4504–4518 *Nucleic Acids Research, 2016, Vol. 44, No. 10* *Published online 21 April 2016*
doi: 10.1093/nar/gkw309

SURVEY AND SUMMARY

Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms

Cyrielle Gasc, Eric Peyretaillade and Pierre Peyret*

EA 4678 CIDAM, Université d'Auvergne, Clermont-Ferrand, 63001, France

Received January 28, 2016; Revised April 7, 2016; Accepted April 12, 2016

ABSTRACT

The recent expansion of next-generation sequencing has significantly improved biological research. Nevertheless, deep exploration of genomes or metagenomic samples remains difficult because of the sequencing depth and the associated costs required. Therefore, different partitioning strategies have been developed to sequence informative subsets of studied genomes. Among these strategies, hybridization capture has proven to be an innovative and efficient tool for targeting and enriching specific biomarkers in complex DNA mixtures. It has been successfully applied in numerous areas of biology, such as exome resequencing for the identification of mutations underlying Mendelian or complex diseases and cancers, and its usefulness has been demonstrated in the agronomic field through the linking of genetic variants to agricultural phenotypic traits of interest. Moreover, hybridization capture has provided access to underexplored, but relevant fractions of genomes through its ability to enrich defined targets and their flanking regions. Finally, on the basis of restricted genomic information, this method has also allowed the expansion of knowledge of nonreference species and ancient genomes and provided a better understanding of metagenomic samples. In this review, we present the major advances and discoveries permitted by hybridization capture and highlight the potency of this approach in all areas of biology.

INTRODUCTION

The emergence of next-generation sequencing (NGS) technologies has radically revolutionized medical, biotechnical, evolutionary and ecological research, providing large amounts of sequencing data at a low cost per base of generated sequence (1). Deep sequencing of human genomes or complex environmental samples in their entirety is therefore feasible but the costs are still high, and the data analysis is complex. For instance, using a HiSeq2000 instrument (2 Gb of data, 2 × 100 bp), 6000 runs must be performed with a global cost of $267 million to produce a dataset representing 1-fold coverage of the microbial community from 1 g of soil containing 10^9 prokaryotic cells, making deep exploration of soil unfeasible (2). Similarly, with the HiSeq X platform (1.8 Tb of data), 16 human genomes can be sequenced per 3-day run with 30× coverage, at slightly more than $1000 per genome (3). However, this sequencing depth remains insufficient for many medical studies, financial costs become substantial for investigations at the population scale and the required data treatment is fastidious and time consuming. Therefore, because exhaustive information on samples is not always necessary and data on specific genomic subsets can be more informative than an overall approach, targeted enrichment methods in which genomic regions of interest are selectively captured before sequencing have been developed (4). These methods, defined as capture, partitioning or enrichment approaches, make studies at a multiple-sample scale feasible, reducing sample complexity to concentrate sequencing efforts on meaningful genomic fractions in a more cost- and time-efficient manner.

PCR has been the most widely used partitioning method for years (4). Nevertheless, numerous limitations linked to PCR bias still exist (5), despite the recent improvements of the methodology in the context of genomic partitioning (6–8). Circularization capture methods, in which the two main approaches are molecular inversion probes (MIPs) (9) and spacer multiplex amplification reaction (SMART) (10), employ highly specific padlock probes and are compatible with a much higher degree of multiplexing (8). Despite their efficiency in targeted enrichment, the PCR and circularization

*To whom correspondence should be addressed. Tel: +33 473178309; Email: pierre.peyret@udamail.fr

capture approaches have some limitations. Indeed, their application is not suited for genomic regions that are several megabases in size because of the multiple primer pairs or probes required to cover large regions and the difficulty of designing them with the necessary high sequence specificity. Similarly, the number of target biomarkers is also limited because of primer design, which quickly causes the application of these strategies to become fastidious and expensive. Finally, neither of these approaches has been developed into commercially available products, making them more expensive and less amenable to a broad range of research applications.

Hybridization capture coupled with high-throughput sequencing has been developed to overcome these limitations, allowing the rapid and simultaneous description of numerous and potentially large targeted DNA regions (Figure 1). The hybridization between the sequencing library and probes can be performed in solution or on a solid support. In solid-phase hybridization capture, herein referred to as array-based hybrid selection (AHS), DNA probes are bound to a solid support such as a glass microarray slide (11,12). DNA libraries are applied to the surface of the support, and targeted DNA fragments hybridize with the immobilized probes. Nonspecific unbound molecules are washed away, and the enriched DNA is eluted and amplified for downstream direct sequencing. In solution hybridization capture, also referred to as solution-phase hybrid selection (SHS), free DNA or RNA probes are biotinylated to enable the selection of targeted fragment-probe heteroduplexes using magnetic streptavidin beads (13). Non-targeted library fragments are removed from the liquid phase through repeated washes, and targeted fragments are then eluted from the beads and amplified for subsequent sequencing. Methods relying on hybridization are the most widely used, with numerous available cost-effective commercial platforms (14–17). Because these approaches employ longer probes than PCR methods, they are generally less specific and show a greater tolerance for the discovery of polymorphisms (13). Furthermore, hybridization capture is highly scalable to targets of different lengths, equally enriching short discontiguous targets or long genomic regions (18), and is highly multiplexable through the use of barcodes to pool many libraries together.

Hybridization capture methods allow the effective enrichment of a wide range of targets and offer a good compromise between specificity, multiplexing and scalability. Additionally, they match the scale of current high-throughput NGS and have become reference methods for answering a variety of questions, from revealing the sources of genetic disease, to elucidating evolutionary histories and evolution stories and the complex exploration of microbial ecosystems (Figure 2). Hybridization capture for DNA resequencing efficiently contributes to the characterization of genetic variations in clinical genetics, agronomical traits and environmental adaptations over time and space. Targeted sequencing directed to the protein-coding portion of the genome (exome) or the captured transcriptome improves the resolution for rare variants detection. Overlooked off-targets sequences obtained during sequence capture by hybridization provide important source to study the unknown flanking regions (new variants, virus and mobile genetics elements integration sites, regulatory signals). These approaches are also very efficient even for degraded samples. Ancient DNA (aDNA) extracted from paleontological remains or museum specimens has been explored for understanding the evolutionary history of numerous lineages (humans, livestock, pets, wild animals, fish, insects, plants, microorganisms) of both extinct and extant species. Hybridization capture is also used to reveal new sequences for a broad range of organisms even distantly related and without reference genomes. Some erroneous relationships in the previously established phylogeny and identifying of new clades have been highlighted. Complex microbial ecosystems could also be efficiently explored allowing to access to complete phylogenetic or functional biomarker diversity as well as providing a first glimpse into the identification of metabolic capabilities through the association of biomarkers. In this review, we present the major hybridization capture experiments and results in the abovementioned fields and illustrate how powerful this approach can be for revealing information that is inaccessible or very difficult to obtain when using other conventional molecular approaches.

RESEQUENCING OF LARGE GENOMIC REGIONS

Based on knowledge of reference genomes, the application of NGS to the genomes of particular individuals gives rise to an exhaustive list of genomic variations. Therefore, DNA resequencing can efficiently contribute to the characterization of genetic variations such as single-nucleotide polymorphisms (SNPs), insertions and deletions (indels), structural variants (inversions, translocations) and copy number variants (CNVs) (8). Nevertheless, the high sequencing depth required to distinguish real variants from sequencing errors, occurring at a rate of 0.1–1% (19), with high reliability, makes this approach difficult and expensive (2). Moreover, complete genome resequencing is not always necessary when only particular genes or genomic regions are of interest. Conventional resequencing approaches based on bidirectional sequencing of targeted PCR amplicons are particularly well suited for the detection of SNPs but are less well suited for the detection of other variations and are not sufficient because of the high levels of multiplexing required when large or numerous genomic regions have to be characterized (20). Consequently, resequencing has evolved toward the hybridization of DNA to resequencing microarrays containing tiling probes designed to detect interesting genomic variants (21). This approach is commonly applied in genome-wide association studies (GWASs) to simultaneously detect thousands of SNPs across a complete genome associated with a studied phenotype (22). However, GWASs are generally restricted to common genomic variants, excluding rare and undiscovered genetic modifications (8), and are principally relevant for the detection of SNPs. Thus, gene capture approaches by hybridization using tiling probes have been developed to specifically enrich and resequence numerous meaningful subsets of the genomes of multiple individuals with a high depth and to better associate genotypes and phenotypes in a cost-effective manner (Figure 2A).

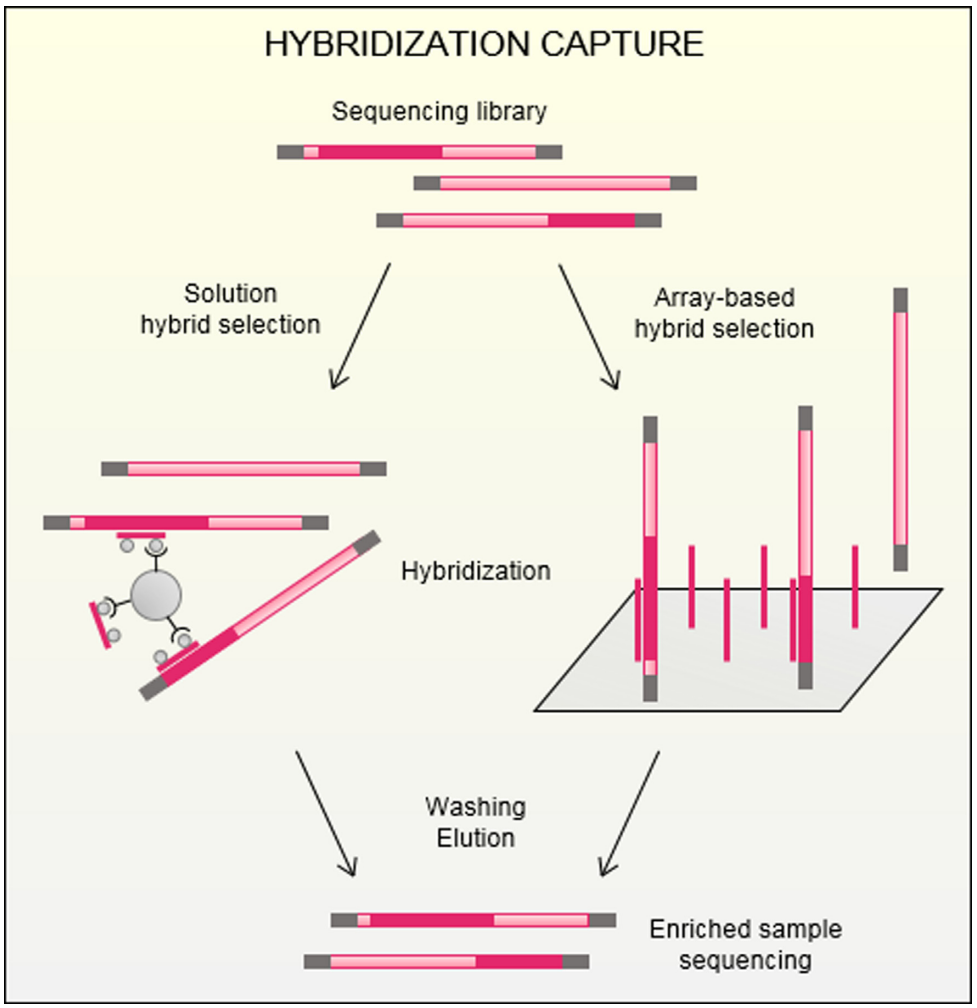

Figure 1. Principle of hybridization capture. A sequencing library containing targeted biomarkers is constructed and hybridized against a set of specific probes. Hybridization can be performed either in solution (solution hybrid selection, SHS) with biotinylated probes captured by streptavidin-coated magnetic beads, or on a solid support (array-based hybrid selection, AHS) on which probes are spotted. After hybridization, nontarget sequences are washed away, and the enriched sample is eluted and sequenced.

Pioneering studies using hybridization capture were carried out by Okou *et al.* (12) and Albert *et al.* (11), who applied high-density microarrays with tiling probes for the enrichment and resequencing of large genomic regions on the order of ten to hundreds of kilobases. Okou *et al.* (12) studied the enrichment of coding and noncoding regions of human genes involved in fragile X syndrome. Similarly, Albert *et al.* (11) resequenced areas of 200 kb, 500 kb, 1 Mb, 2 Mb and 5 Mb surrounding the human BRCA1 gene involved in Burkitt's lymphoma. Likewise, Gnirke *et al.* (13) developed an SHS method that simultaneously targeted four human genomic regions ranging from 220 to 750 kb in size. These approaches have been used by Beaudoin *et al.* (23) to resequence 55 genes known to be associated with ulcerative colitis, by Ruark *et al.* (24) to target 507 genes implicated in DNA repair in association with breast cancer and ovarian cancer, by Clarke *et al.* (25) to discover SNPs associated with traits of agronomical and nutritional importance through 47 *Brassica napus* specific genomic regions, or even by Schiessl *et al.* (26) to explain phenological variations in flowering in *Brassica* through deep resequencing of 29 regulatory genes. Similarly, Tennessen *et al.* (27) used hybridization capture to decipher origins and dynamics of strawberry octoploid subgenomes, Jupe *et al.* (28) discovered and annotated pathogen resistance gene family members, Fu *et al.* (29) evidenced SNPs in maize genome and Faucon *et al.* (30) detected variants involved in *Aedes aegypti* insecticide resistance in more than 760 candidate genes. More recently, the efficacy of capture coupled to third-generation sequencing has been demonstrated by delineating the breakpoint junctions of low copy repeat associated complex structural rearrangements in patients diagnosed with Potocki–Lupski syndrome (31).

The potential of hybridization capture to accurately detect genomic variants in large specific genomic targets, as demonstrated in the congruous studies mentioned above, opened the door for its application to a broader range of biomarkers, rather than a single gene or a limited number of restricted regions. Hybridization capture has therefore been

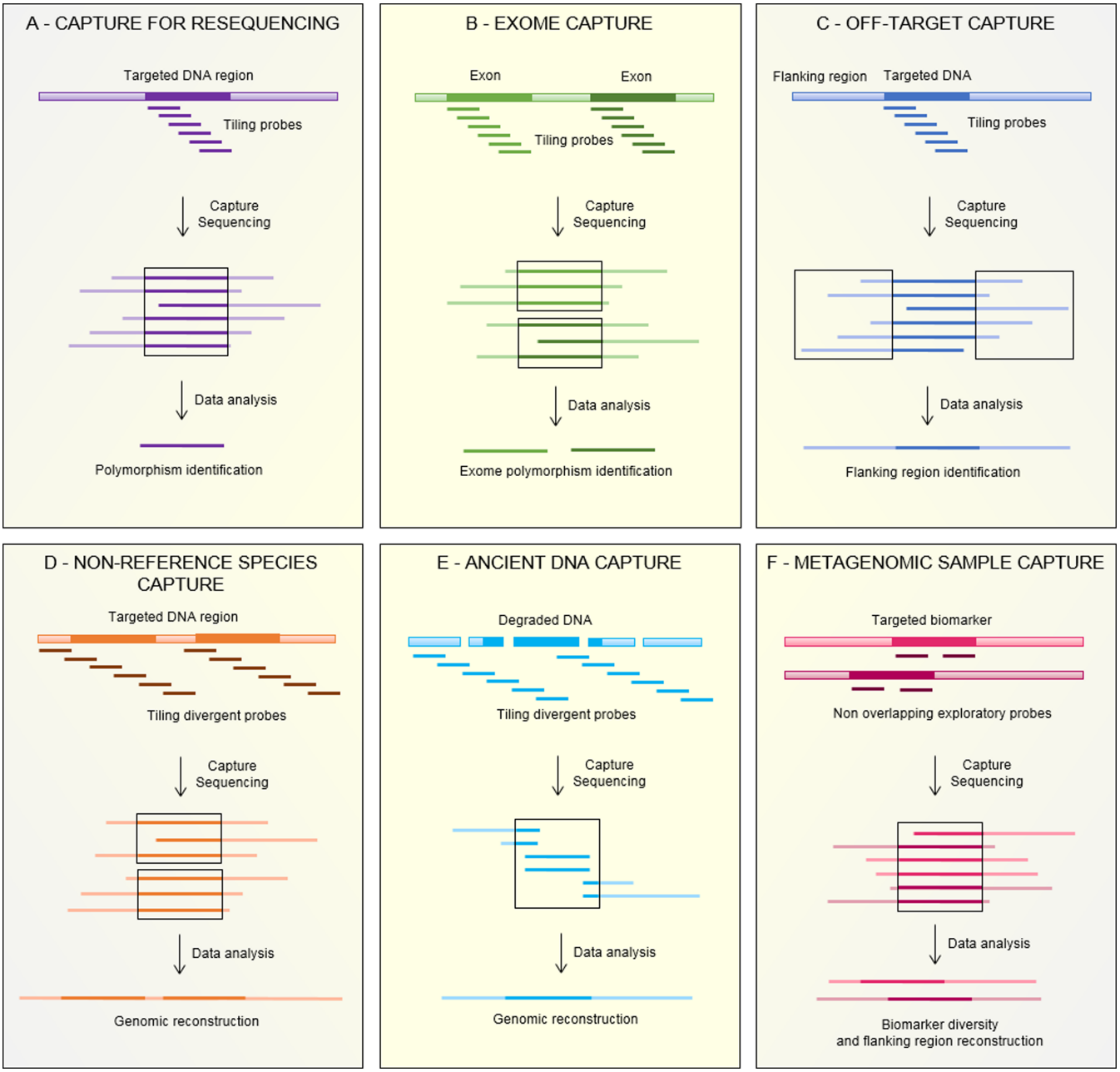

Figure 2. Applications of hybridization capture. (**A**) Capture for resequencing. Specific tiling probes are used to capture genomic regions of interest and to identify sequence polymorphisms. (**B**) Exome capture. Tiling probes targeting the whole exome are used to identify coding mutations in the genome. (**C**) Off-target capture. Tiling probes are used to enrich a biomarker whose indirectly captured flanking regions are of interest. (**D**) Nonreference species capture. Tiling probes designed based on divergent genomes are used to capture biomarkers or genomes of nonreference species. (**E**) Ancient DNA capture. Tiling probes designed based on modern species are used to enrich ancient DNA and reconstruct biomarkers or genomes. (**F**) Metagenomic sample capture. Nonoverlapping exploratory probes are used to enrich all variants of a biomarker from a complex sample to study its diversity and its genomic context.

applied to the resequencing of numerous very informative and widespread short genomic region, gene exons.

EXOME CAPTURE

Exome capture, also known as whole exome sequencing (WES), is targeted sequencing of the protein-coding portion of the genome. This method employs capture by hybridization with exon-specific tiling probes to target the protein-coding variants in the best understood subset of the genome (Figure 2B) (32). Indeed, the exome represents approximately 1% of the human genome, corresponding a total of approximately 30 Mb, but harbors 85% of the known genetic disease-causing mutations (33,34). Moreover, gene exons are more likely to contain pathogenic variants than introns or regions between genes (35–37), and they harbor fewer repetitive elements, making their analysis easier (38).

Exome capture has rapidly become a standard practice in clinical genetics for determining the basis of human dis-

eases (37), especially due to the development of various exome capture platforms (39) that differ in terms of probe design, sensitivity, coverage and their ability to detect different types of variants (e.g. 31–35). WES was first applied for the identification of variants involved in monogenic disorders. Among these genetic diseases arising from single-gene mutations, which are mostly exonic, (40), less than half of the causative genes have been identified for the approximately 10,000 monogenic disorders referenced to date (41). The identification of causative genes is the first step for better understanding the pathology of the diseases and prescribing effective therapies. Exome capture is estimated to be successful in the identification of variant genes in more than half of resequencing projects (42) and has led to the discovery of approximately 800 monogenic disease-causing genes in the last few years (41,43–45). The first illustration of the relevance of exome capture in the realm of Mendelian diseases was the identification of mutations in the DHODH gene, responsible for Miller syndrome, whose cause was previously unknown (43). Similarly, with this approach, the same research group (44) showed previously unidentified mutations in the MLL2 gene that underlies Kabuki syndrome. In other cases, some disease-causative genes are known, but no mutations in these genes can be detected in patients. For example, Peters *et al*. (46) showed that in patients with 3-hydroxyisobutyryl-CoA hydrolase deficiency (HIBCH), no mutations could be found in the HIBCH gene, despite an indicative clinical phenotype and the use of exome sequencing allowed the identification of mutations in the ECHS1 gene, which had not previously been associated with this genetic disorder. Exome capture has also become a proven tool for the identification of mutations underlying complex diseases caused by interactions between genetic mutations and environmental factors (e.g. 16–20). Many mutations in different genes can underlie a given disorder; such mutations often follow the 'common disease-rare variants' hypothesis, suggesting that rare and novel genetic variants are more likely to be detected than common genetic variants (47,48). Consequently, the application of WES is particularly well suited to these cases because of its sensitivity for variant detection and the broad range of targets it explores. Thus, Huyghe *et al.* (49) used WES to explain the remaining undefined genetic contribution to insulin secretion deficiency in type 2 diabetes patients. They demonstrated the presence of low-frequency coding variants associated with five loci related to insulin and proinsulin regulation, and they demonstrated their contribution to the complex traits under study. Similarly, through WES, Do *et al.* (50) identified rare mutations in the LDLR and APOA5 genes involved in myocardial infarction (MI), in addition to the common variants associated with MI risk in the population. Finally, exome capture is also broadly used for the detection of other diseases, cancers (51). In this particular type of genetic disorder, biological samples are composed of a heterogeneous mixture of cell populations, in which cancerous cells can be a minority. The identification of rare carcinogen genetic alterations in these complex samples through conventional molecular approaches is therefore particularly difficult (52). However, the detection of subclonal mutations can be readily achieved by exome capture, as illustrated by Agrawal *et al.* (53) in a survey of head and neck squamous cell carcinoma (HNSCC).

In addition to allowing the detection of causal variants involved in genetic disorders and opening new horizons for medical treatments, WES has been shown to be an effective tool for medical diagnosis (54). Worthey *et al.* (55) successfully used exome capture to diagnose a patient through the identification of gene variants for the first time. In these case, despite a comprehensive clinical evaluation suggesting an immune defect, clinicians were unable to establish a definitive diagnosis, thereby limiting the medical management of the patient. The use of WES revealed a hemizygous missense mutation in the gene responsible for the X-linked inhibitor of apoptosis deficiency. In a different context, through the analysis of mutations detected using WES, Choi *et al.* (56) fortuitously diagnosed congenital chloride diarrhea in a patient thought to be suffering from Bartter syndrome, a renal salt-wasting disease. In addition, a major project emerged 5 years ago: the Deciphering Developmental Disorders (DDD) study (57). The purpose of this study is to determine genetic variants underlying severe undiagnosed developmental disorders in 12 000 children through multiple complementary genome-wide approaches, including exome capture. Thus far, the exomes of more than 1000 children have been sequenced, and more than 30% of the children have been diagnosed, demonstrating the potential of WES in this field (58). This approach benefits from large databases of pathogenic variants of human genes, such as that of the Exome Aggregation Consortium (ExAC), which combines exome sequencing data from more than 60 000 individuals with various diseases (http://exac.broadinstitute.org), and from bioinformatics tools linking mutations discovered through WES back to a medical diagnosis (59,60).

Exome capture has also been used to sequence the messenger RNA (mRNA) fraction as complementary DNA (cDNA) in human medical studies to extend information obtained from DNA-based investigations and reveal information that is inaccessible based on analysis of DNA alone. This method provides an interesting alternative to RNA-sequencing (RNA-Seq) of mRNA, allowing targeted enrichment of the complete coding fraction of the transcriptome. Mercer *et al.* (61) and Halvardson *et al.* (62) demonstrated that, due to the deep coverage of exome capture, this method reveals rare and unannotated transcripts whose expression levels are below the detection limits of conventional sequencing approaches. Moreover, Cabanski *et al.* (63) and Cieslik *et al.* (64) showed that exome capture targeting coding transcripts improves performance compared with RNA-Seq for degraded samples for which either limited amounts of RNA are available or polyA selection is arduous, respectively. They demonstrated that genetic variants and gene fusions are accessible through exome capture and that the obtained absolute and differential gene expression levels are accurate for tumor samples, which has also been shown by Clark *et al.* (65) for long noncoding RNAs. Therefore, exome capture appears to be a complete tool for medical resequencing, providing exhaustive information necessary for determination of the genetic basis of various pathologies and their diagnosis.

Exome capture has also been used in many other model organisms for different purposes. For example, WES has

been employed for to variant identification in medical model species, such as mice (66), pigs (67) and dogs (68,69), providing valuables tools for disease association studies. Moreover, as not all protein-altering variants induce diseases (70), exome capture has also been used for the detection of variants responsible for modifications in phenotypic traits of economic importance in crops and livestock. Mascher et *al.* (71) and Saintenac *et al.* (72) explored natural variation and genetic diversity respectively in barley and in wheat. Muraya *et al.* (73) applied exome capture in maize to detect sequence variation affecting biomass production. Other studies have investigated the cattle exome to determine gene variants associated with milk production (74), fertility (75) and growth (76). Similarly, the *Eucalyptus* exome has been studied, to identify gene mutations responsible for wood property traits (77). This method has also been employed in black cottonwood (78), wheat (72,79), barley (71,80) and switchgrass (81) to elucidate genomic variation in these model species and to better link genotypes and phenotypes for further agronomic improvements. Another similar application of exome capture in this field is for the identification of variants generated through mutagenesis to investigate gene function. Indeed, random mutagenesis approaches induce various unknown mutations affecting phenotypes. WES has been effectively used in such agronomic studies to link phenotypes back to gene mutations, replacing comparative genomic hybridization microarrays (82) and bringing variant detection to the whole genome level as demonstrated by Henry *et al.* (83) and King *et al.* (84). Similarly, Bolon *et al.* (85) successfully used exome capture to detect deletion mutations induced in 120 000 soybean seeds through fast neutron radiation. After phenotypic screening based on traits of interest such as seed composition, maturity, morphology, pigmentation and nodulation, these authors identified candidate genes, allowing the determination of agronomically important markers for soybean crop breeding.

These applications of exome capture in a wide range of studies, from discovering the origins of genetic disorders to determining key agronomic biomarkers, and in a broad spectrum of species, illustrate the potential of hybridization capture for the resequencing of this genomic subset. Nevertheless, the information obtained from the exome remains incomplete. Therefore, many studies have taken advantage of the off-target sequences obtained from hybridization capture to increase the amount of information generated from experiments.

OFF-TARGET CAPTURE

While exome sequencing primarily targets exons, a significant portion of the captured fragments come from outside of these coding regions (86,87) (Figure 2C). Indeed, the hybridization process selects for fragments that contain at least a portion of targeted exons, resulting in enrichment of flanking regions such as introns, untranslated regions (UTRs), promoters and intergenic regions. In a typical exome sequencing experiment, these off-target sequences represent approximately 40–60% of the obtained reads (e.g. 31,32,37,78,79); this percentage can be related to many different factors, such as the exome capture platform, the size of DNA libraries and sequencing technologies (4,14). The off-target sequences generated are usually excluded from subsequent analyses. However, these sequences provide a potentially important source of unknown information that could greatly extend the utility of exome sequencing.

Following the study initiated by Guo *et al.* (88), numerous researcher have devoted attention to these overlooked off-target sequences. Indeed, these studies have demonstrated that reads and SNPs outside of targeted regions coming from different exome capture platforms are of good quality, suggesting that these data can be exploited. They have also shown that through the exploitation of complete sequencing data, it is possible to more than double the number of high-quality SNPs in comparison with those that are made available by focusing on exome regions. Indeed, they found that approximately 50% of the SNPs identified from exome sequencing were in target regions, while 27% were in the 200 bp flanking regions, and the remaining 23% were in regions >200 bp away from exons, illustrating that sequences falling outside of the targeted regions should not be ignored in data processing. A good illustration of a subsequent off-target investigation is a study on the adaptation of residents of the Tibetan Plateau to high altitude through exome capture (89). The authors identified SNPs located in different genes, and showed that the SNPs exhibiting the greatest difference in frequency compared with low-altitude residents were intronic, suggesting the significance of introns and supporting the importance of off-target regions.

In many studies, off-target WES reads have been shown to be a reliable and rich source of human mitochondrial DNA (mtDNA) because of the relative high abundance of the corresponding genome within the total DNA (90). Indeed, each mammalian cell contains approximately 100 mitochondria, which themselves contain between two and 10 copies of the 16.6 kb closed-circle mtDNA genome (91). Thus, even if the mtDNA genome is not the target, it can easily be obtained through indirect methods such as exome capture with an average coverage of 100× (90) and totalling approximately 1–5% of the obtained sequencing reads (86). Therefore, off-target mtDNA represents a precious source of information for many sequencing investigations, as demonstrated by Griffin *et al.* (92) and Zhang *et al.* (93), who showed that the mtDNA obtained through WES allows reliable detection of sequence variants with a low error rate, similar to conventional mitochondrial DNA Sanger sequencing. This capacity of exome capture to reliably enrich both targeted exome and off-target mtDNA has been applied for the detection of mutations involved in mitochondrial diseases. For example, through performing this simultaneous dual-genome analysis and identifying mutations in both exome and mtDNA, Dinwiddie *et al.* (94) successfully diagnosed four patients with distinct mitochondrial disorders for whom conventional molecular testing had previously failed to produce a diagnosis. Such evaluations of mtDNA based on off-target data can be facilitated by dedicated bioinformatics tools such as MitoSeek, which extracts mitochondrial genome information from exome data and analyses associated mutations and structural variants (95).

Based on the potency of this method for expanding knowledge beyond target DNA regions, studies have used

hybridization capture strategies to study the unknown flanking regions of genes or particular genomic regions. For example, Duncavage *et al.* (96) employed a capture strategy to identify the sites of Merker cell polyomavirus (MCPyV) integration in the genomes of Merkel carcinoma patients. Targeting the virus genome, these authors achieved up to 37 000-fold coverage of the MCPyV genome, making detection of viral insertion sites possible through the analysis of off-target flanking regions. Similarly, Ma *et al.* (97) investigated cis-regulatory elements (CREs) surrounding plant microRNA (miRNA) genes via miRNA hybridization capture. They demonstrated the specificity and sensitivity of this enrichment method in *Arabidopsis thaliana* by assembling regions spanning 10–20 kb flanking miRNAs. Finally, Platt II *et al.* (98) identified retrotransposon insertion sites in the bat genus *Myotis* through this approach to construct a robust transposable element insertion-based phylogeny.

Therefore, off-target data appears to be a rich source of diverse information, making the study of nontargeted reads an important field to explore. Nontargeted reads may contain information that is not covered in the current reference genomes and may also provide information about flanking sequences that are yet to be discovered, leading to new insights into genome exploration. These findings show that hybridization capture can expand our knowledge beyond the currently available information on studied organisms, and this potential has been demonstrated through the application of this strategy to enrich genomic regions of interest from organisms without reference sequences or genomes.

NONREFERENCE SPECIES CAPTURE

Despite the intensive use of NGS technologies, there are still no established reference genomes or draft genomes for most nonmodel organisms. Accessing the genetic makeup of these organisms, referred to as nonreference species, can be achieved through *de novo* whole-genome sequencing, but sequencing costs and difficulties in assembly can be prohibitive, without necessarily providing access to complete information (99). Therefore, hybridization capture represents a good alternative for accessing targeted informative subsets of these genomes (Figure 2D). Nevertheless, these methods require *a priori* knowledge of the targeted sequences which is, by definition, unavailable for nonreference species. Different probe design strategies for enriching genomic regions of interest have been developed to overcome this limitation and hybridization capture has been used to reveal new sequence information for a broad range of organisms (99). The first approach for probe design circumventing the need for a reference genome relies on PCR amplification. Due to the large amount of sequence data and numerous primers available in the literature and in databases, obtaining PCR primers for informative loci has become quite easy. These PCR products amplified from a small set of targeted loci can thus be used as capture probes for nonreference species (100). Another strategy leverages partial sequencing data to design probes. Thus, the design can be generated based on low-coverage *de novo* DNA sequencing assemblies (27,101), RNA-Seq transcriptome (94–97,102), expressed sequence tag (EST) data (103–105) or clone libraries (106) which provide core information about the targeted species. Finally, probe design can be based on orthologous genomic regions between divergent reference organisms (107–109). In this case, enrichment relies on the ability of probes designed according to a single reference genome to specifically capture DNA from a broad range of other species. The impact of sequence divergence on the efficiency of hybridization capture has been evaluated in many model and nonmodel species (e.g. 64,73–77). For example, Hedtke *et al.* (110) used exome capture to enrich DNA across frog species spanning approximately 250 million years of evolutionary divergence (up to approximately 10% divergence), thus testing the limits of this approach. They demonstrated that the success of enrichment depends on the divergence time between a given frog species and the reference sequence and that even if there is some missing data, the method appears to be effective between organisms that shared a common ancestor up to 200 million years ago.

Hybridization capture for nonreference species has been proven to be an appealing method addressing various biological issues. It was first used to greatly expand the repertoire of available genomic sequences for nonmodel organisms. Indeed, as illustrated by Neves *et al.* in two different studies (104,111), hybridization capture through exome targeting can efficiently enrich large portions of complex genomes. In these studies, hybridization capture data provided information to guide the assembly of reference loblolly pine genome which was previously largely uncharacterized despite attempts at sequencing, and significantly improved our knowledge of the pine gene space. Similarly, Rosani *et al.* (112) made use of hybridization capture to increase current knowledge of the unknown Mediterranean mussel genome, George *et al.* (113) applied this methodology to generate high-quality, full-length sequences for nonhuman primate species and Seoane-Zonjic *et al.* (114) used it to establish gene structures without reference genomes in the maritime pine mega-genome. Hybridization capture using probes targeting genomic regions from divergent nonreference species has also been broadly employed to define the gene variants at the origin of the process of species divergence. Using this approach, Hebert *et al.* (103) identified SNPs in genes potentially involved in whitefish divergence and reproductive isolation; Nadeau *et al.* (115) captured patterns of divergence in radiating geographical species of *Heliconius* butterflies; and Cosart *et al.* (116) identified thousands of candidate SNPs in livestock and wild cattle species, revealing the detailed genetic basis of adaptive differentiation and speciation. Similarly, with this approach, Schneider *et al.* (117) identified mutations at the origin of polymorphism for melanism in three South American felid species. Finally, this hybridization capture approach for nonreference species can also provide information in the field of phylogenetics. Indeed, it enables the collection of data on hundreds of biomarkers across the genomes of many individuals, which is necessary for the establishment of robust phylogenetic relationships. For example, Stephens *et al.* (118) used this approach to enrich approximately 200 specific nuclear genes from 75 accessions of the North American carnivorous pitcher plant genus *Sarracenia*. Using the multilocus data obtained through hybridization capture, these authors elucidated relationships

within the *Sarracenia* genus, which had not been possible in previous attempts using conventional approaches targeting only a few informative sites. Similarly, based on 259 targeted nuclear loci, Prum *et al*. (119) reconstructed the phylogeny of 198 living bird species, revealing some erroneous relationships in the previously established phylogeny and identifying many new clades. In two analogous studies, Faircloth *et al.* (120,121) designed probes targeting ultraconserved nuclear DNA element (UCEs) that are shared among ray-finned fishes and the insect order Hymenoptera. With these UCEs loci, the authors accurately resolved phylogenetic relationships at both shallow and deep time-scales within the studied organisms. Likewise, in a study conducted by Wang *et al.* (122), hybridization capture was used to sequence mtDNA segments from different accessions of 36 pine species and to gain insights into the mtDNA-based phylogeny of the genus *Pinus*. They revealed that this phylogeny differed from that based on chloroplast and nuclear DNA and, thus, identified a series of mtDNA capture events during pine evolution.

The power of hybridization capture in the field of nonreference species is such that the method has been widely and successfully used to study highly divergent samples from extinct organisms or historical remains.

ANCIENT DNA CAPTURE

Ancient DNA extracted from paleontological remains or museum specimens is a rich source of information for understanding the evolutionary history of lineages of both extinct and extant species (123–125). However, because of postmortem preservation conditions, aDNA samples have particular characteristics that make them difficult to analyse using conventional molecular approaches. Indeed, most ancient samples contain short DNA fragments averaging 30–300 bp (126), resulting from the degradation of genomic DNA by cellular endonucleases and exogenous micro-organisms that fragment DNA and depending strongly on the specific geochemistry of the environment (18). Consequently, ancient samples contain a low percentage of endogenous DNA and are highly contaminated by environmental bacteria and fungi (127–129). For example, one study revealed that 55,000-year-old bone fragments from a Neanderthal individual were composed of a complex mixture of approximately 1% degraded human DNA and 99% contaminant environmental DNA (130). Only ideally preserved permafrost remains exhibit little damage and can yield fragments of nearly 2 kb, with less contamination (131). Another significant source of contamination of aDNA samples is modern human DNA introduced during excavation and museum or laboratory handling (124,132). Finally, postmortem nucleotide changes occur and introduce errors in sequencing reads, thus requiring particular attention during data analysis. The most common changes are cytosine deamination to uracil, which is then converted to thymine by DNA polymerase, and the transition of guanine to adenine, which can be used to reliably differentiate between ancient and contaminating modern DNA sequences (18,133,134).

Until recently, aDNA studies were restricted to PCR amplification of short overlapping DNA fragments due to the length of aDNA fragments. This approach enabled the reconstruction of long DNA sequences and mitochondrial genomes, but hindered complete genome reconstruction (135). The introduction of NGS, based on the sequencing of short sequencing reads that are problematic for PCR amplification, has consequently greatly facilitated aDNA studies (94–96). Nevertheless, because of the very low proportion of endogenous DNA in ancient samples and the considerable sequencing effort required to access it, NGS approaches are usually not sufficiently efficient or financially unfeasible. Therefore, to overcome these methodological limitations, hybridization capture approaches have often been employed (Figure 2E). Due to the ability of hybridization capture to specifically enrich targeted DNA from massive amounts of contaminating environmental DNA with probes designed based on very distant species, this strategy has quickly become a gold standard for ancient aDNA research. Its utility was first demonstrated by Noonan *et al.* (136) in a pioneering study showing that it was possible to reconstruct up to 65 kb of nuclear DNA targets from a Neanderthal library using hybridization capture, compared with only 5 kb when using conventional approaches. The efficiency of hybridization capture has since been widely demonstrated, and was especially well illustrated by Carpenter *et al.* (129) who showed that this method enabled 6- to 159-fold enrichment of ancient human DNA using modern human DNA probes and that targeted DNA represented up to 60% of the obtained reads, versus the average of 1.2% observed before hybridization capture. This revolutionary approach in the field of paleogenomics has been widely employed to effectively target full mitochondrial genomes (e.g. 128,131) as well as nuclear DNA (e.g. 129,132) from human, animal, vegetal or microbial species.

This technological advance in the area of aDNA research has allowed better characterization of ancient extinct organisms and led to more precise definition of phylogenetic links between ancient and modern species. For example, Castellano *et al*. (137) used hybridization capture to enrich the exomes of two Neanderthals from Spain and Croatia. Comparisons with other genomes revealed that the diversity among Neanderthals was remarkably low compared with that found in modern humans and that amino acid substitutions underlie the phenotypic differences between Neanderthals and present-day humans. Similarly, Burbano *et al.* (138) generated data from megabase-size genomic regions from Neanderthal DNA samples. Comparison of these data with sequences of present-day humans at the same positions allowed them to identify 88 amino acid substitutions that have occurred since the population split of modern humans and Neanderthals. In nonhuman species, this approach enabled the resolution of many evolutionary histories through the capture of mitochondrial genomes from aDNA. Llamas *et al.* (139) provided a reliable phylogeny of the Australian megafauna and clarified the phylogenetic position of the extinct giant short-faced kangaroo and giant wallaby, while Vilstrup *et al.* (140) elucidated the taxonomic radiation that gave rise to modern horses, zebras and donkeys by applying a capture strategy to three extinct equid lineages. In addition, hybridization capture provides additional information on the adaptation of species to their environment over time. For instance, Immel *et al*. (141) targeted and recon-

structed the mitochondrial genome of the extinct giant deer *Megaloceros giganteus*. Using these data, they defined the phylogenetic relationship of this species to contemporary taxa and suggested that it survived later in Central Europe than was previously proposed. Similarly, based on the capture of ancient mtDNA, Zhang *et al.* (142) provided evidence of the domestication of cattle in northern China several thousand years before what was previously accepted. In addition, Templeton *et al.* (143) sequenced the whole mitochondrial genomes of human remains from World War Two and 2500-year-old archeological human remains using hybridization capture, demonstrating its efficiency for the identification of individuals compared with standard nuclear short tandem repeat (STR) typing and illustrating its potential for forensic science involving both ancient and modern highly degraded DNA (144). Finally, the introduction of hybridization capture in the study of aDNA has provided information about historical human populations exposed to pathogens and epidemics and facilitated the characterization of numerous pathogen genomes through time and space. This approach has led to the reconstruction of large genomic regions of *Mycobacterium tuberculosis* (145), *Mycobacterium leprae* (146), *Vibrio cholerae* (147) and even *Yersinia pestis* (148,149) providing valuable insights into their dispersal, evolution and phylogenetic affiliations as well as variations influencing their virulence compared with actual pathogens. Hybridization capture has also been applied to nonhuman ancient pathogens, as reported by Tsangaras *et al.* (87), who targeted the koala retrovirus from modern and museum DNA. These studies proved so effective that Bos *et al.* (150) developed an array-based DNA capture screening technique for the simultaneous detection of nearly 100 human paleopathogens from ancient tissues.

Hybridization capture has therefore proven to be a particularly efficient tool for accessing unknown ancient genomic sequences based on the knowledge of a few modern divergent biomarkers. This capability has been utilized for the exploration of complex ecosystems where scarce information is available and diversity is tremendous.

METAGENOME AND METATRANSCRIPTOME CAPTURE

Microbial communities present enormous biological and functional diversity that is crucial for the function of ecosystems (151,152). Understanding these metagenomic samples requires identification of the organisms making up these communities and determination of their functions. The first studies of these complex samples using cultivation approaches restricted the description of the communities to the approximately 1% of bacterial species that were axenically culturable (153,154). To expand the characterization of community structure, molecular cultivation-independent methods such as barcoding have been used. Nevertheless, ecosystems exploration based on one or a few DNA genetic markers limits the comprehension of the studied ecosystems. Indeed, amplifying 16S small subunit ribosomal RNA (16S rRNA) genes provides information regarding the phylogenetic diversity of the studied environment (155), but does not provide information regarding the functions they carry out. The advent of NGS overcame these limitations through the shotgun sequencing of metagenomic samples, which enabled the association of microbial community structure with realized metabolic functions (156,157). However, the establishment of such link based on large genomic sequences or genome reconstruction requires sufficient sequencing coverage to be available for dominant micro-organisms, or a considerable sequencing effort to access less abundant and rare organisms (158). In contrast, Single Cell Genomics (SCG) recovers genomes from targeted individual cells present in complex environments and precisely identifies their metabolic functions; however, this technique is not always practical and is not sufficiently scalable to provide an exhaustive overview of microbial communities (159,160). More innovative, but still anecdotic, targeting enrichment approaches in the field of metagenomics, such as hybridization capture, have been developed to fill this gap in the exploration of microbial communities (Figure 2F). Indeed, although these communities are more complex considering that they are composed of a multitude of organisms, metagenomic samples are similar to nonreference species and aDNA samples in the sense that the genetic information they contain is only suspected and is not well known. The demonstrated potential of hybridization capture to enrich divergent sequence targets in such cases has therefore been applied to metagenomic samples to enrich biomarkers of interest that are known to be present but, whose variability and diversity are not known (158).

Denonfoux *et al.* (161) described the first adaptation of SHS to a complex metagenomic sample. The methodology was used to explore the methanogenic communities present in a lacustrine environment by targeting the methyl coenzyme M reductase subunit A (*mcrA*) gene with a set of nonoverlapping probes, which targeted both known sequences and potential undescribed variants of the *mcrA* gene. The *mcrA* sequences represented more than 40% of the obtained sequences after two cycles of capture, revealing enrichment compared with shotgun sequencing, in which only 0.003% of the sequences corresponded to the target gene. This approach efficiently enriched complete targets but also recovered off-target sequences adjacent to the *mcrA* gene such as parts of the *mcr* operon. The extended *mcrA* flanking regions showed evidence of the association of the *mcr* operon with Fe metabolism genes that are usually located tens or hundreds of kilobases downstream. This finding suggested that this genomic organization, which has never been described previously in methanogens, is a consequence of the adaptation of the methanogen population to the studied environment. In addition, because *mcrA* and 16S gene phylogenies are congruent, this approach allowed the methanogen community to be described and revealed higher diversity than previously observed with other methods. Indeed, hybridization capture recovered sequences from the Methanobacteriales order, belonging to the rare biosphere, which were not detected through direct sample sequencing due to the sequencing depth, or through PCR amplification, due to possible primer bias. Using the same approach, Biderre-Petit *et al.* (162) identified new *Dehalococcoidia* reductive dehalogenase homologous sequences from a metagenomic sample targeting *rdhA* genes or insertion sequence (IS) elements located nearby.

Finally, Manoharan *et al.* (163) applied hybridization capture to metagenomic soil samples. These authors highly enriched functional genes encoding carbohydrate-active enzymes and secretory proteases, while preserving the diversity of targeted biomarkers and the taxonomic distribution of micro-organisms. When applied to other genomic loci, such as 16S or 18S rRNA genes, in other metagenomic samples, the performance of hybridization capture enrichment has been shown to be greater than 90% (unpublished data from our lab). This strategy provides a better taxonomic description of ecosystems because of the more reliable diversity observed compared with conventional approaches, such as barcoding, through access to complete biomarkers and their subsequent accurate affiliation. These promising initial studies offer new strategies for the exploration of ecosystems and allow access to complete phylogenetic or functional biomarker diversity as well as providing a first glimpse into the identification of metabolic capabilities through the association of biomarkers.

To further explore metagenomic samples, we have recently adapted SHS to very large DNA fragments containing targeted biomarkers to reconstruct complete genomic regions and initiate targeted genome reconstruction (unpublished data from our lab). Sequencing data analysis revealed efficient enrichment of sequences from targeted genomes, leading to the assembly of nearly 100 kb DNA fragments including complete plasmids containing the targeted biomarker, with a very high sequence coverage. This allowed the reconstruction of metabolic pathways via biomarker association and linkage between identities and realized metabolic functions, representing a huge step forward in understanding the functions of microbial communities.

Environmental DNA is recognized as a rich source of natural products such as biocatalysts and bioactives, which are of great interest for industrial applications. Commonly, evaluation of the diversity of natural products synthesized by complex microbial communities is achieved more generally through the generation of clone libraries from metagenomic DNA and their subsequent expression in heterologous hosts (164,165). After functional screening for the desired activity, clones of interest are sequenced to determine the gene sequences encoding the metabolites of interest. Although effective, this strategy can be very tedious depending on the complexity of the metagenomic sample because of the screening step in which clones of interest must be identified among thousands of other possible clones (166). To ease this screening step, hybridization capture has been applied to environmental samples to enrich genes involved in specific environmental processes prior to clone library construction and analysis. Thus, Bragalini *et al.* (167) employed SHS to isolate glycoside hydrolase cDNAs from forest soil samples using explorative nonoverlapping capture probes targeting referenced and nonreferenced genes variants. The captured fragments were cloned into an expression vector and transformed into yeast for heterologous cDNA expression. Sequencing of randomly selected clones revealed that, after enrichment, targeted genes represented more than 90% and up to 100% of the obtained sequences, among which approximately 70% were of full-length, allowing their expression in the heterologous system. Hence, up to 25% of the screened recombinant yeasts exhibited secretion of endoxylanases, greatly reducing the depth required for library screening. Finally, as previously observed, hybridization capture also succeeded in selecting phylogenetically diverse representatives of the targeted gene family because none of the captured sequences were identical to the sequences described in databases and used for probe design.

As illustrated here, the success of hybridization capture experiments for the exploration of metagenomic samples strongly depends on probe design, including variant-specific and explorative probes to provide comprehensive diversity. Dedicated software has been developed to meet this specific need, such as PhylArray (168), MetabolicDesign (169), HiSpOD (170) and KASpOD (171). Therefore, hybridization capture for the exploration of metagenomic samples appears to be a particularly promising strategy with diverse applications.

CONCLUSION

In the light of the studies performed to date, hybridization capture appears to be a particularly innovative, useful and effective approach for studying key genomic subsets in numerous biological domains. In resequencing approaches for medical research, hybridization capture through exome targeting has proven to be an extremely useful tool for the identification of mutations underlying Mendelian diseases as well as complex diseases and cancers, which has shed light on the ability of this approach to diagnose pathologies. This partitioning approach has also led to applications in the agronomical field to associate genetic variants with phenotypic traits of agricultural interest in crops and livestock. Moreover, because of the ability of hybridization capture to enrich numerous defined targets as well as their flanking regions, this strategy has broadened our knowledge of relevant underexploited fractions of the genomes of various models. Its capacity to extend what is already known from the genomes of organisms has been demonstrated through its application to nonreference species and paleogenomic samples on the basis of restricted genomic information. Finally, even if the use of hybridization capture to explore metagenomic samples remains anecdotic, its application allows us to better understand these complex ecosystems beyond the limits of conventional molecular approaches.

Hybridization capture strongly depends on NGS technologies and greatly benefits from the massive amounts of enriched sequence data provided. However, as sequencing technologies rapidly improve and the cost per sequenced nucleotide plummets, the price of whole-genome sequencing could quickly become comparable to the cost of exome sequencing, and deep insights into complex metagenomic samples could be achieved with few sequencing runs. Nevertheless, a niche may always remain for this technology because many biological questions do not require exhaustive sequences, but only restricted and targeted information. In addition, decreasing sequencing costs will be associated with an increase in the amount of data requiring processing and storage, with additional costs. Finally, the third-generation sequencing platforms with the possibility of sequencing longer DNA fragments provides real benefits to hybridization capture to span repetitive sequences.

It is therefore suitable for studying complex chromosomal structural variations, especially those involving repeats. But other genomic research applications, such as small insertion and deletion validation, haplotype phasing and microbial genome reconstruction from metagenomics samples could also benefit from this technology. In the latter case, these data will help us to improve our knowledge of the functioning of ecosystems, which remains a less prominent field, but likely one of the richest.

Hybridization capture is a particularly powerful method with a firmly established role in genome research. It has allowed the discovery of many unsuspected findings that would have been difficult to obtain using other approaches. New technical improvements can be made, but the application of this methodology to other biological fields is still underexplored and will provide influential results.

FUNDING

French 'Direction Générale de l'Armement' (DGA); program Investissements d'Avenir AMI 2011 VALTEX; ANR-RF-2015-01 PEROXIDIV; Auvergne Regional Council; European Regional Development Fund (FEDER). Funding for open access charge: Université d'Auvergne.
Conflict of interest statement. None declared.

REFERENCES

1. Shendure,J. and Ji,H. (2008) Next-generation DNA sequencing. *Nat. Biotechnol.*, **26**, 1135–1145.
2. Desai,N., Antonopoulos,D., Gilbert,J.A., Glass,E.M. and Meyer,F. (2012) From genomics to metagenomics. *Curr. Opin. Biotechnol.*, **23**, 72–76.
3. Hayden,E.C. (2014) The $1,000 genome. *Nature*, **507**, 294–295.
4. Mamanova,L., Coffey,A.J., Scott,C.E., Kozarewa,I., Turner,E.H., Kumar,A., Howard,E., Shendure,J. and Turner,D.J. (2010) Target-enrichment strategies for next-generation sequencing. *Nat. Methods*, **7**, 111–118.
5. Edwards,M.C. and Gibbs,R.A. (1994) Multiplex PCR?: advantages, development, and applications. *PCR Methods Appl.*, **3**, S65–S75.
6. Williams,R., Peisajovich,S.G., Miller,O.J., Magdassi,S., Tawfik,D.S. and Griffiths,A.D. (2006) Amplification of complex gene libraries by emulsion PCR. *Nat. Methods*, **3**, 545–550.
7. Tewhey,R., Warner,J.B., Nakano,M., Libby,B., Medkova,M., David,P.H., Kotsopoulos,S.K., Samuels,M.L., Hutchison,J.B., Larson,J.W. *et al.* (2009) Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat. Biotechnol.*, **27**, 1025–1031.
8. Turner,E.H., Ng,S.B., Nickerson,D.A. and Shendure,J. (2009) Methods for genomic partitioning. *Annu. Rev. Genomics Hum. Genet.*, **10**, 263–284.
9. Nilsson,M., Malmgren,H., Samiotaki,M., Kwiatkowski,M., Chowdhary,B.P. and Landegren,U. (1994) Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science*, **265**, 2085–2088.
10. Krishnakumar,S., Zheng,J., Wilhelmy,J., Faham,M., Mindrinos,M. and Davis,R. (2008) A comprehensive assay for targeted multiplex amplification of human DNA sequences. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 9296–9301.
11. Albert,T.J., Molla,M.N., Muzny,D.M., Nazareth,L., Wheeler,D., Song,X., Richmond,T.A., Middle,C.M., Rodesch,M.J., Packard,C.J. *et al.* (2007) Direct selection of human genomic loci by microarray hybridization. *Nat. Methods*, **4**, 903–905.
12. Okou,D.T., Steinberg,K.M., Middle,C., Cutler,D.J., Albert,T.J. and Zwick,M.E. (2007) Microarray-based genomic selection for high-throughput resequencing. *Nat. Methods*, **4**, 907–909.
13. Gnirke,A., Melnikov,A., Maguire,J., Rogov,P., LeProust,E.M., Brockman,W., Fennell,T., Giannoukos,G., Fisher,S., Russ,C. *et al.* (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.*, **27**, 182–189.
14. Asan, Xu,Y., Jiang,H., Tyler-Smith,C., Xue,Y., Jiang,T., Wang,J., Wu,M., Liu,X., Tian,G. *et al.* (2011) Comprehensive comparison of three commercial human whole-exome capture platforms. *Genome Biol.*, **12**, R95.
15. Chilamakuri,C.S.R., Lorenz,S., Madoui,M-A., Vodak,D., Sun,J., Hovig,E., Myklebost,O. and Meza-Zepeda,L. (2014) Performance comparison of four exome capture systems for deep sequencing. *Nat. Protoc.*, **12**, 423–425.
16. Meienberg,J., Zerjavic,K., Keller,I., Okoniewski,M., Patrignani,A., Ludin,K., Xu,Z., Steinmann,B., Carrel,T., Rothlisberger,B. *et al.* (2015) New insights into the performance of human whole-exome capture platforms. *Nucleic Acids Res.*, **43**, e76.
17. Shigemizu,D., Momozawa,Y., Abe,T., Morizono,T., Boroevich,K.A., Takata,S., Ashikawa,K., Kubo,M. and Tsunoda,T. (2015) Performance comparison of four commercial human whole-exome capture platforms. *Sci. Rep.*, **5**, 12742.
18. Marciniak,S., Klunk,J., Devault,A., Enk,J. and Poinar,H.N. (2015) Ancient human genomics: the methodology behind reconstructing evolutionary pathways. *J. Hum. Evol.*, **79**, 21–34.
19. Fox,E.J., Reid-Bayliss,K.S., Emond,M.J. and Loeb,L.A. (2014) Accuracy of next generation sequencing platforms. *Next Gener. Seq. Appl.*, **1**, pii: 1000106.
20. Fan,J-B., Chee,M.S. and Gunderson,K.L. (2006) Highly parallel genomic assays. *Nat. Rev. Genet.*, **7**, 632–644.
21. International HapMap Consortium. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851–861.
22. Zeng,P., Zhao,Y., Qian,C., Zhang,L., Zhang,R., Gou,J., Liu,J., Liu,L. and Chen,F. (2015) Statistical analysis for genome-wide association study. *J. Biomed. Res.*, **29**, 285–297.
23. Beaudoin,M., Goyette,P., Boucher,G., Lo,K.S., Rivas,M.A., Stevens,C., Alikashani,A., Ladouceur,M., Ellinghaus,D., Törkvist,L. *et al.* (2013) Deep resequencing of GWAS Loci Identifies Rare Variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis. *PLoS Genet.*, **9**, e1003723.
24. Ruark,E., Snape,K., Humburg,P., Loveday,C., Bajrami,I., Brough,R., Rodrigues,D.N., Renwick,A., Seal,S., Ramsay,E. *et al.* (2013) Mosaic PPM1D mutations are associated with predisposition to breast and ovarian cancer. *Nature*, **493**, 406–410.
25. Clarke,W.E., Parkin,I.A., Gajardo,H.A., Gerhardt,D.J., Higgins,E., Sidebottom,C., Sharpe,A.G., Snowdon,R.J., Federico,M.L. and Iniguez-Luy,F.L. (2013) Genomic DNA Enrichment Using Sequence Capture Microarrays?: a Novel Approach to Discover Sequence Nucleotide Polymorphisms (SNP) in *Brassica napus* L. *PLoS One*, **8**, e81992.
26. Schiessl,S., Samans,B., Hüttel,B., Reinhard,R. and Snowdon,R.J. (2014) Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. *Front. Plant Sci.*, **5**, 404.
27. Tennessen,J.A., Govindarajulu,R., Ashman,L. and Liston,A. (2014) Evolutionary Origins and Dynamics of Octoploid Strawberry Subgenomes Revealed by Dense Targeted Capture Linkage Maps. *Genome Biol. Evol.*, **6**, 3295–3313.
28. Jupe,F., Witek,K., Verweij,W., Sliwka,J., Pritchard,L., Etherington,G.J., Maclean,D., Cock,P.J., Leggett,R.M., Bryan,G.J. *et al.* (2013) Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.*, **76**, 530–544.
29. Fu,Y., Springer,N.M., Gerhardt,D.J., Ying,K., Yeh,C., Wu,W., Swanson-Wagner,R., D'Ascenzo,M., Millard,T., Freeberg,L. *et al.* (2010) Repeat subtraction-mediated sequence capture from a complex genome. *Plant J.*, **62**, 898–909.
30. Faucon,F., Dusfour,I., Gaude,T., Navratil,V., Boyer,F., Chandre,F., Sirisopa,P., Thanispong,K., Juntarajumnong,W., Poupardin,R. *et al.* (2015) Identifying genomic changes associated with insecticide resistance in the dengue mosquito Aedes aegypti by deep targeted sequencing. *Genome Res.*, **25**, 1347–1359.
31. Wang,M., Beck,C.R., English,A.C., Meng,Q., Buhay,C., Han,Y., Doddapaneni,H. V, Yu,F., Boerwinkle,E., Lupski,J.R. *et al.* (2015) PacBio-LITS: a large-insert targeted sequencing method for

characterization of human disease-associated chromosomal structural variations. *BMC Genomics*, **16**, 214.
32. Hodges,E., Xuan,Z., Balija,V., Kramer,M., Molla,M.N., Smith,S.W., Middle,C.M., Rodesch,M.J., Albert,T.J., Hannon,G.J. *et al.* (2007) Genome-wide in situ exon capture for selective resequencing. *Nat. Genet.*, **39**, 1522–1527.
33. Cheng,J., Kapranov,P., Drenkow,J., Dike,S., Brubaker,S., Patel,S., Long,J., Stern,D., Tammana,H., Helt,G. *et al.* (2005) Transcriptional maps of 10 Human Chromosomes at 5-Nucleotide Resolution. *Science*, **308**, 1149–1154.
34. Wang,Z., Liu,X., Yang,B.-Z. and Gelernter,J. (2013) The Role and Challenges of Exome Sequencing in Studies of Human Diseases. *Front. Genet.*, **4**, 1–8.
35. MacArthur,D.G., Manolio,T.A., Dimmock,D.P., Rehm,H.L., Shendure,J., Abecasis,G.R., Adams,D.R., Altman,R.B., Antonarakis,S.E., Ashley,E.A. *et al.* (2014) Guidelines for investigating causality of sequence variants in human disease. *Nature*, **508**, 469–476.
36. Cirulli,E.T. and Goldstein,D.B. (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.*, **11**, 415–425.
37. Yang,Y., Muzny,D.M., Reid,J.G., Bainbridge,M.N., Willis,A., Ward,P.A., Braxton,A., Beuten,J., Xia,F., Niu,Z. *et al.* (2013) Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.*, **369**, 1502–1511.
38. Meynert,A.M., Ansari,M., FitzPatrick,D.R. and Taylor,M.S. (2014) Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics*, **15**, 247.
39. Warr,A., Robert,C., Hume,D., Archibald,A., Deeb,N. and Watson,M. (2015) Exome Sequencing: Current and Future Perspectives. *G3*, **5**, 1543–1550.
40. Stenson,P.D., Ball,E.V., Howells,K., Phillips,A.D., Mort,M. and Cooper,D.N. (2009) The Human Gene Mutation Database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. *Hum. Genomics*, **4**, 69–72.
41. Stranneheim,H. and Wedell,A. (2016) Exome and genome sequencing: a revolution for the discovery and diagnosis of monogenic disorders. *J. Intern. Med.*, **279**, 3–15.
42. Gilissen,C., Hoischen,A., Brunner,H.G. and Veltman,J.A. (2012) Disease gene identification strategies for exome sequencing. *Eur. J. Hum. Genet.*, **20**, 490–497.
43. Ng,S.B., Buckingham,K.J., Lee,C., Bigham,A.W., Tabor,H.K., Dent,K.M., Huff,C.D., Shannon,P.T., Jabs,E.W., Nickerson,D.A. *et al.* (2010) Exome sequencing identifies the cause of a Mendelian disorder. *Nat. Genet.*, **42**, 30–35.
44. Ng,S.B., Bigham,A.W., Buckingham,K.J., Hannibal,M.C., McMillin,M., Gildersleeve,H., Beck,A.E., Tabor,H.K., Cooper,G.M., Mefford,C. *et al.* (2011) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.*, **42**, 790–793.
45. Yang,Z., Xu,Y., Luo,H., Ma,X., Wang,Q., Wang,Y., Deng,W., Jiang,T., Sun,G., He,T. *et al.* (2014) Whole-exome sequencing for the identification of susceptibility genes of Kashin-Beck disease. *PLoS One*, **9**, e92298.
46. Peters,H., Buck,N., Wanders,R., Ruiter,J., Waterham,H., Koster,J., Yaplito-Lee,J., Ferdinandusse,S. and Pitt,J. (2014) ECHS1 mutations in Leigh disease: a new inborn error of metabolism affecting valine metabolism. *Brain*, **137**, 2903–2908.
47. Marian,A. (2012) Molecular genetic studies of complex phenotypes. *Transl. Res.*, **159**, 64–79.
48. Schork,N.J., Murray,S.S., Frazer,K.A. and Topol,E.J. (2009) Common vs. rare allele hypotheses for complex diseases. *Curr. Opin. Genet. Dev.*, **19**, 212–219.
49. Huyghe,J.R., Jackson,A.U., Fogarty,M.P., Buchkovich,M.L., Stancakova,A., Stringham,H.M., Sim,X., Yang,L., Fuchsberger,C., Cederberg,H. *et al.* (2013) Exome array analysis identifies novel loci and low-frequency variants for insulin processing and secretion. *Nat. Genet.*, **45**, 197–201.
50. Do,R., Stitziel,N.O., Won,H-H., Jørgensen,A.B., Duga,S., Angelica Merlini,P., Kiezun,A., Farrall,M., Goel,A., Zuk,O. *et al.* (2015) Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature*, **518**, 102–106.
51. Rabbani,B., Tekin,M. and Mahdieh,N. (2014) The promise of whole-exome sequencing in medical genetics. *J. Hum. Genet.*, **59**, 5–15.
52. Schmitt,M.W., Prindle,M.J. and Loeb,L.A. (2012) Implications of genetic heterogeneity in cancer. *Ann. N. Y. Acad. Sci.*, **1267**, 110–116.
53. Agrawal,N., Frederick,M.J., Pickering,C.R., Bettegowda,C., Chang,K., Li,R.J., Fakhry,C., Xie,T., Zhang,J., Wang,J. *et al.* (2011) Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science*, **333**, 1154–1157.
54. Ku,C.-S., Cooper,D.N., Polychronakos,C., Naidoo,N., Wu,M. and Soong,R. (2012) Exome sequencing: Dual role as a discovery and diagnostic tool. *Ann. Neurol.*, **71**, 5–14.
55. Worthey,E.A., Mayer,A.N., Syverson,G.D., Helbling,D., Bonacci,B.B., Decker,B., Serpe,J.M., Dasu,T., Tschannen,M.R., Veith,R.L. *et al.* (2011) Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet. Med.*, **13**, 255–262.
56. Choi,M., Scholl,U.I., Ji,W., Liu,T., Tikhonova,I.R., Zumbo,P., Nayir,A., Bakkaloğlu,A., Ozen,S., Sanjad,S. *et al.* (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 19096–19101.
57. Firth,H.V. and Wright,C.F. (2011) The Deciphering Developmental Disorders (DDD) study. *Dev. Med. Child Neurol.*, **53**, 702–703.
58. The Deciphering Developmental Disorders Study (2015) Large-scale discovery of novel genetic causes of developmental disorders. *Nature*, **519**, 223–228.
59. Alemán,A., Garcia-Garcia,F., Medina,I. and Dopazo,J. (2014) A web tool for the design and management of panels of genes for targeted enrichment and massive sequencing for clinical applications. *Nucleic Acids Res.*, **42**, 83–87.
60. Wang,J., Liao,J., Zhang,J., Cheng,W.-Y., Hakenberg,J., Ma,M., Webb,B.D., Ramasamudram-Chakravarthi,R., Karger,L., Mehta,L. *et al.* (2015) ClinLabGeneticist: a tool for clinical management of genetic variants from whole exome sequencing in clinical genetic laboratories. *Genome Med.*, **7**, 77.
61. Mercer,T.R., Gerhardt,D.J., Dinger,M.E., Crawford,J., Trapnell,C., Jeddeloh,J.A., Mattick,J.S. and Rinn,J.L. (2011) Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.*, **30**, 99–104.
62. Halvardson,J., Zaghlool,A. and Feuk,L. (2013) Exome RNA sequencing reveals rare and novel alternative transcripts. *Nucleic Acids Res.*, **41**, e6.
63. Cabanski,C.R., Magrini,V., Griffith,M., Griffith,O.L., McGrath,S., Zhang,J., Walker,J., Ly,A., Demeter,R., Fulton,R.S. *et al.* (2014) cDNA hybrid capture improves transcriptome analysis on low-input and archived samples. *J. Mol. Diagnostics*, **16**, 440–451.
64. Cieslik,M., Chugh,R., Wu,Y., Wu,M., Brennan,C., Lonigro,R., Su,F., Wang,R., Siddiqui,J., Mehra,R. *et al.* (2015) The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res.*, **25**, 1372–1381.
65. Clark,M.B., Mercer,T.R., Bussotti,G., Leonardi,T., Haynes,K.R., Crawford,J., Brunck,M.E., Le Cao,K.-A., Thomas,G.P., Chen,W.Y. *et al.* (2015) Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing. *Nat. Methods*, **12**, 339–342.
66. Fairfield,H., Gilbert,G.J., Barter,M., Corrigan,R.R., Curtain,M., Ding,Y., D'Ascenzo,M., Gerhardt,D.J., He,C., Huang,W. *et al.* (2011) Mutation discovery in mice by whole exome sequencing. *Genome Biol.*, **12**, R86.
67. Robert,C., Fuentes-Utrilla,P., Troup,K., Loecherbach,J., Turner,F., Talbot,R., Archibald,A.L., Mileham,A., Deeb,N., Hume,D.A. *et al.* (2014) Design and development of exome capture sequencing for the domestic pig (*Sus scrofa*). *BMC Genomics*, **15**, 550.
68. Broeckx,B.J.G., Coopman,F., Verhoeven,G.E.C., Bavegems,V., De Keulenaer,S., De Meester,E., Van Niewerburgh,F. and Deforce,D. (2014) Development and performance of a targeted whole exome sequencing enrichment kit for the dog (Canis Familiaris Build 3.1). *Sci. Rep.*, **4**, 5597.
69. Broeckx,B.J.G., Hitte,C., Coopman,F., Verhoeven,G.E.C., De Keulenaer,S., De Meester,E., Derrien,T., Alfoldi,J., Lindblad-Toh,K., Bosmans,T. *et al.* (2015) Improved canine exome designs, featuring ncRNAs and increased coverage of protein coding genes. *Sci. Rep.*, **5**, 12810.

70. MacArthur,D.G. and Tyler-Smith,C. (2010) Loss-of-function variants in the genomes of healthy humans. *Hum. Mol. Genet.*, **19**, R125–R130.
71. Mascher,M., Richmond,T.A., Gerhardt,D.J., Himmelbach,A., Clissold,L., Sampath,D., Ayling,S., Steuernagel,B., Pfeifer,M., D'Ascenzo,M. *et al.* (2013) Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J.*, **76**, 494–505.
72. Saintenac,C., Jiang,D. and Akhunov,E.D. (2011) Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biol.*, **12**, R88.
73. Muraya,M.M., Schmutzer,T., Ulpinnis,C., Scholz,U. and Altmann,T. (2015) Targeted sequencing reveals large-scale sequence polymorphism in maize candidate genes for biomass production and composition. *PLoS One*, **10**, e0132120.
74. Jiang,L., Liu,X., Yang,J., Wang,H., Jiang,J., Liu,L., He,S., Ding,X., Liu,J. and Zhang,Q. (2014) Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. *BMC Genomics*, **15**, 1105.
75. McClure,M.C., Bickhart,D., Null,D., Vanraden,P., Xu,L., Wiggans,G., Liu,G., Schroeder,S., Glasscock,J., Armstrong,J. *et al.* (2014) Bovine exome sequence analysis and targeted SNP genotyping of recessive fertility defects BH1, HH2, and HH3 reveal a putative causative mutation in SMC2 for HH3. *PLoS One*, **9**, e92769.
76. Mullen,M.P., Creevey,C.J., Berry,D.P., McCabe,M.S., Magee,D.A., Howard,D.J., Killeen,A.P., Park,S.D., McGettigan,P.A., Lucy,M.C. *et al.* (2012) Polymorphism discovery and allele frequency estimation using high-throughput DNA sequencing of target-enriched pooled DNA samples. *BMC Genomics*, **13**, 16.
77. Dasgupta,M.G., Dharanishanthi,V., Agarwal,I. and Krutovsky,K. V (2015) Development of genetic markers in Eucalyptus species by target enrichment and exome sequencing. *PLoS One*, **10**, e0116528.
78. Zhou,L. and Holliday,J.A. (2012) Targeted enrichment of the black cottonwood (*Populus trichocarpa*) gene space using sequence capture. *BMC Genomics*, **13**, 703.
79. Allen,A.M., Barker,G.L.A., Wilkinson,P., Burridge,A., Winfield,M., Coghill,J., Uauy,C., Griffiths,S., Jack,P., Berry,S. *et al.* (2013) Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). *Plant Biotechnol. J.*, **11**, 279–295.
80. Mascher,M., Jost,M., Kuon,J.-E., Himmelbach,A., Aßfalg,A., Beier,S., Scholz,U., Graner,A. and Stein,N. (2014) Mapping-by-sequencing accelerates forward genetics in barley. *Genome Biol.*, **15**, R78.
81. Evans,J., Kim,J., Childs,K.L., Vaillancourt,B., Crisovan,E., Nandety,A., Gerhardt,D.J., Richmond,T.A., Jeddeloh,J.A., Kaeppler,S.M. *et al.* (2014) Nucleotide polymorphism and copy number variant detection using exome capture and next-generation sequencing in the polyploid grass *Panicum virgatum*. *Plant J.*, **79**, 993–1008.
82. Carter,N.P. (2007) Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat. Genet.*, **39**, S16–S21.
83. Henry,I.M., Nagalakshmi,U., Lieberman,M.C., Ngo,K.J., Krasileva,K. V., Vasquez-Gross,H., Akhunova,A., Akhunov,E., Dubcovsky,J., Tai,T.H. *et al.* (2014) Efficient genome-wide detection and cataloging of EMS-Induced mutations using exome capture and next-generation sequencing. *Plant Cell*, **26**, 1382–1397.
84. King,R., Bird,N., Ramirez-Gonzalez,R., Coghill,J.A., Patil,A., Hassani-Pak,K., Uauy,C. and Phillips,A.L. (2015) Mutation scanning in wheat by exon capture and next-generation sequencing. *PLoS One*, **10**, e0137549.
85. Bolon,Y.-T., Haun,W.J., Xu,W.W., Grant,D., Stacey,M.G., Nelson,R.T., Gerhardt,D.J., Jeddeloh,J.A., Stacey,G., Muehlbauer,G.J. *et al.* (2011) Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. *Plant Physiol.*, **156**, 240–253.
86. Samuels,D.C., Han,L., Li,J., Quanghu,S., Clark,T.A., Shyr,Y. and Guo,Y. (2013) Finding the lost treasures in exome sequencing data. *Trends Genet.*, **29**, 593–599.
87. Tsangaras,K., Siracusa,M.C., Nikolaidis,N., Ishida,Y., Cui,P., Vielgrader,H., Helgen,K.M., Roca,A.L. and Greenwood,A.D. (2014) Hybridization capture reveals evolution and conservation across the entire Koala retrovirus genome. *PLoS One*, **9**, e95633.
88. Guo,Y., Long,J., He,J., Li,C.-I., Cai,Q., Shu,X.-O., Zheng,W. and Li,C. (2012) Exome sequencing generates high quality data in non-target regions. *BMC Genomics*, **13**, 194.
89. Yi,X., Liang,Y., Huerta-Sanchez,E., Jin,X., Cuo,Z.X.P., Pool,J.E., Xu,X., Jiang,H., Vinckenbosch,N., Korneliussen,T.S. *et al.* (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, **329**, 75–78.
90. Picardi,E. and Pesole,G. (2012) Mitochondrial genomes gleaned from human whole-exome sequencing. *Nat. Methods*, **9**, 523–524.
91. Robin,E.D. and Wong,R. (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J. Cell. Physiol.*, **136**, 507–513.
92. Griffin,H.R., Pyle,A., Blakely,E.L., Alston,C.L., Duff,J., Hudson,G., Horvath,R., Wilson,I.J., Santibanez-Koref,M., Taylor,R.W. *et al.* (2014) Accurate mitochondrial DNA sequencing using off-target reads provides a single test to identify pathogenic point mutations. *Genet. Med.*, **16**, 962–971.
93. Zhang,P., Samuels,D.C., Lehmann,B., Stricker,T., Pietenpol,J., Shyr,Y. and Guo,Y. (2015) Mitochondria sequence mapping strategies and practicability of mitochondria variant detection from exome and RNA sequencing data. *Brief. Bioinform.*, **17**, 224–232.
94. Dinwiddie,D.L., Smith,L.D., Miller,N.A., Atherton,A.M., Farrow,E.G., Strenk,M.E., Soden,S.E., Saunders,C.J. and Kingsmore,S.F. (2013) Diagnosis of mitochondrial disorders by concomitant next-generation sequencing of the exome and mitochondrial genome. *Genomics*, **102**, 148–156.
95. Guo,Y., Li,J., Li,C.-I., Shyr,Y. and Samuels,D.C. (2013) MitoSeek: extracting mitochondria information and performing high-throughput mitochondria sequencing analysis. *Bioinformatics*, **29**, 1210–1211.
96. Duncavage,E.J., Magrini,V., Becker,N., Armstrong,J.R., Demeter,R.T., Wylie,T., Abel,H.J. and Pfeifer,J.D. (2011) Hybrid capture and next-generation sequencing identify viral integration sites from formalin-fixed, paraffin-embedded tissue. *J. Mol. Diagnostics*, **13**, 325–333.
97. Ma,Z. and Axtell,M.J. (2013) Long-range genomic enrichment, sequencing, and assembly to determine unknown sequences flanking a known microRNA. *PLoS One*, **8**, e83721.
98. Platt,R.N., Zhang,Y., Witherspoon,D.J., Xing,J., Suh,A., Keith,M.S., Jorde,L.B., Stevens,R.D. and Ray,D.A. (2015) Targeted Capture of Phylogenetically Informative Ves SINE Insertions in Genus Myotis. *Genome Biol. Evol.*, **7**, 1664–1675.
99. Jones,M.R. and Good,J.M. (2015) Targeted capture in evolutionary and ecological genomics. *Mol. Ecol.*, **25**, 185–202.
100. Peñalba,J. V., Smith,L.L., Tonione,M.A., Sass,C., Hykin,S.M., Skipwith,P.L., McGuire,J.A., Bowie,R.C.K. and Moritz,C. (2014) Sequence capture using PCR-generated probes: a cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Mol. Ecol. Resour.*, **14**, 1000–1010.
101. Weitemier,K., Straub,S.C.K., Cronn,R.C., Fishbein,M., Schmickl,R., McDonnell,A. and Liston,A. (2014) Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. *Appl. Plant Sci.*, **2**, 1400042.
102. Nicholls,J.A., Pennington,R.T., Koenen,E.J.M., Hughes,C.E., Hearn,J., Bunnefeld,L., Dexter,K.G., Stone,G.N. and Kidner,C.A. (2015) Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus Inga (Leguminosae?: Mimosoideae). *Front. Plant Sci.*, **6**, 710.
103. Hebert,F.O., Renaut,S. and Bernatchez,L. (2013) Targeted sequence capture and resequencing implies a predominant role of regulatory regions in the divergence of a sympatric lake whitefish species pair (*Coregonus clupeaformis*). *Mol. Ecol.*, **22**, 4896–4914.
104. Neves,L.G., Davis,J.M., Barbazuk,W.B. and Kirst,M. (2013) Whole-exome targeted sequencing of the uncharacterized pine genome. *Plant J.*, **75**, 146–156.
105. Pootakham,W., Shearman,J.R., Ruang-Areerate,P., Sonthirod,C., Sangsrakru,D., Jomchai,N., Yoocha,T., Triwitayakorn,K., Tragoonrung,S. and Tangphatsornruang,S. (2014) Large-scale SNP discovery through RNA sequencing and SNP genotyping by targeted enrichment sequencing in cassava (*Manihotesculenta* Crantz). *PLoS One*, **9**, e116028.
106. Day,K., Song,J. and Absher,D. (2014) Targeted sequencing of large genomic regions with CATCH-Seq. *PLoS One*, **9**, e111756.

107. Good,J.M., Wiebe,V., Albert,F.W., Burbano,H.A., Kircher,M., Green,R.E., Halbwax,M., André,C., Atencia,R., Fischer,A. *et al.* (2013) Comparative population genomics of the ejaculate in humans and the great apes. *Mol. Biol. Evol.*, **30**, 964–976.
108. Li,C., Hofreiter,M., Straube,N., Corrigan,S. and Naylor,G.J.P. (2013) Capturing protein-coding genes across highly divergent species. *Biotechniques*, **54**, 321–326.
109. Bundock,P.C., Casu,R.E. and Henry,R.J. (2012) Enrichment of genomic DNA for polymorphism detection in a non-model highly polyploid crop plant. *Plant Biotechnol. J.*, **10**, 657–667.
110. Hedtke,S.M., Morgan,M.J., Cannatella,D.C. and Hillis,D.M. (2013) Targeted enrichment: maximizing orthologous gene comparisons across deep evolutionary time. *PLoS One*, **8**, e67908.
111. Neves,L.G., Davis,J.M., Barbazuk,W.B. and Kirst,M. (2014) A high-density gene map of loblolly pine (*Pinus taeda* L.) based on exome sequence capture genotyping. *G3*, **4**, 29–37.
112. Rosani,U., Domeneghetti,S., Pallavicini,A. and Venier,P. (2014) Target capture and massive sequencing of genes transcribed in *Mytilus galloprovincialis*. *Biomed Res. Int.*, **2014**, 538549.
113. George,R.D., McVicker,G., Diederich,R., Ng,S.B., MacKenzie,A.P., Swanson,W.J., Shendure,J. and Thomas,J.H. (2011) Trans genomic capture and sequencing of primate exomes reveals new targets of positive selection. *Genome Res.*, **21**, 1686–1694.
114. Seoane-Zonjic,P., Cañas,R.A., Bautista,R., Gómez-Maldonado,J., Arrillaga,I., Fernández-Pozo,N., Claros,M.G., Cánovas,F.M. and Ávila,C. (2016) Establishing gene models from the *Pinus pinaster* genome using gene capture and BAC sequencing. *BMC Genomics*, **17**, 148.
115. Nadeau,N.J., Whibley,A., Jones,R.T., Davey,J.W., Dasmahapatra,K.K., Baxter,S.W., Quail,M.A., Joron,M., ffrench-Constant,R.H., Blaxter,M.L. *et al.* (2012) Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. *Philos. Trans. R. Soc. B Biol. Sci.*, **367**, 343–353.
116. Cosart,T., Beja-Pereira,A., Chen,S., Ng,S.B., Shendure,J. and Luikart,G. (2011) Exome-wide DNA capture and next generation sequencing in domestic and wild species. *BMC Genomics*, **12**, 347.
117. Schneider,A., Henegar,C., Day,K., Absher,D., Menotti-raymond,M., Barsh,G.S. and Eizirik,E. (2015) Recurrent Evolution of Melanism in South American Felids. *PloS One*, **10**, e1004892.
118. Stephens,J.D., Rogers,W.L., Heyduk,K., Cruse-Sanders,J.M., Determann,R.O., Glenn,T.C. and Malmberg,R.L. (2015) Resolving phylogenetic relationships of the recently radiated carnivorous plant genus Sarracenia using target enrichment. *Mol. Phylogenet. Evol.*, **85**, 76–87.
119. Prum,R.O., Berv,J.S., Dornburg,A., Field,D.J., Townsend,J.P., Lemmon,E.M. and Lemmon,A.R. (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, **562**, 569–573.
120. Faircloth,B.C., Sorenson,L., Santini,F. and Alfaro,M.E. (2013) A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). *PLoS One*, **8**, e65923.
121. Faircloth,B.C., Branstetter,M.G., White,N.D. and Brady,S.G. (2015) Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Mol. Ecol. Resour.*, **15**, 489–501.
122. Wang,B. and Wang,X-R. (2014) Mitochondrial DNA capture and divergence in Pinus provide new insights into the evolution of the genus. *Mol. Phylogenet. Evol.*, **80**, 20–30.
123. Hofreiter,M., Paijmans,J.L.A., Goodchild,H., Speller,C.F., Barlow,A., Fortes,G.G., Thomas,J.A., Ludwig,A. and Collins,M.J. (2014) The future of ancient DNA: technical advances and conceptual shifts. *BioEssays*, **37**, 284–293.
124. Pääbo,S., Poinar,H., Serre,D., Jaenicke-Després,V., Hebler,J., Rohland,N., Kuch,M., Krause,J., Vigilant,L. and Hofreiter,M. (2004) Genetic analyses from ancient DNA. *Annu. Rev. Genet.*, **38**, 645–679.
125. Burrell,A.S., Disotell,T.R. and Bergey,C.M. (2015) The use of museum specimens with high-throughput DNA sequencers. *J. Hum. Evol.*, **79**, 35–44.
126. Green,R.E., Malaspinas,A., Krause,J., Briggs,A.W., Johnson,P.L.F., Uhler,C., Meyer,M., Good,J.M., Maricic,T., Stenzel,U. *et al.* (2008) A complete Neanderthal mitochondrial genome sequence determined by high-throughput sequencing. *Cell*, **134**, 416–426.
127. Green,B.D. and Keller,M. (2006) Capturing the uncultivated majority. *Curr. Opin. Biotechnol.*, **17**, 236–240.
128. Ávila-Arcos,M.C., Cappellini,E., Romero-Navarro,J.A., Wales,N., Moreno-Mayar,J.V., Rasmussen,M., Fordyce,S.L., Montiel,R., Vielle-Calzada,J-P., Willerslev,E. *et al.* (2011) Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Sci. Rep.*, **1**, 74.
129. Carpenter,M.L., Buenrostro,J.D., Valdiosera,C., Schroeder,H., Allentoft,M.E., Sikora,M., Rasmussen,M., Gravel,S., Guillén,S., Nekhrizov,G. *et al.* (2013) Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *Am. J. Hum. Genet.*, **93**, 852–864.
130. Green,R.E., Krause,J., Briggs,A.W., Maricic,T., Stenzel,U., Kircher,M., Patterson,N., Li,H., Zhai,W., Fritz,M.H-Y. *et al.* (2010) A draft sequence of the neandertal genome. *Science*, **328**, 710–722.
131. Lambert,D.M., Ritchie,P.A., Millar,C.D., Holland,B., Drummond,A.J. and Baroni,C. (2002) Rates of evolution in ancient DNA from Adélie penguins. *Science*, **295**, 2270–2273.
132. Barta,J.L., Monroe,C., Teisberg,J.E., Winters,M., Flanigan,K. and Kemp,B.M. (2014) One of the key characteristics of ancient DNA, low copy number, may be a product of its extraction. *J. Archaeol. Sci.*, **46**, 281–289.
133. Brotherton,P., Endicott,P., Sanchez,J.J., Beaumont,M., Barnett,R., Austin,J. and Cooper,A. (2007) Novel high-resolution characterization of ancient DNA reveals C>U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res.*, **35**, 5717–5728.
134. Gansauge,M.-T. and Meyer,M. (2014) Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res.*, **24**, 1543–1549.
135. Knapp,M. and Hofreiter,M. (2010) Next generation sequencing of ancient DNA: requirements, strategies and perspectives. *Genes*, **1**, 227–243.
136. Noonan,J.P., Coop,G., Kudaravalli,S., Smith,D., Krause,J., Alessi,J., Chen,F., Platt,D., Paano,S., Pritchard,J.K. *et al.* (2006) Sequencing and analysis of Neanderthal genomic DNA. *Science*, **314**, 1113–1118.
137. Castellano,S., Parra,G., Sanchez-Quinto,F.A., Racimo,F., Kuhlwilm,M., Kircher,M., Sawyer,S., Fu,Q., Heinze,A., Nickel,B. *et al.* (2014) Patterns of coding variation in the complete exomes of three Neandertals. *Proc. Natl. Acad. Sci.*, **111**, 6666–6671.
138. Burbano,H.A., Hodges,E., Green,R.E., Briggs,A.W., Krause,J., Meyer,M., Good,J.M., Maricic,T., Johnson,P.L.F., Xuan,Z. *et al.* (2010) Targeted investigation of the neandertal genome by array-based sequence capture. *Science*, **723**, 723–725.
139. Llamas,B., Brotherton,P., Mitchell,K.J., Templeton,J.E.L., Thomson,V.A., Metcalf,J.L., Armstrong,K.N., Kasper,M., Richards,S.M., Camens,A.B. *et al.* (2014) Late pleistocene Australian marsupial DNA clarifies the affinities of extinct megafaunal kangaroos and wallabies. *Mol. Biol. Evol.*, **32**, 574–584.
140. Vilstrup,J.T., Seguin-Orlando,A., Stiller,M., Ginolhac,A., Raghavan,M., Nielsen,S.C.A., Weinstock,J., Froese,D., Vasiliev,S.K., Ovodov,N.D. *et al.* (2013) Mitochondrial phylogenomics of modern and ancient equids. *PLoS One*, **8**, e55950.
141. Immel,A., Drucker,D.G., Bonazzi,M., Jahnke,T.K., Münzel,S.C., Schuenemann,V.J., Herbig,A., Kind,C.-J. and Krause,J. (2015) Mitochondrial genomes of giant deers suggest their late survival in Central Europe. *Sci. Rep.*, **5**, 10853.
142. Zhang,H., Paijmans,J.L.A., Chang,F., Wu,X., Chen,G., Lei,C., Yang,X., Wei,Z., Bradley,D.G., Orlando,L. *et al.* (2013) Morphological and genetic evidence for early Holocene cattle management in northeastern China. *Nat. Commun.*, **4**, 2755.
143. Templeton,J.E.L., Brotherton,P.M., Llamas,B., Soubrier,J., Haak,W., Cooper,A. and Austin,J.J. (2013) DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. *Investig. Genet.*, **4**, 26.
144. Budowle,B., Connell,N.D., Bielecka-Oder,A., Colwell,R.R., Corbett,C.R., Fletcher,J., Forsman,M., Kadavy,D.R., Markotic,A., Morse,S.A. *et al.* (2014) Validation of high throughput sequencing and microbial forensics applications. *Investig. Genet.*, **5**, 9.

145. Bos,K.I., Harkins,K.M., Herbig,A., Coscolla,M., Weber,N., Comas,I., Forrest,S.A., Bryant,J.M., Harris,S.R., Schuenemann,V.J. *et al.* (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature*, **514**, 494–497.
146. Schuenemann,V.J., Singh,P., Mendum,T.A., Krause-Kyora,B., Jager,G., Bos,K.I., Herbig,A., Christos,E., Benjak,A., Busso,P. *et al.* (2013) Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science*, **341**, 179–183.
147. Devault,A.M., Golding,G.B., Waglechner,N., Enk,J.M., Kuch,M., Tien,J.H., Shi,M., Fisman,D.N., Dhody,A.N., Forrest,S. *et al.* (2014) Second-Pandemic Strain of *Vibrio cholerae* from the Philadelphia Cholera Outbreak of 1849. *N. Engl. J. Med.*, **370**, 334–340.
148. Schuenemann,V.J., Bos,K., DeWitte,S., Schmedes,S., Jamieson,J., Mittnik,A., Forrest,S., Coombes,B.K., Wood,J.W., Earn,D.J.D. *et al.* (2011) Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proc. Natl. Acad. Sci.*, **108**, E746–E752.
149. Wagner,D.M., Klunk,J., Harbeck,M., Devault,A., Waglechner,N., Sahl,J.W., Enk,J., Birdsell,D.N., Kuch,M., Lumibao,C. *et al.* (2014) *Yersinia pestis* and the Plague of Justinian 541–543 AD: a genomic analysis. *Lancet Infect. Dis.*, **14**, 319–326.
150. Bos,K.I., Jäger,G., Schuenemann,V.J., Vågene,Å.J., Spyrou,M.A., Herbig,A., Nieselt,K. and Krause,J. (2015) Parallel detection of ancient pathogens via array-based DNA capture. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **370**, 20130375.
151. Whitman,W.B., Coleman,D.C. and Wiebe,W.J. (1998) Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 6578–6583.
152. Curtis,T.P., Head,I.M., Lunn,M., Woodcock,S., Schloss,P.D. and Sloan,W.T. (2006) What is the extent of prokaryotic diversity? *Philos. Trans. R. Soc. B Biol. Sci.*, **361**, 2023–2037.
153. Pham,V.H.T. and Kim,J. (2012) Cultivation of unculturable soil bacteria. *Trends Biotechnol.*, **30**, 475–484.
154. Nikolaki,S. and Tsiamis,G. (2013) Microbial diversity in the era of omic technologies. *Biomed. Res. Int.*, **2013**, 958719.
155. Giovannoni,S.J., Britschgi,T.B., Moyer,C.L. and Field,K. (1990) Genetic diversity in Sargasso Sea bacterioplankton. *Nature*, **345**, 60–63.
156. Suenaga,H. (2012) Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environ. Microbiol.*, **14**, 13–22.
157. Riesenfeld,C.S., Schloss,P.D. and Handelsman,J. (2004) Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.*, **38**, 525–552.
158. Gasc,C., Ribière,C., Parisot,N., Beugnot,R., Defois,C., Petit-Biderre,C., Boucher,D., Peyretaillade,E. and Peyret,P. (2015) Capturing prokaryotic dark matter genomes. *Res. Microbiol.*, **166**, 814–830.
159. Blainey,P.C. (2013) The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiol. Rev.*, **37**, 407–427.
160. Stepanauskas,R. (2012) Single cell genomics: an individual look at microbes. *Curr. Opin. Microbiol.*, **15**, 613–620.
161. Denonfoux,J., Parisot,N., Dugat-Bony,E., Biderre-Petit,C., Boucher,D., Morgavi,D.P., Le Paslier,D., Peyretaillade,E. and Peyret,P. (2013) Gene capture coupled to high-throughput sequencing as a strategy for targeted metagenome exploration. *DNA Res.*, **20**, 185–196.
162. Biderre-Petit,C., Dugat-Bony,E., Mege,M., Parisot,N., Adrian,L., Moné,A., Denonfoux,J., Peyretaillade,E., Debroas,D., Boucher,D. *et al.* (2016) Distribution of Dehalococcoidia in the Anaerobic Deep Water of a Remote Meromictic Crater Lake and Detection of Dehalococcoidia-Derived Reductive Dehalogenase Homologous Genes. *PLoS One*, **11**, e0145558.
163. Manoharan,L., Kushwaha,S.K., Hedlund,K. and Ahrén,D. (2015) Captured metagenomics?: large-scale targeting of genes based on 'sequence capture' reveals functional diversity in soils. *DNA Res.*, **22**, 451–460.
164. Trindade,M., van Zyl,L.J., Navarro-Fernández,J. and Abd Elrazak,A. (2015) Targeted metagenomics as a tool to tap into marine natural product diversity for the discovery and production of drug candidates. *Front. Microbiol.*, **6**, 890.
165. Coughlan,L.M., Cotter,P.D., Hill,C. and Alvarez-Ordóñez,A. (2015) Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. *Front. Microbiol.*, **6**, 672.
166. Henne,A., Schmitz,R.A., Bomeke,M., Gottschalk,G. and 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.
167. Bragalini,C., Ribière,C., Parisot,N., Vallon,L., Prudent,E., Peyretaillade,E., Girlanda,M., Peyret,P., Marmeisse,R. and Luis,P. (2014) Solution hybrid selection capture for the recovery of functional full-length eukaryotic cDNAs from complex environmental samples. *DNA Res.*, **21**, 685–694.
168. Militon,C., Rimour,S., Missaoui,M., Biderre,C., Barra,V., Hill,D., Moné,A., Gagne,G., Meier,H., Peyretaillade,E. *et al.* (2007) PhylArray: phylogenetic probe design algorithm for microarray. *Bioinformatics*, **23**, 2550–2557.
169. Terrat,S., Peyretaillade,E., Gonçalves,O., Dugat-Bony,E., Gravelat,F., Moné,A., Biderre-Petit,C., Boucher,D., Troquet,J. and Peyret,P. (2010) Detecting variants with Metabolic Design, a new software tool to design probes for explorative functional DNA microarray development. *BMC Bioinformatics*, **11**, 478.
170. Dugat-Bony,E., Missaoui,M., Peyretaillade,E., Biderre-Petit,C., Bouzid,O., Gouinaud,C., Hill,D. and Peyret,P. (2011) HiSpOD: Probe design for functional DNA microarrays. *Bioinformatics*, **27**, 641–648.
171. Parisot,N., Denonfoux,J., Dugat-Bony,E., Peyret,P. and Peyretaillade,E. (2012) KASpOD–a web service for highly specific and explorative oligonucleotide design. *Bioinformatics*, **28**, 3161–3162.