Accred Qual Assur (2013) 18:77–83
DOI 10.1007/s00769-012-0941-z

REVIEW

Considerations for the development and application of control materials to improve metagenomic microbial community profiling

Jim F. Huggett · Thomas Laver · Sasithon Tamisak · Gavin Nixon · Denise M. O'Sullivan · Ramnath Elaswarapu · David J. Studholme · Carole A. Foy

Received: 1 October 2012 / Accepted: 24 November 2012 / Published online: 28 December 2012

Abstract Advances in DNA sequencing technology provide the possibility to analyse and characterize the genetic material from microbial populations (the microbiome) as a whole. Such comprehensive analysis of a microbiome using these 'metagenomic' approaches offers the potential to understand industrial, clinical and environmental microbiology to a level of detail that is unfeasible using conventional molecular or culture-based methods. However, the complexity offered by metagenomic analysis is also the weakness of this method and poses considerable challenges during analytical standardisation. In this manuscript, we discuss options for developing control materials for metagenomic analysis and describe our preliminary work investigating how such materials can be used to assist metagenomic measurements. The control materials we have developed demonstrate that, when performing 16S rDNA sequencing, different library preparation methods (incorporating adapters before and after the PCR) and small primer mismatches can alter the reported metagenomic profile. These findings illustrate that metagenomic analysis can be heavily biased by the choice of method and underpin the need for control materials that can provide a useful tool in informing choice of protocol for accurate analysis.

Keywords Metagenomics · Microbes · Microbiome · Molecular · Next generation sequencing · Standards · Microbial-profiling · NGS

Introduction

A metagenome can be defined as the genetic material present within an environmental sample. The term environment in this context refers not only to the classic outdoor terrestrial [1] and aquatic [2, 3] domains, but also environments found within other organisms, like the gut [4, 5] and associated matrices like probiotic supplements [6] or linked to other man-made scenarios like water purification [7] or during fermentation [8]. The field of metagenome analysis (metagenomics) has grown rapidly as the last 10 years have seen the development and establishment of next generation (also termed massively parallel) sequencing (NGS), a technique that has the potential to read over a billion sequences in a single run. By combining the power of NGS with microbial population analysis, metagenomics offers huge potential to enable us to understand environmental, industrial and clinical microbiology. By capturing a large proportion of the genetic material, the microbiome can be analysed as the complex population of component organisms present in vivo. This has the potential to offer a considerable breakthrough when compared to existing molecular and culture-based methods that typically focus on individual species or small groups of organisms. There are a number of studies that have already illustrated the power of metagenomics with the Sorcerer II Global Ocean Sampling Expedition [3] and Human Microbiome Project [9] arguably being most prominent.

Presented at BERM-13, June 2012, Vienna, Austria.

Electronic supplementary material The online version of this article (doi:10.1007/s00769-012-0941-z) contains supplementary material, which is available to authorized users.

J. F. Huggett (✉) · S. Tamisak · G. Nixon ·
D. M. O'Sullivan · R. Elaswarapu · C. A. Foy
Molecular Biology, LGC, Queens Road,
Teddington TW11 0LY, UK
e-mail: jim.huggett@lgcgroup.com

T. Laver · D. J. Studholme
Biosciences, University of Exeter, Geoffrey Pope Building,
Stocker Road, Exeter EX4 4QD, UK

Metagenomic analysis does not come without its difficulties; the molecular techniques used to generate the data are not only challenging, but highly disparate, and there are the additional issues associated with how best to manage, store and analyse the huge data sets that are produced from a metagenomics experiment [10]. Dealing with the informatics challenges is important as they represent a new problem when working with NGS due to the size and complexity of the resultant data. This is especially pertinent to metagenomics where sequences of potentially different species of unknown organisms may be being detected and quantified. The informatics challenges are not insurmountable though and considerable achievements have been made to facilitate data analysis [11, 12]. However, while dealing with the challenges associated with informatics is fundamental for metagenomic analysis, it must not overshadow considerations around the more common issues of technical accuracy and standardization. Furthermore, the inclusion of control materials, be they simpler standards for quality control or more complicated reference materials, could provide a valuable tool for understanding and tackling both the technical and the informatics challenges. This leads to the question of how such control materials should best support the different approaches used for metagenomics analysis.

Different approaches

Any standard for metagenomic analysis will need to account for the fact that there are different methods for analysing complex microbial populations following nucleic acid extraction. At the simplest dichotomy, mixed microbial sequencing analysis can be performed by sequencing either the whole metagenome or a targeted subset of it. The simplest approach uses polymerase chain reaction (PCR) to specifically amplify generic target sequences (sequences shared between different microbes) like the bacterial 16S ribosomal RNA gene (16S rDNA) [13]. The PCR products (amplicons) are then sequenced (by an approach termed amplicon sequencing) following their manipulation to prepare them for sequencing (library preparation). Small differences within the amplified deoxyribonucleic acid (DNA) sequences are identified and compared to databases of known sequences to determine which taxonomic groups are present and with what relative abundance. This approach can be very sensitive but is limited to the taxonomic groups defined by the specificity of the PCR primers; for example, the 16S ribosomal DNA gene is only present within bacteria and so a PCR targeting the sequence will not detect eukaryotic microbes or viruses.

Next generation sequencing can also be used to perform a broader assessment of the microbial sequences present. Applying NGS to whole metagenome sequencing allows abundances of eukaryotic, bacterial, archaeal microbes as well as viruses to be measured. A further level of complexity will also be found by measuring ribonucleic acid (RNA), as many viruses have RNA genomes (rather than DNA). RNA may also be a useful proxy for microbial viability. Furthermore, the types of RNA measured may provide an idea of the metabolic challenges facing an environment and assist in explaining why specific microbes predominate in certain environments [14].

Whole metagenome sequencing requires fragmenting the extracted genetic material, library preparation and sequencing of the library fragments. Different NGS instruments provide different data outputs with the current trade-off being an inverse relationship between length and number of sequences reads. The increased number of reads will provide more sequence depth and therefore detect rare sequences, while the longer reads will better facilitate the building of the different microbial genomes present within a sample which is useful when dealing with newly discovered microbes for which reference sequences are unavailable.

Considerations for developing control materials for metagenomic analysis

There are a number of issues that require careful consideration during the development of control materials for metagenomic analysis such as whether they should comprise whole microbes or extracted nucleic acids. The ideal control material would comprise whole microbes. This would allow initial steps required to extract the nucleic acid, as well as those for sequencing, to be controlled during any measurement. The processes of nucleic acid sampling, storage and extraction represent an important, and usually neglected, source of error for molecular measurement that must be considered when measuring metagenomes [15]. However, developing control materials to support this, while being the most desirable, presents considerable challenges.

Assigning values based on whole microbes is not a simple matter; classic microbial culture methods can be used to estimate the number of colony forming units (CFU). However, viable microbe number represents a fundamentally different measure to nucleic acid quantity and the two may or may not be correlated. Large batch production of material offers alternatives, whether production is achieved by mixing known amounts of microbes or following the strategy used for matrix reference materials like water and soil or the world health organization (WHO) clinical international standards [16]. For metagenomics, large batches of real clinical/environmental samples are arguably the most suitable controls. They do, however, present challenges when considering characterization. Furthermore, there are additional problems of ensuring comparable composition and traceability when the material runs out and new batches are required, problems that will be accentuated by the complex nature required from a metagenomic material.

Directing resources into developing, well defined, whole microbial metagenomic materials would not only provide valuable research tools, but the findings from such an exercise would also inform the wider field of the impact of the initial storage and extraction steps on a microbial-profiling experiment. Yet, even with rigorous efforts, the challenges associated with developing whole microbial metagenomic materials are unlikely to be solved in the immediate future. In the absence of whole microbial materials, the availability of simpler process controls, to investigate the specific issues, like those associated with the sequencing steps, could offer considerable benefits. For this reason, early stage materials for metagenomic analysis are likely to comprise nucleic acid extracts, and the remainder of this manuscript will focus on the production of nucleic acid control materials. Types of nucleic acid standards could range from complex environmental extracts, mixes of cultured bacteria or synthetic genomes to small fragments of DNA. Mixtures of cultured microbial genomic DNA (gDNA) offer a simpler metagenomic standard and this is what we and others have been investigating as potential control materials [9, 17].

Technical issues associated with microbial molecular measurement, whether individual microbes or complex communities, can be broadly split into measurement of identity and of quantity. Sequence identity will facilitate the grouping of an organism and depending on the analytical requirement can be down to species level or even identifying different groups (strains) within a species. Frequently, identity is required at the sub-strain level to adequately consider specific genotypes at key genetic loci that may afford resistance to therapies and thus are needed to guide treatment. Microbial quantification requires not only some degree of identity measure, but also estimation of the amount.

While quantifying/assigning relative proportions of the component organisms and/or gene sequences is popular by metagenomic studies [4, 7, 18–20], what has not been addressed are estimations of technical confidence, or uncertainty, with the importance of experimental design to capture repeatability being highlighted in very prominent journals [21]. To confound this, there are few commercially available materials for metagenomic analysis and actual gDNA reference materials are limited to individual microbes [22]. Current methods for preparing metagenomic control materials have relied on combining microbial gDNA extracts together that have been quantified using classical optical methods [17] or using quantitative molecular methods like real time PCR (qPCR) [9]. While further work is required to describe how best to prepare metagenomic control materials, it is already becoming clear that they can offer a valuable tool (see Fig. 1 and discussion below).

Additional considerations for reference materials

When considering the production of microbial genetic reference materials to support both metagenomic analysis and traceability, there are additional biological factors that must be considered in addition to those applied to classical chemical and physical reference materials. Not only is gDNA a large macromolecule, it can also vary in size, complexity and epigenetic status, offering hidden challenges when using conventional methods for value assignment.

The simplest and most frequently used method for quantifying bacterial gDNA is by estimating mass, usually using absorbance or fluorescence [23]. Measurement of absorbance at 260 nm wavelength can be highly variable

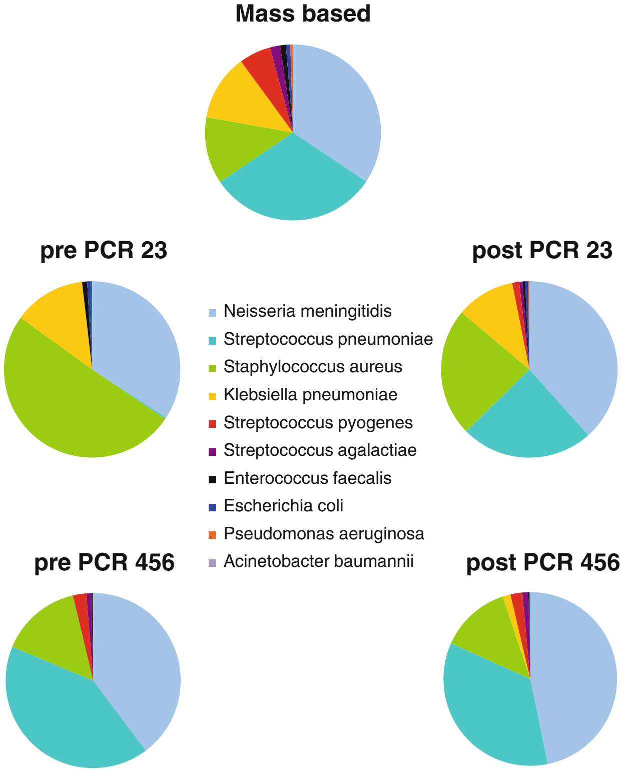

Fig. 1 Proportions of the MCM bacterial gDNA mixtures presented as a *pie chart* based on estimation of mass and analysed using pre-PCR and post-PCR protocols with assays 23 and 456. *Klebsiella pneumoniae* and *Streptococcus pneumoniae* were not measured as a predominant species when pre-PCR protocols are performed using assay 23 and 456, respectively. This effect was less prominent when post-PCR protocols were employed

Table 1 Relative proportions (all values in %) of bacterial gDNA used to make the metagenomic control material based on either mass, gDNA copy number and 16S rDNA copies

Material	Control material based on		
	Mass	gDNA copy number	16S copies
MRSA	0.80	0.75	0.78
MSSA	12.00	11.52	11.96
S. pneumoniae	25.60	31.09	25.83
S. pyogenes	4.00	5.83	7.26
S. galactiae	1.50	1.87	2.72
E. faecalis	1.00	0.81	0.67
P. aeruginosa	1.00	0.43	0.36
K. pneumoniae	24.00	12.18	20.24
A. baumannii	0.10	0.07	0.07
E. coli	2.00	1.03	1.50
N. meningitidis	28.00	34.42	28.60

MRSA methicillin-resistant *Staphylococcus aureus*, *MSSA* methicillin-sensitive *Staphylococcus aureus*

and susceptible to non-nucleic acid impurities. Fluorescence using double-stranded DNA-binding dyes like pico green, while arguably more accurate than absorbance, require comparison to a calibration curve of nucleic acid that must have already been value assigned, thus leaving a circular problem of how to value assign the calibrant. Fluorescence has the additional problem of choice of calibrant; bacterial genomes are fairly large and complex when compared to simpler plasmid and viral sequences that are routinely used for fluorescence quantification, and it is not clear how commutable a plasmid standard is for accurate quantification of different bacterial gDNA.

Using gDNA mass, while simple, is not sufficient when considering bacterial gDNA for metagenomics. Different bacterial species and strains can have very different genome sizes. This means that gDNA mass is not a reliable indicator of biomass or of number of cells. This will impact on a metagenomic reference material (whether made from physically mixed microbial DNA or an extracted environmental sample) as the relative proportion of different bacterial genomes by mass will be different from relative proportions of the actual genome copies (Table 1). Consequently, when quantifying or defining the relative proportion of bacterial DNA, genome copy number must be considered and is arguably the preferred measurand to DNA mass.

Estimation of number of bacterial genome copies is further complicated by the fact that gDNA undergoes complex patterns of chemical modification, such as methylation. These epigenetic modifications will alter the molecular weight of the genome and therefore the number of genome copies in a given mass. The extent of DNA modification may vary among different species as well as among individuals within a species. An additional issue arises from the fact that bacterial gDNA materials are usually made from extracts of bacteria that are replicating their genomes during a logarithmic phase of growth. This further complicates quantification because bacterial DNA replication occurs in a bidirectional fashion from a single point in the genome (the origin of replication). Consequently, within a growing population of bacteria, genes that are proximal to the origin of replication will be more abundant than those that are more distal. This replication-associated difference in gene dosage is well defined [24] and can be seen by NGS analysis [25]. This makes value assignment by enumeration more complicated and further reduces the suitability of using mass alone as a traceable measurand for value assigning reference materials containing bacterial gDNA.

For these reasons, production of a bacterial gDNA reference material would ideally incorporate some measure of sequence homogeneity to account for batch to batch differences in the abundance/dosage of different parts of the genome in question. This and the epigenetic issues will need to be considered when estimating the uncertainty while value assigning microbial materials. Eukaryotic and viral microbes can also contain epigenetic DNA modifications, and there will also be the potential for sequence inhomogeneity, so some measure may also be needed to factor this into any uncertainty estimation.

Ironically, as next generation sequencing technologies improve, they may turn out to be the best methods to estimate the impact of gene dosage and epigenetics on sequence inhomogeneity, thus proving crucial to understanding this increased biological complexity and in turn reducing the uncertainties applied to subsequent metagenomic reference materials. These considerations are particularly important when considering mixes of microbial genomes required to standardize metagenomic experiments because the impact of these biological factors will differ between species and possibly between different preparations of the same species. However, while well-described metagenomic reference materials are likely to be valuable in the future, before any resources are channelled into developing such a complex resource, we must consider what is currently needed by the community.

What is needed?

Whether it was the intention of the initial experiment or not, metagenomic studies frequently publish not only the identities of the microbes present, but also offer some idea of their relative proportions. This is typically presented as a pie chart of the average proportion of organisms, or gene types, with no information on experimental confidence, but which offer the reader striking information highlighting the most abundant group, or groups, of organism(s) (Fig. 1).

Metagenomic control materials will allow laboratories to test not only their experimental procedures for quantitative bias, but also the precision associated with data and give an idea of the error that the technique is providing within a specific laboratory. This is a fundamental consideration to any measurement and should not be neglected within the field of metagenomics.

Initially, the uncertainty associated with metagenomic control materials may not need to be very small. While newer technologies, like digital PCR, are able to produce small uncertainties on minute quantities of a specific piece of DNA [26], current quantitative molecular measurement of individual microbes in clinical scenarios rarely measures changes of less than an order of magnitude and then only with some kind of reference material if inter-laboratory comparability is considered important [27]. Whether this reflects a limitation of the technique that may be improved on, or the true biological variation remains to be determined; however, the precisions of early metagenomic measurements are unlikely to be any better. Consequently, it may be favourable to produce metagenomics control materials now, accepting that not all of the sources of error have been pinned down yet. Such materials would, unavoidably, have large associated uncertainties, and further work is required to define the associated errors. However, these materials could nevertheless be provided with the specific aim of reducing the even larger errors that are likely to occur as a result of numerous different laboratories applying uncontrolled procedures using highly disparate approaches.

Consequently, there is an arguable case for providers of standards and reference materials to produce commercially the types of process controls described below and elsewhere [9, 17].

Example of the application of control materials

To investigate the development of a metagenomic control material, we developed a mixed panel of gDNA from 11 different pathogenic bacteria (Table 1) (Detailed information on the material preparation is available in the Electronic Supplementary Material (ESM)). The bacterial DNA was mixed at defined proportions based on mass; then, we estimated genome copy numbers and numbers of 16S ribosomal gene copies (which can vary between different bacteria species and strains [28]) (ESM, Table 1). This metagenomic control material (MCM) provided approximately 3 orders of magnitude difference in the abundance of the most concentrated gDNA (*N. meningitidis*) compared to the least abundant (*A. baumannii*). The concentrations were chosen to reflect an approximation of what might be expected from a real clinical extract. We prepared 300×50 μl aliquots of the material in TE pH 7; material stability estimation at −20 and −80 °C is ongoing.

In our initial experiments, we used the MCM to compare two amplicon sequencing protocols, with the MCM spiked into 250 ng human gDNA (Promega). The Roche GS Junior sequencing protocol requires the PCR amplicons to have adapter sequences attached during library preparation. These adapters can be added pre- (during primer synthesis) or post- (using ligation) PCR. This experiment investigated how the two approaches differed when measuring the MCM. Both amplicon sequencing protocols used the same two PCR assays to amplify different parts of the bacterial 16S rDNA gene (either variable regions 2 and 3 [assay 23] or variable regions 4, 5 and 6 [assay 456], details in ESM Table 2). Primer sets were designed to bind the conserved regions of the 16S rDNA gene but were selected to not complement perfectly with all 10 MCM species (Table 2). Primer mismatches against specific bacterial species are common when performing 16S analysis as the conserved regions are not perfectly conserved. Such mismatches can impact on the efficiency of the PCR and thus on quantification of individual bacterial gDNA [29].

This experiment aimed to investigate whether the MCM was of value when comparing the use of different amplicon sequencing approaches and for informing whether species-specific primer mismatches could cause measurement issues for metagenomic analysis. Following the respective emulsion PCR, the amplicons were sequenced according to the Roche GS Junior sequencing protocol. Sequence data were analysed using megaBLAST [30] against a custom database of 16S rDNA sequences comprising the 16S rDNA sequences from the ten species included in the MCM. We then used MEGAN [31] to assign sequence reads to bacterial taxa and to infer relative abundances of each sequenced species.

Figure 1 presents the results from different amplicon sequencing approaches using two different PCR assays. The two approaches provide different results with both assays demonstrating better agreement with the initial mass-based estimation of the MCM composition when adapters were ligated onto the amplicon post-PCR. We observed that where the results underestimate the abundance of specific bacteria (compared to the mass-based method) corresponded to where a sequence mismatch occurred (Table 2). This was particularly apparent when presenting data using the pie chart format commonly found in metagenomic publications; when applying the adapters post-PCR, the magnitudes of the predominant organisms differ from the mass-based estimation, but the actual predominant organisms remain the same. When adapters were applied pre-PCR, the predominant organisms change depending on which assay is used.

Table 2 Details of the primer complementarily compared to the 10 organisms used to make the MCM

		Comment
Assay 23		Assay 23 has two mismatches when compared to the 11 organisms comprising the metagenomic control material. These mismatches are present within the forward primer affecting the genus *Streptococcae* and at the 5′ end of the reverse primer affecting *A. baumannii.* This corresponds to underestimation of the three *Streptococcae* in Fig. 1 but did not effect the *A. baumannii* estimated proportion
Forward		
Seven bacteria	AGHGGCGRACGGGTGA	
Three Streptoccocae	AGH**A**GCGRACGGGTGA	
Reverse		
Nine bacteria	CGTATTACCGCGGCTGCT	
A. baumannii	**T**GTATTACCGCGGCTGCT	
Assay 456		Assay 456 has one three mismatches when compared to the 11 organisms comprising the metagenomic control material. These mismatches are present at the 3′ end of the forward primer affecting *A. baumannii* and two errors within the reverse primer effecting *K. pneumoniae* and *E. coli*. These mismatches correspond to underestimation of *A. baumanii, K. pneumonia* and *E. coli* in Fig. 1
Forward		
Nine bacteria	AGCAGCCGCGGTAATACG	
A. baumannii	AGCAGCCGCGGTAATAC**A**	
Reverse		
Eight bacteria	CATCTCACGACACGAGCTGAC	
K. pneumoniae and *E. coli*	CAT**T**TCAC**A**ACACGAGCTGAC	

Sequences denominated as H correspond to A, C or T and R to A or G

The result from this preliminary experiment suggests that amplicon sequencing can differ in its estimation of the abundance of specific bacteria in a metagenomic sample that this can be dependent on choice of assay and appears to be linked to respective sequence mismatches within conserved regions (Table 2). Furthermore, these differences were more pronounced when the sequencing adapters were incorporated onto the primers prior to the PCR suggesting primers which also contain adapter maybe more susceptible to the biases observed. The use of control materials like the LGC MCM demonstrate that findings can differ considerably depending on how the initial experiment is set up and illustrates that they provide a valuable tool for assay development.

While these types of standards have not been extensively used yet, they represent the only means by which technical problems, like those highlighted by Fig. 1, can be identified. Thus, there is a clear need for the availability of carefully prepared and defined materials. This is crucial to facilitate the development of metagenomics capacity in laboratories planning to start performing such analysis and to standardize the measurements between different laboratories performing existing analysis.

Conclusion

Metagenomics offers huge potential to revolutionize our understanding of the microbial world. With the explosion in different high-capacity methodologies to investigate metagenomes, considerable thought is needed to ensure both informatics analysis and initial experimental design are performed using methods that allow inter-experimental comparability. Furthermore, it is vital that the issues associated with metagenomic analysis highlighted here are considered. This is because if results being reported are unrepresentative of reality due to technical biases, our early understanding of the microbiome may be misrepresented. These challenges may not be insurmountable, however, and our early findings suggest that the use of nucleic acid-based materials as process controls could offer a valuable tool for developing metagenomic capacity although it is likely that controls that contain mixes of whole microbes are what will eventually be required. Accurate and comparable metagenomic approaches will facilitate the maximum impact of associated research, ensuring that our understanding of the microbial world in the context of the environment and human health is not misguided by technical bias.

Acknowledgments The authors would like to acknowledge Dr Alison Woolford for critical review of the manuscript. The work described in this manuscript was funded by the UK National Measurement System. Thomas Laver was supported by BBSRC Industrial Case studentship award BB/H016120/1.

References

1. Daniel R (2005) The metagenomics of soil. Nat Rev Microbiol 3(6):470–478
2. Ghai R, Rodriguez-Valera F, McMahon KD, Toyama D, Rinke R, de Oliveira TCS, Garcia JW, de Miranda FP, Henrique-Silva F

(2011) Metagenomics of the water column in the pristine upper course of the Amazon river. PLoS ONE 6(8):e23785

3. Gross L (2007) Untapped bounty: sampling the seas to survey microbial biodiversity. PLoS Biol 5(3):e85
4. The Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. Nature 486 (7402):207–214
5. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331(6016):463–467
6. Gueimonde M, Collado MC (2012) Metagenomics and probiotics. Clin Microbiol Infect 18(Suppl 4):32–34
7. Gomez-Alvarez V, Revetta RP, Santo Domingo JW (2012) Metagenomic analyses of drinking water receiving different disinfection treatments. Appl Environ Microbiol 78(17):6095–6102
8. Smid EJ, Hugenholtz J (2010) Functional genomics for food fermentation processes. Annu Rev Food Sci Technol 1:497–519
9. Human Microbiome Project Consortium (2012) A framework for human microbiome research. Nature 486(7402):215–221
10. National Research Council (US) Committee on Metagenomics: Challenges and Functional Applications (2007) The new science of metagenomics: revealing the secrets of our microbial planet. National Academies Press (US), Washington, DC
11. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7(5):335–336
12. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA (2008) The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinform 9:386
13. Baker GC, Smith JJ, Cowan DA (2003) Review and re-analysis of domain-specific 16S primers. J Microbiol Methods 55(3): 541–555
14. Marchetti A, Schruth DM, Durkin CA, Parker MS, Kodner RB, Berthiaume CT, Morales R, Allen AE, Armbrust EV (2012) Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. Proc Natl Acad Sci USA 109(6):E317–E325
15. Wilner D, Daly J, Whiley D, Grimwood K, Wainwright CE, Hugenholtz P (2012) Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. PloS ONE 7(4):e34605
16. Holden MJ, Madej RM, Minor P, Kalman LV (2011) Molecular diagnostics: harmonization through reference materials, documentary standards and proficiency testing. Expert Rev Mol Diagn 11(7):741–755
17. Huggett JF, O'Sullivan D, Laver T, Temisak S, Elaswarapu R, Nixon G, Studholme DJ, Foy CA (2012) Standards for microbial community molecular profiling. Paper presented at the 13th international symposium on biological and environmental reference materials (BERM 13), Vienna, 25–29 June 2012. http://www.nmschembio.org.uk/ResourcesArticleType2.aspx?m=174&amid=7812
18. Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, Quan PL, Briese T, Hornig M, Geiser DM, Martinson V, vanEngelsdorp D, Kalkstein AL, Drysdale A, Hui J, Zhai J, Cui L, Hutchison SK, Simons JF, Egholm M, Pettis JS, Lipkin WI (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. Science 318(5848):283–287
19. Culley AI, Lang AS, Suttle CA (2006) Metagenomic analysis of coastal RNA virus communities. Science 312(5781):1795–1798
20. Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, Eisen JA, Heidelberg KB, Manning G, Li W, Jaroszewski L, Cieplak P, Miller CS, Li H, Mashiyama ST, Jochimiak MP, van Belle C, Chandonia JM, Soergel DA, Zhai Y, Natarajan K, Lee S, Raphael BJ, Bafna V, Friedman R, Brenner SE, Godzik A, Eisenberg D, Dixon JE, Taylor SS, Strausberg RL, Frazier M, Venter JC (2007) The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. PLoS Biol 5(3):e16
21. Knight R, Jansson J, Field D, Fierer N, Desai N, Fuhrman JA, Hugenholtz P, van der Lelie D, Meyer F, Stevens R, Bailey MJ, Gordon JI, Kowalchuk GA, Gilbert JA (2012) Unlocking the potential of metagenomics through replicated experimental design. Nat Biotechnol 30(6):513–520
22. European Commision Joint Research Centre (2012) Catalogue of Reference Materials 2012–2013
23. De Mey M, Lequeux G, Maertens J, De Maeseneire S, Soetaert W, Vandamme E (2006) Comparison of DNA and RNA quantification methods suitable for parameter estimation in metabolic modeling of microorganisms. Anal Biochem 353(2):198–203
24. Kepes F, Jester BC, Lepage T, Rafiei N, Rosu B, Junier I (2012) The layout of a bacterial genome. FEBS Lett 586(15):2043–2048. doi:S0014-5793(12)00251-7
25. Paszkiewicz K, Studholme DJ (2010) De novo assembly of short sequence reads. Brief Bioinform 11(5):457–472
26. Sanders R, Huggett JF, Bushell CA, Cowen S, Scott DJ, Foy CA (2011) Evaluation of digital PCR for absolute DNA quantification. Anal Chem 83(17):6474–6484
27. Fryer JF, Baylis SA, Gottlieb AL, Ferguson M, Vincini GA, Bevan VM, Carman WF, Minor PD (2008) Development of working reference materials for clinical virology. J Clin Virol 43(4):367–371
28. Kembel SW, Wu M, Eisen JA, Green JL (2012) Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance. PLoS Comput Biol 8(10):e1002743
29. Bru D, Martin-Laurent F, Philippot L (2008) Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example. Appl Environ Microbiol 74(5):1660–1663
30. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schaffer AA (2008) Database indexing for production MegaBLAST searches. Bioinformatics 24(16):1757–1764
31. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. Genome Res 21(9):1552–1560