic diseases is nowadays one of the most important concerns in global economies and human health (Daszak et al., 2000; Chomel et al., 2007; Thompson et al., 2009; Rhyan and Spraker, 2010). Another factor influencing the spread of parasites and therefore affecting the conservation status of mammal species is the interaction of indigenous populations with alien taxa. Alien species can indeed carry along with them non-indigenous parasites and these may be transmitted to native species usually lacking an appropriate defense mechanism (Dunn and Hatcher, 2015; Romeo et al., 2015).

Knowledge of the exact species of parasite and/or of the mammal that is carrying harmful pathogens is fundamental to shed light on the factors influencing the occurrence, proliferation, and transmission mediated by animal vectors of such agents (Besansky et al., 2003; Criscione et al., 2005; Kent, 2009). In this framework, molecular methods and in the last decade the DNA barcoding approach, have been playing a key role to understand the complex relationships occurring among mammal hosts, parasites and their intermediate vectors. Most parasites are indeed difficult to discriminate based on morphology, for different reasons (lack of discriminating features, very different life stages, recovery of damaged or partial specimens, see for instance Ferri et al., 2009). For example in the case of endoparasites, their identification is often based on post-mortem examination of the hosts, because less-invasive approaches (e.g., the collection of eggs, larvae or pieces in host blood, tissue samples or faeces) cannot permit an easier identification owing to the loss of many diagnostic tracts (Ondrejicka et al., 2014). DNA barcoding approach contributed to overcome these limits and successful protocols have been developed to identify the principal classes of parasites affecting mammals such as filarioid nematodes (Ferri et al., 2009), cestodes (Galimberti et al., 2012a), ticks (Zhang and Zhang, 2014) and mosquitoes (Cywinska et al., 2006). In other cases, DNA barcoding has been largely applied to identify the mammal hosts of important parasites / pathogens. These case studies especially involved rodent species complexes characterized by a high number of cryptic taxa inhabiting poorly studied areas of the planet. Specifically, in 2012, Lu and co-workers, studied the relationships between *Rickettsia* bacteria (i.e., the agent responsible for the spotted fever) and ten rodent hosts of China (Lu et al., 2012). DNA barcoding was used to differentiate host species and the values of molecular divergence highlighted the need for further taxonomic investigations on some species groups. Similarly, in 2013, Müller and co-workers used *coxI* barcode sequences to recognize members of Sigmodontinae subfamily in Brazil which are reservoirs of zoonoses including arenaviruses, hantaviruses, Chagas disease and leishmaniasis (Müller et al., 2013).

One of the most innovative applications of DNA barcoding in the study of host-parasite interactions is the characterization of insect bloodmeals. As a matter of fact, most zoonoses are likely to be vector-borne by blood-feeding arthropods (Jones et al., 2008) which dictate the relationship between host and pathogen (Thiemann et al., 2012). Blood feeding vectors may transmit agents responsible for emerging diseases such as malaria, viral encephalitis, West Nile virus, Chagas disease, Lyme disease or African sleeping sickness (Kent, 2009). By studying arthropods behaviour, it has been possible to understand the evolution of host specificity between vertebrates and their ectoparasites, how the host choice drives pathogen transmission, and the economic and demographic impacts of ectoparasite infestations on wildlife and domestic livestock (Kent, 2009). A deep knowledge of these factors can help improving reliable disease risk models to be used in veterinary and public health contingency plans (Kent, 2009; Gomez-Diaz and Figuerola, 2010; Collini et al., 2015). Several DNA barcoding-based surveys have been conducted in the last years to fill the gaps in the comprehension of such dynamics. Published studies involved a specific group of blood-feeding arthropods such as *Culex* spp. mosquitoes (Muños et al., 2012; Thiemann et al., 2012), ticks (Gariepy et al., 2012; Collini et al., 2015), biting midges (Lassen et al., 2011), tsetse flies (Muturi et al., 2011) as well as the simultaneous analysis of a wide range of vectors (Alcaide et al., 2009).

In all of these case studies, the analysis of *coxI* barcode sequences obtained from the bloodmeal consumed by hematophagous vectors allowed to trace the identity of the "last supper" (i.e., the vertebrate host – often a mammal) on which the vector fed before being collected. Finally, in a recent study conducted in Peninsular Malaysia, a biodiversity hotspot, Lee and colleagues (2015) proposed the DNA barcoding analysis of the stomach content of the saprophagous / coprophagous blow-

flies (Calliphoridae) as a suitable, fast and economic tool to characterize the mammal biodiversity of a study area.

In conclusion, the analysis of parasites is a complex matter and molecular tools, like DNA barcoding, are really welcomed.

Massive DNA sequencing

In the last decade, there has been a great revolution in DNA sequencing technologies. The introduction of the so-called "Next Generation Sequencing", NGS, also better defined as "High Throughput DNA Sequencing", HTS, expanded the universe of DNA sequencing. The rise of DNA barcoding took place in the same years and it was only a matter of time to assist to the encounter of these two worlds. The DNA metabarcoding is the result of this marriage (Taberlet et al., 2012). HTS has revolutionized DNA-based research, especially biodiversity assessment in complex biological matrix (i.e. comprising many species contemporaneously) (Shokralla et al., 2012). In HTS, DNA sequences are accumulated at an unprecedented rate and it is now possible to analyze simultaneously several samples (through multiplexing) identified by custom-designed oligonucleotide tags.

The idea is simple: DNA is everywhere, and this molecule is relatively stable and durable in dry, but even wet conditions (Dejean et al., 2012; Yoccoz et al., 2012). This DNA represents the so-called "environmental DNA" or eDNA (Shokralla et al., 2012; Thomsen and Willerslev, 2015). eDNA is formed by short DNA molecules (i.e., free, cellular debris or particle-bound), which are released by living or dead organisms. eDNA is typically defined by the process used to collect it, and this makes its definition in a some way foggy. Much more clear is the use of eDNA: the living beings present in the environmental sample are not known and HTS allows to identify them. In addition, even if DNA in the environment is relatively stable, it is also usually degraded. In such a condition the classic DNA barcoding approach is often useless, conversely to metabarcoding, due to the possibility of generating a huge amount of data. The first application in mammals was aimed at uncovering the diets composition of elusive animals (Valentini et al., 2009). This approach was successfully adopted in the last 5 years with some group being very well represented, such as Chiroptera (Bohmann et al., 2011; Alberdi et al., 2012; Vesterinen et al., 2013; Krüger et al., 2014a,b; Burgar et al., 2014; Clare et al., 2014a,b; Hope et al., 2014; Sedlock et al., 2014).

Although it is now relatively simple to characterize the diets of herbivorous and insectivorous mammals, the analysis of diets of carnivores is really challenging because predator DNA can be simultaneously amplified with prey DNA (Symondson, 2002; King et al., 2008; Symondson and Harwood, 2014; Boyer et al., 2015). To avoid this problem an interesting approach was the introduction of blocking primers in the analysis of snow leopard (*Panthera uncia*) diet (Shehzad et al., 2012). This molecular approach prevents the amplification of predator DNA allowing the amplification of the other vertebrate groups.

HTS techniques can also be used to identify elusive mammal species from the faeces found on the ground (Michalski et al., 2011; Chaves et al., 2012; Rodgers and Janecka, 2013) or as a general method to identify mammals in complex mixtures (Foote et al., 2012; Galan et al., 2012; Deagle et al., 2013; Tillmar et al., 2013). Noteworthy, the possibility of better defining the areas of distribution of some species with such non-invasive sampling is of particular interest to increase the knowledge of mammals biology and conservation.

In spite of these practical approaches, HTS techniques in mammals have also been used to characterize population structure (Rasgour et al., 2011; Botero-Castro et al., 2013). The rapid developments of these technologies have created new possibilities to build quickly and cost-efficiently reference libraries for whole mitochondrial genomes in a wide range of animal lineages. The accumulation of whole mitogenomes in the public domain covering the Tree of Animal Life will improve our knowledge on evolutionary history of animals and global patterns in genomic features of mitochondria as a sort of future next comprehensive barcode marker.

In conclusion, HTS and the DNA metabarcoding approaches are expanding fields of research that will likely be very fertile for several years to come, particularly considering the rapid increase of reference databases that allows a better characterization of complex cases.

The integrative role of DNA barcoding

As described in the previous sections, DNA barcoding can be successfully involved as a supporting tool for both theoretical and applicative necessities. The presented case studies highlighted the versatility of the approach, and the aptitude of being integrated with other sources of taxonomic information in a highly interconnected environment.

As a matter of fact, species are not unequivocally defined and their designations based on a single category of taxonomic features (morphological, ecological, molecular, or biogeographic) is questionable. In this context, molecular techniques and more recently the DNA barcoding, triggered a small revolution inside taxonomy: the process of identifying biological entities opened the doors to a real integration of knowledge to improve practical purposes (Unit of Conservation *sensu* Dodson et al., 1998) or theoretical approaches (Unit of Evolution or Evolutionarily Significant Unit, ESU, *sensu* Ryder, 1986).

In a framework of integration, divergent molecular lineages do not necessarily reflect distinct species but, in many cases, molecular data remains at the core of current taxonomic approaches. However, the future of taxonomy cannot rely only on molecular markers. Rather, it is more and more oriented towards the definition of the best way to integrate molecular data into multidisciplinary taxonomic approaches.

In an attempt of providing a better understanding of the possible taxonomic outcomes deriving from an integrative DNA barcoding-based approach, Galimberti and colleagues recently proposed a schematization using echolocating bats as a model (Galimberti et al., 2012b). In this schematic view, the taxonomic ranks are grouped based on their information content: from individuals (i.e., the less informative level), to species (i.e., the more informative level), passing through intermediate categories defined by the adoption of a single (i.e., morphotype, Molecular Operational Taxonomic Unit - MOTU and unconfirmed candidate species) or an integrative approach (i.e., Integrative Operational Taxonomic Units - IOTU, deep conspecific lineage and confirmed candidate species).

Such schematization, tested on Italian bats species, confirmed the risk of erroneous taxonomic interpretations when molecular entities (MOTUs) are used as the only criterion (see the case of *Eptesicus* species in Galimberti et al., 2012b). The authors also proposed a new entity, the IOTU, defined by molecular lineages that have further support from at least one additional line of evidence. This concept links different data sources in taxonomy, allowing morphological, ecological, geographical and other characteristics of living beings to be better combined with molecular data. The application of IOTU concept to the study of echolocating bats showed for example the occurrence a new undescribed species of *Myotis nattereri* inhabiting the southern part of the Italian peninsula.

Known problems of DNA barcoding of mammals

DNA barcoding generated huge controversies, but like any other diagnostic technique it has pros and cons. Since its launch, the practicalities of a universal barcode for all the living beings showed pitfalls, as firstly dependent on the group of organisms under examination (see Casiraghi et al., 2010 and Collins and Cruicsshank, 2013 and references therein). Concerning mammals, three main categories of problems should be taken into account when DNA barcoding is applied to their study. The first concerns the availability of public and well populated reference archives of DNA barcodes and related specimens (see the dedicated paragraph above). Reference sequences constitute the main core of the DNA barcoding initiative and their absence or the lack of control of the correct identification of the source specimens by expert taxonomists, can irremediably affect the assignment of newly generated query sequences. The second problem category is directly related to the processes of molecular evolution, such as the occurrence of NUMTs (i.e., nuclear copies of mitochondrial DNA). NUMTs are usually considered a challenge in those case studies based on mtDNA

due to the fact that they can be inadvertently amplified, thus causing bias in the barcode dataset and in the accuracy of subsequent analyses (e.g., overestimating intra and interspecific variability levels) (Bensasson et al., 2001; Song et al., 2008; Ermakov et al., 2015). Recently, Ermakov and co-workers (2015), described the amplification of NUMTs in a species of Eurasian ground squirrel. This is only one of the multiple documented examples of this problem. NUMTs have been found in over 20 mammalian species belonging to seven different orders (see (Triant and DeWoody, 2007) for more details). To overcome the risk of NUMTs interference, Song et al. (2008) and Buhay (2009) suggested step-by-step procedures in order to identify possible pseudogenes. BOLD itself provides a quality control tool to check sequences for the presence of stop codons and verify that they derive from *coxI* by comparing them against a Hidden Markov Model (Ratnasingham and Hebert, 2007). To avoid NUMTs interference, Triant and DeWoody (2007) suggested three alternative strategies: i) the isolation of the entire mtDNA genome, ii) the use of tissue sources naturally rich in mtDNA (e.g., liver and muscle), and iii) the use of PCR primers that amplify substantial portions of the mtDNA molecule (i.e., > 1 kb). In other cases, the re-extraction of gDNA and the reamplification of the barcode region can help resolving the matter (Ermakov et al., 2015). The last group of issues causing failure of DNA barcoding identification are mainly due to the essence of biological species, rather than in the method, and relies on the criteria adopted to discriminate species. As well as in many other cases, species delimitation in mammals is based almost completely on two strategies: the genetic distance and the reciprocal monophyly (Dávalos and Russell, 2014). However, when dealing with mtDNA, attention is needed when automatically associating divergence values (which are often useful "hypothesis generator") with the extent of gene flow. As discussed by Dávalos and co-workers (2014), such way of thinking can lead to false-positive errors in which distances or monophyly diagnose species despite ongoing gene flow, and false-negative errors when gene flow is taken into account despite its absence. Mitochondrial DNA barcode markers, are indeed prone to problems such as introgression, incomplete lineage sorting and hybridization and this may generate misleading results particularly in mammals (Heckman et al., 2007; Godinho et al., 2011; Melo-Ferreira et al., 2012).

In a DNA barcoding study conducted on the whole panel of species of Eurasian ground squirrels, Ermakov and colleagues (2015), documented the occurrence of mtDNA introgression in four cases due to ancient hybridization events followed by divergence. Similar conditions have been also detected in other groups of mammals such as bears (Hailer et al., 2012), marmots (Brandler et al., 2010) and bats (Berthier et al., 2006; Artyushin et al., 2009).

Moreover, mammals are often characterized by sex-biased gene flow in which males disperse widely and females exhibit natal philopatry (Greenwood, 1980). Such condition also shape the genetic structure of species and populations when maternally-inherited mitochondrial markers are analysed (Clare, 2011; Dávalos and Russell, 2014). To overcome this limit of mtDNA, the selection of complementary loci with independent evolutionary histories can help depicting a more realistic schematization of the divergences at both the intra and interspecific level. For example, in 2011, Clare published a study in which she successfully compared the phylogeographic patterns revealed through the maternally inherited mitochondrial *coxI* and the paternally inherited 7th intron region of the *Dby* gene on the Y-chromosome in eight common Neotropical bat species (Clare, 2011). The combined approach proposed by Clare allowed the author to validate patterns of gene flow and also to find previously unrecognized species.

Similarly, Silva and coauthors (2014) developed a method based on polymorphism of the mitochondrial *cytb* and the nuclear *KCAS* gene to identify nine ungulate species occurring in North Africa.

As a final consideration, it is important to underline that when DNA barcoding investigations reveal the occurrence of new intraspecific lineages, they should be integrated with alternate lines of evidence such as ecological data, morphology and geography to avoid misinterpretation of genetic variability (Galimberti et al., 2012b). DNA barcoding problems are well known, but as underlined above, we do not have to stop at them, and consider the whole system created by this technique.

The future of DNA barcoding of mammals

In spite of an apparent decreasing trend in the rate of publication on the topic "mammals DNA barcoding" (see Fig. 1), this molecular approach is still alive and healthy. Probably, this apparent reduction is due to the fact that the modern taxonomic system is now a matter of fact, and the DNA barcoding approach is often integrated even without naming it. Indeed, DNA barcoding does not rely on the use of a monospecific marker only, as often stated, but is currently referred to as a way of thinking rather than a name of a technique.

In the case of mammals, DNA barcoding is alive and proactive, because these animals represent the principal group in which the scientific community moved from a *sensu stricto* approach to broader applications. Indeed, DNA barcoding *sensu stricto* is designed for not specialized operators in a certain taxonomic field. Generally speaking, the specialist does not have real problems to discriminate among the living beings he/she is studying, because in most cases, he/she himself/herself is the one who created the classification system (hopefully using a robust integrated approach). Consequently, the specialist is the principal actor who has to work to create a solid DNA barcoding system to help other users in achieving a correct identification for purposes ranging from wildlife management, to conservation, eco-ethological studies and so on.

As we underlined in our essay, in many cases DNA barcoding in mammals has already reached this level and we foresee that in the next future this approach will move towards two main branches of application. The first branch (the molecular one) is that of taxonomic studies to fully uncover the hidden biodiversity within this animal group. On the other side, even if strictly connected, there will be the branch of "taxonomic services" in which DNA barcoding is one of the more correct, easier and more sparing (both in terms of money and time) solutions.

References

Agrizzi J., Loss A.C., Farro A.P.C., Duda R., Costa L.P., Leite Y.L. 2012. Molecular diagnosis of Atlantic forest mammals using mitochondrial DNA sequences: didelphid marsupials. Open Zool. J. 5: 2–9.

Alberdi A., Garin I., Aizpurua O., Aihartza J., 2012. The foraging ecology of the mountain long-eared bat *Plecotus macrobullaris* revealed with DNA mini-barcodes. PLoS One 7(4): e35692.

Alcaide M., Rico C., Ruiz S., Soriguer R., Muñoz J., Figuerola J., 2009. Disentangling vector-borne transmission networks: a universal DNA barcoding method to identify vertebrate hosts from arthropod bloodmeals. PLoS One 4(9): e7092.

Alfonsi E., Méheust E., Fuchs S., Carpentier F.G., Quillivic Y., Viricel A., Hassani S., Jung J.L., 2013. The use of DNA barcoding to monitor the marine mammal biodiversity along the French Atlantic coast. Zookeys 365: 5–24.

Ardura A., Planes S., Garcia-Vazquez E., 2013. Applications of DNA barcoding to fish landings: authentication and diversity assessment. ZooKeys 365: 49–65.

Artyushin I.V., Bannikova A.A., Lebedev V.S., Kruskop S.V., 2009. Mitochondrial DNA relationships among North Palaearctic *Eptesicus* (Vespertilionidae, Chiroptera) and past hybridization between Common Serotine and Northern Bat. Zootaxa 2262: 40–52.

Baker C.S., 2008. A truer measure of the market: the molecular ecology of fisheries and wildlife trade. Mol. Ecol. 17(18): 3985–3998.

Baker R.J., Bradley R.D., 2006. Speciation in mammals and the genetic species concept. J Mammal. 87: 643–662.

Bannikova A.A., Abramov A.V., Borisenko A.V., Lebedev V.S., Rozhnov V.V., 2011. Mitochondrial diversity of the white-toothed shrews (Mammalia, Eulipotyphla, *Crocidura*) in Vietnam. Zootaxa 2812: 1–20.

Barbuto M., Galimberti A., Ferri E., Labra M., Malandra R., Galli P., Casiraghi M., 2010. DNA barcoding reveals fraudulent substitutions in shark seafood products: the Italian case of "palombo" (*Mustelus* spp.). Food Res. Int. 43(1): 376–381.

Becker S., Hanner R., Steinke D. 2011., Five years of FISH-BOL: Brief status report. Mitochondr. DNA 22(S1): 3–9.

Bensasson D., Zhang D.X., Hartl D.L., Hewitt G.M., 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol. Evol. 16: 314–321.

Berthier P., Excoffier L., Ruedi M., 2006. Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*. Proc. R. Soc. Lond. [Biol] 273: 3101–3109.

Bertolino S., 2009. Animal trade and non-indigenous species introduction: the world-wide spread of squirrels. Diversity and Distributions 15(4): 701–708.

Besansky N.J., Severson D.W., Ferdig M.T., 2003. DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. Trends Parasitol. 19(12): 545–546.

Bitanyi S., Bjørnstad G., Ernest E.M., Nesje M., Kusiluka L.J., Keyyu J.D., Mdegela R.H., Røed K.H., 2011. Species identification of Tanzanian antelopes using DNA barcoding. Mol. Ecol. Resour. 11(3): 442–449.

Bitanyi S., Bjørnstad G., Nesje M., Ernest E.M., Mdegela R.H., Røed K.H., 2013. Molecular identification versus local people's information for accurate estimates of bushmeat utilization from the Serengeti ecosystem, Tanzania. Afr. J. Biotechnol. 11(1): 243–252.
Bogdanowicz W., Hulva P., Cerná Bolfíková B., Bus M.M., Rychlicka E., Sztencel-Jablonka A., Cistrone L., Russo D., 2015. Cryptic diversity of Italian bats and the role of the Apennine refugium in the phylogeography of the western Palaearctic. Zool. J. Linnean Soc. (Early View) doi:10.1111/zoj.12248
Bohmann K., Monadjem A., Noer C.L., Rasmussen M., Zeale M.R., Clare E., Jones G., Willerslev E., Gilbert M.T.P., 2011. Molecular diet analysis of two African free-tailed bats (Molossidae) using high throughput sequencing. PLoS One 6(6): e21441.
Bomford M., Kraus F., Barry S.C., Lawrence E., 2009. Predicting establishment success for alien reptiles and amphibians: a role for climate matching. Biol. Invasions 11(3): 713–724.
Borgerson C., 2015. The Effects of Illegal Hunting and Habitat on Two Sympatric Endangered Primates. Int. J. Primatol. 36(1): 74–93.
Borisenko A.V., Lim B.K., Ivanova N.V., Hanner R.H., Hebert P.D., 2008. DNA barcoding in surveys of small mammal communities: a field study in Suriname. Mol. Ecol. Resour. 8(3): 471–479.
Botero-Castro F., Tilak M.K., Justy F., Catzeflis F., Delsuc F., Douzery E.J., 2013. Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf-nosed bats (Phyllostomidae). Mol. Phylogenet. Evol. 69(3): 728–739.
Boyer S., Cruickshank R.H., Wratten S.D., 2015. Faeces of generalist predators as "biodiversity capsules": A new tool for biodiversity assessment in remote and inaccessible habitats. Food Webs 3: 1–6.
Boykin L.M., Armstrong K.F., Kubatko L., De Barro P., 2012. Species delimitation and global biosecurity. Evol. bioinform. Online 8: 1–37.
Brandler O.V., Lyapunova E.A., Bannikova A.A., Kramerov D.A., 2010. Phylogeny and systematics of marmots (*Marmota*, Sciuridae, Rodentia) inferred from inter-SINE PCR data. Russ. J. Genet. 46: 283–292.
Buhay J.E., 2009. "COI-like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. J. Crust. Biol. 29(1): 96–110.
Burgar J.M., Murray D.C., Craig M.D., Haile J., Houston J., Stokes V., Bunce M., 2014. Who's for dinner? High-throughput sequencing reveals bat dietary differentiation in a biodiversity hotspot where prey taxonomy is largely undescribed. Mol. Ecol. 23(15): 3605–3617.
Cai Y., Zhang L., Shen F., Zhang W., Hou R., Yue B., Li J., Zhang Z., 2011. DNA barcoding of 18 species of Bovidae. Chin. Sci. Bull. 56(2): 164–168.
Cai Y., Zhang L., Wang Y., Liu Q., Shui Q., Yue B., Zhang Z., Li J., 2015. Identification of deer species (Cervidae, Cetartiodactyla) in China using mitochondrial cytochrome c oxidase subunit I (mtDNA COI). Mitochondr. DNA 12: 1–4.
Cao M., Wang J., Yao L., Xie S., Du J., Zhao X., 2014. Authentication of animal signatures in traditional Chinese medicine of Lingyang Qingfei Wan using routine molecular diagnostic assays. Mol. Biol. Rep. 41(4): 2485–2491.
Casiraghi M., Labra M., Ferri E., Galimberti A., De Mattia F., 2010. DNA barcoding: a six-question tour to improve users' awareness about the method. Brief. Bioinform. 11(4): 440–453.
Cerboncini R.A.S., Rubio M.B.G., Bernardi I.P., Braga T.V., Roper J.J., Passos F.C., 2014. Small mammal community structure and vertical space use preferences in non-fragmented Atlantic Forest. Mammalia 78(4): 429–436.
Cervantes F.A., Arcangeli J., Hortelano-Moncada Y., Borisenko A.V., 2010. DNA barcodes effectively identify the morphologically similar Common Opossum (*Didelphis marsupialis*) and Virginia Opossum (*Didelphis virginiana*) from areas of sympatry in Mexico. Mitochondr. DNA 21(S1): 44–50.
Chaves P.B., Graeff V.G., Lion M.B., Oliveira L.R., Eizirik E., 2012. DNA barcoding meets molecular scatology: short mtDNA sequences for standardized species assignment of carnivore noninvasive samples. Mol. Ecol. Res. 12(1): 18–35.
Chen J., Jiang Z., Li C., Ping X., Cui S., Tang S., Chu H., Liu B., 2015. Identification of ungulates used in a traditional Chinese medicine with DNA barcoding technology. Ecol. Evol. 5(9): 1818–1825.
Chomel B.B., Belotto A., Meslin F.X., 2007. Wildlife, exotic pets, and emerging zoonoses. Emerg. Infect. Diseases 13(1): 6–11.
Clare E.L., 2011. Cryptic species? Patterns of maternal and paternal gene flow in eight Neotropical bats. PLoS One 6(7): e21460.
Clare E.L., Adams A.M., Maya-Simões A.Z., Eger J.L., Hebert P.D.N., Fenton M.B., 2013. Diversification and reproductive isolation: cryptic species in the only New World high-duty cycle bat, *Pteronotus parnellii*. BMC Evol. Biol. 13(1): 26.
Clare E.L., Lim B.K., Engstrom M.D., Eger J.L., Hebert P.D., 2007. DNA barcoding of Neotropical bats: species identification and discovery within Guyana. Mol. Ecol. Notes 7(2): 184–190.
Clare E.L., Lim B.K., Fenton M.B., Hebert P.D., 2011. Neotropical bats: estimating species diversity with DNA barcodes. PLoS One 6(7): e22648.
Clare E.L., Symondson W.O., Broders H., Fabianek F., Fraser E.E., MacKenzie A., Boughen A., Hamilton R., Willis C.K.R., Martinez-Nuñez F., Menzies A.K., Norquay K.J.O., Brigham M., Poissant J., Rintoul J., Barclay R.M.R., Reimer, J.P., 2014a. The diet of *Myotis lucifugus* across Canada: assessing foraging quality and diet variability. Mol. Ecol. 23(15): 3618–3632.
Clare E.L., Symondson W.O., Fenton M.B., 2014b. An inordinate fondness for beetles? Variation in seasonal dietary preferences of night-roosting big brown bats (*Eptesicus fuscus*). Mol. Ecol. 23(15): 3633–3647.
Coghlan M.L., Haile J., Houston J., Murray D.C., White N.E., Moolhuijzen P., Bellgard M.B., Bunce M., 2012. Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns. PLoS Genet 8(4): e1002657.
Collini M., Albonico F., Hauffe H.C., Mortarino M., 2015. Identifying the last bloodmeal of questing sheep tick nymphs (*Ixodes ricinus* L.) using high resolution melting analysis. Vet. Parasitol. 210(3–4): 194–205.
Collins R.A., Cruickshank R.H., 2013. The seven deadly sins if DNA barcoding. Mol. Ecol. Resour. 13(6): 969–975.
Comtet T., Sandionigi A., Viard F., Casiraghi M., 2015. DNA (meta) barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. Biol. Invasions 17(3): 905–922.
Conteh A., Gavin M.C., Solomon J., 2015. Quantifying illegal hunting: A novel application of the quantitative randomised response technique. Biol. Cons. In press doi: 10.1016/j.biocon.2015.02.002
Criscione C.D., Poulin R., Blouin M.S., 2005. Molecular ecology of parasites: elucidating ecological and microevolutionary processes. Mol. Ecol. 14(8): 2247–2257.
Cywinska A., Hunter F.F., Hebert P.D., 2006. Identifying Canadian mosquito species through DNA barcodes. Med. Vet. Entomol. 20(4): 413–424.
D'Amato M.E., Alechine E., Cloete K.W., Davison S., Corach D., 2013. Where is the game? Wild meat products authentication in South Africa: a case study. Investig. Genet. 4(1): 6.
Dalton D.L., Kotze A., 2011. DNA barcoding as a tool for species identification in three forensic wildlife cases in South Africa. Forensic Sci. Int. 207(1): e51–e54.
Daszak P., Cunningham A.A., Hyatt A.D., 2000. Emerging infectious diseases of wildlife-threats to biodiversity and human health. Science 287(5452): 443–449.
Dávalos L.M., Russell A.L., 2014. Sex-biased dispersal produces high error rates in mitochondrial distance-based and tree-based species delimitation. J. Mammal. 95(4): 781–791.
Dawnay N., Ogden R., McEwing R., Carvalho G.R., Thorpe R.S., 2007. Validation of the barcoding gene COI for use in forensic genetic species identification. Forensic Sci. Int. 173(1): 1–6.
De Matteo K.E., Rinas M.A., Argueelles C.F., Holman B.E., Di Bitetti M.S., Davenport B., Parker P.G., Eggert L.S., 2014. Using detection dogs and genetic analyses of scat to expand knowledge and assist felid conservation in Misiones, Argentina. Integr. Zool. 9(5): 623–639.
Deagle B.E., Thomas A.C., Shaffer A.K., Trites A.W., Jarman, S.N., 2013. Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count?. Mol. Ecol. Resour. 13(4): 620–633.
Dejean T., Valentini A., Miquel C., Taberlet P., Bellemain E., Miaud C., 2012. Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog *Lithobates catesbeianus*. J. Appl. Ecol. 49: 953–959.
De Pasquale P.P., Galimberti A., 2014. New records of the Alcathoe bat, *Myotis alcathoe* (Vespertilionidae) for Italy. Barbastella 7: 1.
Dodson J.J., Gibson R.J., Cunjak R.A., Friedland K.D., de Leaniz C.G., Gross M.R., Newbury R., Nielsen J.L., Power M.E., Roy S., 1998. Elements in the development of conservation plans for Atlantic salmon (*Salmo salar*). Can. J. Fish. Aquat. Sci. 55(Suppl. 1): 312–323.
Douangboubpha B., Bumrungsri S., Satasook C., Wanna W., Soisook P., Bates P.J., 2015. Morphology, genetics and echolocation calls of the genus *Kerivoula* (Chiroptera: Vespertilionidae: Kerivoulinae) in Thailand. Mammalia (Ahead of Print) 10.1515/mammalia-2014-0004
Dunn A.M., Hatcher M.J., 2015. Parasites and biological invasions: parallels, interactions, and control. Trends Parasitol. 31(5): 189–199.
Eaton M.J., Meyers G.L., Kolokotronis S.O., Leslie M.S., Martin A.P., Amato G., 2010. Barcoding bushmeat: molecular identification of Central African and South American harvested vertebrates. Conserv. Genet. 11(4): 1389–1404.
Ehrenfeld J.G., 2010. Ecosystem consequences of biological invasions. Annu. Rev. Ecol. Evol. Syst. 41: 59–80.
Ermakov O.A., Simonov E., Surin V.L., Titov S.V., Brandler O.V., Ivanova N.V., Borisenko, A.V., 2015. Implications of Hybridization, NUMTs, and Overlooked Diversity for DNA Barcoding of Eurasian Ground Squirrels. PLoS One 10(1): e0117201.
Faria P.J., Kavembe G.D., Jung'a J.O., Kimwele C.N., Estes L.D., Reillo P.R., Mwangi A.G., Bruford M.W., 2011. The use of non-invasive molecular techniques to confirm the presence of mountain bongo *Tragelaphus eurycerus isaaci* populations in Kenya and preliminary inference of their mitochondrial genetic variation. Conserv. Genet. 12(3): 745–751.
Ferri E., Barbuto M., Bain O., Galimberti A., Uni S., Guerrero R., Ferté H., Bandi C., Martin C., Casiraghi M., 2009. Integrated taxonomy: traditional approach and DNA barcoding for the identification of filarioid worms and related parasites (Nematoda). Front. Zool. 6(1): 1–12.
Foote A.D., Thomsen P.F., Sveegaard S., Wahlberg M., Kielgast J., Kyhn L.A., Salling A.B., Galatius A., Orlando L., Gilbert M.T.P., 2012. Investigating the potential use of environmental DNA (eDNA) for genetic monitoring of marine mammals. PLoS One 7(8): e41781.
Francis C.M., Kingston T., Zubaid A., 2007. A new species of *Kerivoula* (Chiroptera: Vespertilionidae) from peninsular Malaysia. Acta Chiropterologica 9(1): 1–12.
Francis C.M., Borisenko A.V., Ivanova N.V., Eger J.L., Lim B.K., Guillén-Servent A., Kruskop S.V., Mackie I., Hebert P.D.N., 2010. The role of DNA barcodes in understanding and conservation of mammal diversity in Southeast Asia. PLoS One 5(9): e12575.
Gabrielli M., Cardoso Y.P., Benitez V., Gozzi A.C., Guichón M.L., Lizarralde M.S., 2014. Genetic characterization of *Callosciurus* (Rodentia: Sciuridae) Asiatic squirrels introduced in Argentina. Ital. J. Zool. 81(3): 328–343.
Galan M., Pagès M., Cosson J.F., 2012. Next-generation sequencing for rodent barcoding: species identification from fresh, degraded and environmental samples. PLoS One 7(11): e48374.
Galimberti A., De Mattia F., Losa A., Bruni I., Federici S., Casiraghi M., Martellos S., Labra M., 2013. DNA barcoding as a new tool for food traceability. Food Res. Int. 50(1): 55–63.
Galimberti A., Romano D.F., Genchi M., Paoloni D., Vercillo F., Bizzarri L., Sassera D., Bandi C., Genchi C., Ragni B., Casiraghi M., 2012. Integrative taxonomy at work: DNA barcoding of taeniids harboured by wild and domestic cats. Mol. Ecol. Resour. 12(3): 403–413.
Galimberti A., Spada M., Russo D., Mucedda M., Agnelli P., Crottini A., Ferri E., Martinoli A., Casiraghi M., 2012. Integrated operational taxonomic units (IOTUs) in echolocating bats: a bridge between molecular and traditional taxonomy. PLoS One 7(6): e40122.
Gariepy T.D., Lindsay R., Ogden N., Gregory T.R., 2012. Identifying the last supper: utility of the DNA barcode library for bloodmeal identification in ticks. Mol. Ecol. Resour. 12(4): 646–652.
Gaubert P., Njiokou F., Olayemi A., Pagani P., Dufour S., Danquah E., Nutsuakor M.E.K., Ngua G., Missoup A.D., Tedesco P., Dernat R., Antunes A., 2014. Bushmeat genetics: setting up a reference framework for the DNA typing of African forest bushmeat. Mol. Ecol. Resour. 15(3): 633–651.
Gavin M.C., Solomon J.N., Blank S.G., 2010. Measuring and monitoring illegal use of natural resources. Conserv. Biol. 24(1): 89–100.

Genovesi P., Carnevali L., Alonzi A., Scalera R., 2012. Alien mammals in Europe: updated numbers and trends, and assessment of the effects on biodiversity. Integr. Zool. 7(3): 247–253.

Godinho R., Llaneza L., Blanco J.C., 2011. Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. Mol. Ecol. 20(24): 5154–5166.

Gomez-Diaz E., Figuerola J., 2010. New perspectives in tracing vector-borne interaction networks. Trends Parasitol. 26(10): 470–476.

Greenwood P.J., 1980. Mating systems, philopatry and dispersal in birds and mammals. Anim. Behav. 28(4): 1140–1162.

Hailer F.V., Kutschera E., Hallstrom B.M., Klassert D., Fain S.R., Leonard J.A., Arnason U., Janke A., 2012. Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. Science 336: 344–347.

Hajibabaei M., Singer G.A., Hickey D.A., 2006. Benchmarking DNA barcodes: An assessment using available primate sequences. Genome 49: 851–854.

Hanner R., 2009. Data Standards for BARCODE Records in INSDC (BRIs). Available from http://barcoding.si.edu/PDF/DWG_data_standards-Final.pdf Accessed 15.05.15.

Harrison R.D., Tan S., Plotkin J.B., Slik F., Detto M., Brenes T., Itoh A., Davies S.J., 2013. Consequences of defaunation for a tropical tree community. Ecol. Lett. 16(5): 687–694.

Hatcher M.J., Dunn A.M., 2011. Parasites in ecological communities: from interactions to ecosystems. Cambridge University Press, Cambridge.

Hebert P.D.N., Cywinska A., Ball S.L., deWaard J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. [Biol] 270(1512): 313–321.

Hebert P.D.N., Stoeckle M.Y., Zemlak T.S., Francis C.M., 2004. Identification of birds through DNA barcodes. PLoS Biol. 2(10): e312. 10.1371/journal.pbio.0020312

Heckman K.L., Mariani C.L., Rasoloarison R., Yoder A.D., 2007. Multiple nuclear loci reveal patterns of incomplete lineage sorting and complex species history within western mouse lemurs (*Microcebus*). Mol. Phylogenet. Evol. 43(2): 353—367.

Hernández-Dávila A., Vargas J.A., Martínez-Méndez N., Lim B.K., Engstrom M.D., Ortega J., 2012. DNA barcoding and genetic diversity of phyllostomid bats from the Yucatán Peninsula with comparisons to Central America. Mol. Ecol. Resour. 12(4): 590–597.

Hope P.R., Bohmann K., Gilbert M.T.P., Zepeda-Mendoza M.L., Razgour O., Jones G., 2014. Second generation sequencing and morphological faecal analysis reveal unexpected foraging behaviour by *Myotis nattereri* (Chiroptera, Vespertilionidae) in winter. Front. Zool. 11(1): 39.

Huang S., Stephens P.R., Gittleman J.L., 2012. Traits, trees and taxa: global dimensions of biodiversity in mammals. Proc. R. Soc. Lond. [Biol] 279(1749): 4997–5003.

Huffman J.E., Wallace J.R. 2012. Wildlife forensics: methods and applications (Vol. 6). John Wiley & Sons.

Inoue E., Akomo-Okoue E.F., 2015. Application of DNA barcoding techniques to mammal inventories in the African rain forest: droppings may inform us of the owners. Tropics 23(4): 137–150.

Iyengar A., 2014. Forensic DNA analysis for animal protection and biodiversity conservation: A review. J. Nat. Conserv. 22(3): 195–205.

Jacquet F., Nicolas V., Bonillo C., Cruaud C., Denys C., 2012. Barcoding, molecular taxonomy, and exploration of the diversity of shrews (Soricomorpha: Soricidae) on Mount Nimba (Guinea). Zool. J. Linnean Soc. 166(3): 672–687.

Jenkins R.K., Keane A., Rakotoarivelo A.R., Rakotomboavonjy V., Randrianandrianina F. H., Razafimanahaka H.J., Ralaiarimalala S.R., Jones J.P., 2011. Analysis of patterns of bushmeat consumption reveals extensive exploitation of protected species in eastern Madagascar. PLoS ONE 6(12): e27570. 10.1371/journal.pone.0027570

Jinbo U., Kato T., Ito M., 2011. Current progress in DNA barcoding and future implications for entomology. Entomol. Sci. 14(2): 107–124.

Jones K.E., Patel N.G., Levy M.A., Storeyard A., Balk D., Gittleman J.L., Daszak P., 2008. Global trends in emerging infectious diseases. Nature 451: 990–994.

Johnson R.N., Wilson-Wilde L., Linacre A., 2014. Current and future directions of DNA in wildlife forensic science. Forensic Sci. Int-Gen 10: 1–11.

Kent R.J., 2009. Molecular methods for arthropod bloodmeal identification and applications to ecological and vector-borne disease studies. Mol. Ecol. Resour. 9(1): 4–18.

Khan F.A.A., Solari S., Swier V.J., Larsen P.A., Abdullah M.T., Baker R.J., 2010. Systematics of Malaysian woolly bats (Vespertilionidae: *Kerivoula*) inferred from mitochondrial, nuclear, karyotypic, and morphological data. J. Mammal. 91(5): 1058–1072.

King R.A., Read D.S., Traugott M., Symondson W.O.C., 2008. Molecular analysis of predation: a review of best practice for DNA-based approaches. Mol. Ecol. 17(4): 947–963.

Krüger F., Clare E.L., Greif S., Siemers B.M., Symondson W.O.C., Sommer R.S., 2014a. An integrative approach to detect subtle trophic niche differentiation in the sympatric trawling bat species *Myotis dasycneme* and *Myotis daubentonii*. Mol. Ecol. 23(15): 3657–3671.

Krüger F., Clare E.L., Symondson W.O., Keišs O., Petersons G., 2014b. Diet of the insectivorous bat *Pipistrellus nathusii* during autumn migration and summer residence. Mol. Ecol. 23(15): 3672–3683.

Kruskop S.V., Borisenko A.V., 2013. A new species of South-East Asian *Myotis* (Chiroptera: Vespertilionidae), with comments on Vietnamese “whiskered bats”. Acta Chiropterologica 15(2): 293–305.

Kruskop S.V., Borisenko A.V., Ivanova N.V., Lim B.K., Eger J.L., 2012. Genetic diversity of northeastern Palaearctic bats as revealed by DNA barcodes. Acta Chiropterologica 14(1): 1–14.

Kvist S., 2013. Barcoding in the dark? A critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. Mol. Phylogenet. Evol. 69(1): 39–45.

Larsen R.J., Knapp M.C., Genoways H.H., Khan F.A.A., Larsen P.A., Wilson D.E., Baker R.J., 2012. Genetic diversity of Neotropical *Myotis* (Chiroptera: Vespertilionidae) with an emphasis on South American species. PLoS One 7(10): e46578. doi:10.1371/journal.pone.0046578

Lassen S.B., Nielsen S.A., Skovgård H., Kristensen M., 2011. Molecular identification of bloodmeals from biting midges (Diptera: Ceratopogonidae: *Culicoides* Latreille) in Denmark. Parasitol. Res. 108(4): 823–829.

Lee P.S., Sing K.W., Wilson J.J., 2015. Reading Mammal Diversity from Flies: The Persistence Period of Amplifiable Mammal mtDNA in Blowfly Guts (*Chrysomya megacephala*) and a New DNA Mini-Barcode Target. PLoS One 10(4): e0123871. doi: 10.1371/journal.pone.0123871

Li S., Sun K., Lu G., Lin A., Jiang T., Jin L., Hoyt J.R., Feng J., 2015. Mitochondrial genetic differentiation and morphological difference of *Miniopterus fuliginosus* and *Miniopterus magnater* in China and Vietnam. Ecol. Evol. 5(6): 1214–1223.

Li J., Zheng X., Cai Y., Zhang X., Yang M., Yue B., 2015. DNA barcoding of Murinae (Rodentia: Muridae) and Arvicolinae (Rodentia: Cricetidae) distributed in China. Mol. Ecol. Resour. 15(1): 153–167.

Lim B.K., 2012. Preliminary assessment of Neotropical mammal DNA barcodes: an underestimation of biodiversity. Open Zool. J. 5: 10–17.

Linacre A., Tobe S.S., 2011. An overview to the investigative approach to species testing in wildlife forensic science. Investig. Genet. 2: 2.

Linacre A., Gusmao L., Hecht W., Hellmann A.P., Mayr W.R., Parson W., Prinz M., Schneider P.M., Morling N., 2011. ISFG: recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations. Forensic Sci. Int-Genet. 5(5): 501–505.

Long J.L., 2003. Introduced mammals of the world. Their History, Distribution and Influence. Wallingford, United Kingdom.

López-Wilchis R., Guevara-Chumacero L.M., Pérez N.Á., Juste J., Ibáñez C., Barriga-Sosa I.D., 2012. Taxonomic status assessment of the Mexican populations of funnel-eared bats, genus *Natalus* (Chiroptera: Natalidae). Acta Chiropterologica 14(2): 305–316.

Lorenz J.G., Jackson W.E., Beck J.C., Hanner R., 2005. The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. Philos. T. Roy. Soc. B 360: 1869–1878.

Lu L., Chesters D., Zhang W., Li G., Ma Y., Ma H., Song X., Wu H., Meng F., Zhu C., Liu, Q., 2012. Small mammal investigation in spotted fever focus with DNA-barcoding and taxonomic implications on rodents species from Hainan of China. PLoS One 7(8): e43479. doi:10.1371/journal.pone.0043479

Luo J.Y., Yan D., Song J.Y., Zhang D., Xing X.Y., Han Y.M., Yang M.H., Dong X.P., Peng C., Chen S.L., Xiao X.H., 2013. A strategy for trade monitoring and substitution of the organs of threatened animals. Sci. Rep. 3: 3108.

Mack R.N., Simberloff D., Lonsdale M.W., Evans H., Clout M., Bazzaz F.A., 2000. Biotic invasions: causes, epidemiology, global consequences, and control. Ecol. Appl. 10(3): 689–710.

Martinoli A., Bertolino S., Preatoni D.G., Balduzzi A., Marsan A., Genovesi P., Tosi G., Wauters L.A., 2010. Headcount 2010: the multiplication of the grey squirrel populations introduced to Italy. Hystrix 21(2): 127-136. doi:10.4404/hystrix-21.2-4463

McDonough M.M., Ferguson A.W., Ammerman L.K., Granja-Vizcaino C., Burneo S.F., Baker R.J., 2011. Molecular verification of bat species collected in Ecuador: Results of a country-wide survey. Occ. Pap. The Museum of Texas Tech University 301: 1–28.

Melo-Ferreira J., Boursot P., Carneiro M., Esteves P.J., Farelo L., Alves P.C., 2012. Recurrent introgression of mitochondrial DNA among Hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulation. Syst. Biol. 61: 367–381.

Michalski F., Valdez F.P., Norris D., Zieminski C., Kashivakura C.K., Trinca C.S., Smith H.B., Vynne C., Wasser S.K., Metzger S.P., Eizirik E., 2011. Successful carnivore identification with faecal DNA across a fragmented Amazonian landscape. Mol. Ecol. Resour. 11(5): 862–871.

Minhós T., Wallace E., Da Silva M.J.F., Sá R.M., Carmo M., Barata A., Bruford M.W., 2013. DNA identification of primate bushmeat from urban markets in Guinea-Bissau and its implications for conservation. Biol. Conserv. 167: 43–49.

Mitchell A., 2015. Collecting in collections: a PCR strategy and primer set for DNA barcoding of decades-old dried museum specimens. Mol Ecol. Resour. (Early View) doi:10.1111/1755-0998.12380

Moore J.E., 1988. A key for the identification of animal hairs. Sci. Justice 28(5): 335–339.

Müller L., Gonçalves G.L., Cordeiro-Estrela P., Marinho J.R., Althoff S.L., Testoni A.F., González E.M., Freitas T.R., 2013. DNA barcoding of sigmodontine rodents: identifying wildlife reservoirs of zoonoses. PLoS One 8(11): e80282. doi:10.1371/journal.pone.0080282

Muñoz J., Ruiz S., Soriguer R., Alcaide M., Viana D.S., Roiz D., Vázquez A., Figuerola J., 2012. Feeding patterns of potential West Nile virus vectors in South-West Spain. PLoS One 7(6): e39549. doi:10.1371/journal.pone.0039549

Muturi C.N., Ouma J.O., Malele I.I., Ngure R.M., Rutto J.J., Mithöfer K.M., Enyaru J., Masiga D.K., 2011. Tracking the feeding patterns of tsetse flies (*Glossina* genus) by analysis of bloodmeals using mitochondrial cytochromes genes. PLoS One 6(2): e17284. doi:10.1371/journal.pone.0017284

Nasi R., Brown D., Wilkie D., Bennett E., Tutin C., van Tol G., Christophersen T., 2008. Conservation and use of wildlife based resources: the bushmeat crisis. In: Technical Series no. 33. Secretariat de la Convention sur la diversite biologique, Montreal et Centre pour la recherche forestiere (CIFOR), Bogor, Indonesia. 50.

Nicolas V., Schaeffer B., Missoup A.D., Kennis J., Colyn M., Denys C., Tatard C., Cruaud C., Laredo C., 2012. Assessment of three mitochondrial genes (16S, Cytb, CO1) for identifying species in the Praomyini tribe (Rodentia: Muridae). PLoS One 7(5): e36586. doi:10.1371/journal.pone.0036586

Nijman V., Aliabadian M., 2010. Performance of distance-based DNA barcoding in the molecular identification of Primates. C. R. Biol. 333(1): 11–16.

Ondrejicka D.A., Locke S.A., Morey K., Borisenko A.V., Hanner R.H., 2014. Status and prospects of DNA barcoding in medically important parasites and vectors. Trends Parasitol. 30(12): 582–591.

Pedersen A.B., Jones K.E., Nunn C.L., Altizer S., 2007. Infectious diseases and extinction risk in wild mammals. Conserv. Biol. 21(5): 1269–1279.

Pisanu B., Obolenskaya E.V., Baudry E., Lissovsky A.A., Chapuis J.-L., 2013. Narrow phylogeographic origin of five introduced populations of the Siberian chipmunk *Tamias (Eutamias) sibiricus* (Laxmann, 1769)(Rodentia: Sciuridae) established in France. Biol. Invasions 15(6): 1201–1207.

Price P.W. 1980. Evolutionary biology of parasites. Princeton University Press, Princeton, New Jersey.

Puillandre N., Bouchet P., Boisselier-Dubale M.C., Brisset J., Buge B., Castelin M., Chagnoux S., Corbari T.C.L., Lambourdière J., Louzet P., Marani G., Rivasseau A., Silva N., Terryn Y., Tillier S., Utge J., Samadi S., 2012. New taxonomy and old collections: integrating DNA barcoding into the collection curation process. Mol. Ecol. Resour. 12(3): 396–402.

Ratnasingham S., Hebert P.D.N., 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). Mol. Ecol. Notes 7: 355–364.

Ratnasingham S., Hebert P.D.N., 2013 A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. PLoS ONE 8(8): e66213. doi10.1371/journal.pone.0066213

Razgour O., Clare E.L., Zeale M.R., Hanmer J., Schnell I.B., Rasmussen M., Gilbert T.P., Jones G., 2011. High-throughput sequencing offers insight into mechanisms of resource partitioning in cryptic bat species. Ecol. Evol. 1(4): 556–570.

Reeder D.M., Helgen K.M., Wilson D.E., 2007. Global trends and biases in new mammal species discoveries. Occ. Pap. The Museum of Texas Tech University 269: 1–35.

Rhyan J.C., Spraker T.R., 2010. Emergence of diseases from wildlife reservoirs. Vet. Pathol. 47(1): 34–39.

Roca A.L., Georgiadis N., Pecon-Slattery J., O'Brien S.J., 2001. Genetic Evidence for Two Species of Elephant in Africa. Science 293: 1473–1477.

Rodgers T.W., Janecka J.E. 2013. Applications and techniques for non-invasive faecal genetics research in felid conservation. Eur. J. Wildl. Res. 59(1): 1–16.

Rodrigues A.S., Grenyer R., Baillie J.E., Bininda-Emonds O.R., Gittlemann J.L., Hoffmann M., Safi K., Schipper J., Stuart S.N., Brooks T., 2011. Complete, accurate, mammalian phylogenies aid conservation planning, but not much. Philos. Trans. R. Soc. London [Biol] 366(1578): 2652–2660.

Romeo C., Ferrari N., Lanfranchi P., Saino N., Santicchia F., Martinoli A., Wauters L.A., 2015. Biodiversity threats from outside to inside: effects of alien grey squirrel (Sciurus carolinensis) on helminth community of native red squirrel (Sciurus vulgaris). Parasitol. Res. 114(7): 2621-2628. doi:10.1007/s00436-015-4466-3

Ruiz-García M., Pinedo-Castro M., Shostell J.M., 2014. How many genera and species of woolly monkeys (Atelidae, Platyrrhine, Primates) are there? The first molecular analysis of *Lagothrix flavicauda*, an endemic Peruvian primate species. Mol. Phylogenet. Evol. 79: 179–198.

Ryder O.A., 1986. Species conservation and systematics: the dilemma of subspecies. Trends Ecol. Evol. 1: 9–10.

Sanches A., Tokumoto P.M., Peres W.A., Nunes F.L., Gotardi M.S., Carvalho C.S., Pelizzon C., Godoi T.G., Galetti M., 2012. Illegal hunting cases detected with molecular forensics in Brazil. Investig. Genet. 3(1): 1–5.

Sedlock J.L., Krüger F., Clare E.L., 2014. Island bat diets: does it matter more who you are or where you live?. Mol. Ecol. 23(15): 3684–3694.

Shokralla S., Spall J.L., Gibson J.F., Hajibabaei M., 2012. Next generation sequencing technologies for environmental DNA research. Mol. Ecol. 21:1794–1805.

Savolainen V., Cowan R.S., Vogler A.P., Roderick G.K., Lane R. 2005. Towards writing the encyclopedia of life: an introduction to DNA barcoding. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360(1462): 1805–1811.

Shehzad W., Riaz T., Nawaz M.A., Miquel C., Poillot C., Shah S.A., Pompanon F., Coissac E., Taberlet P., 2012. Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. Mol. Ecol. 21(8): 1951–1965.

Silva T.L., Godinho R., Castro D., Abáigar T., Brito J.C., Alves P.C., 2014. Genetic identification of endangered North African ungulates using noninvasive sampling. Mol. Ecol. Resour. 15(3): 652–661.

Song H., Buhay J.E., Whiting M.F., Crandall K.A., 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proc. Natl. Acad. Sci. USA 105: 13486–13491.

Sousa L.C.C., Gontijo C.M.F., Lacorte G.A., Meireles S.N., Silva A.P., Fonseca C.G., 2012. Molecular characterization of an opossum *Didelphis albiventris* (Marsupialia, Didelphidae) population in an urban fragment of Brazilian Atlantic Rain Forest and support to species barcode identification. Genet. Mol. Res. 11(3): 2497–2496.

Stanton D.W., Hart J., Vosper A., Kumpel N.F., Wang J., Ewen J.G., Bruford M.W., 2014. Non-invarive generic identification confirms the presence of the endangered okapi *Okapia johnstoni* south-west of the Congo River. Oryx (First View). doi:10.1017/S0030605314000593

Strayer D.L., 2012. Eight questions about invasions and ecosystem functioning. Ecol. Lett. 15(10): 1199–1210.

Symondson W.O.C., 2002. Molecular identification of prey in predator diets. Mol. Ecol. 11(4): 627—641.

Symondson W.O.C., Harwood J.D., 2014. Special issue on molecular detection of trophic interactions: unpicking the tangled bank. Introduction. Mol. Ecol. 23(15): 3601–3604.

Taberlet P., Coissac E., Pompanon F., Brochmann C., Willerslev E., 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol. Ecol. 21(8): 2045—2050.

Taylor L.H., Latham S.M., Mark E.J., 2001. Risk factors for human disease emergence. Philos. Trans. R. Soc. B 356(1411): 983–989.

Thiemann T.C., Brault A.C., Ernest H.B., Reisen W.K., 2012. Development of a high-throughput microsphere-based molecular assay to identify 15 common bloodmeal hosts of *Culex* mosquitoes. Mol. Ecol. Resour. 12(2): 238–246.

Thoisy B.D., Pavan A.C., Delaval M., Lavergne A., Luglia T., Pineau K., Ruedi M., Rufray V., Catzeflis F., 2014. Cryptic diversity in common mustached bats *Pteronotus* cf. *parnellii* (Mormoopidae) in French Guiana and Brazilian Amapa. Acta Chiropterologica 16(1): 1–13.

Thompson R.C., Kutz S.J., Smith A., 2009. Parasite zoonoses and wildlife: emerging issues. Int. J. Environ. Res. Public Health 6(2): 678–693.

Thomsen P.F., Willerslev E., 2015. Environmental DNA –– An emerging tool in conservation for monitoring past and present biodiversity. Biol. Conserv. 183: 4–18.

Tillmar A.O., Dell'Amico B., Welander J., Holmlund G., 2013. A universal method for species identification of mammals utilizing next generation sequencing for the analysis of DNA mixtures. PLoS One 8(12): e83761. doi:10.1371/journal.pone.0083761

Triant D.A., DeWoody J.A., 2007. The occurrence, detection, and avoidance of mitochondrial DNA translocations in mammalian systematics and phylogeography. J. Mammal. 88(4): 908–920.

Valentini A., Miquel C., Nawaz M.A., Bellemain E., Coissac E., Pompanon F., Gielly L., Cruaud C., Nascetti G., Wincker P., Swenson J.E., Taberlet P., 2009. New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. Mol. Ecol. Resour. 9(1): 51–60.

Vesterinen E.J., Lilley T., Laine V.N., Wahlberg N., 2013. Next generation sequencing of fecal DNA reveals the dietary diversity of the widespread insectivorous predator Daubenton's Bat (*Myotis daubentonii*) in Southwestern Finland. PLoS One 8(11): e82168. doi:10.1371/journal.pone.0082168

Viricel A., Rosel P.E., 2012. Evaluating the utility of *cox1* for cetacean species identification. Mar. Mam. Sci. 28(1): 37–62.

Wilson D.E., Reeder D.M., 2005. Mammal species of the world: a taxonomic and geographic reference. Baltimore, Maryland: Johns Hopkins University Press.

Wilson J.J., Sing K.W., Halim M.R.A., Ramli R., Hashim R., Sofian-Azirun M., 2014. Utility of DNA barcoding for rapid and accurate assessment of bat diversity in Malaysia in the absence of formally described species. Genet. Mol. Res. 13(1): 920–925.

Yan D., Luo J.Y., Han Y.M., Peng C., Dong X.P., Chen S.L., Sun L.G., Xiao X.H., 2013. Forensic DNA barcoding and bio-response studies of animal horn products used in traditional medicine. PLoS One 8(2): e55854. doi:10.1371/journal.pone.0055854

Yoccoz N.G., Brathen K.A., Gielly L., Haile J., Edwards M.E., Goslar T., Von Stedingk H., Brysting A.K., Coissac E., Pompanon F., Sønstebø J.H., Miquel C., Valentini A., De Bello F., Chave J., Thuiller W., Wincker P., Cruaud C., Gavory F., Rasmussen M., Gilbert M.T.P., Orlando L., Brochmann C., Willerslev E., Taberlet P., 2012. DNA from soil mirrors plant taxonomic and growth form diversity. Mol. Ecol. 21(15): 3647–3655.

Zhang R.L., Zhang B., 2014. Prospects of using DNA barcoding for species identification and evaluation of the accuracy of sequence databases for ticks (Acari: Ixodida). Ticks Tick Borne Dis. 5(3): 352–358.

Associate Editor: D.G. Preatoni

4

Towards a Universal Molecular Approach for the Quality Control of New Foodstuffs

Andrea Galimberti, Anna Sandionigi, Antonia Bruno, Ilaria Bruni, Michela Barbuto, Maurizio Casiraghi, and Massimo Labra

ZooPlantLab®, Department of Biotechnology and Biosciences, University of Milano-Bicocca, P.za della Scienza 2, 20126 Milan (MI), Italy

4.1 Food Quality and Safety Assessment in the Era of Genomics

The growing globalization of the food market has led to a corresponding increase in issues concerning the assessment of authenticity and safety of imported foods. Consumers are susceptible to any form of food alteration that may occur during the standard manufacturing processes and pay attention to food ingredients as these can influence nutritional and health conditions (Galimberti *et al.* 2013). The consumer is becoming more and more demanding in terms of food quality and safety, seeking products with exhaustive labelling containing information about the original raw materials and assurances that the product is free from harmful chemical and microbial contaminants.

Historically, food authenticity involved efforts to ensure human safety by preventing the spread of specific diseases that were thought to be transmitted through food products of animal or plant origin (Myers 2011). Several studies have since been undertaken with the aim of understanding the role and association of microorganisms, especially bacteria, in food. These investigations have not been limited to pathogenic microorganisms, but also regard those bacteria or yeasts involved in food transformation (e.g. fermentation) or preservation (Ross *et al.* 2002). In the last years, sensitive and fast laboratory methods to isolate and identify microorganisms in food products have been developed (Elmerdahl Olsen 2000; Ray & Bhunia 2013; Ceuppens *et al.* 2014; Sohier *et al.* 2014). Thanks to these technical advances, most of the commercial foods are routinely subjected to microbiological analyses to exclude harmful biological contamination. This condition represented an essential step for the spread of studies and protocols to guarantee the highest food safety for the consumer. As well as microbiological analysis, the development of food authentication tests was also required by the market to identify food substitutions and economic frauds (Myers 2011). As a consequence, over the past 25 years there have been considerable advances in the development and use of molecular techniques for the rapid detection of food components and food microorganisms (both beneficial and pathogenic). These approaches are normally based on detecting specific DNA or RNA target sequences using amplification processes, in particular the polymerase chain reaction (PCR). In many instances, molecular techniques have almost completely replaced conventional culture detection methods (Galimberti *et al.* 2013; Ceuppens *et al.* 2014). In the last years the increasing demand for molecular traceability systems has led to the adoption of universal DNA-based approaches. One of the most-used methods is DNA barcoding: a standardized method providing species identification through the analysis of the variability in

Advances in Food Biotechnology, First Edition. Edited by Ravishankar Rai V.
© 2016 John Wiley & Sons, Ltd. Published 2016 by John Wiley & Sons, Ltd.

a short DNA gene region (the 'barcode'; Casiraghi *et al.* 2010; Galimberti *et al.* 2013).

4.2 DNA Barcoding: General Characteristics and Applications for the Analysis of Modern Foodstuffs

The basic idea of DNA barcoding, first proposed by Hebert *et al.* (2003), is quite simple: through the analysis of variability in a single or in a few standard molecular marker(s), it is possible to discriminate biological entities (hopefully belonging to the taxonomic rank of species). DNA sequence(s) can be used to identify different species in the same way a supermarket scanner uses the black stripes of the UPC barcode to identify any purchase. This method relies on the assumption that the genetic variation between species exceeds that within species. Consequently, the ideal DNA barcoding analysis mirrors the distributions of intra- and interspecific variabilities separated by a distance called 'DNA barcoding gap' (Meyer & Paulay 2005; Wiemers & Fiedler 2007). The principal barcode regions for animals, plants and bacteria are fragments of the mitochondrial *coxI*, the plastidial *rbcL* + *matK*, and the ribosomal 16s rRNA genes, respectively (Hebert *et al.* 2003; Hollingsworth *et al.* 2011; Chakraborty *et al.* 2014). These short sequences, referred to as 'barcodes', should be amplified using universal primers (Hebert *et al.* 2003; Hajibabaei *et al.* 2007). Efforts in DNA barcoding development and management are coordinated by the International Barcode of Life project (iBOL; http://ibol.org/).

In the case of unidentified samples, the molecular barcode is compared with a library of reference sequences (e.g. the BOLD system, Ratnasingham & Hebert 2007, 2013), encompassing taxonomically defined species (Casiraghi *et al.* 2010; Sandionigi *et al.* 2012). The final goal of DNA barcoding pipeline is to build a robust, efficient and standardized system for species identification (Fig. 4.1).

In the modern context of food traceability, DNA barcoding allows the characterization not only of food raw materials but also of the associated microbial communities. A correct evaluation of the origin and safety of food components is also essential for new foodstuffs (e.g. the modern functional foods), where the microbial component plays a key role in enhancing their nutritional value.

In this chapter we introduce a new concept of food traceability through the analysis of all biological components of food products. Each aliment should be considered as a sort of complex ecosystem, where one or more raw materials are combined with one or more microorganisms to obtain a commercial product with suitable nutritional value and safety for the consumer.

Before introducing the advantages and fields of application of these molecular approaches, it is necessary to describe first the concept of food as 'biological ecosystem' (Giraffa & Neviani 2001; Montville & Matthews 2013) and the effects of biotechnological processes on the development and production of the new foodstuffs.

4.3 Microbiological Composition of Foodstuffs

Biotechnology procedures have been involved in food production since time immemorial, mainly by taking advantage of environmental microorganisms and their metabolisms such as bacteria and yeasts transforming raw materials into enriched foodstuffs. Well-known examples include: wine and alcohol, where biotransformation increases their nutritional value and extends their shelf life; yogurt and dairy products, where microorganisms transform milk into products exhibiting peculiar sensory characteristics (e.g. cheese); and bread and other bakery products obtained by the fermentation activity of selected yeasts. Pools of microorganisms can modify chemical and physical characteristics of raw materials to derive new metabolites and materials and therefore influence the organoleptic, safety and nutritional properties of the final transformed food products (Steinkraus 1997; Caplice & Fitzgerald 1999; van Hylckama Vlieg *et al.* 2011; Bull *et al.* 2013).

With the exception of traditionally biotransformed foods and beverages, an astounding number of edible products (including the emerging 'functional foods') involves microbial activities during at least one step of their production. An example is the large use of probiotics in dairy products. Microbes therefore play an important role in human food production and this trend is increasing with the manufacture of modern food products. Most sterile foods harbour one or more types of microorganisms with bacteria, yeasts, moulds and viruses being the most relevant. Some of these can potentially cause food spoilage because of their ability to grow in foods. At the same time, others have a key role in food transformation processes such as fermentation and biopreservation.

Moreover, in some cases the same microorganisms can mediate beneficial transformations of food raw material or trigger spoilage phenomena such as moulds in cheese. This condition could be related to the relative abundance of the microorganisms, their growth rate and the time of exposition

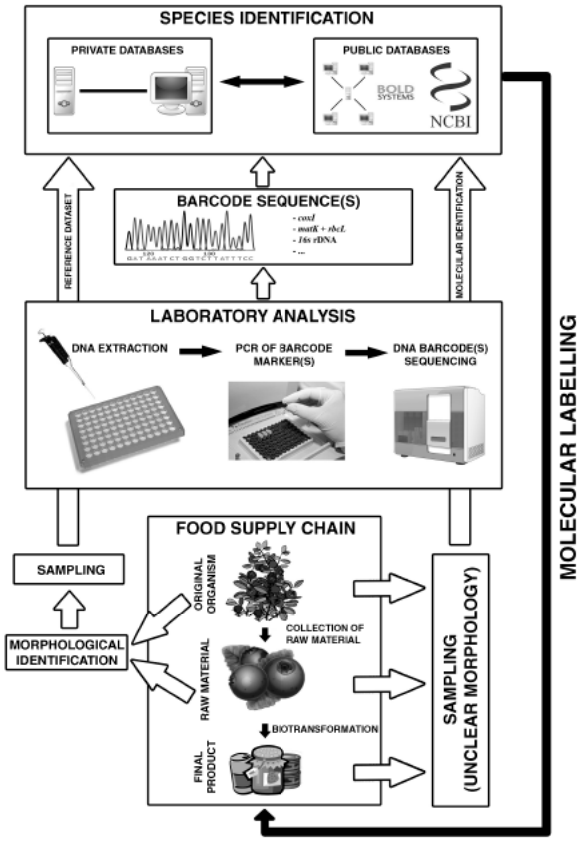


Figure 4.1 *Flowchart of DNA barcoding approach. In order to provide a proper labelling of any foodstuff, each element in the food supply chain should be subjected to identification procedures (from the original organism and the microorganisms used during biotransformation to the final shelf-product). After sampling at any stage of food production, standard laboratory techniques are used to obtain DNA barcoding sequences. When morphological data of the original organism (or raw material) are available, DNA barcodes can be stored in private and/or public databases and used as a 'reference' to compare and identify other barcodes from unknown samples (e.g. processed raw material, fermented products). Identification results are used to assign a 'molecular label' to the commercialized products and to assess their quality and safety.*

of microorganisms to the raw/food material. For these reasons, biotransformation processes often require a final step of microbial inactivation as in the case of wine production, where sodium bisulphite is introduced to arrest alcoholic fermentation (Amerine *et al.* 1980).

Currently, there are three main roles played by microorganisms involved in food manufacturing at the industrial level: (1) fermentation; (2) biopreservation; and (3) the production of new metabolites or functions (i.e. functional foods).

4.3.1 Fermentation

The fermentation process consists of the oxidation of carbohydrates to obtain a range of products with organic acids, alcohol and carbon dioxide as major end-products (Ray & Daeschel 1992). In addition a number of secondary metabolites, including vitamins, polyols or anti-oxidants, can be produced during fermentation thanks to the alternative metabolic pathways of microorganisms. These products may bring specific health benefits to humans (van Hyklckama Vlieg *et al.* 2011).

Food fermentation has been practised for millennia, being one of the earliest technologies developed by humans. The period during which humans began deliberately to use yeast or bacteria to start fermentation is still unknown. However, a shared assumption is that household production of fermented dairy products and alcoholic beverages began with the introduction of agriculture and animal husbandry approximately 10,000 years ago in Asia and the Middle East (Cordain *et al.* 2005). Littoral foragers in Asia are believed to have fermented vegetables prior to the development of crop-based agriculture from 8000 to 3000 BC (Lee 2009). At the same period, dairy fermentations were first adopted for cheese-making in the Middle East (Mesopotamia), likely following the domestication of cattle (Fox 1993). Many historical sources seem to indicate alcohol from fermented fruits as the first product of fermentation to be discovered (Breidt *et al.* 2013). Indeed, archaeologists have found molecular evidence for the production of fermented beverages dated back to 7000 BC and 5400–5000 BC from China and Mesopotamia, respectively (McGovern *et al.* 1996, 2004). Later, more sophisticated fermentation skills for the production of alcoholic beverages were developed around 4000 BC by Egyptians and Sumerians (Sicard & Legras 2011). While in Asia alcoholic beverages were mainly obtained from rice fermentation, in Egypt and Mesopotamia they were mostly made from fruits (wine), honey (mead) and cereals (beer). At all these sites, naturally occurring yeasts were probably used to start the transformation reactions (Sicard & Legras 2011).

The first evidence of the production of leavened bread from dough fermentation is also attributed to Egyptians (Samuel 1996) at around 1300–1500 BC. Coarsely milled emmer wheat (*Triticum dicoccum* Schübl.) was used to prepare dough. Starting from those times, fermentation skills and techniques expanded throughout the Old World. Documentation for the history of some food products (e.g. bread) is scarce; in other cases however, the main steps of diffusion are better known as in the case of wine and beer (This *et al.* 2006). Wine production spread first in Mediterranean countries such as Greece (5000 BC), Italy (900 BC), France (600 BC), Northern Europe (100 AD) and finally to the Americas (1500 AD) (Grassi *et al.* 2002; De Mattia *et al.* 2009; Zecca *et al.* 2010). Brewing history followed two separate routes, depending on the characteristics of fermentation procedures (i.e. temperature and type of yeast). Ale beer technology was acquired from the Middle East by Germanic and Celtic tribes around the 1st century AD, whereas lager beer technology was introduced later during the Middle Ages in Europe (Hornsey 2003; Kodama *et al.* 2005).

Although fermentation has been exploited as a food processing method for thousands of years, it was only in the last two centuries that bacteria, yeasts and other microorganisms were recognized as key points in the fermentation process (Ross *et al.* 2002). Pasteur made the first significant contribution to the development of modern food microbiology in 1857, when he demonstrated that alcohol produced through fermentation in beer and wine resulted from microbial activity rather than 'abiotic' chemical processes. A few decades later, in Denmark, Hansen developed the first technique using pure cultures of selected yeast strains for brewing (Carlsberg Yeast No. 1 *Saccharomyces carlsbergensis*, now classified as a strain of *Saccharomyces cerevisiae*).

These theoretical and experimental advances emerged at the time of the industrial revolution: a period of population growing and concentration in larger cities. This phenomenon resulted in a substantial shift of food production from the artisanal level to large food industries to satisfy the increasing demand of expanding and globalized markets (Ross *et al.* 2002). Industrialization increased large-scale processes for the massive production of fermented foods and beverages using yeast to obtain beer and wine and lactic acid bacteria (LAB) for a great number of dairy, vegetable and many other foodstuffs.

Subsequent advances in the sector of industrial fermentation arrived in the first half of the twentieth century and encompassed, among others, the employment of *Aspergillus niger* (a mould) for the manufacturing of citric acid and *Penicillium chrysogenum* for the production of the antibiotic penicillin. Due to the continuous discoveries in biotechnology and genetic engineering in the last 20 years, fermentation is definitively recognized as an industrialized and life-science-driven technology (Waites *et al.* 2009). Fermentations today play the primary role in the manufacture of many different foods such as: production of vitamin supplements (Sybesma *et al.* 2004; Santos *et al.* 2008); the introduction of specific functional attributes in food (Pham & Shah 2009); the removal of unwanted compounds (Amoa-Awua *et al.* 1997); and the delivery of probiotics (Bull *et al.* 2013).

Thanks to this long history and the advancement of biotechnologies, today there is an astonishing variety of

fermented foods deriving from a broad range of food substrates including plants, meat, milk and many others (Ross *et al.* 2002). In modern society, fermented foods substantially contribute to enrich the human diet providing a large amount of flavours, aromas and textures (van Hyklckama Vlieg *et al.* 2011). Fermentation plays a key role in several food industrial compartments (Bourdichon *et al.* 2012) including: (1) food preservation (production of organic acid, ethanol and many others acting as inhibitory metabolites; Ross *et al.* 2002; Breidt *et al.* 2013); (2) development of desired sensory properties (i.e. organoleptic quality; Singh *et al.* 2012; Dueñas-Sánchez *et al.* 2014); (3) improvement of nutritional value (van Boekel *et al.* 2010; Zhang *et al.* 2010); and (4) improvement of food safety through pathogen inhibition (Smaoui *et al.* 2010; Cizeikiene *et al.* 2013).

Currently, fermented foods represent approximately one-third of the human diet (Campbell-Platt 1994). For this reason, suitable tools for the control of fermented food are needed to characterize the composition of final transformed products.

Food fermentations are classified today by categories, classes or commodity (see Steinkraus 1997), although the boundaries between the classification criteria are often hazy. For example, the principal categories are: alcohol (e.g. beer, wine); acetic acid (e.g. apple cider, vinegars); lactic acid (dairy products); carbon dioxide (e.g. bread); and amino acids or peptides from protein (fish fermentations and others). This general classification relies on the prediction of the product of the reaction (Steinkraus 1996; Johnson & Steele 2013). Although these categories include most of the food and beverages subjected to fermentation, this classification is not exhaustive because it does not take into account which microorganisms are involved in the reaction. Moreover, different microorganisms can produce a number of different secondary metabolites from the same raw material. An exhaustive molecular characterization of microorganisms involved in fermentation processes could help to clarify the characteristics of food (chemical, nutritional, etc.) and the effectiveness of the biotransformation process. A large portion of the industrial market is devoted to the production of new fermenting microorganisms that show peculiar properties in terms of growth rate, stability and yield of secondary metabolites production. A reliable molecular characterization system allows these strains to be differentiated and the intellectual property of biotechnological companies preserved.

4.3.2 Biopreservation

Being perishable, food and beverages require treatments that elongate their shelf life in order to maintain an acceptable level of quality and safety from manufacturing to consumption. Although there is evidence for early historical approaches to preserve food and beverages, the first scientific contribution was developed by Pasteur during the nineteenth century (Bulloch 1938; Ross *et al.* 2002). Pasteur introduced a food conservation approach consisting of a simple heating treatment, originally used as a way of preventing wine and beer from souring due to microorganisms. This treatment takes his name: pasteurization.

Modern food preservation approaches act directly on the biological activity of microorganisms and/or their metabolites (Ross *et al.* 2002). These approaches are generally known as biopreservation. One of the earliest biopreservation technologies is once again fermentation (Caplice & Fitzgerald 1999). Preservation by fermentation relies on the fact that the end-products of oxidation of carbohydrates (e.g. acids, alcohol and CO_2) can control the growth of food spoilage microorganisms (Ross *et al.* 2002). Antimicrobial compounds, as well as proteinaceous substances able to inhibit or contrast food spoilage, are produced by starter microorganisms that were traditionally involved in naturally occurring fermentations. The exploitation of such naturally produced antagonists holds great potential, especially in recent years as consumer awareness of so-called 'green technologies' (i.e. minimally processed foods, free from chemical and harmful preservatives) has grown. This scenario has resulted in a large number of new biopreservatives (e.g. bacteriocins) which can also be combined to elongate shelf life and enhance quality of foodstuffs (Ross *et al.* 2002; Montville & Chikindas 2013).

Modern industrial production now exploits the use of specific strain starter systems for biopreservation purposes. In western countries, a considerable number of starter strains (used mainly in fermented foods) are derived from the activity of lactic acid bacteria (LAB). LAB are among the best-studied microorganisms in biopreservation approaches and their biochemical and metabolic characteristics in mediating antibiosis are well known (Cizeikiene *et al.* 2013; Crowley *et al.* 2013; Johnson & Steele 2013; Sohier *et al.* 2014). The carbohydrate catabolism of LAB produces a wide panel of compounds. Some of these contribute to the flavour, aroma and texture of the final fermented products (see Johnson and Steele 2013 for a review), whereas others include molecules with antimicrobial functions such as organic acids, ethanol, hydrogen peroxide and bacteriocins (Caplice & Fitzgerald 1999). For example, lactic acid, acetic acid and other organic acids have low pH environmental values, therefore impeding the growth of several pathogenic and spoilage microorganisms (e.g. bacteria, yeasts and moulds). LAB and other microorganisms can also produce bacteriocins, peptides and

proteins with antagonistic activity against contaminant bacteria (Ross *et al.* 2002). More than 170 bacteriocins have been described, and they are usually classified according to their chemistry and molecular weight (Klaenhammer 1993; Hammami *et al.* 2010).

The last frontier of biopreservation is the use of microbial antagonistic molecules produced by other microorganisms to functionalize food packages. These types of packaging interact with the product or the headspace between the package and the food, achieving the reduction or inhibition of microorganism growth (Appendini & Hotchkiss 2002; Sofos *et al.* 2013). Active packaging systems include natural antimicrobials as additives including nisin, one of the most studied and commercialized bacteriocins (O'Grady & Kerry 2008). As an example, bacteriocins applied to food packaging materials were found to inhibit *Listeria monocytogenes* on meats (Ming *et al.* 1997; Gálvez *et al.* 2007).

The widespread use of biopreservation approaches has largely reduced the adoption of physical techniques and chemical substances with antimicrobial activity, therefore enhancing food quality and safety. However, biopreservation techniques involve the introduction of new microorganisms and/or biomolecules into raw materials and final food products, which should be taken into account in a context of food traceability. An element of additional concern is that in some cases producers, with the intent of reducing the food alteration risk, use a mix of microorganisms without fully considering the characteristics of each biological component. In this case, a proper traceability system should address the identification of both the original raw materials and the microorganisms involved in the biopreservation process.

4.3.3 Functionalization

The continuous progresses in biotechnology, alongside the considerable change in consumer demands, are driving a new trend in food production towards the research and development of so-called 'functional foods' (Roberfroid 2000). This term was first proposed and legally approved in Japan in terms of Foods for Specified Health Use (FOSHU). A recent working definition proposed by the European Commission on Concerted Action on Functional Food Science in Europe (FUFOSE; European Commission 2010) is: 'A food can be regarded as functional if it is satisfactorily demonstrated to affect beneficially one or more target functions in the body, beyond adequate nutritional effects, in a way that is relevant to either an improved state of health and well-being and/or reduction of risk of disease. It is consumed as part of a normal food pattern. It is not a pill, a capsule or any form of dietary supplement.'

Functional foods must meet three basic conditions. They must be (1) derived from naturally occurring ingredients; (2) consumed as a part of daily diet; and (3) directly involved in the regulation of specific processes such as delaying the aging process, preventing the risk of disease and improving immunological ability (Betoret *et al.* 2011). These types of food also showed several synonyms: designer foods, medicinal foods, nutraceuticals, therapeutic foods, superfoods, foodiceuticals and medifoods (Shah 2007). Consumers are increasingly aware that foods contribute directly to their health (Jones & Jew 2007; Siro *et al.* 2008). Indeed, in the current perspective, foods are intended not only as the primary source of nutrients but also to boost the immune system, reduce the risk of disease and enhance both physical and mental well-being (Stanton *et al.* 2005; Nöthlings *et al.* 2007; Sharma & Devi 2014). In this context, functional foods play an outstanding role (Betoret *et al.* 2011).

To improve the quality of functional foods, microbiologists have recently focused on those microorganisms and related compounds that have significant benefits for human health (Pfeiler & Klaenhammer 2013). This research had led to the spread of an emergent category of functional foods, including probiotics, prebiotics and synbiotics (Stanton *et al.* 2005). A probiotic (from the Latin and Greek words meaning 'for life') is a live microorganism which, when administered in adequate amounts, confers a health benefit on the host (Joint FAO/WHO Working Group). Differently from the antagonistic action of antibiotics, the term 'probiotics' was initially adopted to describe substances produced by one or more microorganisms that stimulate the growth of other microorganisms in a host (Lilly & Stillwell 1965). Prebiotics are non-digestible food ingredients that stimulate growth and/or activity of other bacteria, with positive effects on the health of the host (Gibson *et al.* 2004; Pfeiler & Klaenhammer 2013). As an example, if the amount of prebiotics in the diet increases, healthy bacteria in the gut of the host also increase (Gibson *et al.* 2003). Food can naturally contain prebiotics as in the case of fermented foods; alternatively, they can be fortified with them during the manufacturing process, with the final aim of increasing probiotic efficacy in the host (Ranadheera *et al.* 2010). Finally, when both prebiotics and probiotics are present in the same food product, those functional foods are referred to as synbiotics (Sharma & Devi 2014).

Although the beneficial properties of some fermented foods have been known of since the Roman era (Stanton *et al.* 2005), the concept of probiotic is usually attributed to the work of Metchnikoff at the beginning of the twentieth

century. He observed that the consumption of fermented milk could reverse putrefactive effects on the gut microflora of patients (Metchnikoff 2004). Metchnikoff theorized that the microbial flora of the human gut could be related to the occurrence of infections and other problems. He isolated a *Lactobacillus* culture from Bulgarian fermented milk (known for its healthy properties) and successfully implanted the strain in the intestine of patients, observing beneficial effects. Today, there is growing scientific evidence that the maintenance of healthy gut microflora may provide protection against gastrointestinal disorders, pathologies and even cancer (see Sharma and Devi 2014 for a review of the principal findings). Moreover, further studies indicate that the use of probiotics not only leads to health benefits in the gastrointestinal tract, but also helps to maintain the natural balance of the autochthonous microbial population of the respiratory and urogenital tracts (Hao *et al.* 2011; Maldonado *et al.* 2012; De Gregorio *et al.* 2014). As a direct consequence of this trend in the use of probiotics, a wide range of functional foods including probiotics and/or prebiotics became suitable for large-scale industrial production, with additional improvements to maintain good viability of microorganisms during storage (Stanton *et al.* 2005). Due to their health benefits, probiotic bacteria have been increasingly included in yoghurts and fermented dairy products. Today, they are also delivered through cereals, infant formulas, freeze-dried supplements (capsules, pills, liquid suspensions and sprays) and fruit juices (Sharma & Devi 2014).

Another important research trend concerns the development of probiotic cultures for use in both agricultural and pets (Pfeiler & Klaenhammer 2013). The principal objectives of this strategy are the enhancement of animal growth and the reduction in the transfer of human enteric pathogens to the consumer. This issue is of primary importance because control of enteric pathogens at farm level can reduce the risk of subsequent food-borne illness.

A great number of genera of bacteria are used as probiotics, but the main species showing probiotic characteristics are *Lactobacillus acidophilus*, *Bifidobacterium* spp. and *L. casei* (Bull *et al.* 2013). An example is the Italian product known as Enterogermina®, registered and distributed since 1958.

Yeasts also play an important role as probiotics, with *Saccharomyces boulardii* as the most known probiotic fungus which has been successfully used for curing intestinal diseases (Guslandi *et al.* 2000; Czerucka *et al.* 2007). Several applications of probiotics and/or prebiotics have been studied, from the enhancement of immune response to positive effects in contrasting allergies and even AIDS or other pathologies. A complete list of references describing these applications can be found in Haller *et al.* (2010) and Sharma & Devi (2014).

In functional foods microorganisms might also indirectly promote food health by producing bioactive metabolites during fermentation processes; these are known as biogenics (Takano 2002). The most important of these are the B vitamins and bioactive peptides (see Stanton *et al.* 2005 for detailed references). In most cases, probiotic strains are covered by patents and the development of specific primers or probes for their rapid detection could be challenging.

As a general rule, the health or biopreservation benefits imparted by probiotic microorganisms are very strain specific (Sharma & Devi 2014). A reliable traceability system is therefore essential to guarantee their quality and safety.

4.4 Pathogenic Microorganisms and Food Spoilage

Microorganisms are able to alter raw materials and final food products, posing serious risks to the health of consumers and industrial economy. It has been estimated that about 30% of people in industrialized countries suffer from a food-borne disease each year and about 25% of global food production is lost due to microbial contamination (Bondi *et al.* 2014).

In better cases, microbial spoilage leads to an alteration of food organoleptic characteristics, whereas a wide panel of exogenous microorganisms are responsible for serious threats to human health (Settanni & Corsetti 2007; Newell *et al.* 2010). The main cause of food spoilage is the growth and metabolism of bacteria which form volatile substances that cause, for example, off-odours (Nattress & Jeremiah 2000; Gram & Dalgaard 2002). The microbial load in a food product is closely related to: (1) the conditions of the growing/farming environment; (2) the initial microbial concentration; and (3) the preservation method. Food-borne pathogens, as well as spoilage microorganisms, can already be present in the indigenous microbiota, or are introduced to the final food product by contamination during manufacturing (Newell *et al.* 2010). There are over 200 known microbial, chemical or physical agents that can cause illness when ingested (Acheson 1999). Among microorganisms, there is a great number of species and genera traditionally associated with human diseases and for which every food product should be tested in order to ensure their absence. *Salmonella* spp. is one of the major pathogens responsible for food-borne disease outbreaks throughout the world in humans and animals. *Salmonella* spp. typically causes intestinal infection, fever, abdominal cramps and diarrhoea. *S. enterica* is the most frequently isolated species from food-borne outbreaks (Jackson *et al.* 2013).

Other important and frequently reported food-borne pathogens belong to the following genera: *Campylobacter*, *Yersinia*, *Shigella*, *Vibrio*, *Clostridium*, *Bacillus*, *Listeria* and *Staphylococcus* (Settanni & Corsetti 2007; Doyle & Buchanan 2013). Both culture-dependent and DNA-based tests were developed for the detection of these microorganisms in food matrices, and most of these tests are routinely used by monitoring food centres (e.g. Scheu *et al.* 1998).

Although all detection methods show a consistent analytical sensitivity, their universality is far from complete. For this reason, each food should be subjected to multiple analyses; this unfortunately involves long analytical periods and high costs. Microorganism detection and unequivocal identification is an important issue for the food industry: a rapid assessment of a potential microbial risk can predict and enhance the shelf life of foodstuffs, avoiding health hazards and economic losses, through the application of preventive measures. A rapid and correct detection could enable appropriate medical care decisions to tackle an outbreak of pathogens, as well as can improve understanding of the epidemiology of food-borne infections. For all these reasons, inexpensive and reliable molecular identification methods are necessary to detect and identify food microorganisms associated with infections and diseases.

In most cases, species identification is not sufficient. An example is provided by *Escherichia coli* strains: most of them are harmless (Donnenberg & Whittam 2001), but many others are effectively or potentially harmful including the Shiga toxin-(STEC), the verocytotoxin-producers (VTEC) and the enterohaemorrhagic *E. coli* (EHEC). In these cases, molecular traceability systems are able to distinguish different strains.

Similarly, many microbial pathogens belong to the same genera of bacteria and yeasts used for food transformation (fermentation, biopreservation) and production of functional foods. For example, some proposed probiotic strains include genera *Bacillus* and *Enterococcus*, which contain species that are identified as major food-borne pathogens (Pfeiler & Klaenhammer 2013). Also in this context, molecular tools used for microorganism identification are the best choice for detection, capably distinguishing dangerous microorganisms from harmless ones.

4.5 Towards a Molecular Identification of Food-Related Microorganisms

The present taxonomy of bacteria is the starting point to identify a universal molecular approach for microorganism identification. This is a complex topic for biologists as well as an area of growing interest, partly because the definition of microbial species as a taxonomic unit lacks a commonly accepted theoretical basis (Felis & Dellaglio 2007). Microbial taxonomy directly influences a number of basic scientific and applied fields where microorganisms are involved, including food production, conservation and probiotic activity (Tautz *et al.* 2003). It has practical usefulness, for example in: (1) characterizing new isolates based on similarity to known taxa; (2) assessing and monitoring the use of industrial strains for food production; and (3) communicating to consumers which beneficial microbial ingredients are included or were involved during manufacturing. It should be underlined that microbiologists work with strains, as the strain is the microbial individual. Strains can show different functional or metabolic characteristics; however, when a large number of different strains are homogeneous under different criteria, it is possible to assign peculiar properties directly to the species.

Several analytical methods could have a different resolution power when studying a microorganism. To better support the identification of a microorganism, results from a large number of techniques should be compared: this practice was also known as 'polyphasic' (Colwell 1970; Vandamme *et al.* 1996).

An updated inventory of microorganisms commonly used in food/beverages production includes about 200 bacteria and 69 yeasts and moulds species, often including dozens of strains (Bourdichon *et al.* 2012).

Progress in bacterial taxonomy has always been dependent on advances in technology (Felis & Dellaglio 2007). To date, due to the shortcomings of conventional culture-based methods, genetic techniques have become increasingly important in food microbiology being the most rapid and accurate in identifying bacteria and other microorganisms, either as a complement or alternative to classical methods (Ceuppens *et al.* 2014; Chakraborty *et al.* 2014).

As well as increasing the sensitivity and specificity of the detection process, molecular approaches are less subjective in interpreting morphological and physiological or biochemical data (Settanni & Corsetti 2007). This tendency has made bacterial species identification by sequencing and phylogenetic analysis commonplace (Felis & Dellaglio 2007). Genomics now underlies a renaissance in food microbiology, therefore accelerating food safety monitoring (Ceuppens *et al.* 2014). Molecular approaches to taxonomy demonstrate several advantages in comparison to the conventional culture-based methods, such as:

1. The possibility of investigating microorganisms and strains, which are difficult (if not impossible) to culture *in vitro* (e.g. *Campylobacter*; Denis *et al.* 2001) or for

which a selective medium is unavailable (e.g. *Bacillus cereus*; Fricker *et al.* 2008). In order to increase sensitivity, traditional isolation methods include a selective enrichment and a facultative pre-enrichment step, both of which are laborious and time-consuming (Elmerdhal Olsen 2000).

2. The time required for a molecular detection system is generally much shorter than that for conventional culture-based methods. PCR-based approaches permit the detection of food-borne pathogens in a few hours with high sensitivity and reliability (Cunningham *et al.* 2010). Rapid detectability is of utmost importance when the target microorganisms are particularly slow growing: *Salmonella*, *Shigella*, *Yersinia* and *Campylobacter* required up to 5 days to grow and be identified (Cunningham *et al.* 2010).
3. DNA-based detection methods can specifically detect and quantify species (or strain), also overcoming the problem of co-occurrence of other dominant populations which could mask the target organisms (Solieri & Giudici 2010; Soler *et al.* 2012; Herbel *et al.* 2013).

Similarly to other detection systems, DNA-based techniques have some limitations (Ceuppens *et al.* 2014). In the context of food quality and safety assessment, the principal drawback is the detection of DNA traces from dead microorganisms, which can lead to false positives cases of contamination. However, several solutions now exist to assist for a correct interpretation of results (e.g. sample pretreatments and PCR selective protocols; Ceuppens *et al.* 2014).

The integration of DNA-based methods in the identification of microorganisms has led to a growing importance of molecular information in describing new species. The results of this innovative perspective in the study of biodiversity are well described under the concept of 'DNA taxonomy' (Tautz *et al.* 2003; Blaxter 2004). One of the principal aspects of DNA taxonomy is a strict standardization, which allows the taxonomic approach to be extended to vast groups of organisms not deeply related (Casiraghi *et al.* 2010). It also provides a framework for routine identification and represents the primary database for DNA barcoding (Vogler & Monaghan 2007). Thanks to the advancement of sequencing technologies and bioinformatics, the scientific community has standardized and ameliorated DNA sequencing approaches (Chakraborty *et al.* 2014).

Short unique sequences may help to discriminate microorganisms due to the existence of genetic variation in the closely related taxa. In this context, many scientists used PCR-amplified 16S rRNA gene for species-level typing of microorganisms (Schmidt & Relman 1994; Janda & Abbott 2007). The use of this region as barcode reveals a fast, reproducible and inexpensive method for discrimination due to its peculiar properties (Patel 2001). First, 16s rRNA gene is present in all the bacterial species. Secondly, the function of the gene is conserved in almost taxa. Finally, being approximately 1500 base pairs (bp) long, it contains sufficient information to discriminate species and in some cases strains (Heilig *et al.* 2002; Muñoz-Quezada *et al.* 2013). Last but not least, the 16s rRNA relies upon an impressive archive of reference sequences such as Greengenes (De Santis *et al.* 2006, http://greengenes.lbl.gov/cgi-bin/nph-index.cgi) and SILVA (Pruesse et al. 2007, http://www.arb-silva.de/Silva). For these reasons the 16s rRNA gene can be considered a sort of universal DNA barcode for microorganism identification. It can be evaluated for food traceability purposes because it can identify both pathogens and beneficial microorganisms (see also Table 4.1).

4.6 Towards a Standardized Molecular Identification of Food Raw Materials

4.6.1 From Molecular-Based Approaches to DNA Barcoding

Besides biotransformation processes mediated by microorganisms, the quality of food (e.g. nutritional value, sensory characteristics) is strongly related to the quality level of the raw materials (Konczak & Roulle 2011; Pereira *et al.* 2011). The analysis of raw materials is performed by various laboratory tests which represent the mandatory starting point for a proper food traceability system. In the last decades, the demand for efficient systems of food traceability has influenced the scientific research, leading to the introduction of a wide range of analytical approaches to the problem (Mafra *et al.* 2008; Fajardo *et al.* 2010; Bottero & Dalmasso 2011; Hellberg & Morrisey 2011). As in the case of microorganism detection and characterization, DNA markers have become the most effective instrument in the analysis of plant cultivars and animal breeds, and are also used to track raw materials in food industry processes (Woolfe & Primrose 2004; Mafra *et al.* 2008; Kumar *et al.* 2009). In the last 20 years, discontinuous molecular marker techniques such as RAPDs, AFLPs and their variants (i.e. ISSR, SSAP, SAMPL) have been used to characterize different kinds of raw material (Nijman *et al.* 2003; Grassi *et al.* 2006; Mafra *et al.* 2008; De Mattia *et al.* 2009; Fajardo *et al.* 2010; Chuang *et al.* 2011). The selection of the most suitable molecular approach depends on the sample

Table 4.1 *Selection of reference studies concerning the use of standardized molecular approaches (e.g. DNA barcoding) to assess quality, safety and traceability of foodstuffs and food-related microorganisms.*

DNA barcoding application	Raw material/food product	Notes	References
Raw materials tracebility	Aromatic plants	Identification of spices from fresh samples to manufactured or processed products	Kojoma *et al.* 2002; De Mattia *et al.* 2011; Theodoridis *et al.* 2012; Federici *et al.* 2013; Gismondi *et al.* 2013; Wang *et al.* 2013
	Legumes	Legume seeds traceability	Ganopoulos *et al.* 2012; Madesis *et al.* 2012
	Herbal infusions	Traceability of tea products	Stoeckle *et al.* 2011
	Fruit	Identification and traceability of mango	Hidayat *et al.* 2012
		Identification of *Citrus* species	Yu *et al.* 2011
		Identification of Goji	Xin *et al.* 2013
		Identification of berries	Jaakola *et al.* 2010
	Vegetables	Identification of *Capsicum* cultivars	Jarret 2008
	Medical plant and food supplements	Traceability of medicinal plants	Zuo *et al.* 2011
	Seafood	Identification of Amazonian commercial fish	Ardura *et al.* 2010
		Authentication and diversity assessment of commercially exploited fish stocks	Ardura *et al.* 2013
		Identification of tuna and other scombrid species in food products	Botti & Giuffra 2010
		Identification of smoked fish products	Smith *et al.* 2008
		Identification of crab meat products	Haye *et al.* 2012
	Meat	Identification of Bovidae species	Cai *et al.* 2011
		Meat identification	Lin *et al.* 2014
Microorganisms traceability	Dairy products	Determination of cheese origin by using 16S rDNA fingerprinting of bacteria communities	Arcuri *et al.* 2013
		Yeast populations associated with the artisanal cheese	Borelli *et al.* 2006; El-Sharoud *et al.* 2009; Gori *et al.* 2013
		Identification of lactic acid bacteria in cheese	Gala *et al.* 2008; Carraro *et al.* 2011
		Identification of potential risk factors of Gram-negative isolates in cheese	Coton *et al.* 2012
		Identification of moulds from the Taleggio cheese	Panelli *et al.* 2012
	Beverages	Characterization of contaminant bacteria in Iranian Kefir type drink	Hosseini *et al.* 2012
		Bacterial community during grape marc storage for the production of grappa	Maragkoudakis *et al.* 2014
	Functional foods	Identification of probiotics in fermented dairy products	Gueimonde *et al.* 2004; Tabasco *et al.* 2007; Raeisi *et al.* 2013
Commercial frauds and dangerous substitutions	Aromatic products	Identification of spices adulterants	Parvathy *et al.* 2014; Federici *et al.* 2014
	Vegetal flour	Identification of buckwheat in commercial foodstuffs	Hirao *et al.* 2005

Table 4.1 Continued

DNA barcoding application	Raw material/food product	Notes	References
	Legumes	Seed admixture and adulteration	Bosmali *et al.* 2012; Ganopoulos *et al.* 2012
	Fruit	Identification and adulteration of fruit-based products	Ng *et al.* 2005
	Oil	Oil adulteration	Ganopoulos *et al.* 2013
	Medicinal plants/ food supplements	Dangerous substitution of *Solanum lyratum* with *Aristolochia mollissima*	Li *et al.* 2012a
		Adulteration of herbal products	Newmaster *et al.* 2013
	Tea	Contamination of tea products	Dhiman & Singh 2003
	Seafood	Mislabeling in a commercial freshwater catfish from Brazil	Carvahlo *et al.* 2011
		Fraudulent substitutions in shark seafood products	Barbuto *et al.* 2010
		Puffer fish (tetrodotoxin) poisoning from mislabelled product	Cohen *et al.* 2009
		Mislabelling of commercial fish products	Filonzi *et al.* 2010; Galal-Kahallaf *et al.* 2014
	Dairy products	Contamination of raw milk	Kagkli *et al.* 2007
Molecular characterization of complex food matrices	Natural health products	Identification of pharmaceutical plants in commercial products	Wallace *et al.* 2012
	Juice and vegetal beverages	Juice authentication	Han *et al.* 2012; Li *et al.* 2012*b*; Faria *et al.* 2013
	Honey	Identification of pollen and plant residuals	Valentini *et al.* 2010
	Jams or yogurt	Identification of fruit in commercial products	Ortola-Vidal *et al.* 2007; Arleo *et al.* 2012
	Food supplements	Identification of allergenic plants	Yano *et al.* 2007; Madesis *et al.* 2013
	Dairy products	Microbiome characterization of raw milk and cheese	Delbès *et al.* 2007; Ercolini *et al.* 2009, 2012; Giannino *et al.* 2009; Masoud *et al.* 2011; Quigley *et al.* 2012; Callon *et al.* 2014
	Meat	Species identification of mixed meat specimens	Colombo *et al.* 2011
	Beverages	Characterization of wine and must microorganisms	Bokulich *et al.* 2012; Campanaro *et al.* 2014; David *et al.* 2014
		Characterization of kefir microbiome	Dobson *et al.* 2011

composition, the time needed for the analysis, the cost/effectiveness ratio and the expertise of laboratories. Furthermore, genomic techniques require high-quality DNA to work successfully because their efficiency can be negatively influenced by fragmented or inhibitor-rich DNA (Pafundo *et al.* 2007; Hellberg & Morrisey 2011).

Regarding sequencing-based systems, single-nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs) are the most frequently used because of their high level of polymorphism and reproducibility (Kumar *et al.* 2009). These approaches are used both in the identification of plant cultivars (Pasqualone *et al.* 1999; Labra *et al.* 2003) and animal breeds (Nijman *et al.* 2003) and in fraudulent commercial activities prevention (Chuang *et al.* 2011). However, being highly species specific, SSR and SNP approaches require access to a thorough knowledge of an organism's genome (strains/varieties or ecotypes) and their application is often limited to a single or a few closely related taxa.

Low levels of standardization and universality are the most relevant problems of DNA-based identification approaches. The introduction of DNA barcoding offers an innovative solution to this issue because it combines two important aspects dealing with modern taxonomy: standardization of the analytic procedure (i.e. amplifying a universal genomic region with universal primer pairs) and computerization (i.e. development and continuous improvement of a universal DNA database for all organisms) (Casiraghi *et al.* 2010). The 5′-end portion of mitochondrial *cox1* gene was suggested by Hebert and colleagues (2003) as standard DNA barcode region for metazoans. Based on preliminary results on *cox1* discriminatory power, specimens have been correctly identified at the species level with a success rate ranging from 98 to 100% in fish (Ward *et al.* 2005) and in several other animal groups (Hajibabaei *et al.* 2006; Ferri *et al.* 2009; Galimberti *et al.* 2012).

In terrestrial plants, mitochondrial DNA has slower substitution rates than in metazoans and shows intra-molecular recombination (Mower *et al.* 2007), therefore limiting its resolution in identification. For this reason, in 2009 the CBoL (Consortium for the Barcode of Life) Plant Working Group (Hollingsworth *et al.* 2009) suggested the combination of two plastidial loci (*rbcL* and *matK*) as core-barcode regions, because of the straightforward recovery rate of *rbcL* and the high resolution of *matK*. Among other potential barcodes, the *trnh-psbA* intergenic spacer is easily amplified and has a high genetic variability among closely related taxa (Bruni *et al.* 2010, 2012; De Mattia *et al.* 2012). The nuclear ITS region was also indicated as a supplementary DNA barcode region (Li *et al.* 2011) due to its higher evolution rate (Hollingsworth *et al.* 2011). The strength of this method relies on the availability of an international platform. Coordinated by the International Barcode of Life Project (iBOL), BOLD (barcode of life database) is a repository supporting the collection of DNA barcodes with the aim of creating a reference library for all living species (Ratnasingham & Hebert 2007, 2013).

4.6.2 Advantages and Limitations of DNA Barcoding in Food Traceability

The most important advantages of DNA barcoding comes from the rapid acquisition of molecular data with relatively low analyses costs. In the field of food traceability and safety, DNA barcoding has nowadays gained a role of primary importance (Galimberti *et al.* 2013).

Current technical advances have made DNA barcoding a sensitive, fast, cheap and reliable method for identifying and tracking a wide panel of raw materials and deriving food commodities (even in case of strongly processed food products), with several implications in the fields of forensic sciences, food traceability, diet analyses, monitoring of illegal trade of endangered species and use of foodstuffs or microorganisms for biotransformation (potentially harmful to human health). Due to its universality, DNA barcoding can be used in different contexts and by different operators. The cost- and time-effectiveness of DNA barcoding and the recent development of high-throughput sequencing technologies allow a certain degree of automation in species identification, which is particularly useful in simultaneous monitoring activities of different food batches and of the microorganisms used for their fermentation or preservation (Galimberti *et al.* 2015).

The growing use of DNA barcoding derives from a combination of some advantageous characteristics: (1) the falling cost of molecular analyses; (2) the increasing availability of equipped laboratories and skilled personnel; (3) the presence of freely available web-based resources to share and molecular reference data to consult; and (4) the increasing amount of informed consumers requiring high standards of quality in food products. This scenario has increased demand for a technique based on molecularization, standardization and computerization and, in this context, DNA barcoding is not only up to date but is the natural product of the 2000s (Galimberti *et al.* 2013).

International agencies or institutions responsible for quality control of raw materials or food commodities can co-operate by exchanging their data, hence creating population reference databases; the lack of such databases is the main limit of the method. While some groups of organisms

(e.g. fish) are well represented, a lot of work is required to provide a reliable source of reference DNA barcoding data for groups which have been poorly investigated. As a diagnostic method, DNA barcoding approach can be more or less fallacious, and it should be taken into account that failures are mainly in the essence of biological species rather than in the method (see Casiraghi *et al.* 2010 for a review). As an example, the method cannot yet be easily applied to the differentiation of GM (genetically modified) food raw materials, based on the standard molecular markers. The modified genomic tracts usually do not involve the plastidial or nuclear regions analysed in a classical DNA barcoding approach (DNA barcoding *sensu strictu*; Casiraghi *et al.* 2010). However, given the increasing demand for a fast and reliable traceability system for these kinds of products, a panel of additional markers (i.e. promoters, reporter genes) could be applied in combination with classical DNA barcodes (DNA barcoding *sensu lato*; Casiraghi *et al.* 2010).

4.6.3 DNA Barcoding and Food Traceability: An Overview

The identification of raw materials is fundamental to ensure high standards of quality for the food industry and market (Novak *et al.* 2007; Myers 2011). DNA barcoding is effective in certifying both origin and quality of food raw materials and detecting food adulteration or species substitution in the industrial food chain (see Table 4.1).

A clear example of DNA barcoding usefulness is the traceability of seafood (Becker *et al.* 2011) because: (1) classical identification approaches, based on morphological analyses, are not reliable in many cases (with processed food in particular); (2) in comparison to other animal sources (e.g. cattle, sheep, goat, horse), the number of species is higher so there is the need for a rapid identification technique; and (3) in seafood, more than for other living groups, molecular identification can go further than the species level, allowing in several cases the identification of local varieties and hence identifying the origin of a certain product. Moreover, the mitochondrial *coxI* shows good discriminatory power in the identification of fish species (98% of probed marine species and 93% of freshwater species were successfully identified; Ward *et al.* 2009). Successful results were also obtained from a starting small portion of fresh or processed material, thanks to the use of few universal primer combinations (Steinke & Hanner 2011). To date, more than 100,000 barcode sequences from 10,672 species (33% of the total) have been stored in the Fish Barcode of Life Initiative (FISH-BOL; www.fishbol.org), a framework of an international collaborative research. For this reason DNA barcoding was proposed by the US Food and Drug Administration for the authentication of fish-based commercial products (Yancy *et al.* 2008). In particular, the US FDA planned to include DNA barcode data into the Regulatory Fish Encyclopaedia, in order to help investigation of mislabelling and fish species substitution.

DNA barcoding has also proven effective in tracking seafood after industrial processing, when a complex manufacturing process is required or in the case of fish sold in parts (e.g. steaks, blocks, surimi, fish sticks and fins; Barbuto *et al.* 2010).

In the case of plant raw materials, DNA barcoding has been successfully used in several cases (for a review see Galimberti *et al.* 2014), such as for the recognition and traceability of spices. In this case study, De Mattia *et al.* (2011) analysed some of the most important groups of traded spices belonging to the genera *Mentha*, *Ocimum*, *Origanum*, *Salvia*, *Thymus* and *Rosmarinus*. The selected DNA barcode regions were the core-barcode (*matK*+*rbcL*) and the plastidial *trnH-psbA* intergenic spacer. Results suggest that many common spices can be identified; the only exclusions were marjoram and oregano (both belonging to the genus *Oregano*) which exhibited an intraspecific diversity higher than the interspecific diversity. Moreover, DNA barcoding showed high performances in discriminating basil species: indeed, *matK* and *trnH-psbA*7 were able to distinguish commercial basil (*Ocimum basilicum* L) from other *Ocimum* species, as well as differentiate among the principal basil cultivars.

In another case related to plant traceability, Bruni *et al.* (2010) evaluated the effectiveness of DNA barcoding in separating toxic from edible species, demonstrating a clear molecular distinction between cultivated species of the genera *Solanum* (*Solanum tuberosum* L., *Solanum lycopersicum* L. group) and *Prunus* (*Prunus armeniaca* L., *Prunus avium* L., *Prunus cerasus* L., *Prunus domestica* L.) and their toxic congenerics. This study suggests that DNA barcoding is also useful to assess food safety, being able to distinguish edible species from their non-edible or toxic congenerics as confirmed by Jaakola *et al.* (2010).

After the analysis of several case studies published in the field of food traceability, it is possible to state that the main limit of universal barcode markers is the low level of genetic variability of some cultivars and animal breeds. Most crops species and farmed animals derive from complicated breeding programmes; genetic differences among them and the parental taxa could therefore be extremely reduced and are not detectable if analysed with DNA barcode markers.

In an attempt to bypass this shortcoming, Kane and Cronk (2008) proposed ultra-barcoding methodology which is based on the sequencing of the whole plastidial genome, together with large portions of the nuclear genome (Kane *et al.* 2012; Li *et al.* 2015). This combination provides enough information to highlight genetic diversity below the level of species, allowing hybrids to be differentiated from pure lines and showing higher sensitivity than traditional DNA barcoding (Parks *et al.* 2009; Nock *et al.* 2011; Steele & Pires 2011). Kane *et al.* (2012) evaluated the effectiveness of ultra-barcoding on cocoa (*Theobroma cacao* L.), and found several plastidial and nuclear SNPs which were useful in discriminating several cultivars. This technique is promising and technically accessible thanks to the recent advances in the field of DNA sequencing (high-throughput sequencing devices). This will allow conventional barcode markers to be extended from a gene fragment to the entire chloroplast genome.

4.7 Next-Generation Technologies to Characterize Complex Food Matrices and their Microbiome

DNA barcoding has moved fast in the last decade; universal DNA barcode regions have been identified for all of the most relevant group of organisms and PCR protocols and universal primers sets have also been defined (Chakraborty *et al.* 2014; Joly *et al.* 2014). These technical advances allow the identification of organisms starting from a small portion of tissue, without any morphological information available. In this context, DNA barcoding represents a suitable food traceability system, able to assess the quality and safety of a product during the different steps of the food supply chain from the field to the consumer's table (Barbuto *et al.* 2010). However, most food products are composed of a mix of organisms and a complex microbiome. In this case, the use of universal primers produces several DNA barcode fragments, corresponding to the different species in the analysed food (raw material and microorganisms). Sanger-based DNA sequencing, although being effective when used for DNA barcoding, is therefore only a feasible approach in the case of food mixtures if preceded by a pre-treatment to separate or isolate amplicons. Several techniques, such as digestion with specific restriction enzymes (i.e. RFLP) or electrophoretic analysis (Mane *et al.* 2009; Teletchea 2009; Colombo *et al.* 2011), were used to separate different DNA fragments before the sequencing process. However, these methods are effective only when the food matrix comprises only a few species and when they have relevant differences in their DNA barcodes (i.e. different target regions for restriction enzymes and sequences of different length). In other cases, amplicons should be cloned into plasmid vectors and introduced into bacterial competent cells (Zeale *et al.* 2011) in order to obtain single fragments.

With the ultimate goal of characterizing the complete spectrum of ingredients in complex food matrices, as well as alterations or peculiarities in their microbial composition, a cloning approach is inadequate to uncover this huge diversity. As a result, the requirement for high-throughput sequencing techniques grew by an unpredicted extent (Solieri *et al.* 2013). Several novel approaches evolved to replace the traditional Sanger sequencing method; these modern advances have been referred to as 'next-generation sequencing' (NGS) and, more recently, 'high-throughput sequencing' (HTS). HTS techniques are able to provide sequence data around a hundred times faster and cheaper than the conventional Sanger approach. Sequencers from 454 Life Sciences/Roche (producing about a million sequences of length 800–1000 base pairs), Solexa/Illumina and Applied Biosystems SOLiD technology (producing over a billion sequences of length 50–500 base pairs) were produced as second-generation technologies and other competitive instruments appeared on the market such as the Ion Torrent and PacBio. Prior to reaching a taxonomic assignment of the whole biological content of a food ecosystem (i.e. including raw materials and food-borne microorganisms and viruses), sequences generated with HTS have to be filtered, denoised and analysed using bioinformatic tools. Another advantage of the use of HTS technologies concerns the possibility of preparing several DNA samples, from different extracts and marked with different DNA tags, at the same time. Thanks to these practical advantages, it is possible to analyse in parallel a very high number of samples and hence lower the analysis cost (Madesis *et al.* 2014). The reduction in cost and time for generating DNA sequence data has resulted in a range of new successful applications, including food traceability and especially food microbiology (Madesis *et al.* 2014; Galimberti *et al.* 2015; see also Table 4.1).

As an example, pyrosequencing has been used to identify fruit species in yogurts (Ortola-Vidal *et al.* 2007) and pollen composition in processed honeys (Valentini *et al.* 2010). The main limit of this identification approach is the reduced length of the sequenced barcode region, ranging in length from 50 to 500 base pairs (depending on the adopted technology). This issue has been partially resolved using minibarcodes, shorter fragments (about 150 bp) targeting a hypervariable part of the conventional DNA barcode region (Meusnier *et al.* 2008; Hellberg & Morrisey 2011; Little 2014) suitable for HTS devices. The mini-barcode

approach provides enough information to identify species from different matrices (Hajibabaei *et al.* 2006, 2007, 2011). The length issue is not fully resolved, as it does not always allow the discrimination of closely related species.

Concerning food microbiology topics, information on the microbial taxonomy and/or diversity of culturable and unculturable bacteria in the food sample can be obtained from the generated sequences, usually based on 16s rDNA. The primary field of application of HTS in food microbiology is the study of microbiota composition from an ecological perspective (e.g. to elucidate the molecular mechanisms and the interactions of microorganisms in food ecosystems). Such approaches could also provide information on the presence of beneficial or undesirable microorganisms in food matrices and could be used to trace fermented foods, especially concerning dairy products (Ercolini 2013). The HTS-driven advances have been exploited mainly to resequence strains and individuals, aiming at a better sampling of genomic diversity within microbial species (i.e. bacterial strains, yeast and filamentous fungi), opening the possibility of constructing 'personalized genomics' for microbial active elements of food interest (see Solieri *et al.* 2013 for a review).

4.8 Conclusions

DNA barcoding can be used as a universal tool to identify foodstuff components and their related microorganisms. The growing importance of DNA barcoding is based on a combination of factors: (1) the falling cost of molecular analyses; (2) the increasing availability of equipped laboratories and skilled personnel; (3) the presence of freely available web-based resources; and (4) the increasing amount of informed consumers who require high standards of quality in food products.

A huge number of case studies and technical advancements clearly indicate that DNA barcoding is a sensitive, fast, cheap and reliable method for identifying and tracking a wide panel of raw materials, their derived food commodities (even in the case of strongly processed food products) and pathogenic microorganisms potentially occurring in food matrices and generating food spoilage (or, at worst severe outbreaks).

Due to its universality, DNA barcoding can be used in different contexts and by different operators. International agencies or institutions responsible for quality control of raw materials or food commodities can co-operate by exchanging their data, hence creating population reference databases. However, a lot of work is required to provide a reliable source of reference DNA barcoding data for groups of animals, plants, fungi and microorganisms involved in the food supply chain, which have been poorly investigated. For this reason DNA barcoding is likely to become a routine test in many fields in the near future, in particular for assessing quality, safety and traceability of the modern biotransformed foodstuffs.

References

Acheson, D.W. 1999. Foodborne infections. *Current Opinion in Gastroenterology* **15**(6), 538.

Amerine, M.A., Berg, H.W., Kunkee, R.E., Ough, C.S., Singleton, V.L. & Webb, A.D. (eds) 1980. *The Technology of Wine Making*. AVI Publishing, Westport, CT.

Amoa-Awua, W.K., Frisvad, J.C., Sefa-Dedeh, S. & Jakobsen, M. 1997. The contribution of moulds and yeasts to the fermentation of agbelima cassava dough. *Journal of Applied Microbiology* **83**(3), 288–296.

Appendini, P. & Hotchkiss, J.H. 2002. Review of antimicrobial food packaging. *Innovative Food Science & Emerging Technologies* **3**(2), 113–126.

Arcuri, E.F., ElSheikha, A.F., Rychlik, T., Piro-Métayer, I. & Montet, D. 2013. Determination of cheese origin by using 16S rDNA fingerprinting of bacteria communities by PCR-DGGE: preliminary application to traditional Minas cheese. *Food Control* **30**(1), 1–6.

Ardura, A., Linde, A.R., Moreira, J.C. & Garcia-Vazquez, E. 2010. DNA barcoding for conservation and management of Amazonian commercial fish. *Biological Conservation* **143**(6), 1438–1443.

Ardura, A., Planes, S. & Garcia-Vazquez, E. 2013. Applications of DNA barcoding to fish landings: authentication and diversity assessment. *ZooKeys* **365**, 49–65.

Arleo, M., Ruibal, F., Pereyra, J., Miquel, E., Fernández, M. & Martínez, C. 2012. A DNA-based approach to discriminate between quince and apple in quince jams. *International Food Research Journal* **19**(4), 1471–1477.

Barbuto, M., Galimberti, A., Ferri, E., Labra, M., Malandra, R., Galli, P. & Casiraghi, M. 2010. DNA barcoding reveals fraudulent substitutions in shark seafood products: the Italian case of "palombo"(*Mustelus* spp.). *Food Research International* **43**(1), 376–381.

Becker, S., Hanner, R. & Steinke, D. 2011. Five years of FISH-BOL: brief status report. *Mitochondrial DNA* **22**(S1), 3–9.

Betoret, E., Betoret, N., Vidal, D. & Fito, P. 2011. Functional foods development: trends and technologies. *Trends in Food Science & Technology* **22**(9), 498–508.

Blaxter, M. L. 2004. The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **359**(1444), 669–679.

Bokulich, N.A., Joseph, C.L., Allen, G., Benson, A.K. & Mills, D.A. 2012. Next-generation sequencing reveals significant bacterial diversity of botrytized wine. *PloS One* **7**(5), e36357.

Bondi, M., Messi, P., Halami, P.M., Papadopoulou, C. & de Niederhausern, S. 2014. Emerging microbial concerns in food safety and new control measures. *BioMed Research International* 6 July, 1–3.

Borelli, B.M., Ferreira, E.G., Lacerda, I.C.A., Franco, G.R. & Rosa, C.A. 2006. Yeast populations associated with the artisanal cheese produced in the region of Serra da Canastra, Brazil. *World Journal of Microbiology and Biotechnology* **22**(11), 1115–1119.

Bosmali, I., Ganopoulos, I., Madesis, P. & Tsaftaris, A. 2012. Microsatellite and DNA-barcode regions typing combined with high resolution melting (HRM) analysis for food forensic uses: A case study on lentils (*Lens culinaris*). *Food Research International* **46**(1), 141–147.

Bottero, M.T. & Dalmasso, A. 2011. Animal species identification in food products: evolution of biomolecular methods. *Veterinary Journal* **190**(1), 34–38.

Botti, S. & Giuffra, E. 2010. Oligonucleotide indexing of DNA barcodes: Identification of tuna and other scombrid species in food products. *BMC Biotechnology* **10**(60), 1–7.

Bourdichon, F., Casaregola, S., Farrokh, C. *et al*. 2012. Food fermentations: microorganisms with technological beneficial use. *International Journal of Food Microbiology* **154**(3), 87–97.

Breidt, F., Mc Feeters, R.F., Perez-Diaz, R. & Lee, C.H. 2013. Fermented vegetables. In *Food Microbiology: Fundamentals and Frontiers*, 4th edn (eds M.P. Doyle & R.L. Buchanan), pp 841–855. ASM Press, Washington, DC.

Bruni, I., De Mattia, F., Galimberti, A., Galasso, G., Banfi, E., Casiraghi, M. & Labra, M. 2010. Identification of poisonous plants by DNA barcoding approach. *International Journal of Legal Medicine* **124**(6), 595–603.

Bruni, I., De Mattia, F., Martellos, S., Galimberti, A., Savadori, P., Casiraghi, M., Nimis, P.L. & Labra, M. 2012. DNA barcoding as an effective tool in improving a digital plant identification system: a case study for the area of Mt. Valerio, Trieste (NE Italy). *PloS One* **7**(9), 1–8.

Bull, M., Plummer, S., Marchesi, J. & Mahenthiralingam, E. 2013. The life history of *Lactobacillus acidophilus* as a probiotic: a tale of revisionary taxonomy, misidentification and commercial success. *FEMS Microbiology Letters* **349**(2), 77–87.

Bulloch, W. (ed.) 1938. *The History of Bacteriology*. Oxford University Press, Oxford.

Cai, Y.S., Zhang, L.A., Shen, F.J., Zhang, W.P., Hou, R., Yue, B.S., Li, J. & Zhang, Z.H. 2011. DNA barcoding of 18 species of Bovidae. *Chinese Science Bulletin* **56**, 164–168.

Callon, C., Retureau, E., Didienne, R. & Montel, M.C. 2014. Microbial biodiversity in cheese consortia and comparative *Listeria* growth on surfaces of uncooked pressed cheeses. *International Journal of Food Microbiology* **174**, 98–109.

Campanaro, S., Treu, L., Vendramin, V., Bovo, B., Giacomini, A. & Corich, V. 2014. Metagenomic analysis of the microbial community in fermented grape marc reveals that *Lactobacillus fabifermentans* is one of the dominant species: insights into its genome structure. *Applied Microbiology and Biotechnology* **98**, 6015–6037.

Campbell-Platt, G. 1994. Fermented foods – a world perspective. *Food Research International* **27**(3), 253–257.

Caplice, E. & Fitzgerald, G.F. 1999. Food fermentations: role of microorganisms in food production and preservation. *International Journal of Food Microbiology* **50**(1), 131–149.

Carraro, L., Maifreni, M., Bartolomeoli, I., Martino, M.E., Novelli, E., Frigo, F., Marino, M. & Cardazzo, B. 2011. Comparison of culture-dependent and -independent methods for bacterial community monitoring during Montasio cheese manufacturing. *Research in Microbiology* **162**(3), 231–239.

Carvalho, D.C., Neto, D.A., Brasil, B.S. & Oliveira, D.A. 2011. DNA barcoding unveils a high rate of mislabeling in a commercial freshwater catfish from Brazil. *Mitochondrial DNA* **22**(S1), 97–105.

Casiraghi, M., Labra, M., Ferri, E., Galimberti, A. & De Mattia, F. 2010. DNA barcoding: a six-question tour to improve users awareness about the method. *Briefings in Bioinformatics* **11**(4), 440–453.

Ceuppens, S., Li, D., Uyttendaele, M., Renault, P., Ross, P., Van Ranst, M., Cocolin, L. & Donaghy, J. 2014. Molecular methods in food safety microbiology: interpretation and implications of nucleic acid detection. *Comprehensive Reviews in Food Science and Food Safety* **13**(4), 551–577.

Chakraborty, C., Doss, C.G.P., Patra, B.C. & Bandyopadhyay, S. 2014. DNA barcoding to map the microbial communities: current advances and future directions. *Applied Microbiology and Biotechnology* **98**(8), 3425–3436.

Chuang, H., Lur, H., Hwu, K. & Chang, M. 2011. Authentication of domestic Taiwan rice varieties based on fingerprinting analysis of microsatellite DNA markers. *Botanical Studies* **52**(4), 393–405.

Cizeikiene, D., Juodeikiene, G., Paskevicius, A. & Bartkiene, E. 2013. Antimicrobial activity of lactic acid bacteria against pathogenic and spoilage microorganism isolated from food and their control in wheat bread. *Food Control* **31**(2), 539–545.

Cohen, N.J., Deeds, J.R., Wong, E.S., Hanner, R.H., Yancy, H.F., White, K.D., Thompson, T.M., Wahl, M., Pham, T.D., Guichard, F.M., Huh, I., Austin, C., Dizikes, G. & Gerber, S.I. 2009. Public health response to puffer fish (tetrodotoxin) poisoning from mislabeled product. *Journal of Food Protection* **72**(4), 810–817.

Colombo, F., Chess, S., Cattaneo, P. & Bernardi, C. 2011. Polymerase chain reaction products (PCR) on 'DNA barcode zone' resolved by temporal temperature gradient electophoresis: a tool for species identification of mixed meat specimens – a technical note on preliminary results. *Food Control* **22**(8), 1471–1472.

Colwell, R.R. 1970. Polyphasic taxonomy of the genus *Vibrio*: numerical taxonomy of *Vibrio cholerae*, *Vibrio parahaemolyticus*, and related *Vibrio* species. *Journal of Bacteriology* **104**(1), 410–433.

Cordain, L., Eaton, S.B., Sebastian, A., Mann, N., Lindeberg, S., Watkins, B.A., OKeefe, J.H. & Brand-Miller, J. 2005. Origins and evolution of the western diet: health implications for the 21st century. *The American Journal of Clinical Nutrition* **81**(2), 341–354.

Coton, M., Delbés-Paus, C., Irlinger, F., Desmasures, N., Le Fleche, A., Stahl, V., Montel, M.C. & Coton, E. 2012. Diversity and assessment of potential risk factors of Gram–negative isolates associated with French cheeses. *Food Microbiology* **29**(1), 88–98.

Crowley, S., Mahony, J. & van Sinderen, D. 2013. Current perspectives on antifungal lactic acid bacteria as natural biopreservatives. *Trends in Food Science & Technology* **33**(2), 93–109.

Cunningham, S.A., Sloan, L.M., Nyre, L.M., Vetter, E.A., Mandrekar, J. & Patel, R. 2010. Three-hour molecular detection of *Campylobacter, Salmonella, Yersinia,* and *Shigella* species in feces with accuracy as high as that of culture. *Journal of Clinical Microbiology* **48**(8), 2929–2933.

Czerucka, D., Piche, T. & Rampal, P. 2007. Review article: yeast as probiotics – *Saccharomyces boulardii*. *Alimentary Pharmacology & Therapeutics* **26**(6), 767–778.

David, V., Terrat, S., Herzine, K., Claisse, O., Rousseaux, S., Tourdot-Maréchal, R., Masneuf-Pomarede, I., Ranjard, L. & Alexandre, H. 2014. High-throughput sequencing of amplicons for monitoring yeast biodiversity in must and during alcoholic fermentation. *Journal of Industrial Microbiology & Biotechnology* **41**(5), 811–821.

De Gregorio, P.R., Tomás, M.S.J., Terraf, M.C.L. & Nader-Macías, M.E.F. 2014. In vitro and in vivo effects of beneficial vaginal lactobacilli on pathogens responsible for urogenital tract infections. *Journal of Medical Microbiology* **63**(5), 685–696.

De Mattia, F., Lovicu, G., Tardaguila, J., Grassi, F., Imazio, S., Scienza, A. & Labra, M. 2009. Genetic relationships between Sardinian and Spanish viticulture: the case of Cannonau and Garnacha. *Journal of Horticultural Science and Biotechnology* **84**(1), 65–71.

De Mattia, F., Bruni, I., Galimberti, A., Cattaneo, F., Casiraghi, M. & Labra, M. 2011. A comparative study of different DNA barcoding markers for the identification of some members of Lamiaceae. *Food Research International* **44**(3), 693–702.

De Mattia, F., Gentili, R., Bruni, I., Galimberti, A., Sgorbati, S., Casiraghi, M. & Labra, M. 2012. A Multi-marker DNA barcoding approach to save time and resources in vegetation surveys. *Biological Journal of the Linnean Society* **169**(3), 518–529.

De Santis, T.Z., Hugenholtz, P., Larsen, N., Rojas, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. & Andersen, G.L. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* **72**(7), 5069–5072.

Delbès, C., Ali-Mandjee, L. & Montel, M.C. 2007. Monitoring bacterial communities in raw milk and cheese by culture-dependent and -independent 16S rRNA gene-based analyses. *Applied and Environmental Microbiology* **73**(6), 1882–1891.

Denis, M., Refrégier-Petton, J., Laisney, M.J., Ermel, G. & Salvat, G. 2001. Campylobacter contamination in French chicken production from farm to consumers. Use of a PCR assay for detection and identification of Campylobacter jejuni and Camp. coli. *Journal of Applied Microbiology* **91**(2), 255–267.

Dhiman, B. & Singh, M. 2003. Molecular detection of cashew husk (*Anacardium occidentale*) adulteration in market samples of dry tea (*Camellia sinensis*). *Planta Medica* **69**(09), 882–884.

Dobson, A., OSullivan, O., Cotter, P.D., Ross, P. & Hill, C. 2011. High-throughput sequence-based analysis of the bacterial composition of kefir and an associated kefir grain. *FEMS Microbiology Letters* **320**(1), 56–62.

Donnenberg, M.S. & Whittam, T.S. 2001. Pathogenesis and evolution of virulence in enteropathogenic and enterohemorrhagic Escherichia coli. *Journal of Clinical Investigation* **107**(5), 539–548.

Doyle, M.P. & Buchanan, R.L. (eds) 2013. *Food Microbiology: Fundamentals and Frontiers*, 4th edn. ASM Press, Washington, DC.

Dueñas-Sánchez, R., Pérez, A.G., Codón, A.C., Benítez, T. & Rincón, A.M. 2014. Overproduction of 2-phenylethanol by industrial yeasts to improve organoleptic properties of bakers products. *International Journal of Food Microbiology* **180**, 7–12.

El-Sharoud, W.M., Belloch, C., Peris, D. & Querol, A. 2009. Molecular identification of yeasts associated with traditional Egyptian dairy products. *Journal of Food Science*, **74**(7), 341–346.

Elmerdahl Olsen, J. 2000. DNA-based methods for detection of food-borne bacterial pathogens. *Food Research International* **33**(3), 257–266.

Ercolini, D. 2013. High-throughput sequencing and metagenomics: moving forward in the culture-independent analysis of food microbial ecology. *Applied Environmental Microbiology* **79**(10), 3148–3155.

Ercolini, D., Russo, F., Ferrocino, I. & Villani, F. 2009. Molecular identification of mesophilic and psychrotrophic bacteria from raw cows milk. *Food Microbiology* **26**(2), 228–231.

Ercolini, D., De Filippis, F., La Storia, A. & Iacono, M. 2012. "Remake" by high-throughput sequencing of the microbiota involved in the production of water buffalo mozzarella cheese. *Applied and Environmental Microbiology* **78**(22), 8142–8145.

European Commission. 2010. Functional Foods: Studies and Reports. Available at: ftp://ftp.cordis.europa.eu/pub/fp7/kbbe/docs/functional-foods_en.pdf (accessed 11 June 2015).

Fajardo, V., Gonzàlez, I., Rojas, M., Garcìa, T. & Martîn R. 2010. A review of current PCR-based methodologies for the authentication of meats from game animal species. *Trends in Food Science Technology* **21**(8), 408–421.

Faria, M.A., Magalhães, A., Nunes, M. E. & Oliveira, M.B.P.P. 2013. High resolution melting of *trnL* amplicons in fruit juices authentication. *Food Control* **33**(1), 136–141.

Federici, S., Galimberti, A., Bartolucci, F., Bruni, I., Cortis, P. & Labra, M. 2013. DNA barcoding to analyse taxonomically complex groups in plants: the case of *Thymus* (Lamiaceae). *Botanical Journal of the Linnean Society* **171**(4), 687–699.

Federici, S., Fontana, D., Galimberti, A., Bruni, I., De Mattia, F., Cortis, P., Galasso, G. & Labra, M. 2014. A rapid diagnostic approach to identify poisonous plants using DNA barcoding data. *Plant Biosystems*, published online 30 July 2014, doi: 10.1080/11263504.2014.941031.

Felis, G.E. & Dellaglio, F. 2007. Taxonomy of lactobacilli and bifidobacteria. *Current Issues in Intestinal Microbiology* **8**(2), 44–61.

Ferri, E., Barbuto, M., Bain, O., Galimberti, A., Uni, S., Guerrero, R., Ferté, H., Bandi, C., Martin, C. & Casiraghi, M. 2009. Integrated taxonomy: traditional approach and DNA barcoding for the identification of filarioid worms and related parasites (Nematoda). *Frontiers in Zoology* **6**(1), 1–12.

Filonzi, L., Chiesa, S., Vaghi, M. & Nonnis Marzano, F. 2010. Molecular barcoding reveals mislabelling of commercial fish products in Italy. *Food Research International* **43**(5), 1383–1388.

Fox, P.F. 1993. Cheese: an overview. In *Cheese: Chemistry, Physics and Microbiology* (ed. P.M. Fox), pp. 1–36. Chapman & Hall, London, England.

Fricker, M., Reissbrodt, R. & Ehling-Schulz, M. 2008. Evaluation of standard and new chromogenic selective plating media for isolation and identification of Bacillus cereus. *International Journal of food Microbiology* **121**(1), 27–34.

Gala, E., Landi, S., Solieri, L., Nocetti, M., Pulvirenti, A. & Giudici, P. 2008. Diversity of lactic acid bacteria population in ripened Parmigiano Reggiano cheese. *International Journal of Food Microbiology* **125**(3), 347–351.

Galal-Khallaf, A., Ardura, A., Mohammed-Geba, K., Borrell, Y.J. & Garcia-Vazquez, E. 2014. DNA barcoding reveals a high level of mislabeling in Egyptian fish fillets. *Food Control* **46**, 441–445.

Galimberti, A., Spada, M., Russo, D., Mucedda, M., Agnelli, P., Crottini, A., Ferri, E., Martinoli, A. & Casiraghi, M. 2012. Integrated operational taxonomic units (IOTUs) in echolocating bats: a bridge between molecular and traditional taxonomy. *PloS One* **7**(6), 1–11.

Galimberti, A., De Mattia, F., Losa, A., Bruni, I., Federici, S., Casiraghi, M., Martellos, S. & Labra, M. 2013. DNA barcoding as a new tool for food traceability. *Food Research International* **50**(1), 55–63.

Galimberti, A., Labra, M., Sandionigi, A., Bruno, A., Mezzasalma, V. & De Mattia, F. 2014. DNA barcoding for minor crops and food traceability. *Advances in Agriculture* **2014**, 1–8.

Galimberti, A., Bruno, A., Mezzasalma, V., De Mattia, F., Bruni, I. & Labra, M. 2015. Emerging DNA-based technologies to characterize food ecosystems. *Food Research International* **69**, 424–433.

Gálvez, A., Abriouel, H., López, R.L. & Omar, N.B. 2007. Bacteriocin-based strategies for food biopreservation. *International Journal of Food Microbiology* **120**(1), 51–70.

Ganopoulos, I., Madesis, P., Darzentas, N., Argiriou, A. & Tsaftaris, A. 2012. Barcode high resolution melting (Bar–HRM) analysis for detection and quantification of PDO "Fava Santorinis" (*Lathyrus clymenum*) adulterants. *Food Chemistry* **133**(2), 505–512.

Ganopoulos, I., Bazakos, C., Madesis, P., Kalaitzis, P. & Tsaftaris, A. 2013. Barcode DNA high-resolution melting (Bar-HRM) analysis as a novel close-tubed and accurate tool for olive oil forensic use. *Journal of the Science of Food and Agriculture* **93**(9), 2281–2286.

Giannino, M.L., Marzotto, M., Dellaglio, F. & Feligini, M. 2009. Study of microbial diversity in raw milk and fresh curd used for Fontina cheese production by culture-independent methods. *International Journal of Food Microbiology* **130**(3), 188–195.

Gibson, G.R., Rastall, R.A. & Fuller, R. 2003. The health benefits of probiotics and prebiotics. In *Gut Flora, Nutrition, Immunity and Health*, 1st edn (eds R. Fuller & G. Perdigon), pp. 52–76. Blackwell Publishing, Oxford, UK.

Gibson, G.R., Probert, H.M., Van Loo, J., Rastall, R.A. & Roberfroid, M.B. 2004. Dietary modulation of the human colonic microbiota: updating the concept of prebiotics. *Nutrition Research Reviews* **17**(2), 259–275.

Giraffa, G. & Neviani, E. 2001. DNA-based, culture-independent strategies for evaluating microbial communities in food-associated ecosystems. *International Journal of Food Microbiology* **67**(1), 19–34.

Gismondi, A., Fanali, F., Labarga, J.M.M., Caiola, M.G. & Canini, A. 2013. *Crocus sativus* L. genomics and different DNA barcode applications. *Plant Systematics and Evolution* **299**(10), 1859–1863.

Gori, K., Ryssel, M., Arneborg, N. & Jespersen, L. 2013. Isolation and identification of the microbiota of Danish farmhouse and industrially produced surface-ripened cheeses. *Microbial Ecology* **65**(3), 602–615.

Gram, L. & Dalgaard, P. 2002. Fish spoilage bacteria – problems and solutions. *Current Opinion in Biotechnology* **13**(3), 262–266.

Grassi, F., Labra, M., Scienza, A. & Imazio, S. 2002. Chloroplast SSR markers to assess DNA diversity in wild and cultivated grapevines. *Vitis* **41**(3), 157–158.

Grassi, F., Labra, M. & Minuto, L. 2006. Molecular diversity in Ligurian local races of common bean (Phaseolus vulgaris L.). *Plant Biosystems* **140**(1), 17–20.

Gueimonde, M., Delgado, S., Mayo, B., Ruas-Madiedo, P., Margolles, A. & de los Reyes-Gavilán, C.G. 2004. Viability and diversity of probiotic *Lactobacillus* and *Bifidobacterium* populations included in commercial fermented milks. *Food Research International* **37**(9), 839–850.

Guslandi, M., Mezzi, G., Sorghi, M. & Testoni, P.A. 2000. *Saccharomyces boulardii* in maintenance treatment of Crohns disease. *Digestive Diseases and Sciences* **45**(7), 1462–1464.

Hajibabaei, M., Smith, M.A., Janzen, D.H., Rodriguez, J.J., Whitfield, J.B. & Hebert, P.D.N. 2006. A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes* **6**(4), 959–964.

Hajibabaei, M., Singer, G.A., Clare, E.L. & Hebert, P.D.N. 2007. Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC Biotechnology* **5**(24), 1–7.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A.C. & Baird, D.J. 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One* **6**(4), e17497, 1–7.

Haller, D., Antoine, J.M., Bengmark, S., Enck, P., Rijkers, G.T. & Lenoir-Wijnkoop, I. 2010. Guidance for substantiating the evidence for beneficial effects of probiotics: probiotics in chronic inflammatory bowel disease and the functional disorder

irritable bowel syndrome. *The Journal of Nutrition* **140**(3), 690S–697S.

Hammami, R., Zouhir, A., Le Lay, C., Hamida, J.B. & Fliss, I. 2010. BACTIBASE second release: a database and tool platform for bacteriocin characterization. *BMC Microbiology* **10**(22), 1–5.

Han, J., Wu, Y., Huang, W., Wang, B., Sun, C., Ge, Y. & Chen, Y. 2012. PCR and DHPLC methods used to detect juice ingredient from 7 fruits. *Food Control* **25**(2), 696–703.

Hao, Q., Lu, Z., Dong, B.R., Huang, C.Q. & Wu, T. 2011. Probiotics for preventing acute upper respiratory tract infections. *Cochrane Database of Systematic Reviews* **9**(9), 1–51.

Haye, P.A., Segovia, N.I., Vera, R., Gallardo, M.D.L.Á. & Gallardo-Escárate, C. 2012. Authentication of commercialized crab-meat in Chile using DNA barcoding. *Food Control* **25**(1), 239–244.

Hebert, P.D.N., Ratnasingham, S. & deWaard, J.R. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Science* **270**, S96–S99.

Heilig, H.G., Zoetendal, E.G., Vaughan, E.E., Marteau, P., Akkermans, A.D. & de Vos, W.M. 2002. Molecular diversity of *Lactobacillus* spp. and other lactic acid bacteria in the human intestine as determined by specific amplification of 16S ribosomal DNA. *Applied and Environmental Microbiology* **68**(1), 114–123.

Hellberg, R.S. & Morrisey, M.T. 2011. Advances in DNA-based techniques for the detection of seafood species substitution on the commercial market. *Journal of Laboratory Automation* **16**(4), 308–321.

Herbel, S.R., Lauzat, B., Nickisch-Rosenegk, M., Kuhn, M., Murugaiyan, J., Wieler, L.H. & Guenther, S. 2013. Species-specific quantification of probiotic lactobacilli in yoghurt by quantitative real-time PCR. *Journal of Applied Microbiology* **115**(6), 1402–1410.

Hidayat, T., Pancoro, A. & Kusumawaty, D. 2012. Utility of *matK* gene as DNA barcode to assess evolutionary relationship of important tropical forest tree genus *Mangifera* (Anacardiaceae) in Indonesia and Thailand. *Biotropia* **18**(2), 74–80.

Hirao, T., Imai, S., Sawada, H., Shiomi, N., Hachimura, S. & Kato, H. 2005. PCR method for detecting trace amounts of buckwheat (*Fagopyrum* spp.) in food. *Bioscience, Biotechnology, and Biochemistry* **69**(4), 724–731.

Hollingsworth, M.L., Clark, A.A., Forrest, L.L., Richardson, J., Pennington, R.T., Long, D.G., Cowan, R., Chase, M.W., Gaudeul, M. & Hollingsworth, P.M. 2009. Selecting barcoding loci for plants: evaluation of seven candidate loci with species level sampling in three divergent groups of land plants. *Molecular Ecology Resources* **9**(2), 439–457.

Hollingsworth, P.M., Graham, S.W. & Little, D.P. 2011. Choosing and using a plant DNA barcode. *PloS One* **6**(5), 1–13.

Hornsey, I.S. (ed.) 2003. *A History of Beer and Brewing*. RSC Paperbacks (Book 34), London.

Hosseini, H., Hippe, B., Denner, E., Kollegger, E. & Haslberger, A. 2012. Isolation, identification and monitoring of contaminant bacteria in Iranian Kefir type drink by 16S rDNA sequencing. *Food Control* **25**(2), 784–788.

Jaakola, L., Suokas, M. & Häggman, H. 2010. Novel approaches based on DNA barcoding and high-resolution melting of amplicons for authenticity analyses of berry species. *Food Chemistry* **123**(2), 494–500.

Jackson, B.R., Griffin, P.M., Cole, D., Walsh, K.A. & Chai, S.J. 2013. Outbreak-associated *Salmonella enterica* serotypes and food commodities, United States, 1998–2008. *Emerging Infectious Diseases* **19**(8), 1239–1244.

Janda, J.M. & Abbott, S.L. 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of Clinical Microbiology* **45**(9), 2761–2764.

Jarret, R.L. 2008. DNA Barcoding in a crop genebank: the *Capsicum annuum* species complex. *Open Biology Journal* **1**, 35–42.

Johnson, M.E. & Steele, J.L. 2013. Fermented dairy products. In *Food Microbiology: Fundamentals and Frontiers*, 4th edn (eds M. P. Doyle & R. L. Buchanan), pp. 825–839. ASM Press, Washington, DC.

Joly, S., Davies, T.J., Archambault, A., Bruneau, A., Derry, A., Kembel, S.W., Peres-Neto, P., Vamosi, J. & Wheeler, T.A. 2014. Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. *Molecular Ecology Resources* **14**(2), 221–232.

Jones, P.J. & Jew, S. 2007. Functional food development: concept to reality. *Trends in Food Science & Technology* **18**(7), 387–390.

Kagkli, D.M., Vancanneyt, M., Vandamme, P., Hill, C. & Cogan, T.M. 2007. Contamination of milk by enterococci and coliforms from bovine faeces. *Journal of Applied Microbiology* **103**(5), 1393–1405.

Kane, N.C. & Cronk, Q. 2008. Botany without borders, barcoding in focus. *Molecular Ecology* **17**(24), 5175–5176.

Kane, N., Sveinsson, S., Dempewolf, H., Yong Yang, J., Zhang, D., Engels, J.M.M. & Cronk, Q. 2012. Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *American Journal of Botany* **99**(2), 320–329.

Klaenhammer, T.R. 1993. Genetics of bacteriocins produced by lactic acid bacteria. *FEMS Microbiology Reviews* **12**(1–3), 39–86.

Kodama, Y., Kielland-Brandt, M.C., & Hansen, J. 2006. Lager brewing yeast. In *Comparative Genomics* (ed. H. Nicholas), pp. 145–164. Springer, Berlin Heidelberg.

Kojoma, M., Kurihara, K., Yamada, K., Sekita, S., Satake, M. & Iida, O. 2002. Genetic identification of cinnamon (*Cinnamomum* spp.) based on the trnL-trnF chloroplast DNA. *Planta Medica* **68**(1), 94–96.

Konczak, I. & Roulle, P. 2011. Nutritional properties of commercially grown native Australian fruits: Lipophilic antioxidants and minerals. *Food Research International* **44**(7), 2339–2344.

Kumar, P., Gupta, V.K., Misra, A.K., Modi, D.R. & Pandey, B.K. 2009. Potential of Molecular Markers in Plant Biotechnology. *Plant Omics* **2**(4), 141–162.

Labra, M., Imazio, S., Grassi, F., Rossoni, M., Citterio, S., Scienza, A. & Failla, O. 2003. Molecular approach to assess the origin of cv. Marzemino. *Vitis* **42**(3), 137–140.

Lee, C.H. 2009. Food biotechnology. In *Food Science and Technology* (ed. G. Campbell-Platt), pp. 85–114. Wiley–Blackwell, West Sussex, United Kingdom.

Li, D.Z., Gao, L.M., Li, H.T., Wang, H., Ge, X.J., Liu, J.Q., Chen, Z.D., Chen, S.L., Tang, J.B., Fu, C.X., Zeng, C.X., Yan, H.F., Zhu, Y.J., Sun, Y.S., Chen, S.Y., Zhao, L., Wang, K., Yang, T. & Duan, G.W. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences of the United States of America* **108**(49), 19641–19646.

Li, M., Au, K.Y., Lam, H., Cheng, L., Jiang, R.W., But, P.P. & Shaw, P.C. 2012a. Identification of Baiying (*Herba solani lyrati*) commodity and its toxic substitute Xungufeng (*Herba aristolochiae mollissimae*) using DNA barcoding and chemical profiling techniques. *Food Chemistry* **135**(3), 1653–1658.

Li, M., Wong, K.L., Chan, W.H., Li, J.X., But, P.P.H., Cao, H. & Shaw, P.C. 2012b. Establishment of DNA barcodes for the identification of the botanical sources of the Chinese cooling beverage. *Food Control* **25**(2), 758–766.

Li, X., Yang, Y., Henry, R.J., Rossetto, M., Wang, Y. & Chen, S. 2015. Plant DNA barcoding: from gene to genome. *Biological Reviews* **90**(1), 157–166.

Lilly, D.M. & Stillwell, R.H. 1965. Probiotics: growth-promoting factors produced by microorganisms. *Science* **147**(3659), 747–748.

Lin, C.C., Fung, L.L., Chan, P.K., Lee, C.M., Chow, K.F. & Cheng, S.H. 2014. A rapid low-cost high-density DNA-based multi-detection test for routine inspection of meat species. *Meat Science* **96**(2), 922–929.

Little, D.P. 2014. A DNA mini-barcode for land plants. *Molecular Ecology Resources* **14**(3), 437–446.

Madesis, P., Ganopoulos, I., Anagnostis, A. & Tsaftaris, A. 2012. The application of Bar-HRM (Barcode DNA-high resolution melting) analysis for authenticity testing and quantitative detection of bean crops (Leguminosae) without prior DNA purification. *Food Control* **25**(2), 576–582.

Madesis, P., Ganopoulos, I., Bosmali, I. & Tsaftaris, A. 2013. Barcode high resolution melting analysis for forensic uses in nuts: A case study on allergenic hazelnuts (*Corylus avellana*). *Food Research International* **50**(1), 351–360.

Madesis, P., Ganopoulos, I., Sakaridis, I., Argiriou, A. & Tsaftaris, A. 2014. Advances of DNA-based methods for tracing the botanical origin of food products. *Food Research International* **60**, 163–172.

Mafra, I., Ferreira, I.M. & Oliveira, M.B.P. 2008. Food authentication by PCR-based methods. *European Food Research and Technology* **227**(3), 649–665.

Maldonado, J., Cañabate, F., Sempere, L., Vela, F., Sànchez, A.R, Narbona, E., Lòpez-Huertas, E., Geerlings, A., Valero, A.D., Olivares, M. & Lara-Villoslada, F. 2012. Human milk probiotic *Lactobacillus fermentum* CECT5716 reduces the incidence of gastrointestinal and upper respiratory tract infections in infants. *Journal of Pediatric Gastroenterology and Nutrition* **54**(1), 55–61.

Mane, B.G., Mendiratta, S.K. & Tiwari, A.K. 2009. Polymerase chain reaction assay for identification of chicken in meat and meat products. *Food Chemistry* **116**(3), 806–810.

Maragkoudakis, P.A., Nardi, T., Bovo, B., DAndrea, M., Howell, K.S., Giacomini, A. & Corich, V. 2013. Biodiversity, dynamics and ecology of bacterial community during grape marc storage for the production of grappa. *International Journal of Food Microbiology* **162**(2), 143–151.

Masoud, W., Takamiya, M., Vogensen, F.K., Lillevang, S., Al-Soud, W.A., Sorensen, S.J. & Jakobsen, M. 2011. Characterization of bacterial populations in Danish raw milk cheeses made with different starter cultures by denaturating gradient gel electrophoresis and pyrosequencing. *International Dairy Journal* **21**(3), 142–148.

McGovern, P.E., Glusker, D.L., Exner, L.J. & Voigt, M.M. 1996. Neolithic resinated wine. *Nature* **381**(6582), 480–481.

McGovern, P.E., Zhang, J., Tang, J., Zhang, Z., Hall, G.R. & Moreau, R.A. 2004. Fermented beverages of pre- and proto-historic China. *Proceedings of the National Academy of Sciences of the United States of America* **101**(51), 17593–17598.

Meusnier, I., Singer, G.A., Landry, J.F., Hickey, D.A., Hebert, P.D. & Hajibabaei, M. 2008. A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* **9**(1), 1–4.

Metchnikoff, L.E. (ed.) 2004. *The Prolongation of Life: Optimistic Studies*. Springer Publishing Company, New York, USA.

Meyer, C. P. & Paulay, G. 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS biology* **3**(12), 2229–2238.

Ming, X., Weber, G.H., Ayres, J.W. & Sandine, W.E. 1997. Bacteriocins applied to food packaging materials to inhibit *Listeria monocytogenes* on meats. *Journal of Food Science* **62**(2), 413–415.

Montville T.J. & Chikindas, M.L. 2013. Biological control of foodborne bacteria. In *Food Microbiology: Fundamentals and Frontiers*, 4th edn (eds M. P. Doyle & R. L. Buchanan), pp. 803–822. ASM Press, Washington, DC.

Montville T.J. & Matthews K.R. 2013. Physiology, growth, and inhibition of microbes in foods. In *Food Microbiology: Fundamentals and Frontiers*, 4th edn (eds M.P. Doyle & R.L. Buchanan), pp. 3–18. ASM Press, Washington, DC.

Mower, J.P., Touzet, P., Gummow, J., Delph, L.F. & Palmer, J.D. 2007. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evolutionary Biology* **7**(1), 1–14.

Muñoz-Quezada, S., Chenoll, E., Vieites, J.M., Genovés, S., Maldonado, J., Bermùdez-Brito, M., Gomez-Llorentem C., Matencio, E., Bernal, M.J., Romero, F., Suàrez, A., Ràmon, D. & Gil, A. 2013. Isolation, identification and characterisation of three novel probiotic strains (*Lactobacillus paracasei* CNCM I-4034, *Bifidobacterium breve* CNCM I-4035 and *Lactobacillus rhamnosus* CNCM I-4036) from the faeces of exclusively breast-fed infants. *British Journal of Nutrition* **109**(S2), S51–S62.

Myers, M.J. 2011. Molecular identification of animal species in food: transition from research laboratories to the regulatory laboratories. *The Veterinary Journal* **190**(1), 7–8.

Nattress, F.M. & Jeremiah, L.E. 2000. Bacterial mediated off-flavours in retail-ready beef after storage in controlled atmospheres. *Food Research International* **33**(9), 743–748.

Newell, D.G., Koopmans, M., Verhoef, L., Duizer, E., Aidara-Kane, A., Sprong, H., Opsteegh, M., Langelaar, M., Threfall, J., Scheultz, F., van der Giessen, J. & Kruse, H. 2010. Food-borne diseases – the challenges of 20 years ago still persist while new ones continue to emerge. *International Journal of Food Microbiology* **139**(Suppl.), S3–S15.

Newmaster, S.G., Grguric, M., Shanmughanandhan, D., Ramalingam, S. & Ragupathy, S. 2013. DNA barcoding detects contamination and substitution in North American herbal products. *BMC Medicine* **11**(222), 1–13.

Ng, C.C., Lin, C.Y., Tzeng, W.S., Chang, C.C. & Shyu, Y.T. 2005. Establishment of an internal transcribed spacer (ITS) sequence-based differentiation identification procedure for mei (Prunus mume) and plum (Prunus salicina) and its use to detect adulteration in preserved fruits. *Food Research International* **38**(1), 95–101.

Nijman, I.J., Otsen, M., Verkaar, E.L., de Ruitjer, C., Hanekamp, E., Ochieng, J.W., Shamshad, S., Reqe, J.E., Hanotte, O., Barwengen, M.W., Sulawati, T. & Lenstra, J.A. 2003. Hybridization of banteng (*Bos javanicus*) and zebu (*Bos indicus*) revealed by mitochondrial DNA, satellite DNA, AFLP and microsatellites. *Heredity* **90**(1), 10–16.

Nock, C.J., Waters, D.L., Edwards, M.A., Bowen, S.G., Rice, N., Cordeiro, G.M. & Henry, R.J. 2011. Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnology Journal* **9**(3), 328–333.

Nöthlings, U., Murphy, S.P., Wilkens, L.R., Henderson, B.E. & Kolonel, L.N. 2007. Flavonols and pancreatic cancer risk the multiethnic cohort study. *American Journal of Epidemiology* **166**(8), 924–931.

Novak, I., Gruber-Gréger, G. & Lukas, B. 2007. DNA-based authentication of plant extracts. *Food Research International* **40**(3), 388–392.

O'Grady, M.N. & Kerry, J.P. 2008. Smart packaging technologies and their application in conventional meat packaging systems. In *Meat Biotechnology* (ed. F. Toldra), pp. 425–451. Springer, New York, USA.

Ortola-Vidal, A., Schnerr, H., Rojmyr, M., Lysholm, F. & Knight, A. 2007. Quantitative identification of plant genera in food products using PCR and Pyrosequencing® technology. *Food Control* **18**(8), 921–927.

Pafundo, S., Agrimonti, C., Maestri, E. & Marmiroli, N. 2007. Applicability of SCAR markers to food genomics: olive oil traceability. *Journal of Agricultural and Food Chemistry* **55**(15), 6052–6059.

Panelli, S., Buffoni, J. N., Bonacina, C. & Feligini, M. 2012. Identification of moulds from the Taleggio cheese environment by the use of DNA barcodes. *Food Control* **28**(2), 385–391.

Parks, M., Cronn, R. & Liston, A. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology* **7**(84), 1–17.

Parvathy, V.A., Swetha, V.P., Sheeja, T.E., Leela, N.K., Chempakam, B. & Sasikumar, B. 2014. DNA barcoding to detect chilli adulteration in traded black pepper powder. *Food Biotechnology* **28**(1), 25–40.

Pasqualone, A., Lotti, C. & Blanco, A. 1999. Identification of durum wheat cultivars and monovarietal semolinas by analysis of DNA microsatellites. *European Food Research and Technology* **210**(2), 144–147.

Patel, J.B. 2001. 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Molecular Diagnosis* **6**(4), 313–321.

Pereira, C., Barros, L., Carvalho, A.M. & Ferreira, I.C.F.R. 2011. Nutritional composition and bioactive properties of commonly consumed wild greens: potential sources for new trends in modern diets. *Food Research International* **44**(9), 2634–2640.

Pfeiler, E.A. & Klaenhammer, T.R. 2013. Probiotics and prebiotics. In *Food Microbiology: Fundamentals and Frontiers*, 4th edn (eds M.P. Doyle & R.L. Buchanan), pp. 949–971. ASM Press, Washington, DC.

Pham, T.T. & Shah, N.P. 2009. Performance of starter in yogurt supplemented with soy protein isolate and biotransformation of isoflavones during storage period. *Journal of Food Science* **74**(4), M190–M195.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwing, W., Peplies, J. & Glöckner, F.O. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* **35**(21), 7188–7196.

Quigley, L., O'Sullivan, O., Beresford, T.P., Ross, R.P., Fitzgerald, G.F. & Cotter, P.D. 2012. High-throughput sequencing for detection of subpopulations of bacteria not previously associated with artisanal cheeses. *Applied and Environmental Microbiology* **78**(16), 5717–5723.

Raeisi, S.N., Ouoba, L.I.I., Farahmand, N., Sutherland, J. & Ghoddusi, H.B. 2013. Variation, viability and validity of bifidobacteria in fermented milk products. *Food Control* **34**(2), 691–697.

Ranadheera, R.D.C.S., Baines, S.K. & Adams, M.C. 2010. Importance of food in probiotic efficacy. *Food Research International* **43**(1), 1–7.

Ratnasingham, S. & Hebert, P.D.N. 2007. BOLD: the Barcode of Life Datasystem (www.barcodinglife.org). *Molecular Ecology Notes* **7**(3), 355–364.

Ratnasingham, S. & Hebert, P.D.N. 2013. A DNA-based registry for all animal species: the barcode index number (BIN) system. *PLoS One* **8**(8), 1–13.

Ray, B. & Bhunia, A. (eds) 2013. *Fundamental Food Microbiology*, 4th edn. CRC press, USA.

Ray, B. & Daeschel, M. (eds) 1992. *Food Biopreservatives of Microbial Origin*. CRC Press, USA.

Roberfroid, M.B. 2000. An European consensus of scientific concepts of functional foods. *Nutrition* **16**(7), 689–691.

Ross, R.P., Morgan, S. & Hill, C. 2002. Preservation and fermentation: past, present and future. *International Journal of Food Microbiology* **79**(1–2), 3–16.

Samuel, D. 1996. Investigation of ancient Egyptian baking and brewing methods by correlative microscopy. *Nature* **273**, 488–490.

Sandionigi, A., Galimberti, A., Labra, M., Ferri, E., Panunzi, E. & Casiraghi, M. 2012. Analytical approaches for DNA barcoding

data – how to find a way for plants? *Plant Biosystems* **146**(4), 805–813.

Santos, F., Wegkamp, A., de Vos, W.M., Smid, E.J. & Hugenholtz, J. 2008. High-level folate production in fermented foods by the B12 producer *Lactobacillus reuteri* JCM1112. *Applied and Environmental Microbiology* **74**(10), 3291–3294.

Scheu, P.M., Berghof, K. & Stahl, U. 1998. Detection of pathogenic and spoilage micro-organisms in food with the polymerase chain reaction. *Food Microbiology* **15**(1), 13–31.

Schmidt, T.M. & Relman, D.A. 1994. Phylogenetic identification of uncultured pathogens using ribosomal RNA sequences. *Methods in Enzymology* **235**, 205–222.

Settanni, L. & Corsetti, A. 2007. The use of multiplex PCR to detect and differentiate food- and beverage-associated microorganisms: a review. *Journal of Microbiological Methods* **69**(1), 1–22.

Shah, N.P. 2007. Functional cultures and health benefits. *International Dairy Journal* **17**(11), 1262–1277.

Sharma, M. & Devi, M. 2014. Probiotics: a comprehensive approach toward health foods. *Critical Reviews in Food Science and Nutrition* **54**(4), 537–552.

Sicard, D. & Legras, J.L. 2011. Bread, beer and wine: yeast domestication in the Saccharomyces sensu stricto complex. *Comptes Rendus Biologies* **334**(3), 229–236.

Singh, V.P., Pathak, V. & Verma, A.K. 2012. Fermented meat products: organoleptic qualities and biogenic amines – a review. *American Journal of Food Technology* **7**(5), 278–288.

Siro, I., Kapolna, E., Kapolna, B. & Lugasi, A. 2008. Functional food. Product development, marketing and consumer acceptance – a review. *Appetite* **51**(3), 456–467.

Smaoui, S., Elleuch, L., Bejar, W., Karray-Rebai, I., Ayadi, I., Jaouadi, B., Mathieu, F., Chouayekh, H., Bejar, S. & Mellouli, L. 2010. Inhibition of fungi and gram-negative bacteria by bacteriocin BacTN635 produced by *Lactobacillus plantarum* sp. TN635. *Applied Biochemistry and Biotechnology* **162**(4), 1132–1146.

Smith, P.J., McVeagh, S.M. & Steinke, D. 2008. DNA barcoding for the identification of smoked fish products. *Journal of Fish Biology* **72**(2), 464–471.

Sofos, J.N., Flick, G., Nychas, G.J., OBryan, C.A., Ricke, S.C. & Crandall, P.G. 2013. Meat, poultry, and seafood. In *Food Microbiology: Fundamentals and Frontiers*, 4th edn (eds M.P. Doyle & R.L. Buchanan), pp. 111–167. ASM Press, Washington, DC.

Sohier, D., Pavan, S., Riou, A., Combrisson, J. & Postollec, F. 2014. Evolution of microbiological analytical methods for dairy industry needs. *Frontiers in Microbiology* **5**(16), 1–10.

Soler, M., Ruiz-Rueda, O., Lopez-Siles, M., Calvo, L., Kaclikova, E. & Garcia-Gil, J.L. 2012. A new validated real-time PCR-based method for the specific and fast detection of *Cronobacter* spp. in infant formula. *Food Analytical Methdods* **5**(2), 179–187.

Solieri, L. & Giudici, P. 2010. Development of a sequence-characterized amplified region marker-targeted quantitative PCR assay for strain-specific detection of *Oenococcus oeni* during wine malolactic fermentation. *Applied and Environmental Microbiology* **76**(23), 7765–7774.

Solieri, L., Dakal, T.C. & Giudici, P. 2013. Next-generation sequencing and its potential impact on food microbial genomics. *Annals of Microbiology* **63**(1), 21–37.

Stanton, C., Ross, R.P., Fitzgerald, G.F. & Sinderen, D.V. 2005. Fermented functional foods based on probiotics and their biogenic metabolites. *Current Opinion in Biotechnology* **16**(2), 198–203.

Steele, P.R. & Pires, J.C. 2011. Biodiversity assessment: State-of-the-art techniques in phylogenomics and species identification. *American Journal of Botany* **98**(3), 415–425.

Steinke, D. & Hanner, R. 2011. The FISH-BOL collaborators protocol. *Mitochondrial DNA* **22**(S1), 10–14.

Steinkraus, K.H. (ed.) 1996. *Handbook of Indigenous Fermented Foods*, 2nd edn. Marcel Dekker, New York, USA.

Steinkraus, K.H. 1997. Classification of fermented foods: worldwide review of household fermentation techniques. *Food Control* **8**(5), 311–317.

Stoeckle, M.Y., Gamble, C.C., Kirpekar, R., Young, G., Ahmed, S. & Little, D.P. 2011. Commercial teas highlight plant DNA barcode identification successes and obstacles. *Scientific Reports* **1**(42), 1–7.

Sybesma, W., Burgess, C., Starrenburg, M., Sinderen, D.V. & Hugenholtz, J. 2004. Multivitamin production in *Lactococcus lactis* using metabolic engineering. *Metabolic Engineering* **6**(2), 109–115.

Tabasco, R., Paarup, T., Janer, C., Peláez, C. & Requena, T. 2007. Selective enumeration and identification of mixed cultures of *Streptococus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus*, *L. acidophilus*, *L. paracasei* subsp. *paracasei* and *Bifidobacterium lactis* in fermented milk. *International Dairy Journal* **17**(9), 1107–1114.

Takano, T. 2002. Anti-hypertensive activity of fermented dairy products containing biogenic peptides. *Antonie van Leeuwenhoek* **82**(1–4), 333–340.

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H. & Vogler, A.P. 2003. A plea for DNA taxonomy. *Trends in Ecology and Evolution* **18**(2), 70–74.

Teletchea, F. 2009. Molecular identification methods of fish species: reassessment and possible applications. *Reviews in Fish Biology and Fisheries* **19**(3), 265–293.

Theodoridis, S., Stefanaki, A., Tezcan, M., Aki, C., Kokkini, S. & Vlachonasios, K.E. 2012. DNA barcoding in native plants of the Labiatae (Lamiaceae) family from Chios Island (Greece) and the adjacent Çeşme-Karaburun Peninsula (Turkey). *Molecular Ecology Resources* **12**(4), 620–633.

This, P., Lacombe, T. & Thomas, M.R. 2006. Historical origins and genetic diversity of wine grapes. *Trends in Genetics* **22**(9), 511–519.

Valentini, A., Miquel, C. & Taberlet, P. 2010. DNA barcoding for honey biodiversity. *Diversity* **2**(4), 610–617.

van Boekel, M., Fogliano, V., Pellegrini, N., Stanton, C., Scholz, G., Lalljie, S., Somoza, V., Knorr, D., Rao Jasti, P. & Eisenbrand, G. 2010. A review on the beneficial aspects of food

processing. *Molecular Nutrition & Food Research* **54**(9), 1215–1247.

van Hylckama Vlieg, J.E., Veiga, P., Zhang, C., Derrien, M. & Zhao, L. 2011. Impact of microbial transformation of food on health – from fermented foods to fermentation in the gastrointestinal tract. *Current Opinion in Biotechnology* **22**(2), 211–219.

Vandamme, P., Pot., B., Gillis, M., De Vos, P., Kersters, K. & Swings, J. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiological Reviews* **60**(2), 407–438.

Vogler, A.P. & Monaghan, M. T. 2007. Recent advances in DNA taxonomy. *Journal of Zoological Systematics and Evolutionary Research* **45**(1), 1–10.

Waites, M.J., Morgan, N.L., Rockey, J.S. & Higton, G. (eds) 2009. *Industrial Microbiology: An Introduction*. Wiley-Blackwell, Oxford, UK.

Wallace, L.J., Boilard, S.M., Eagle, S.H., Spall, J.L., Shokralla, S. & Hajibabaei, M. 2012. DNA barcodes for everyday life: routine authentication of natural health products. *Food Research International* **49**(1), 446–452.

Wang, M., Zhao, H.X., Wang, L., Wang, T., Yang, R.W., Wang, X.L., Zhou, Y.H., Ding, C.B. & Zhang, L. 2013. Potential use of DNA barcoding for the identification of Salvia based on cpDNA and nrDNA sequences. *Gene* **528**(2), 206–215.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R. & Hebert, P.D.N. 2005. DNA barcoding Australias fish species. *Philosophical Transactions of the Royal Society B* **360**(1462), 1847–1857.

Ward, R.D., Hanner, R. & Hebert, P.D.N. 2009. The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology* **74**(2), 329–356.

Wiemers, M. & Fiedler, K. 2007. Does the DNA barcoding gap exist? A case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology* **4**(8), 1–16.

Woolfe, M. & Primrose, S. 2004. Food forensics: using DNA technology to combat misdescription and fraud. *Trends in Biotechnology* **22**(5), 222–226.

Xin, T., Yao, H., Gao, H., Zhou, X., Ma, X., Xu, C., Chen, J., Han, J., Pang, X., Xu, R., Song, J. & Chen, S. 2013. Super food *Lycium barbarum* (Solanaceae) traceability via an internal transcribed spacer 2 barcode. *Food Research International* **54**(2), 1699–1704.

Yancy, H.F., Zemlak, T.S., Mason, J.A., Washington, J.D., Tenge, B.J., Nguyen, N.L., Barnett, J.D., Savary, W.E., Hill, W.E., Moore, M.M., Fry, F.S., Randolph, S.C. Rogers, P.L. & Hebert, P.D.N. 2008. Potential use of DNA barcodes in regulatory science: applications of the Regulatory Fish Encyclopedia. *Journal of Food Protection* **71**(1), 210–217.

Yano, T., Sakai, Y., Uchida, K., Nakao, Y., Ishinata, K., Nakano, S., Yamada, T., Sakai, S., Urisu, A., Akiyama, H. & Maitani, T. 2007. Detection of walnut residues in processed foods by polymerase chain reaction. *Bioscience, Biotechnology, and Biochemistry* **71**(7), 1793–1796.

Yu, J., Yan, H.X., Lu, Z.H. & Zhou, Z.Q. 2011. Screening potential DNA barcode regions of chloroplast coding genome for *Citrus* and its related genera. *Scientia Agricultura Sinica* **44**, 341–348.

Zeale, M.R.K., Butlin, R.K., Barker, G.L.A., Lees, D.C. & Jones, G. 2011. Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources* **11**(2), 236–244.

Zecca, G.A., De Mattia, F., Lovicu, G., Labra, M., Sala, F. & Grassi, F. 2010. Wild grapevine: silvestris, hybrids or cultivars that escaped from vineyards? Molecular evidence in Sardinia. *Plant Biology* **12**(3), 558–562.

Zhang, W., Xiao, S., Samaraweera, H., Lee, E.J. & Ahn, D.U. 2010. Improving functional value of meat products. *Meat Science* **86**(1), 15–31.

Zuo, Y., Chen, Z., Kondo, K., Funamoto, T., Wen, J. & Zhou, S. 2011. DNA barcoding of *Panax* species. *Planta Medica* **77**(2), 182–187.