Ann Microbiol (2013) 63:21–37
DOI 10.1007/s13213-012-0478-8

REVIEW ARTICLE

Next-generation sequencing and its potential impact on food microbial genomics

Lisa Solieri · Tikam Chand Dakal · Paolo Giudici

Received: 23 January 2012 / Accepted: 24 April 2012 / Published online: 31 May 2012

Abstract Recent efforts of researchers to elucidate the molecular mechanisms of biological systems have been revolutionized greatly with the use of high throughput and cost-effective techniques such as next generation sequencing (NGS). Application of NGS to microbial genomics is not just limited to predict the prevalence of microorganisms in food samples but also to elucidate the molecular basis of how microorganisms respond to different food-associated conditions, which in turn offers tremendous opportunities to predict and control the growth and survival of desirable or undesirable microorganisms in food. Concurrently, NGS has facilitated the development of new genome-assisted approaches for correlating genotype and phenotype. The aim of this review is to provide a snapshot of the various possibilities that these new technologies are opening up in area of food microbiology, focusing the discussion mainly on lactic acid bacteria and yeasts associated with fermented food. The contribution of NGS to a system level understanding of food microorganisms is also discussed.

Keywords Next generation sequencing · Food microbiology · Community profiling · Metagenomics · Genome sequencing

Introduction

DNA sequencing has a profound impact on the advancement of molecular biology (Gilbert 1981). For the past 30 years, the Sanger dideoxy chain-termination sequencing method has been relied on for determining gene/DNA sequences. This technology reached its zenith in the development of single tube chemistry with fluorescently marked termination bases, heat stable polymerases, and automated capillary electrophoresis; but then reached a plateau in terms of technical development. With the ultimate goal of deciphering complete genomes, the requirement for high-throughput sequencing grew by an unpredicted extent. Several novel approaches evolved to replace the Sanger method as the dominant provider of sequencing data. These approaches fall under a broad definition of "next generation sequencing" (NGS) and provide sequence data around a hundred times faster and cheaper than the conventional Sanger approach. Sequencers from 454 Life Sciences/Roche, Solexa/Illumina and Applied Biosystems (SOLiD technology) are already in production as second generation technologies, and other competitive technologies are likely to appear on the market soon (so-called 3rd generation NGS). The reduction in cost and time for generating DNA sequence data has resulted in a range of new successful applications, such as whole genome sequencing, resequencing, RNA-sequencing to profile the cellular transcriptome, ChIP-sequencing to identify binding sites of DNA-associated proteins, as well as in fields as diverse as ecology (Angly et al. 2006; Edwards et al. 2006; Sogin et al. 2006) and the study of ancient DNA (Poinar et al. 2006).

Food microbiology deals with the study of microorganisms that have both beneficial and deleterious effects on the quality and safety of food products. The fast and low-cost NGS approaches have revolutionized microbial taxonomy and classification and have changed the landscape of genome sequencing projects for food-associated microbial species (Coenye et al. 2005). The NGS-driven advances have been exploited mainly to re-sequence strains and individuals

L. Solieri (✉) · T. C. Dakal · P. Giudici
Department of Agricultural and Food Sciences,
University of Modena and Reggio Emilia,
via Amendola 2,
42100 Reggio Emilia, Italy
e-mail: lisa.solieri@unimore.it

for which reference genome sequences are available in order to sample genomic diversity within microbial species. Besides this, NGS technologies have also been employed for de novo sequence assemblage where pre-existing information related to sequence is not required. Such studies have identified individual genome variation in bacterial strains, yeast and filamentous fungi, opening the possibility of constructing "personalized genomics" also for microbial cells. In fact, effort in the direction of developing personalized genomics for humans to predict disease susceptibility and genetic risk factors have already begun (Mardis 2008a; Boguski et al. 2009). Moreover, NGS approaches have also greatly increased the ability of researchers to profile food microbial communities, as well as to elucidate the molecular mechanisms of interesting functionalities in food ecosystems. These applications enable the culture-independent sequencing of collective sets of DNA or RNA molecules obtained from mixed microbial communities to determine their content. An overview of the main applications of NGS technologies, and their integration, in the field of food microbiology is shown in Fig. 1.

In this article, we review various aspects of the impact of NGS technologies on food microbial genomics, with attention focused on the two main microbial groups involved in fermented food production: lactic acid bacteria (LAB) and yeasts. We also address the contribution of NGS in providing DNA/RNA sequence data for constructing a system level understanding of microbial behavior in food ecosystems.

What sets NGS apart from conventional sequencing technology?

The landmark publication in 1977 by Sanger and colleagues described a means of sequencing genes using a chain-termination method, often called 'Sanger 'or 'dideoxy'

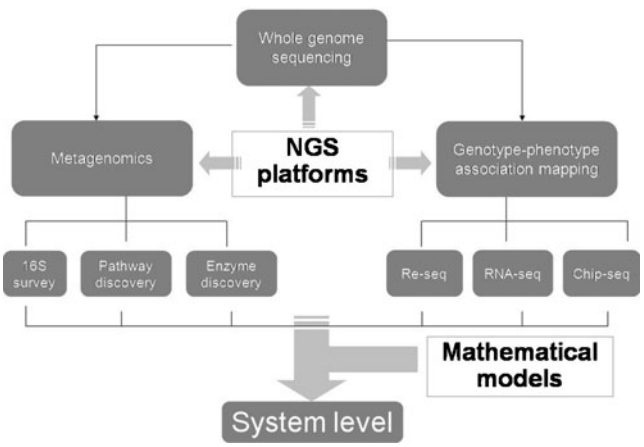

Fig. 1 Main applications of "next generation sequencing" (NGS) to address food microbiology questions

sequencing (Sanger et al. 1977). Based on capillary electrophoresis using the ABI 3730xL platform, this method remained the most commonly used DNA sequencing technique for more than three decades. This technique was used to sequence the bacterial genome of *Haemophilus influenzae*, which was published in 1995 (Fleischmann et al. 1995). It has been also used to accomplish complete human genome sequencing initiatives undertaken by a group of publicly funded researchers (International Human Genome Sequencing Consortium) in 1990. The 3 billion bases of the entire euchromatic human genome was sequenced in around 13 years at a cost of approximately $3 billion (Venter et al. 2001) (Fig. 2).

Even if Sanger sequencing is regarded as the foundation of genomics, research in the field of microbial genomics, transcriptomics, and metagenomics was greatly limited by the unavailability of an efficient technology for high-throughput screening and sequencing large sets of genome data, which conventional Sanger sequencing technology was unable to process efficiently and cost-effectively. The limitations associated with Sanger sequencing technology drove the research for more scalable and lower-cost sequencing solutions. These alternative approaches can be grouped in four different categories: (1) microelectrophoretic methods (microfabricated capillary electrophoretic sequencing), (2) sequencing by hybridization (ChIP and Microarray), (3) real-time observation of single molecules (nanopore sequencing), (4) cyclic-array sequencing (discussed below in detail) (reviewed by Shendure et al. 2004; Shendure and Ji 2008). In this review, we will use 'next-generation sequencing' (NGS) to refer to the various recently commercialized implementations of cyclic-array sequencing. This method accomplishes sequencing of a dense array of spatially separated nucleotide features by iterative cycles of enzymatic manipulation where each cycle queries only one or a few bases, but thousands to billions of features are processed in parallel (Mitra and Church 1999; Mitra et al. 2003). In order to present a comprehensive overview of cyclic array or NGS, we use the convention of 2nd generation to indicate platforms that require PCR amplification of the template molecules prior to sequencing, and 3rd generation to indicate sequencing platforms that rely on use of a single DNA molecule for sequencing without any prior PCR amplification step. We focus our attention on the main applications of 2nd generation NGS (called simply NGS for brevity).

Second generation NGS technologies were born at the dawn of twenty-first century in the year 2000 with the foundation of 454 Life Sciences (originally 454 Corporation) by Jonathan Rothberg. Concurrently, other sequencing platforms, such as Solexa (Illumina) and SOLid (ABI/Life Technologies), were also introduced into the market (Hert et al. 2008). While 454/Roche pyrosequencing is based on detection of pyrophosphates release during DNA synthesis, Solexa/

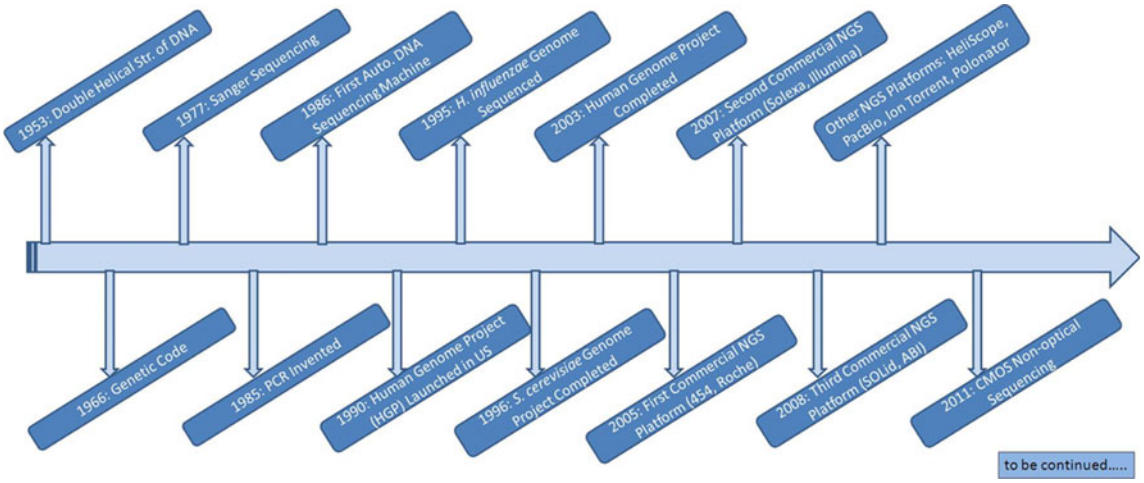

Fig. 2 Timeline of the progress and evolution of sequencing technologies

Illumina relies on detection of single bases as they are incorporated into growing DNA strands and SOLid /ABI relies on multiple cycles of ligation, detection and cleavage. Despite their technological differences, these three platforms are based on conceptually similar work flows for the production and analysis of sequencing libraries (Shendure and Ji 2008). First, the sample nucleic acids have to be sheared in order to reach a size compatible with sequencing (typically <500 bp). Second, DNA adapters containing unique sequences are attached at both ends of the sheared DNA molecules. These adapters subsequently allow the DNA fragments to be singled out, either on beads or on a slide ("flowcell"), enabling them to be sequenced in parallel.

As most of the imaging systems in NGS instruments were not designed to detect single fluorescent events, amplification of templates is necessary (Metzker 2010). There are many ways by which the clonally clustered amplicons that are to serve as sequencing features can be achieved. These methods include in situ polonies, emulsion PCR (emPCR) or bridge PCR (Table 1). As PCR amplification steps are often associated with PCR bias, the possibilities of introducing base sequence errors or favoring certain sequences over others exist. These biases can have a significant impact on the relative frequency and abundance of the various DNA fragments that existed before amplification. These potential biasses can be avoided only if a single DNA molecule is used directly for sequencing without undergoing PCR amplification steps. This category of technologies (referred to as 3rd generation NGS technology) hold great promise in terms of offering rapid and cost-effective sequencing of gene/genome from a single DNA molecule (Schadt et al. 2010), and includes sequencing platforms available commercially from Helicos Biosciences (HeliScope) (Braslavsky et al. 2003) and Pacific Biosciences (PacBio) (Ansorge 2009).

Second and third generation platforms have shown great diversity in terms of sequencing biochemistry, configuration, as well as in generation of array (Table 1). For more comprehensive information on technical aspects, sample preparation methods and data analysis of the sequence reads generated from NGS technologies, interested readers are directed to some excellent reviews (Ansorge 2009; Mardis 2008a, b, 2009, 2010). One plausible reason for the co-existence of a wide variety of NGS platforms on the market is the distinctive chemistries and functions they possess, and having obvious advantages for particular applications (Table 1). Unlike conventional capillary-based sequencing, which reads only 96 samples at a time (Margulies et al. 2005), NGS technologies rely heavily on automation and high-throughput technologies that are capable of processing millions of sequence reads in parallel fashion in very short time-frames without significant loss of accuracy (Metzker 2010). This massive parallel throughput may require only very few (one or two) instrument runs to accomplish a sequencing experiment (Mardis 2008a; Morozova and Marra 2008; Riesenfeld et al. 2004). Additionally, conventional capillary-based sequencing generates sequence reads produced from fragment 'libraries' that have been subjected to vector-based cloning and *Escherichia coli*-based amplification stages that are often associated with cloning biases (Farris and Olson 2007; Mardis 2008a). Owing to absence of vector-based cloning and *E. coli*-based amplification stages, NGS reads are free from cloning-associated biases (Liu 2009; Mardis 2008a).

However, all currently available NGS technologies are associated with inherent weaknesses: they are computationally expensive and produce short read lengths with error rates that prevent assembly software to resolve large structural rearrangements (insertions, deletions, inversions) in re-sequencing and to disambiguate repeat regions in de novo sequencing (Table 1). The implementation of paired-end and mate-paired sequencing into NGS platforms has helped to overcome these problems. In addition to sequence information, these methodologies give information about the

Table 1 A comparison of features, chemistry and performance of first, second and third generation sequencing platforms. *CMOS* Complementary metal-oxide semiconductor, *TH* throughput, *SM* single molecule, *SMRT* single molecule real time

Platform	Chemistry	PCR amplification	Starting DNA	Read length	Reads/run	TH/run	Run time	Disadvantages	Applications	References
Sanger sequencing	Asynchronous with base-specific terminator	Standard PCR	0.5-1 mg	700	Few 1,000 bp	1 Mb	2 h	PCR biases; low degree of parallelism; high cost of sequencing	Gene/genome sequencing	Sanger et al. 1977; Margulies et al. 2005
Roche 454	Sequencing-by-synthesis (Pyrosequencing)	EmPCR	1 μg for shotgun library and 5 μg for pair-end	> 400	1,000,000	0.4-0.6Gb	7-10 h	PCR biases; asynchronous synthesis; homopolymer run; base insertion and deletion errors; emPCR is cumbersome and technically challenging	De novo genome sequencing, RNA-seq, resequencing/targeted re-sequencing	http://www.454.com/; Mardis 2008a; Metzker 2010
Illumina	Polymerase-based sequencing-by-synthesis	Bridge amplification	<1 μg for single or pair-end	75/2 × 100[a]	40,000,000	3-6/200* Gb	3-4 days	PCR biases; low multiplexing capability of samples	De novo genome sequencing, RNA-seq, resequencing/ targeted re-sequencing, metagenomics, ChIP	http://www.illumina.com/technology/sequencing_technology.ilmn; Mardis 2008a; Metzker 2010
SOLiD	Ligation-based sequencing	Em PCR	<2 μg for shotgun library and 5-20 μg for pair-end	35-40	85,000,000	10-20Gb	7 days	EmPCR is cumbersome and technically challenging PCR biases; long run time	Transcript counting, mutation detection, ChIP, RNA-seq etc.	http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing.html; Mardis 2008a; Metzker 2010
HeliScope	Polymerase (asynchronous extension)	SM; no PCR	<2 μg, single end only	25-50	1,000,000,000	28Gb	8 days	Asynchronous synthesis; homopolymer run; high instrument cost; short read lengths; high error rates compared with other reversible terminator chemistries	Resequencing, transcript counting, ChIP, RNA-seq	http://www.helicosbio.com/Products/HelicosregGeneticAnalysis System/HeliScopetrade Sequencer/tabid/87/Default.aspx Metzker 2010
Polonator	Synchronous controlled synthesis	Em PCR	—	26	160,000,000	4.5Gb	4 days	Low read length; emPCR is cumbersome and technically challenging	Bacterial genome, resequencing, SNPs and structural variants detection	Metzker 2010
PacBio	Phospho-linked fluorescent Nucleotides	SMRT	~1.5 μg (ideally 2-3 μg)	1,000-1,200	100,000,000	100Gb/Hr	8 h	High instrument cost; low number of sequence read per run; highest error rates compared with other NGS chemistries	De novo genome sequencing, RNA-seq, resequencing/targeted re-sequencing, metagenomics, SNPs and structural variants detection	http://www.pacificbiosciences.com/applications/overview http://www.pacificbiosciences.com/products/smrt-technology/smrt-sequencing-advantage http://www.pacificbiosciences.com/partner_products/PartekGenomicsSuite_2.7.12.pdf; Travers et al. 2010 Metzker 2010
CMOS non-optical sequencing	Template-directed DNA polymerase synthesis	—[b]	—	—	—	—	—	—	De novo genome sequencing	Rothberg et al. 2011

[a] Considering new Illumina HiSeq 2000 sequencing platform

[b] Not reported in this review

physical distance between the two reads in the reference genome. Paired ends can be obtained from the ends of random, usually small, DNA fragments and the resulting data allow the scaffolding of contigs (contiguous sequences) in the absence of contiguous coverage of intervening sequences (Bentley 2006). In mate-pair sequencing, random DNA fragments are circularized, thereby combining previously distant ends. This DNA is then sheared to generate linear fragments as templates for sequencing (Korbel et al. 2007). The difference between paired-end and mate-paired is typically that the mate-paired method generates a longer insert size compared to paired-end, with insert sizes measuring between 2 and 20 kb. Besides template preparation, new algorithms and computational analysis tools increase the possibility of handling the large amount of NGS data, but often impose additional costs (Liu 2009). Comprehensive lists of relevant software can be found on the SEQanswers website (http://seqanswers.com/). Finally, all the DNA sequencing technologies described here are limited by their requirement for imaging technology, electromagnetic intermediates (either X-rays or light) and specialized nucleotides or other reagents that increase the cost of sequencing still further. For each NGS platform, an accurate costs evaluation has been provided by Glenn (2011). Recently, novel and ground-breaking NGS platforms such as those based on non-optical sequencing technology (Ion Torrent technology) have emerged to overcome these drawbacks and are expected to provide scalable methods for genome sequencing at substantially reduced cost (Rothberg et al. 2011).

De novo sequencing

An important application of NGS is aimed at de novo sequencing of microbial genomes. The total number of completed microbial genome sequences has more than doubled over the past 2 years and, at the time of writing, there were approximately 2,878 publicly listed bacterial and archaeal and 168 eukaryal genome projects in various stages of progress (http://www.genomesonline.org). In addition to new species, multiple strains of the same bacterial species are being sequenced (re-sequencing). The huge amount of genomic sequences from closely related organisms has led to significant advances in microbial phylogeny. The impact of genomic on prokaryotic systematics has been explained in several in-depth reviews (Coenye et al. 2005; Klenk and Göker 2010). From an evolutionary point of view, we would highlight here that the significant expansion of the tree of life provided by whole genome sequencing is opening the possibility of assessing the taxonomic relationships between prokaryotic species based on complete genome sequences. A major step in the progress towards a genome-based classification of microorganisms was the creation of a set of reference genomes that more broadly covers the evolutionary diversity of Bacteria and Archaea (GEBA). This project covers the current lack of completely sequenced genomes for many of the major lineages of prokaryotes and for most type strains (http://jgi.doe.gov/programs/GEBA/; Wu et al. 2009).

The availability of complete genome sequences of closely related organisms presents an opportunity to reconstruct events of genome evolution. Using comparative genomics approaches, mobile genetic elements (MGE) and horizontal gene transfer (HGT) were found to have a key role in prokaryotic genome plasticity, adaptation and speciation. For example, comparative genomics has reconstructed the ancestral gene sets of the LAB and highlighted that the origin of *Lactobacillales* involved extensive loss of ancestral genes (600–1,200 genes) during their transition to life in a nutritionally rich medium, which allowed a reduction in catabolic capacity and increased stress resistance (Makarova et al. 2006). Similarly, comparing close and more distant yeast species has led to the reconstruction of the ancestral genome in *Saccharomyces* sensu stricto complex (Sipiczki 2011). This ancestral progenitor has been subjected to whole-genome duplication, followed by massive sequence loss, divergence, and segmental duplication. In addition, subtelomeric regions have been subjected to further duplications and rearrangements via ectopic exchanges (as reviewed by Liti and Louis 2005).

With its long read lengths and high accuracy, capillary electrophoresis-based sequencing has been the gold standard for de novo genome sequencing projects in past decades. Although, the shotgun Sanger approach has been used more frequently to sequence many cultured microorganisms, the application of NGS in the de novo sequencing and assembling of genomes is becoming widespread. The first NGS-sequenced genomes were from bacteria *Mycoplasma genitalium* (580 kbp), *Streptococcus pneumoniae* (2.1 Mbp) (Margulies et al. 2005), *Pseudomonas syringae* pv. *oryzae* (Reinhardt et al. 2009) and *Pseudomonas syringae* pv. *syringae* (6 Mbp) (Farrer et al. 2009), and from yeast *Saccharomyces cerevisiae* (13.1 Mbp) (Qi et al. 2009), due to the small size of their genomes and reduced repeat regions as compared to higher eukaryotic genomes. Currently, the application of next-generation technologies for de novo sequencing of considerably larger, more complex, and often repeat-rich genomes is still laborious and requires elaborate assembly strategies and sophisticated hardware resources. NGS reads contain a high error rate, which can be overcome with high coverage because the higher number of reads effectively "quenches" errors in single reads and leads to overall high accuracy in the final assembly (Ilie et al. 2011; Yang et al. 2010). Another concern is related to short sequence reads (e.g., 35–50 bases), which need 25- to 30- fold coverage of the genome to capture all the genetic

information. Since the mapping process is based on unique short sequence reads within the genome, increasing the number of reads alone would not be able to overcome this problem because of the presence of repeat regions, which cannot been assembled from reads that are shorter than the lengths of the repeats.

Considering that the Illumina/Solexa platform relies on millions of small reads with a high coverage, whereas the 454 generates larger fragments over 400 nucleotides (see Table 1), the 454 has been generally considered more adapted to genomes containing abundant repeated regions. This situation is being improved by (1) the longer read length using better extension reagents and chemistries; (2) the development of paired-end tag (PET) sequencing approaches also for NGS platforms, in which both ends of a fragment of defined size are sequenced to provide more information about the fragment; and (3) and novel assembly algorithms that can deal with large numbers of short reads. Due to this continuous upgrading, NGS was applied also for eukaryotic de novo genome sequencing. NGS was first used in combination with Sanger sequencing, for sequencing the genomes of plant *Vitis vinifera* (Velasco et al. 2007), the filamentous fungus *Grosmannia clavigera* (Diguistini et al. 2009) and the cucumber *Cucumis sativus* (Huang et al. 2009). The first two eukaryotic genomes to be assembled solely from NGS reads are those of the giant panda (Li et al. 2010), assembled from Solexa reads, and the filamentous fungus *Sordaria macrospora*, assembled from a combination of Solexa and 454 reads (Nowrousian et al. 2010). Until now, no de novo genome sequencing project from yeasts has been completed using NGS technology alone. However, successful sequencing projects in other lineages have demonstrated the feasibility of NGS for accurate, cost-effective, and rapid de novo assembly of eukaryotic genomes and it is only a matter of time until NGS-based approaches are applied to yeasts on a broad scale (Imelfort and Edwards 2009; Turner et al. 2009).

Genotype-phenotype association mapping

The genetic basis of phenotypes have traditionally been identified by genetic selections and/or screens followed by Sanger sequencing. These approaches include the analysis of plasmid-based genomic libraries and gene-disruption libraries. Other traditional methods have been based on whole-genome array, comparative genome hybridization or single-nucleotide polymorphism (SNPs) arrays to capture causal loci. In these experiments, the ability to probe every gene present in an organism is limited by the number of cells that can be screened and the number of targets that can be sequenced. Therefore, in addition to being labor intensive and costly, these traditional methods are not suitable for identifying all the relevant genes produce underling complex traits.

Advances in NGS technologies have lessened the costs of DNA sequencing significantly down to the level that genotype-phenotype association mapping is now feasible also for complex microbial phenotypes. NGS platforms can be applied at three levels to elucidate genotype-phenotype relationships (Fig. 3). The first is the identification of individual genetic variation within a population for which a reference genome is available (re-sequencing). Sequence reads that are mapped to a reference genome were used to detect SNPs, small insertions or deletions (indels), and large-scale structural variations, such as copy number variations (CNV), thereby improving our knowledge on evolutive mechanisms shaping genome and phenotype diversity within a species (Nowrousian et al. 2010). The second and third levels are the adoptions of NGS technologies for high-throughput studies that were previously performed mostly by hybridization-based methods such as microarrays. In this context, it is worth mentioning the use of NGS for transcriptomics (RNA-seq) or the genome-wide analysis of DNA/protein interactions (ChIP-seq) (Fig. 3).

Re-sequencing: getting to individual variants

One genome sequence is inadequate to describe the complexity of species, genera and their inter-relationships. Multiple genome sequences are needed to describe the pan-genome, which represents approximately the genetic variability of a microbial species. Re-sequencing implies sequencing of genomes from species for which a reference genome is already available in public databases, and is currently one of the major areas of application of NGS. NGS-produced short sequence reads can be aligned efficiently to reference genomes. Consequently, mapped sequence reads are very useful in order to detect and evaluate individual genetic variations with a high level of confidence.

As the 454 platform has a read length appropriate for sequence assembly, and a library construction approach that avoids cloning bias, several studies using this method have been published. The most ambitious one involves sequencing over 1,000 individual human genomes in order to map human genetic variation at a fine scale and to support genome-wide phenotypic and disease association studies (http://www.1000genomes.org/). For yeasts (Carter 2009), filamentous fungi (e.g., http://www.jgi.doe.gov/genome-projects/; http://www.broadinstitute.org/science/projects/fungal-genome-initiative) and several LAB species, there is already more than one genome available or sequencing is still in progress (Table 2). Liti et al. (2009) compared the whole genomes of several isolates of *S. cerevisiae* and its closest relative, *Saccharomyces paradoxus*, to find

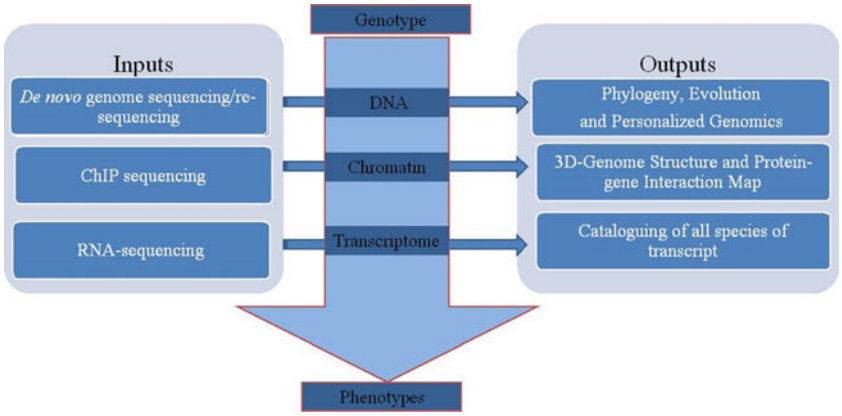

Fig. 3 Three-levels of contributions of NGS to genotype-phenotype association mapping

a high correlation between phenotypic variation and global genome-wide phylogenetic relationships. This study demonstrated that *S. cerevisiae* wine strains are usually polyclonic and differ significantly in enological performance and genotype. The extent of genetic differences ranges from single-nucleotide substitutions to whole-genome duplication (Sipiczki 2011).

Information collected by resequencing is used to identify individual genetic variations responsible for mutated phenotypes or functional properties involved in food quality, safety and shelf life. Whole genome mutational profiling can be applied, for instance, to mapping the genetic differences between toxin producers and non-toxinogenic variants of food-borne disease pathogens, or between wild-type strains and strains optimized for food and industrial starter applications. Using the Illumina/Solexa platform, Studholme et al. (2010) highlighted virulence factors genes in *Xanthomonas campestris*. Smith et al. (2008) resequenced a mutant strain of *Pichia stipitis* with three NGS technologies: 454 Life Sciences (Roche), Illumina (formerly Solexa), and SOLiD (Applied Biosystems). This yeast has been evolved adaptively over a period of 7 years in order to obtain a more efficient xylose fermenting mutant. With 10- to 15-fold redundant genome coverage, all three sequencing technologies identified the same 17-point mutations, 10 of which resulted in an amino acid change within protein-coding genes.

NGS-driven mutational profiling can also address open questions such as mutation rates in evolving populations and their correlation to adaptation. This topic is of particular interest, considering those microorganisms living in foods are constantly exposed to fluctuating environmental conditions, and many of these conditions are potentially detrimental and stressful. DNA instability was demonstrated to sharply elevate spontaneous mutation rates in a *Lb. plantarum* strain in order to transiently enhance its ability to adapt to environmental changes (Machielsen et al. 2010). Whole-genome sequencing efforts were performed for tracking the regions of the genome that contain genetic variants that affect yeast phenotypes evolved under controlled laboratory conditions (Araya et al. 2010; Lynch et al. 2008). Finally, NGS technologies have been combined efficiently with traditional genetic techniques for mapping complex quantitative traits in *Drosophila simulans* (Andolfatto et al. 2011) and yeasts (Birkeland et al. 2010).

RNA-seq

Analysis of gene expression has been a primary tool to study cellular mechanisms. For profiling mRNA populations, microarrays have dominated for more than a decade, providing gene expression information at relatively low cost and increased throughput. Although microarrays are now used widely for monitoring transcript expression, hybridization-based technologies have several important limitations. First, low-abundance transcripts cannot be measured accurately. Second, discovery of novel transcripts is limited. Third, direct comparison of transcripts within an individual sample is inaccurate because hybridization kinetics for individual mRNAs are sequence dependent, necessitating ratiometric comparison between paired samples (Croucher and Thomson 2010).

RNA-sequencing transcriptomics (RNA-seq) is an emerging NGS platforms-based approach for comprehensive identification and quantification of transcripts independent of any annotated sequence feature (Pachter 2011). This methodology allows us to (1) catalogue and improve the annotation of all species of transcripts, including protein encoding mRNAs as well as (small) non-coding regulatory, structural or catalytic RNAs; (2) determine the transcriptional structure of genes in terms of start sites and 3′ ends, splicing patterns and other maturation processes; and (3) quantify the changing expression level of each transcript under different conditions over the full dynamic range of cellular RNA expression (Zhou et al. 2010).

In microbiology, an integrated RNA-seq and tiling array approach was first applied to characterize transcripts of *Mycoplasma pneumoniae* (Züell et al. 2009). Subsequently, RNA-seq was applied to map transcription start sites in *Helicobacter pylori* (Sharma et al. 2010) and to analyze the transcriptome in the typhoid bacillus *Salmonella typhi*

Table 2 Overview of the major whole genome sequencing projects generated using NGS technologies for some food microorganisms

Species (strain)	Origin	GenBank accession no.	Size (Mb)	Sequencing platform	Biological role	Reference
LAB						
Lactobacillus casei (BD-II)	Homemade Koumiss, China	CP002618 CP002619	3.1	Roche 454 and Solexa (Illumina)	Probiotics	Ai et al. 2011
Lactobacillus delbrueckii subsp. *bulgaricus* (ND02)	Fermented Yak milk, China.	CP002341 CP002342	2.1	Roche 454 and Solexa (Illumina)	Dairy starter for yoghurt production	Sun et al. 2011
Lactobacillus plantarum (ST-III)	Kimichi	CP002222	3.3	Roche 454 and Solexa (Illumina)	Probiotic supplement for dairy and food products	Wang et al. 2011
Lactobacillus casei (LC2W)	Traditional dairy product, Mongolia	CP002616 CP002617	3.0	Roche 454 and Solexa (Illumina)	Probiotics	Chen et al. 2011
Lactobacillus buchneri (NRRL B-30929)	Ethanol plant	CP002653 CP002654 CP002655	2.5	Whole-genome shotgun strategy using Roche 454 Titanium pyrosequencing	Prevention of silage spoilage by yeast and molds and anti-bacterial property	Lui et al. 2011
Lactobacillus ruminis (ATCC 27782)	Bovine and human intestine	XXYYZZ123[a]	2.1	Roche 454 and Solexa (Illumina)	Probiotics	Forde et al. 2011
Lactobacillus farciminis (KCTC 3681)	Kimichi, Korea	AEOT00000000 AEOT01000000	2.5	Whole-genome shotgun strategy using Roche 454 Titanium pyrosequencing	Kimichi fermentation	Nam et al. 2011a
Lactobacillus coryniformis subsp. *coryniformis* (KCTC 3167)	Kimichi, Korea	AELK00000000 AELK01000000	3.0	Whole-genome shotgun strategy using Roche 454 Titanium pyrosequencing	Kimichi fermentation	Nam et al. 2011b
Weissella koreensis (KACC 15510)	Kimichi, Korea	CP002899	1.4	Roche 454 GS FLX Titanium pyrosequencing	Kimichi fermentation	Lee et al. 2011
Other bacteria						
Salmonella enterica serovar *typhi* (CT18 and Ty2)	Contaminated food and water	AL513382 AE014613 AE006468 CP000026	—[b]	Roche 454 and Solexa (Illumina)	Food spoilage	Holt et al. 2008
Yeasts						
Saccharomyces cerevisiae: S288C and RMII-Ia	—	—	4.4/5.2	Solexa (Illumina)	Bakery and beverages	Qi et al. 2009
Lalvin QA23	Lallemand Inc.	ADVV00000000	11.6	Roche 454 GS FLX Titanium	Wine	Borneman et al. 2011
AWRI796	Maurivin	ADVS00000000	11.6	Roche 454 GS FLX Titanium	Wine	
Vin13	Anchor Bio-Technologies	ADXC00000000	11.5	Roche 454 GS FLX Titanium	Wine	
FostersO	Fosters Group Ltd.	AEEZ00000000	11.4	Roche 454 GS FLX Titanium	Brewing	
FostersB	Fosters Group Ltd.	AEHH00000000	11.5	Roche 454 GS FLX Titanium	Brewing	
VL3	Laffort	AEJS00000000	11.4	Roche 454 GS FLX Titanium	Wine	
13 strains involved in fermentation and 3 in baking	—	—	—	ABI 3730 and Solexa (Illumina)	Bakery and beverages	Liti et al. 2009
Saccharomyces paradoxus 35 strains from different geographical regions	—	—	—	ABI 3730 and Solexa (Illumina)	Bakery and beverages	Liti et al. 2009

[a] Finished genome sequence of ATCC 27782

[b] Not reported in this review

(Perkins et al. 2009) and archeon *Sulfolobus solfataricus* (Wurtzel et al. 2010). These studies revealed a high transcriptome complexity and redundancy in both Bacteria and Archaea, where riboswitch elements and chromosomally encoded cis-antisense transcripts are common forms of regulation. Moreover, different promoters appear to be driving expression of the same genes under different conditions, leading to the division of genes into "suboperons" (Croucher and Thomson 2010; Sorek and Cossart 2010; Van Vliet 2010).

Similarly, RNA-seq has led to the discovery of the high complexity of the eukaryotic transcriptional landscape (Nagalakshmi et al. 2008; Wilhelm et al. 2008). The existence of various classes of non-coding RNAs originating from pervasive transcription of eukaryotic genomes suggested new levels of regulation of gene expression and genome plasticity (Berretta and Morillon 2009). Recently, tiling arrays and sequencing were employed in the discovery of widespread bidirectional promoters in *S. cerevisiae* (Neil et al. 2009; Xu et al. 2009). These studies revealed that transcription of the two main classes of non-coding RNAs in yeast initiates predominantly from the promoter regions of protein-coding genes, suggesting that bidirectionality is an intrinsic feature of eukaryotic promoters.

ChIP-seq

NGS technologies offer the potential to substantially accelerate the study of heritable gene regulation that does not involve the DNA sequence itself but its modifications and higher-order structures, including posttranslational modifications of histones, interaction between transcription factors and their direct targets, nucleosome positioning on a genome-wide scale and characterization of DNA methylation patterns (Chung et al. 2010; Fous et al. 2010; Pareek et al. 2011). In particular, architectural proteins drive compaction and organization of genomic material into chromatin, modulating DNA accessibility and consequently numerous cellular processes, including transcription, replication and repair. Understanding how various factors regulate transcription elongation in living cells has been aided greatly by chromatin immunoprecipitation (ChIP) studies, which can provide spatial and temporal resolution of protein-DNA binding events. The coupling of ChIP and high-throughput sequencing technologies (ChIP-seq) has significantly facilitated the whole-genome mapping of DNA-binding protein sites.

Although, ChIP-seq gained rapid support in eukaryotic systems, it remained underused in the mapping of bacterial transcriptional regulator-binding sites. To date, very little is known about DNA folding in Bacteria and Archaea. Even though no relevant literature exists yet for food-related microorganisms, it is expected that ChIP-seq will soon be applied to this field. In comparison to its array-based predecessor, i.e., ChIP-chip technology, ChIP-seq offers higher resolution, lower noise and better coverage. With the ever-decreasing cost of sequencing, ChIP-seq has become an indispensable tool for studying gene regulation and epigenetic mechanisms (Schmid and Bucher 2007).

Community profiling

The basic goal of community profiling in relation to food microbiology is to understand the relationships among community composition, microbial functions, and their impact on food sensorial properties. Exploring microbial diversity is critical in order to evaluate each introduced or indigenous species in food processing, as well as to ascertain the role of microorganisms as food pathogens, in food spoilage, or as potential starter cultures. Fermented food in particular comprises complex and diverse communities of microorganisms, which are vital in catalyzing desired biochemical transformations and in maintenance of food quality. Since shifts in these populations determine changes in the final composition, in past decades food microbial diversity was explored by both cultivation-dependent and independent methods.

Generally, it is hard to determine entire microbial populations in food using culture-based techniques owing to their resistance to culture under standard laboratory conditions (Riesenfeld et al. 2004). Even traditional culture-independent approaches, including denaturing gradient gel electrophoresis (DGGE), temperature gradient gel electrophoresis (TGGE), terminal restriction fragment length polymorphism (T-RFLP), fluorescence in situ hybridization (FISH), single stranded conformation polymorphisms (SSCP), quantitative PCR (qPCR), and DNA microarray, have inherent pitfalls (as reviewed by Amann et al. 1995; Giraffa and Neviani 2001; Hugenholtz et al. 1998; Jany and Barbier 2008; Wintzingerode et al. 1997).

Alternatively, sequencing of PCR-amplified libraries targeted to phylogenetic barcodes (like 16S rRNA genes) is a culture-free approach that basically relies on (1) PCR amplification of 16S rRNA from DNA bulk-extracted from an environmental or food sample, using universal primers for Eukarya, Bacteria and Archaea; (2) construction of DNA clone libraries; and (3) Sanger sequencing. The recent introduction of NGS techniques eliminates the time-consuming steps of in vivo cloning, colony picking, and capillary electrophoresis, allowing the study of complex microbial communities from global soil (Fierer et al. 2007; Leininger et al. 2006), deep mines (Edwards et al. 2006), ocean (Angly et al. 2006; Sogin et al. 2006), human microbiome (Qin et al. 2010), and fermented food (Table 3). To date, 454 pyrosequencing has been the platform most used for 16S amplicon (16S pyrotags) surveys (Tringe and Hugenholtz 2008),

Table 3 Application of NGS technologies to food microbial community surveys

Product	Product description	Biological role/ application	Origin/country	Platform	Microbes identified	Reference
Nukaoko	Naturally fermented rice bran mash	Pickling vegetables	Japan	16S-Pyrosequencing	*Lactobacillus namurensis* and *Lactobacillus acetotolerans*	Sakamoto et al. 2011
Jeotgal 'fermented seafood'	Jeotgal is made of fish or shellfish, such as shrimp, oysters, fish roes, and fish tripes with lots of salt	Consumed as side food or used as ingredient to improve the flavor of other food such as Kimichi	Korea	Multiplex barcoded pyrosequencing	Archaea and Bacteria	Roh et al. 2010
Soft drinks	Beverages	Food spoilage	United Kingdom	26S rDNA Sequencing	*Candida davenportii* sp. nov.	Stratford et al. 2002
Danish raw milk cheese	Dairy product	Food product	Denmark	16S rRNA gene sequencing	*Lactococcus, Lactobacillus* and *Streptococcus* spp.	Masoud et al. 2011
Ben-saalga	Fermented food made by cooking pearl millet slurries	Consumed by adults and infants as complementary food to breast feeding	Burkina Faso, West Africa	Tagged 16S rRNA gene	*Firmicutes* (*Lactobacillus*, *Pediococcus*, *Leuconostoc*, and *Weissella* genera), *Proteobacteria*, *Actinobacteria*, and *Bacteroidetes* spp.	Humblot and Guyot 2009
Meju	Traditional Korean fermented soybean bricks	For preparation of Korean soy sauce (Ganjang) and soybean paste (Doenjang)	Korea	Barcoded 16S rRNA gene pyrosequencing	*Firmicutes* (*Bacillus* and Lactic acid bacteria of *Lactococcus Leuconostoc*, and *Enterococcus* genera), *Proteobacteria*, and *Bacteroidetes* spp.	Kim et al. 2011
Fermented shrimp, kimchi, and sauerkraut	Fermented Food	Food products	Asia and Europe (Fermented Shrimp), China, Korea (Kimchi), Eastern Europe and North America (Sauerkraut)	Barcoded 16S rRNA gene pyrosequencing	*Pseudomonas, Cupriavidus* and *Propionibacterium* spp.	Park et al. 2011
Fen liquor	Fermented light-fragranced distilled liquor	Beverages	China	16S rRNA gene sequencing and ITS region pyrosequencing	*Lactobacillus acetolerans, S. cerevisiae, Issatchenkia orientalis* and of *Saccharomycopsidaceae* family	Li et al. 2011
Beef meat	Meat product	Food product	Italy	Pyrosequencing	*Brochothrix Thermosphacta, Pseudomonas spp., Carnobacterium divergens,* and LAB	Ercolini et al. 2011
Kefir	Fermented milk beverages	Health benefits	Tibet and China Ireland	Pyrosequencing	*Lactobacillus kefiranofaciens* ZW3 *Lactobacillus* spp.	Wang et al. 2011; Dobson et al. 2011

although the Illumina/Solexa sequencing platform has been applied recently for sequencing paired-end libraries of 16S amplicons (iTags) from a microbial community (Degnan and Ochman 2011). However, NGS-assisted and clone-and-sequencing 16S surveys are not mutually exclusive. The clone-and-sequence approach provides a valuable reference base or 'gold standard' for the high-throughput NGS technologies in order to identify novel and phylogenetically accurate lineages in the life tree (Tringe and Hugenholtz 2008).

Despite their huge potential in microbial community profiling, NGS techniques encounter some pitfalls. Universal primers have demonstrated a bias in the detection of high GC content bacteria (Farris and Olson 2007). Being ubiquitous in nature, 16S rRNA genes may not have enough variability in order to assign taxonomy to genus and species level (Juste et al. 2008). Furthermore, the presence of multiple copies of 16S rRNA genes may also negatively affect microbial identification and characterization (Riesenfeld et al. 2004). Another disadvantage of pyrosequencing-assisted microbial profiling is the short sequence reads ranging from 100 bp on an average for the GS 20 (Genome Sequencer 20 DNA Sequencing System, 454 Life Sciences, Bradford, CT) to 200–300 bp for the newer GS FLX (Genome Sequencer FLX System, 454 Life Sciences). To make pyrosequencing technology suitable for 16S survey, Sogin et al. (2006) first PCR-amplified the short 16S rRNA V6 variable region from eight distinct environments using universal primers, then processed them separately in a single 454 run, producing 118,000 16S pyrotags. A follow-up study, also using GS20 technology, generated more than 900,000 bacterial and archaeal 16S pyrotags (Huber et al. 2007). Subsequently, NGS-assisted community profiling has been employed successfully to different food ecosystems (Table 3), allowing the comparative analysis of population dynamics in relation to altered environmental and processing parameters. For example 16S pyrosequencing and quantitative RT-PCR highlighted the presence of some uncultured mesophilic Crenarchaeota phlyotypes other than *Firmicutes* and *Proteobacteria* in some fermented seafood (Roh et al. 2010). To increase the number of samples that can be analyzed at a time, sample-specific key sequences of few nucleotides, called 'barcodes' or 'tags', were implemented recently in a pyrosequencing platform to evaluate multiple related samples (e.g., spatial and temporal series) in parallel (Binladen et al. 2007; Hamady et al. 2008; Parameswaran et al. 2007). This multiplex barcoded pyrosequencing was used successfully to simultaneously profile a large number of seafood samples (Roh et al. 2010). Another study revealed that different processing conditions have huge impacts on microbial profile during the course of fermentation of pearl millet slurries (Humblot and Guyot 2009).

Metagenomics: beyond community profiling

Metagenomics (also referred to as environmental and community genomics) is a term first introduced by Handelsman et al. (1998) to describe sequence-based analysis of the collective microbial genomes of uncultured microbes contained in an environmental sample in order to understand their diversity, function, and cooperation/behavior in an ecosystem and how they evolved during the course of evolution (Mitra et al. 2011; Ghosh et al. 2011). In contrast to community profiling, metagenomics efforts aim at sequencing all genes present in a habitat rather than just 16S, thereby providing clues regarding the functionalities of a community rather than just its phylogenetic composition.

Traditionally, metagenomic approaches enabled the cloning of total microbial DNA directly into a large-insert library, without prior amplification of particular genes, to avoid PCR-associated bias. Alternatively, small-insert libraries and Sanger sequencing-derived shotgun sequencing were successfully applied to metagenomic studies (Tyson et al. 2004). Constructing food-based libraries involves the same methods as the cloning of genomic DNA of individual microorganisms; that is, fragmentation of food DNA by restriction-enzyme digestion or mechanical shearing, insertion of DNA fragments into an appropriate vector system, and transformation of the recombinant vectors into a suitable host. Although the generation of food libraries is conceptually simple, the enormous size of food metagenome and the large number of clones that are to be screened by functional and sequence-based approaches make this task daunting. In 2006, the first sequences of two different soil samples generated using pyrosequencing were published (Edwards et al. 2006). Since then, several metagenomic outcomes have relied on 454 pyrosequencing to analyze environmental samples, increasing throughput and avoiding Sanger sequencing-associated cloning steps.

Compared to environmental microbiology, few studies to date have been carried out to identify the pathways or enzymes responsible for significant food processes. Functional analysis has elucidated metabolic pathways (e.g., antibiotic and vitamin biosynthesis) (Entcheva et al. 2001; Eschenfeldt et al. 2001; Rondon et al. 2000), and identified novel antibiotics, degradative enzymes, and bioactive compounds (Henne et al. 2000), and also led to the discovery of biocatalysts (e.g., lipolytic genes and polysaccharide degrading/modifying enzymes) (Rondon et al. 2000). More recently, Jung et al. (2011) used pyrosequencing to analyze metabolic potential of the fermenting microbial community from kimchi—a traditional Korean food produced by fermentation of vegetables such as Chinese cabbage and radish. Similarly, the Illumina/Solexa sequencing technique was employed to assemble a genome map of the human gut

microbiome (Qin et al. 2010). It can be predicted that NGS-assisted food metagenomics will have great relevance to the identification of the genes responsible for characteristic properties and functionalities such as probiotic activity, flavor formation and taste development.

Finally, complementary approaches to metagenomics, e.g., meta-metabolomics (the study of all the naturally occurring molecules in a biological sample) and metatranscriptomic (deep sequence surveys of expressed genes from overwhelmingly complex metagenomes), contribute significantly to clarify microbial behaviors in food ecosystems. An in-depth discussion of the technical and methodological aspects of these meta-"omics" approaches is beyond the scope of this review and can be found in the literature (Morales and Holben 2011; Raes and Bork 2008). Integration among metagenomic, transcript and metabolic information is critical to investigate relationships among metabolite production, metabolic potential, and ecological composition of the food microbial community, connecting this "omic" information in the context of space and time (Raes and Bork 2008). The analysis of microorganisms in different environments and the quantification of metabolic fluxes can be of crucial importance to better understand microbial roles in food.

Practical applications of NGS to food microbiology

Food safety and process optimization

A principal challenge for the food industry is to produce safe foods with the desired functionalities using minimal processing technologies. Whole genome characterization of undesirable microorganisms in food stuff is the first step towards prevention of food spoilage. NGS could have important applications in reducing the risks of food-borne diseases due to the huge improvements in the rate at which the whole genome of food microorganisms from different species and from strains belonging to the same species can be generated. Bacterial whole-genome sequencing was applied to typing *Lysteria monocytogenes* strains associated to meat products (Gilmour et al. 2010), methicillin-resistant *Staphylococcus aureus* (Harris et al. 2010), non-typhoidal *Salmonella* spp. (Andrews-Polymenis et al. 2009) and *E. coli* strains (Brzuszkiewicz et al. 2006). Moreover, the comparison of whole genome data collected by NGS studies has elucidated the role of SNPs in different pathogenic phenotypes and the evolutive mechanisms in emerging pathogenicity (He et al. 2010). Recent outbreaks in peanut butter and peanut paste products associated with *Salmonella* across the United States were monitored by pyrosequencing (Liu 2009). In this regard, inclusion of high-pressure processing steps in peanut product manufacturing reduced the population of *Salmonella* as confirmed by pyrosequencing (Liu 2009). Pyrosequencing of the variable V3 region of 16S rRNA gene from fermented foods in Nigeria (such as Kuna-zaki and Ogi) revealed the presence of phylotypes corresponding to potential pathogenic microorganisms such as *Clostridium perfringens* and *Bacillus cereus*, and consequently highlighted the need for more controlled and optimized fermentation conditions to ensure the good health and well being of consumers (Oguntoyinbo et al. 2011).

Environmental and processing parameters such as the quality of raw material, selection of bacterial combinations to be used as starter cultures, and controlled fermentation conditions, shape population structures and dynamics in food. These fluctuations should be considered properly during fermentation processes and should be monitored by fast and cost-effective 16S tagged NGS approaches. Accordingly, 16S tagged pyrosequencing in conjunction with qPCR showed that bacteria grew during fermentation of the traditional Chinese fragranced-liquor called fen liquor, while fungi remained stable (Li et al. 2011). This latter study suggested that simultaneous quantification of bacteria and fungi during food fermentation processes can be used for tracking the corresponding variations in biochemical composition. Other NGS-assisted studies have linked microbial composition to environmental and process parameters. For example, meat quality is strictly dependent on storage: the complex shift in microbiota and secreted metabolites (butanoic acid and acetoin) under different conditions was checked by pyrosequencing and gas chromatography/mass spectrometry, respectively (Ercolini et al. 2011). With the increasingly widespread use of NGS, it is reasonable that data generated by NGS and other "omics" techniques (i.e., transcriptomic and metabolomics) will be integrated by mathematical algorithms into a system model at the species and "meta"-species levels, so that environmental and processing parameters will be predictive of species composition in food (Fig. 1).

NGS-assisted starter optimization

Selection and dominance of a starter culture on indigenous population in fermented food can speed up fermentation significantly and increase sensorial properties. Phenotypic investigations compare the metabolic behaviors of different strains to select the most promising strains for further targeted starter optimization (Wittmann and Hinzle 2002). The engineering of microbial cells possessing desired functions for industrial and food production can be greatly improved by massive sequencing coupled with computational techniques to identify genes and genomes related to relevant phenotypes. A rational and whole genome-assisted choice of starter cultures can have a big impact on food safety, quality and recently also on health benefits, for example in

the case of probiotics, which could be rationally designed and developed so as to maintain the correct balance of the microbial community and to ensure good health and well being in humans (O'flaherty and Klaenhammer 2011; Van Hylckama Vlieg et al. 2011). Direct pyrosequencing of oral metagenomes revealed the presence of certain microbes playing a role against cariogenic microbes and suggested the use of these microbes in the formulation of probiotics to prevent dental caries and promote oral health (Rademaker et al. 2006; Belda-Ferre et al. 2011). Moreover, pyrosequencing was used to unravel the genetic basis of improvement in colitis inflammation upon consumption of fermented milk product containing *Bifidobacterium animalis* subsp. *lactis* (Veiga et al. 2010).

Another field in which NGS-assisted starter optimization is maturing is the study of wine yeast. An annotated genome sequence for *S. cerevisiae* is available, which provides a framework for genome-scale metabolic network reconstruction (Borneman et al. 2007). Given the wealth of experimental and computational data available for *S. cerevisiae*, several studies have begun to integrate biological and computational data sources to provide a holistic view of yeast cellular processes (Borneman et al. 2011; Herrgård et al. 2010). Such reconstructions offer biochemical models describing the formation and depletion of each metabolite, and provide simulations of how the metabolic network operates at different conditions on the basis of mass-balance boundary conditions. Thanks to these stoichiometric models, relationships between gene functions can be predicted. For this purpose, whole genome Illumina/Solexa sequencing was used to identify SNPs between strain S288c and an evolved strain obtained after rounds of directed evolution (Otero et al. 2010). This genome-assisted selection can be applied to wine strains in order to understand how differences in fermentation behavior and wine flavor are related to genomic and transcriptional profile differences and, furthermore, how these characteristics can be modulated logically to tailor wine composition.

Future prospects and conclusions

Raes and Bork (2008) wrote that 'to be successful, however, any systems-biology study requires data on three important aspects of the system: the 'parts list'; the connectivity between the parts; and the placement of connectivity in the context of time and space'. This holistic knowledge can be translated to food ecosystems to predict the behaviour of microbial cells in silico. By allowing DNA/RNA to be assayed more rapidly than previously possible, NGS technologies promise a deeper understanding of microbial genomes and biology by providing a "part list". Meanwhile NGS technologies have improved our ability to routinely profile food microbial ecosystems and to make genotype–phenotype correlations, identifying the genetic basis of complex phenotypes, engineering new phenotypes, and combining beneficial phenotypes in industrial hosts. Hence, high-throughput sequencing can play a key role in whole genome-assisted optimizing of food starter cultures. It is reasonable that future challenges will be aimed at achieving connectivity between data generated by NGS and other "omics" techniques in the context of time and space. This integration will provide comprehensive genetic maps of important food traits, as well as predictive models of the contribution of individual microorganisms in the development of food quality and safety.

References

Ai L, Chen C, Zhou F, Wang L, Zhang H, Chen W, Guo B (2011) Complete genome sequence of the probiotic strain *Lactobacillus casei* BD-II. J Bacteriol 192:3160–3161

Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169

Andolfatto P, Davison D, Erezyilmaz D, Hu TT, Mast J, Sunayama-Morita T, Stern DL (2011) Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res 21:610–617

Andrews-Polymenis HL, Santiviago CL, McClelland M (2009) Novel genetic tools for studying food borne Salmonella. Curr Opin Biotechnol 20:149–157

Angly FE, Felts B, Breitbarts M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R, Rayhawk S, Suttle CA, Rohwer F (2006) The marine viromes of four oceanic regions. PLoS Biol 4:2121–2131

Ansorge WJ (2009) Next-generation DNA sequencing techniques. Nat Biotechnol 25:195–203

Araya CL, Payen C, Dunham MJ, Fields S (2010) Whole-genome sequencing of a laboratory-evolved yeast strain. BMC Genomics 11:1–10

Belda-Ferre P, Alcaraz LD, Cabrera-Rubio R, Romero H, Simon-Soro A, Pignatelli M, Mira A (2011) The oral metagenome in health and disease. ISME J 6:46–56. doi:10.1038/ismej.2011.85

Bentley DR (2006) Whole-genome re-sequencing. Curr Opin Genet Dev 16:545–552, Epub 2006 Oct 18

Berretta J, Morillon A (2009) Pervasive transcription constitutes a new level of eukaryotic genome regulation. EMBO 10:973–982

Binladen J, Gilbert MT, Bollback JP, Panitz F, Bendixen C, Nielsen R, Willerslev E (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS One 2:1–10

Birkeland R, Jin N, Ozdemir AC, Lyons RH Jr, Weisman LS, Wilson TE (2010) Discovery of mutations in *Saccharomyces cerevisiae* by pooled linkage analysis and whole-genome sequencing. Genetics 186:1127–1237

Boguski MS, Arnaout R, Hill C (2009) Customized care 2020: how medical sequencing and network biology will enable personalized medicine. f-1000 Biol Rep 1: 73

Borneman AR, Chambers PJ, Pretorius IS (2007) Yeast system biology: modelling the winemaker's art. Trends Biotechnol 25:349–355

Borneman AR, Desany BA, Riches D, Affourtit JP, Forgan AH, Pretorius IS, Egholm M, Chambers PJ (2011) Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. PLoS Genet 7:1–10

Braslavsky I, Hebert B, Kartalov E, Quake SR (2003) Sequence information can be obtained from single DNA molecules. Proc Natl Acad Sci USA 100:3960–3964

Brzuszkiewicz E, Bruggemann H, Liesegang H (2006) How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. Proc Natl Acad Sci USA 103:12879–12884

Carter D (2009) *Saccharomyces* genome resequencing project, user manual. http://www.sanger.ac.uk/Teams/Team118//sgrp/. Accessed 4 September 2009 (Original publication: 2008)

Chen C, Ai L, Zhou F, Wang L, Zhang H, Chen W, Guo B (2011) Complete genome sequence of the probiotic bacterium *Lactobacillus casei* LC2W. J Bacteriol 193:3419–3420

Chung CAB, Boyd VL, McKernan KJ, Fu Y, Monighetti C, Peckham HE, Barker M (2010) Whole methylome analysis by ultradeep sequencing using two-base encoding. PLoS One 5:1–8

Coenye T, Gevers D, Van de Peer Y, Vandamme P, Swings J (2005) Towards a prokaryotic genomic taxonomy. FEMS Microbiol Rev 29:147–167

Croucher NJ, Thomson NR (2010) Studying bacterial transcriptomes using RNA-seq. Curr Opin Microbiol 13:619–624

Degnan PH, Ochman H (2011) Illumina-based analysis of microbial community diversity. ISME J. doi:10.1038/ismej.2011.74

Diguistini S, Liao N, Platt D, Robertson G, Seidel M, Chan SK, Docking TR, Birol I, Holt RA, Hirst M, Mardis E, Marra MA, Hamelin RC, Bohlmann J, Breuil C, Jones SJM (2009) De novo genome sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data. Genome Biol 10:R94.1–R94.12

Dobson A, O'Sullivan O, Cotter PD, Ross P, Hill C (2011) High-throughput sequence-based analysis of the bacterial composition of kefir and an associated kefir grain. FEMS Microbiol Lett 320:56–62

Edwards RA, Rodrizues-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, Saar MO, Alexanser S, Alexander EC Jr, Rohwer F (2006) Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genomics 7:1–13

Entcheva P, Liebl W, Johann A, Hartsch T, Streit WR (2001) Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. Appl Environ Microbiol 67:89–99

Ercolini D, Ferrocino I, Nasi A, Ndagijimana M, Verrocchi P, La Storia A, Laghi L, Mauriello G, Villiani F (2011) Microbial metabolites and bacterial diversity in beef stored in different packaging conditions monitored by pyrosequencing, PCR-DGGE, SPME-GC/MS and HNMR. Appl Environ Microbiol 77:7372–7381. doi:10.1128/AEM.05521-11

Eschenfeldt WH, Stols L, Rosenbaum H, Khambatta ZS, Quaite-Randall E, Wu S, Kilgore DC, Trent JD, Donnelly MI (2001) DNA from uncultured organisms as a source of 2,5-diketo-D-gluconic acid reductases. Appl Environ Microbiol 67:4206–4214

Farrer RA, Kemen E, Jones JDG, Studholme DJ (2009) De novo assembly of the *Pseudomonas syringae* pv. *Syringae* B728a genome using Illumina/Solexa short sequence reads. FEMS Microbiol Lett 291:103–111

Farris MH, Olson JB (2007) Detection of Actinobacteria cultivated from environmental samples reveals bias in universal primers. Lett Appl Microbiol 45:376–381

Fierer N, Breitbart M, Nultons J (2007) Metagenomic and small-subunit rRNA analyses of the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Appl Environ Microbiol 73:7059–7066

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–512

Forde BM, Neville BA, O'Donnell MM, Riboulet-Bisson E, Claesson MJ, Coghlan A, Ross RP, O'Tolle PW (2011) Genome sequences and comparative genomics of two *Lactobacillus ruminis* strains from the bovine and human intestinal tracts. Microb Cell Factories 10:1–15

Fous SD, Nagarajan RP, Costello JF (2010) Genome-scale DNA methylation analysis. Epigenetics 2:105–117

Ghosh TS, Mohammed MH, Rajasingh H, Chadaram S, Mande SS (2011) HabiSign: a novel approach for comparison of metagenomes and rapid identification of habitat-specific sequences. BMC Bioinforma 12(Suppl 13):S9

Gilbert W (1981) DNA sequencing and gene structure Nobel lecture, 8 December 1980. Biosci Rep 1:353–375

Gilmour MW, Graham M, Van Domselaar G, Tyler S, Kent H, Trout-Yakel KM, Larios O, Allen V, Lee B, Nadon C (2010) High-throughput genome sequencing of two *Listeria monocytogenes* clinical isolates during a large foodborne outbreak. BMC Genomics 11:1–15

Giraffa G, Neviani E (2001) DNA-based, culture independent strategies for evaluating microbial communities in food-associated ecosystems. Int J Food Microbiol 67:19–34

Glenn TC (2011) Field guide to next-generation DNA sequencers. Mol Ecol Resour 11:759–769

Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods 5:235–237

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol 5: R245–R249

Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, Edgeworth JD, De Lencastre H, Parkhill J, Peacock SJ, Bentley SD (2010) Evolution of MRSA during hospital transmission and intercontinental spread. Science 327:469–474

He H, Sebaihia M, Lawley TD et al (2010) Evolutionary dynamics of *Clostridium difficile* over short and long time scales. Proc Natl Acad Sci USA 107:7527–7532

Henne A, Schmitz A, Bömeke M, Gottschalk G, Daniel R (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. Appl Environ Microbiol 66:3113–3116

Herrgård MJ, Swainston N, Dobson P et al (2010) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat Biotechnol 26:1155–1160

Hert DG, Fredlake CP, Barron AE (2008) Advantages and limitations of next generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. Electrophoresis 29:4618–4626

Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill FX, Goodhead I, Rance R, Baker S, Maskell DJ, Wain J, Dolecek C, Achtman M, Dougan G (2008) High-throughput sequencing provides insights into genome variation and evolution in *Salmonella Typhi*. Nat Genet 40:987–993

Huang S, Li R, Zhang Z et al (2009) The genome of the cucumber, *Cucumis sativus*. Nat Genet 41:1275–1281

Huber JA, Welch DBM, Morrison HG, Huse SM, Neal PR, Buutterfield DA, Sogin ML (2007) Microbial population structures in the deep marine biosphere. Science 318:97–100

Hugenholtz P, Goebel BM, Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180:4765–4774

Humblot C, Guyot JP (2009) Pyrosequencing of tagged 16S rRNA gene amplicons for rapid deciphering of the microbiomes of fermented foods such as pearl millet slurries. Appl Environ Microbiol 75:4354–4361

Ilie M, Fazayeli F, Ilie S (2011) HiTEC: accurate error correction high-throughput sequencing data. Bioinformatics 27:295–302

Imelfort M, Edwards D (2009) De novo sequencing of plant genomes using second-generation technologies. Brief Bioinform 10:609–618

Jany JL, Barbier G (2008) Culture-independent methods for identifying microbial communities in cheese. Food Microbiol 25:839–848

Jung JY, Lee SH, Kim JM, Park MS, Bae JW, Hahn Y, Madsen EL, Jeon CO (2011) Metagenomic analysis of kimchi, a traditional Korean fermented food. Appl Environ Microbiol 77:2264–2274

Juste A, Thomma BPHJ, Lievens B (2008) Recent advances in molecular techniques to study microbial communities in food-associated matrices and processes. Food Microbiol 25:745–761

Kim YS, Kim MC, Kwon SW, Kim SJ, Park IC, Ka JO, Weon HY (2011) Analysis of bacterial communities in Meju, a Korean traditional fermented soybean bricks, by cultivation-based and pyrosequencing methods. J Microbiol 49:340–348

Klenk HP, Göker M (2010) En route to a genome-based classification of Archaea and Bacteria? Syst Appl Microbiol 33:175–182

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders AC, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M (2007) Paired-end mapping reveals extensive structural variation in the human genome. Science 318:420–426

Lee SH, Jung JY, Lee SH, Jeon CO (2011) Complete genome sequence of *Weissella koreensis* KACC 15510, isolated from Kimchi. J Bacteriol 193:5534

Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW, Prosser JI, Schuster SC, Schleper C (2006) Archaea predominate among ammoniaoxidizing prokaryotes in soils. Nature 442:806–809

Li R, Fan W, Tian G et al (2010) The sequence and de novo assembly of the giant panda genome. Nature 463:311–317

Li XR, Ma EB, Yan LZ, Meng H, Du XW, Zhang SW, Quan ZX (2011) Bacterial and fungal diversity in the traditional Chinese liquor fermentation process. Int J Food Microbiol 146:31–37

Liti G, Louis EJ (2005) Yeast evolution and comparative genomics. Annu Rev Microbiol 59:135–153

Liti G, Carter DM, Mosses AM et al (2009) Population genomics of domestic and wild yeasts. Nature 458:337–341

Liu GE (2009) Applications and case studies of the next-generation sequencing technologies in food, nutrition and agriculture. Recent Patents Food Nutr Agric 1:75–79

Lui S, Leathers TD, Copeland A (2011) Complete genome sequence of *Lactobacillus buchneri* NRRL B-30929, a novel strain from a commercial ethanol plant. J Bacteriol 193:4019–4020

Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, Thomas WK (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl Acad Sci USA 105:9272–9277

Machielsen R, van Alen-Boerrigter IJ, Koole LA, Bongers RS, Kleerebezem M, Van Hylckama Vlieg JET (2010) Indigenous and environmental modulation of frequencies of mutation in *Lactobacillus plantarum*. Appl Environ Microbiol 76:1587–1595

Makarova K, Slesarev A, Wolf W et al (2006) Comparative genomics of lactic acid bacteria. Proc Natl Acad Sci USA 103:15611–15616

Mardis ER (2008a) The impact of next-generation sequencing technology on genetics. Trends Genet 24:133–141

Mardis ER (2008b) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387–402

Mardis ER (2009) New strategies and emerging technologies for massively parallel sequencing: applications in medical research. Genome Med 1:40.1–40.4

Mardis ER (2010) The $1,000 genome, the $100,000 analysis? Genome Med 2:84.1–84.3

Margulies M, Egholm M, Altman WE et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380

Masoud W, Takamiya M, Vogensen FK (2011) Characterization of bacterial populations in Danish raw milk cheeses made with different starter cultures by denaturating gradient gel electrophoresis and pyrosequencing. Int Dairy J 21:142–148

Metzker ML (2010) Sequencing technologies-the next generation. Nat Rev Genet 11:31–46

Mitra RD, Church GM (1999) In situ localized amplification and contact replication of many individual DNA molecules. Nucleic Acids Res 27:1–6

Mitra RD, Shendure J, Olejnik J, Krzymanska OE, Church GM (2003) Fluorescent in situ sequencing on polymerase colonies. Anal Biochem 320:55–65

Mitra S, Stärk M, Huson DH (2011) Analysis of 16S rRNA environmental sequences using MEGAN. BMC Genomics 12(Suppl 3):S17

Morales SE, Holben WE (2011) Linking bacterial identities and ecosystem processes: can 'omic' analyses be more than the sum of their parts? FEMS Microbiol Ecol 75:2–16

Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. Genetics 92:255–264

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Synder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344–1349

Nam SH, Choi SH, Kang A, Kim DW, Kim RN, Kim A, Kim DS, Park HS (2011a) Genome sequence of *Lactobacillus farciminis* KCTC 3681. J Bacteriol 193:1790–1791

Nam SH, Choi SH, Kang A, Kim DW, Kim RN, Kim A, Kim DS, Park HS (2011b) Genome sequence of *Lactobacillus coryniformis* subsp. *coryniformis* KCTC 3167. J Bacteriol 193:1014–1015

Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature 457:1038–1042

Nowrousian M, Stajich JE, Chu M, Engh I, Espagne E, Halliday K, Kamerewerd J, Kempken F, Knab B, Kuo HC, Osiewacz HD, Poggeler S, Read ND, Seiler S, Smith KM, Zickler D, Kuck U, Freitag M (2010) De novo assembly of a 40 Mb eukaryotic genome from short sequence reads: *Sordaria macrospora*, a model organism for fungal morphogenesis. PLoS Genet 6:1–22

O'Flaherty S, Klaenhammer TR (2011) The impact of omic technologies on the study of food microbes. Annu Rev Food Sci Biotechnol 2:353–371

Oguntoyinbo FA, Tourlomousis P, Gasson MJ, Narbad A (2011) Analysis of bacterial communities of traditional fermented West African cereal foods using culture independent methods. Int J Food Microbiol 145:205–210

Otero JM, Vongsangnak W, Asadollahi MA, Olivares-Hernandes R, Maury J, Farinelli L, Barlocher L, Osteras M, Schalk M, Clark A, Neilsen J (2010) Whole genome sequencing of *Saccharomyces cerevisiae*: from genotype to phenotype for improved metabolic engineering applications. BMC Genet 11:1–17

Pachter L (2011) Models for transcript quantification from RNA-Seq. arXiv:1104.3889v2

Parameswaran P, Jalili R, Tao L, Shokralla S, Gharizadeh B, Ronaghi M, Fire AZ (2007) A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. Nucleic Acids Res 35:1–9

Pareek CS, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. J Appl Genet 52:413–435

Park EJ, Kim KH, Abell GCJ, Kim MS, Roh SW, Bae JW (2011) Metagenomic analysis of the viral communities in fermented foods. Appl Environ Microbiol 77:1284–1289

Perkins TT, Kingsley RA, Fookes MC, Gardner PP, James KD, Yu L, Assefa SA, He M, Croucher NJ, Pickard DJ, Maskell DJ, Parkhill J, Choudhary J, Thomas NR, Dougan G (2009) A strand-specific RNA-seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. PLoS Genet 5:1–13

Poinar HN, Schwarz C, Qi J, Shapiro B, MacPhee RDE, Buigues B, Thikonov A, Huson DH, Tomsho LP, Auch A, Rampp M, Miller W, Schuster SC (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science 311:392–394

Qi J, Wijeratne AJ, Tomsho LP, Hu Y, Schuster SC, Ma H (2009) Characterization of meiotic crossovers and gene conversion by whole-genome sequencing in *Saccharomyces cerevisiae*. BMC Genet 10:475–486

Qin J, Li R, Raes J et al (2010) A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464:59–67

Rademaker JLW, Hoolwerf JD, Wagendorp AA, Te Giffel MC (2006) Assessment of microbial population dynamics during yoghurt and hard cheese fermentation and ripening by DNA population fingerprinting. Int Dairy J 16:457–466

Raes J, Bork P (2008) Molecular eco-systems biology: towards an understanding of community function. Nat Rev 6:693–699

Reinhardt JA, Baltrus DA, Nishimura MT, Jeck WR, Jones CD, Dangl JL (2009) De novo assembly using low-coverage short read sequence data from the rice pathogen *Pseudomonas syringae* pv. *oryzae*. Genome Res 19:294–305

Riesenfeld CS, Schloss PD, Handelsman J (2004) Metagenomics: genomic analysis of microbial communities. Annu Rev Genet 38:525–552

Roh SW, Kim KH, Nam YD, Chang HW, Park EJ, Bae JW (2010) Investigation of archaeal and bacterial diversity in fermented seafood using barcoded pyrosequencing. ISME J 4:1–16

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J, Goodman RM (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol 66:2541–254

Rothberg JM, Hinz W, Rearick TM et al (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature 475:348–352

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. Int J Food Microbiol 144:352–359

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schadt EE, Turner S, Kasarskis A (2010) A window into third generation sequencing. Hum Mol Genet 19:227–240

Schmid CD, Bucher P (2007) ChIP-Seq data reveal nucleosome architecture of human promoters. Cell 131:831–832

Sharma CM, Hoffmann S, Darfeuille F et al (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature 464:250–255

Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26:1135–1145

Shendure J, Mitra RD, Varma C, Church GM (2004) Advanced sequencing technologies: methods and goals. Nat Rev Genet 5:335–345

Sipiczki M (2011) Diversity, variability and fast adaptive evolution of the wine yeast (*Saccharomyces cerevisiae*) genome—a review. Ann Microbiol 61:85–93

Smith DR, Quinlan AR, Peckham HE et al (2008) Rapid whole-genome mutational profiling using next generation sequencing technologies. Genome Res 18:1638–1642

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere. Proc Natl Acad Sci USA 103:12115–12120

Sorek R, Cossart P (2010) Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat Rev Genet 11:9–16

Stratford M, Bond CJ, James SA, Steels H (2002) *Candida davenportii* sp. nov., a potential soft drinks spoilage yeast isolated from a wasp. Int J Syst Evol Microbiol 52:1369–1375

Studholme DJ, Kemen E, MacLean D, Schornack S, Aritua V, Thwaites R, Grant M, Smith J, Jones JD (2010) Genome-wide sequencing data reveals virulence factors implicated in banana *Xanthomonas wilt*. FEMS Microbiol Lett 310:182–192

Sun Z, Chen X, Wang J, Zhao W, Shao Y, Guo Z, Zhang X, Zhuo Z, Sun T, Wang L, Meng H, Zhang H, Chen W (2011) Complete genome sequence of *Lactobacillus delbrueckii* subsp. *bulgaricus* strain ND02. J Bacteriol 193:3426–3427

Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW (2010) A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res 38:e159. doi:10.1093/nar/gkq543

Tringe SG, Hugenholtz P (2008) A renaissance for the pioneering 16S rRNA gene. Curr Opin Microbiol 11:442–446

Turner DJ, Keane TM, Sudbery I, Adams DJ (2009) Next generation sequencing of vertebrate experimental organisms. Mamm Genome 20:327–338

Tyson GW, Chapman J, Hugenholtz P et al (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428:37–43

Van Hylckama Vlieg JET, Veiga P, Zhang C, Derrien M, Zhao L (2011) Impact of microbial transformation of food on health-from fermented foods to fermentation in the gastro-intestinal tract. Curr Opin Biotechnol 22:211–219

Van Vliet AH (2010) Next generation sequencing of microbial transcriptomes: challenges and opportunities. FEMS Microbiol Lett 302:1–7

Veiga P, Gallini CA, Beal C, Michaud M, Dubois A, Khlebnikov A, Van Hylckama Vlieg JET, Punit S, Glickman JN, Onderdonk A, Glimcher LH, Garrett WS (2010) *Bifidobacterium animalis* subsp. *lactis* fermented milk product reduces inflammation by altering a niche for colitogenic microbes. Proc Natl Acad Sci USA 107:18132–18137

Velasco R, Zharkikh A, Troggio M et al (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS One 2:1–18

Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. Science 291:1304–1351

Wang Y, Wang J, Ahmed Z, Bai X, Wang J (2011) Complete genome sequence of *Lactobacillus kefiranofaciens* ZW3. J Bacteriol 193:4280–4281

Wilhelm BT, Marguerat S, Watt S, Schubart F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bahmer J (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 453:1239–1243

Wintzingerode FV, Göbel UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol Rev 21:213–229

Wittmann C, Hinzle E (2002) Genealogy profiling through strain improvement by using metabolic network analysis: metabolic flux genealogy of several generations of lysine-producing corynebacteria. Appl Environ Microbiol 68:5843–5859

Wu D, Hugenholtz P, Mavromatis K et al (2009) A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature 462:1056–1060

Wurtzel R, Sapra F, Chen Y, Xu Y, Simmons BA, Sorek R (2010) A single-base resolution map of an archaeal transcriptome. Genome Res 20:133–141

Xu Z, Wei W, Gagneur J, Perrochi F, Clauder-Munster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM (2009) Bidirectional promoters generate pervasive transcription in yeast. Nature 457:1033–1037

Yang X, Dorman KS, Aluru S (2010) Reptile: representative tiling for short read eroor correction. Bioinformatics 26:2526–2533

Zhou X, Ren L, Meng Q, Li Y, Yu Y, Yu J (2010) The next-generation sequencing technology and application. Protein Cell 1:520–536

Züell M, van Noort V, Yus E, Chen WH, Leigh-Bell J, Michalodimitrakis K, Yamada T, Arumugam M, Doerks T, Kuhner S, Rode M, Suyama M, Schmidt S, Gavin AC, Bork P, Serrano L (2009) Transcriptome complexity in a genome-reduced bacterium. Science 326:1268–1271