Identification of Species with DNA-Based Technology: Current Progress and Challenges

Filipe Pereira*, João Carneiro and António Amorim

Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), Portugal, Faculdade de Ciências da Universidade do Porto, Portugal

Received: March 17, 2008; Accepted: May 16, 2008; Revised: May 16, 2008

Abstract: One of the grand challenges of modern biology is to develop accurate and reliable technologies for a rapid screening of DNA sequence variation. This topic of research is of prime importance for the detection and identification of species in numerous fields of investigation, such as taxonomy, epidemiology, forensics, archaeology or ecology. Molecular identification is also central for the diagnosis, treatment and control of infections caused by different pathogens. In recent years, a variety of DNA-based approaches have been developed for the identification of individuals in a myriad of taxonomic groups. Here, we provide an overview of most commonly used assays, with emphasis on those based on DNA hybridizations, restriction enzymes, random PCR amplifications, species-specific PCR primers and DNA sequencing. A critical evaluation of all methods is presented focusing on their discriminatory power, reproducibility and user-friendliness. Having in mind that the current trend is to develop small-scale devices with a high-throughput capacity, we briefly review recent technological achievements for DNA analysis that offer great potentials for the identification of species.

Keywords: Species identification, DNA hybridization, RFLP, RAPD, AFLP, real-time PCR, species-specific PCR primers, DNA sequencing, DNA microarray.

INTRODUCTION

The concept of "species" is perhaps the most debated subject in evolutionary biology as demonstrated by the existence of more than twenty definitions founded on different methods and criteria [1]. The difficulty in assigning an organism to a biologically meaningful category should be well considered before the use of any molecular identification tool. Researchers should be aware of the evolutionary history and taxonomic position of the specimen under study. Ideally, they should understand the order of branching and ages of divergence (phylogeny) of the organisms in examination and be familiarized with the nomenclature used in previous studies. Terms such as "strain", "variant", "subspecies" or "breed" could be highly subjective in some circumstances and be used as synonyms by different investigators to describe the same biological entity.

All methods for the identification of species that rely on DNA or protein sequence analysis presuppose the neutral theory of molecular evolution, in which different lineages diverge over evolutionary times by the accumulation of molecular changes (most of them neutral) [2]. These methods are based on the assumption that individuals from a same species carry specific DNA (or protein) sequences that are different from those found in individuals from other species. However, the distribution of a given molecular variant in time and in space will be influenced by the reproductive success of individuals, migratory events and random genetic drift. Therefore, it should be realized that a continuous genetic variability does exist among individuals of a species. The level of intraspecies diversity in the locus under study has to be properly assessed before undertaking any taxonomic identification in order to guarantee that there is no overlap between intraspecies variation and interspecies divergence. Furthermore, different loci have variable rates of evolution owing to the action of processes such as mutation and recombination [3]. Therefore, to choose the appropriate loci is vital to the success of the identification.

If a new method or genetic marker will be employed for the first time, the genetic composition of all species of that taxon should be determined. If possible, a representative sample of individuals should be genotyped, preferentially from different geographic locations (or from different hosts, in cases of internal parasites). The preservation of voucher specimens to serve as a future reference is also highly recommended.

It is also important to keep in mind that there is no perfect DNA-typing method and that the choice of a particular technique is often a compromise that depends on a number of factors, including: the resources of the laboratory, financial constraints, available expertise, time limitations and, more importantly, the research question pursued. All points should be scrutinized carefully to avoid an inappropriate choice.

In this review, we have compared some of the most classical DNA-based methods for the detection and identification of species with regard to the principles of the techniques and their most important attributes (reproducibility, discriminatory power, user-friendliness, etc). We also highlight emerging technologies for the screening of DNA and speculate on their potential application in the field of species identification.

It should be noted that this review is not meant to provide a comprehensive list of all available techniques capable of

*Address correspondence to this author at the IPATIMUP, Rua Dr. Roberto Frias s/n, 4200-465 Porto. Portugal; Tel: +351 22 5570700; Fax: +351 22 5570799; E-mail: fpereira@ipatimup.pt

1872-2156/08 $100.00+.00

analysing genetic variation. Other techniques have been described in numerous reviews [3-6], patent applications [7-10] and published manuscripts [11-13]. Moreover, this review does not embrace the particularities of assigning organisms to the species category in different taxa. A series of published works are recommended for detailed information on particular groups (e.g. microbial organisms [14], invasive species [15], insects [16] or parasites [17,18]).

"PRE-DNA WORLD" IN SPECIES IDENTIFICATION PROCEDURES

The term "morphology" is used in biology to refer to the form and structure of an organism as a whole or its component parts. It is unquestionable that the use of external or internal features of an individual is still the most applied process in both identification and taxonomy. Occasionally, the identification procedure is supported by behavioural characteristics (difference in habitat preferences, breeding seasons, epidemiology, etc) or physiological features (growth rates, biochemical composition, etc). Although extremely useful in several cases to assign organisms to well-defined categories, the use of anatomical characters for species identification procedures has several disadvantages. First, there is a considerable morphological plasticity between organisms of the same species. For instance, coloration in some species of birds and fishes are known to vary due to different nutritional regimes. For that reason, a reliable diagnostic procedure can be time-consuming and require the expertise of different taxonomists (when available).

The use of morphology is also complicated by the existence of sibling species - species that are morphologically nearly identical but are nonetheless reproductively isolated from one another. Groups of closely related species in this condition can form large cryptic species complexes. A recent report demonstrates that these complexes are almost evenly distributed among major metazoan taxa [19]. Moreover, these methods are hampered by the existence of convergent evolution, in which the same phenotypic feature can emerge independently in phylogenetic unrelated organisms. This fact can led to erroneous identifications if a small number of morphological features are considered in the analysis.

Finally, most morphology-based approaches cannot be applied in cases where there is just a small amount of biological material available for examination. For instance, forensic laboratories often have to deal with extremely low quantities of biological material, most of the times highly degraded after a prolonged exposition to harsh environmental conditions.

It was only in the second half of last century that, with the convergence of new ideas from genetics and biochemistry and a set of new technological developments, the field of species identification started to rely on information from the molecular components of cell. The first molecular methods successfully employed were based on the analysis of proteins: protein sequencing, protein electrophoresis, isoenzyme analysis, immunological reactions, etc (e.g. [20-22]). Although each method has its own advantages, a number of features are known to limit the use of proteins: its rapid degradation in samples under stress conditions, the risk of cross-reactions with proteins from closely related species, the differential expression of proteins in specific tissues, the scarcity of available antibodies for immunological reactions, among others [18,23]. As described in the next section, most limitations of protein-based methods have been circumvented with the advent of DNA-based procedures.

THE "DNA REVOLUTION" IN MOLECULAR SPECIES IDENTIFICATION

Three major characteristic of the DNA molecule makes it an extremely useful tool for molecular species identification. First, DNA is an extremely stable and long-lived biological molecule that can be recovered from biological material that has been under stress conditions (processed food products, coprolites, mummified plant tissues, blood stains, etc). A variety of methods have also been developed to make the collection and storage of DNA samples very simple and efficient [24-26]. Second, DNA is found in all biological tissues or fluids with nucleated cells (or non-nucleated cells with plastids and/or mitochondria), enabling its analysis from almost all kinds of biological substrates (saliva, faeces, plant seeds, milk, etc). Finally, DNA can provide more information than proteins due to the degeneracy of the genetic code and the presence of large non-coding stretches.

The exception to the universal rule of DNA as the genetic material of a species is found in certain virus families that possess an RNA genome (for instance, the human immunodeficiency virus). Nevertheless, the identification of these viruses can be achieved by conventional methods of DNA analysis, after the use of a reverse transcriptase enzyme that transcribes single-stranded RNA into DNA [27].

In the following sections we describe some of the most used methods for identification of species based on the analysis of DNA. In Table (**1**) is a summary of the main attributes of the various DNA-based typing approaches. A schematic representation of the most important procedural steps of each assay is displayed in Fig. (**1**).

DNA hybridization. The hybridization of complementary DNA oligonucleotides is a basic principle of molecular biology used in a variety of methods with possible applications in species identification. Some of the early assays were based on solid-phase hybridizations conducted on nitrocellulose or nylon membranes between whole genomic or synthetic DNA probes of known origin and DNA extracted from the target sample (e.g. [28-30]). The probe or target DNA was usually labelled with fluorescent or radioactive molecules. A positive hybridization indicates the presence of biological material from the species used to construct the probe.

A number of factors are known to limit the widespread application of traditional DNA-DNA hybridization methods: a) they cannot be applied to degraded samples because high amounts of good quality or undegraded DNA are usually required; b) the comparison of results between different laboratories is extremely difficult because small changes in experimental conditions originate different results; c) non-discrimination between closely related species may occur due to cross-hybridizations and d) it is a time-consuming procedure. However, these basic methods have the advantage of enabling the simultaneous detection of multiple species in

Table 1. Attributes of the Various DNA-Based Typing Methods*

Criterion	DNA hybridiz.	RFLPs	AFLPs	RAPD	Conventional PCR	Real-time PCR	Sequencing	Microarray
Quantity of information	Low	Moderate	High	Moderate	Moderate	High	High	High
Requirement of prior information	No	Yes	No	No	Yes	Yes	Yes	Yes
Suitability for the detection of mixtures	No	Yes	Variable	Variable	Yes	Yes	No	Yes
Inter-laboratory reproducibility	Moderate	Moderate	Moderate	Poor	Good	Good	Good	Good
Cost of equipments and reagents	Moderate	Moderate	Moderate	Moderate	Moderate	High	High	High
Throughput capacity	Low	Moderate	Moderate	Moderate	Moderate	High	High	High
Easy of use	Easy	Easy	Moderate	Moderate	Easy	Moderate	Moderate	Difficult

*This table is based on the most frequent laboratorial application of each method and does not represent the diversity of available procedures.

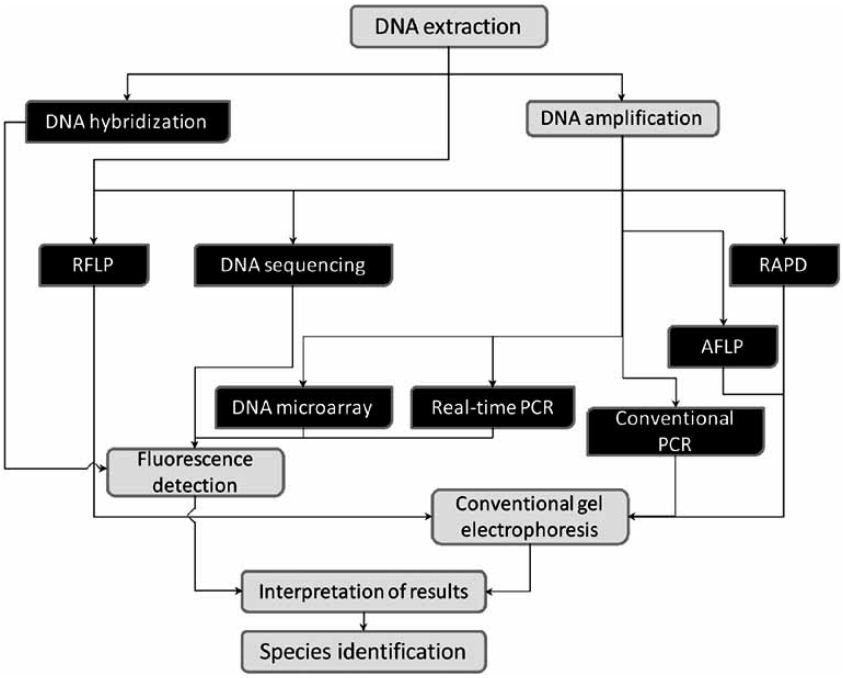


Fig. (1). Schematic representation of most important procedural steps involved in the various DNA-based methods described in the text. The scheme is based on the most common laboratorial application of each method and does not represent the total diversity of available procedures.

a sample with the use of two or more specific probes either tested in separate reactions or labelled with unique fluorescent dyes (see reference [31] for an example in food samples).

A widely known DNA hybridization-based approach is the fluorescence in situ hybridization (FISH) technique [32-34]. This technique uses fluorescently labelled probes to detect nucleic acid sequences in whole cells, allowing the direct detection of organisms in complex microbial communities (e.g. environmental samples, gastro-intestinal flora or oral cavity samples [33]). A higher stability and affinity in the hybridization with FISH assays can be achieved by using peptide nucleic acid (PNA) probes. A PNA is a synthetic molecule in which the sugar phosphate backbone of the natural nucleic acid has been replaced by an uncharged pseudopeptide backbone [35]. PNA/DNA hybrids

are more stable than the corresponding nucleic acid complexes due to the lack of electrostatic repulsion between PNA and DNA strands (in comparison to that existing between two negatively-charged complementary DNA oligomers). PNA-FISH assays have been described for the identification of multiple species (see [33,36-38] and references therein). Unfortunately, PNA probes are not widely available and the cost of their production is still higher than that of DNA oligonucleotides.

Recently, a novel class of nucleic acid analogues (locked nucleic acid - LNA) has been used to improve the efficiency of DNA and RNA hybridizations [39-42]. A LNA is a RNA derivative in which the ribose ring is constrained by a methylene linkage between the 2′-oxygen and the 4′-carbon. The restriction in the conformational arrangement of the RNA derivative substantially increases the binding affinity for complementary sequences. The incorporation of LNA bases in DNA probes can be used to enhance the efficiency of in situ hybridization procedures [43]. LNAs can be produced using standard reagents and automated synthesizers due to their structural resemblance to native nucleic acids [39,40].

A high-throughput detection of bacterial, viral and fungal pathogens is also possible with the use of multiplexed direct hybridization assays using a microsphere-based suspension array platform (Luminex® xMAP™) (for an overview of principles and applications see references [44-47]). The Luminex xMAP system incorporates microspheres embedded with two fluorophores that have a characteristic emission wavelength. The use of different intensities of each fluorophore permits the construction of an array with 100 different microsphere sets with unique spectral signatures. On the surface of each microsphere is possible to attach species-specific oligonucleotide probes that will hybridize directly with the target DNA (such as fluorescently labelled amplified DNA). After hybridization with the target fluorescent product, suspended microspheres pass through a detection chamber based on the principles of flow cytometry. Two separate lasers coupled with a high-speed digital signal processing equipment classify each microsphere based on its spectral address (measuring the fluorescence of both fluorophores) and quantify the reaction on the surface (measuring the fluorescence associated with the captured product). This technique has the advantage of permitting the discrimination between closely related species in a multiplex format [46].

The principles underling the hybridization between DNA oligonucleotides are at the basis of several diagnostic assays (for example, line probe assay (LiPA) [48] or hybridization protection assays (HPA) [49]) and are a crucial component of modern analytic techniques, such as DNA microarrays and real-time PCR (discussed below).

Restriction Fragment Length Polymorphisms (RFLPs). The RFLP analysis is widely used for the detection of interspecies variation at the DNA sequence level. It consists in the generation of species-specific band profiles through the digestion of DNA with one or more restriction endonucleases [50] (Fig **2**). These restriction enzymes cleave the DNA molecule at specific 4-6 base pair (bp) recognition sites, originating a set of fragments with different lengths that could be separated according to their molecular size by conventional gel electrophoresis. The RFLP banding pattern could be visualized by hybridizing restriction fragments with a labelled probe in a solid support (for instance, by Southern blotting) or by treating the electrophoretic gel with ethidium bromide or silver staining (Fig **2**). The distinctive RFLP profile of each species is the result of the unique genomic distribution of recognition sites (generated or removed by single-base substitutions) and the distance between them (that varies due to large genomic rearrangements, such as translocations, transposable elements or tandem duplications).

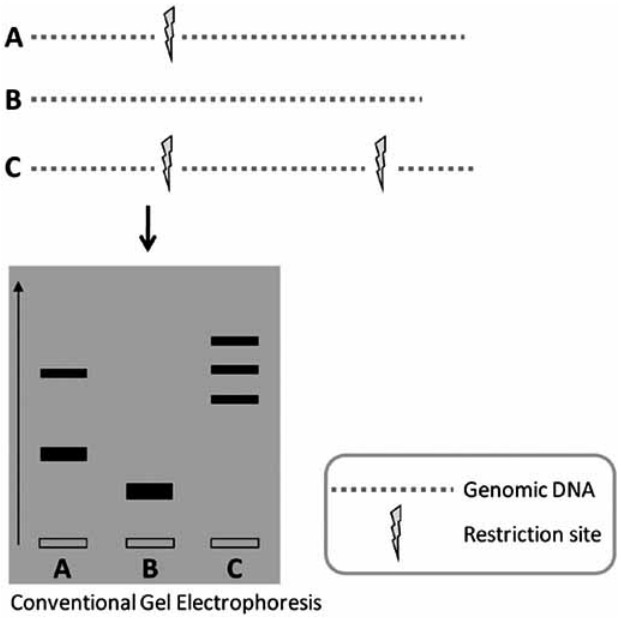


Fig (2). Schematic representation of the Restriction Fragment Length Polymorphisms (RFLPs) method. Genomic DNA is digested by restriction enzymes, originating a set of fragments with different lengths. Species (A, B and C) are identified by running the restriction fragments on a conventional electrophoretic gel.

Initial RFLP assays were performed without any amplification protocol on whole genomic DNA or isolated mitochondrial DNA (mtDNA) [see Box **1** for details on mtDNA)]. Most of the time, these assays were based on the use of an infrequently cutting restriction enzyme followed by a pulsed-field gel electrophoresis to separate large genomic fragments [51]. RFLP assays usually do not require any sophisticated equipment and no prior sequence information about the species. However, it was only with the advent of the polymerase chain reaction (PCR) technique [52] [Box **2**] that RFLP analysis (known as PCR-RFLP) has become routinely used for species detection. Several protocols have been developed for the identification of species in a myriad of taxonomic groups [53-56]. Most PCR-RFLP approaches focus on mtDNA cytochrome b [55,56] or ribosomal RNA (rRNA) genes [Box **3**] [57,58].

A major disadvantage of the RFLP technique is the possible existence of intraspecies mutations at restriction sites that can lead to false results due to the gain or loss of restriction fragments. This method relies on just a few

Box (1). Mitochondrial DNA (mtDNA)

The mitochondrial genome consists of a double-stranded DNA molecule devoted to the coding of key subunits of the electron transport chain found in mitochondria (the powerhouses of eukaryotic cells). With few exceptions, all eukaryotic species have mitochondria. The mitochondrial genome of animals and plants are known to evolve at different rates. The typical animal mtDNA has a high mutation rate and an exceptional organizational economy, with rare non-coding segments. In contrast, mitochondrial genomes found in plants have large amounts of non-coding segments and a low accumulation of diversity. The accelerated evolutionary rate of animal mtDNA (and also of certain fungi and protists species) implies that significant amounts of sequence variation could be found in closely related species – a useful feature for species identification procedures. Moreover, in most species, mtDNA is uniparentally inherited without recombination, a fact that greatly simplifies the interpretation of results. The mtDNA is also easier to retrieve from low-quantity and/or degraded DNA samples since it is present in many copies per cell, providing a clear advantage over nuclear genome-based methods.

The most important limitation of using mtDNA information in the definition of species is the putative occurrence of male-biased gene flow between species (in cases where the mtDNA is maternally inherited).

Box (2)

Polymerase chain reaction (PCR)

The development of the polymerase chain reaction (PCR) technique [52] has significantly improved the efficiency of laboratorial diagnostic procedures by allowing the *in vitro* formation of a large number of DNA copies (amplification) using a specific genomic region as template. Since it only requires a small amount of template DNA, the PCR method could be particularly useful for the identification of species in suboptimal DNA samples (processed food products, forensic samples, archaeological remains, etc). The amplification of a short amplicon (DNA fragments produced by PCR amplification) is more likely to generate conclusive results in degraded samples due to the inverse correlation between amplicon length and amplification efficiency. Major drawbacks of PCR-based techniques are the risk of contamination and the requirement of sequence information for the design of primers.

In order to conserve template DNA, minimize expenses and save time it is possible to simultaneous amplify two or more different DNA target regions in a single reaction tube by using more than one pair of primers (multiplex PCR).

Nucleic acid isothermal amplification

A large number of non-PCR based methods have been developed to *in vitro* synthesize significant amounts of DNA molecules under isothermal conditions. The most important advantages of isothermal amplification techniques are the limited risk of contamination and the tolerances to some inhibitory materials that affect the PCR efficiency. Moreover, they eliminate the need for an expensive and cost-intensive thermocycler. Alongside these advantages is important to be aware of a number of disadvantages of using isothermal methods. For instance, some techniques can be quite lengthy and require ~8 to 16 hours for amplification. Some isothermal systems are not adapted to work with low quality DNA and have a low rate of success in amplifying the target area. An excess of undesired amplifications can also occur and compromise the assay sensitivity.

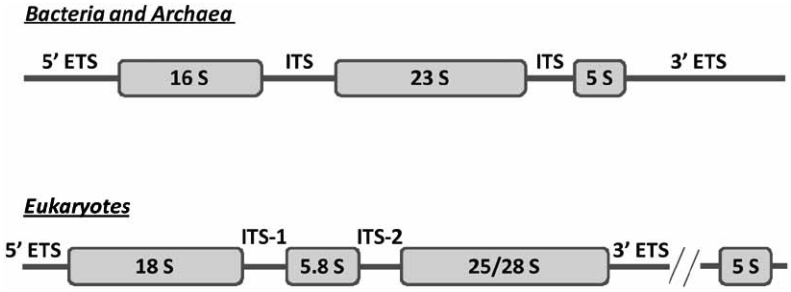
Examples of isothermal amplifications are: nucleic acid sequence-based amplification (NASBA), rolling circle amplification, strand displacement amplification (SDA), transcription-mediated amplification, loop-mediated isothermal amplification of DNA, helicase-dependent amplification, among others (see reference [112] for a review).

informative DNA sequence positions, meaning that several restriction enzymes are usually required to achieve a correct identification. In those situations, the use of different enzymes generates highly complex RFLP patterns of difficult interpretation. Moreover, it is not amenable for automation and standardization because it requires a substantial amount of high quality DNA (unless a whole genome amplification [Box **4**] has been performed prior to the RFLP analysis).

An additional problem may occur when analysing genes that are found in multiple copies within a genome. The intragenomic sequence heterogeneity of certain multigene families (such as rRNA genes) can lead to unexpected RFLP patterns and ambiguous results. However, it has been shown that multiple-copy rRNA genes undergo concerted evolution such that sequences of all gene copies are usually very similar within an organism [59]. The putative presence of two or more different mtDNA sequences in the same cell or individual (heteroplasmy) should also be taken into account to avoid misclassifications.

Amplified Fragment Length Polymorphisms (AFLPs). The AFLP method combines the reproducibility of restriction fragment analysis with the power of PCR. It is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA [60] (Fig **3**). The method usually works by digesting a small amount of purified genomic DNA with two or more restriction enzymes (such as *Eco*RI and *Mse*I). Double-stranded oligonucleotide adapters (10-30 bp long) are ligated to the sticky ends of DNA fragments (both 5´ and 3´ ends) generated during the restriction digestion. The ligated DNA fragments are then amplified twice under highly stringent conditions by PCR using primers complementary to the adapter and restriction site sequence. These selective primers include additional nucleotides at their 3´ end to reduce the complexity of the mixture of fragments. For instance, a selective primer with

Box (3). Ribosomal RNA genes



The genomic organization of ribosomal RNA (rRNA) genes - responsible for the synthesis of RNA species (the core of ribosomes) - is slightly different in prokaryotes and eukaryotes. Most Bacteria and Archaea contain either a single or multiple copies of rDNA clusters dispersed in the genome. Each cluster includes the 16S, 23S and 5S rRNAs separated by the internal transcribed spacer (ITS) regions and flanked by the 5´ and 3´ external transcribed spacers (5´-ETS and 3´-ETS). Almost all eukaryotes have several copies of each rDNA cluster organized in tandem repeats. In this case, each cluster contains the 18S, 5.8S and 25/28S rRNAs, while the 5S gene is present in separate repeat arrays in the majority of eukaryotes.

Box (4) Whole genome amplification (WGA)

Whole genome amplification methods are used to generate large amounts of genomic DNA from small or precious samples. In the field of species identification, the ability to rapidly amplify genomes from limited DNA samples can significantly increase the classification accuracy by enabling the large-scale screening of several loci using high-throughput technologies. Several WGA strategies have been developed either using PCR or isothermal amplifications, each one with their own specific strengths and limitations [122-124]. Ideally, a WGA method should provide good genome coverage and preserve the sequence representation. However, it has been shown that WGA has a limited forensic utility unless the samples are of a very high quality (which is not the case in most casework studies). In low quality DNA samples, WGA generates small amplicons with incomplete coverage of loci throughout the genome (sometimes referred to as allelic dropout). Template-independent background DNA synthesis has also been observed with some WGA methodologies [125].

The choice of a particular WGA methodology must take into account the quality of starting DNA, technical difficulties and cost of reagents (for an overview on different WGA approaches see references [122-124]).

the sequence GAATTCA (GAATTC is the *Eco*RI restriction site) at their 3´ end will only amplify restriction fragments with T nucleotide immediately after the *Eco*RI restriction site (CTTAAGT). Polymorphisms are revealed by running the amplified fragments on a denaturing polyacrylamide gel or similar technique [61,62] (Fig. **3**).

The AFLP technique permits the simultaneous screening of different loci randomly distributed throughout the genome. However, it is technically demanding in the laboratory, labour consuming and the interpretation of results may need automated computer analysis. Additionally, the AFLP method can be a costly technique since it requires an expensive software package to analyze a large number of AFLP patterns.

Random Amplified Polymorphic DNA (RAPD). RAPD profiles are generated by the random PCR amplification of DNA segments using short primers of arbitrary nucleotide sequence of usually 9 or 10 nucleotides long [63,64]. These primers hybridize with sufficient affinity to different genomic regions at low annealing temperatures. Amplification products are generated when two RAPD primers anneal within a few thousand bases of each other in the proper orientation. Each species is identified by a specific banding pattern in an electrophoretic gel or similar technique resulting from the different genomic location of primer-binding sites [63,64] (Fig. **4**). This technology is also known as arbitrarily primed-polymerase chain reaction (APPCR) and has been successfully used in a number of studies (e.g. [65-69]).

Various fingerprinting strategies are based on the use of primers that are specifically designed to hybridize within repetitive genomic regions. The differential genomic distribution of repetitive elements between species is responsible for species-specific profiles. Examples of target repetitive regions are: mammalian-wide interspersed repeats (MIR) [70-72], repetitive extragenic palindromic sequences (REP) [73] and enterobacterial repetitive intergenic consensus sequences (ERIC) [74].

The RAPD method does not require prior sequence information for PCR primer design but is extremely dependent on variations in laboratorial conditions (such as template DNA concentration, PCR and electrophoretic settings, etc), needing carefully developed laboratory protocols to be reproducible. An imperfect hybridization between the primer and the target site may result in a completely different banding profile. The RAPD method, as well as other fingerprinting techniques, generates results that can be difficult to interpret in cases where biological materials from different species are present in the sample (for instance, in some food products or in biological material from an individual infected with parasites). Another disadvantage is the need of purified DNA of high molecular weight.

Conventional PCR. In recent years, a number of approaches based on conventional PCR techniques have

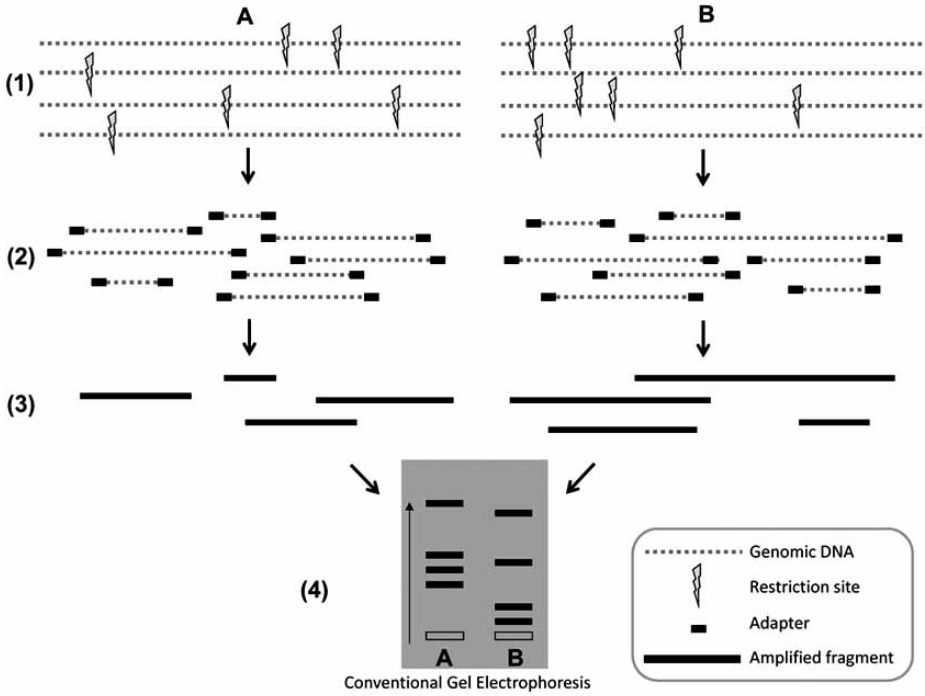


Fig. (3). Schematic representation of the Amplified Fragment Length Polymorphisms (AFLPs) method. Genomic DNA is digested by restriction enzymes (1) and adapters are ligated to the restriction fragments (2). By using primers with selective nucleotides at the 3´-end, only a subset of the ligated fragments is amplified (3). Species (A and B) are identified by running the amplified products on a conventional electrophoretic gel (4).

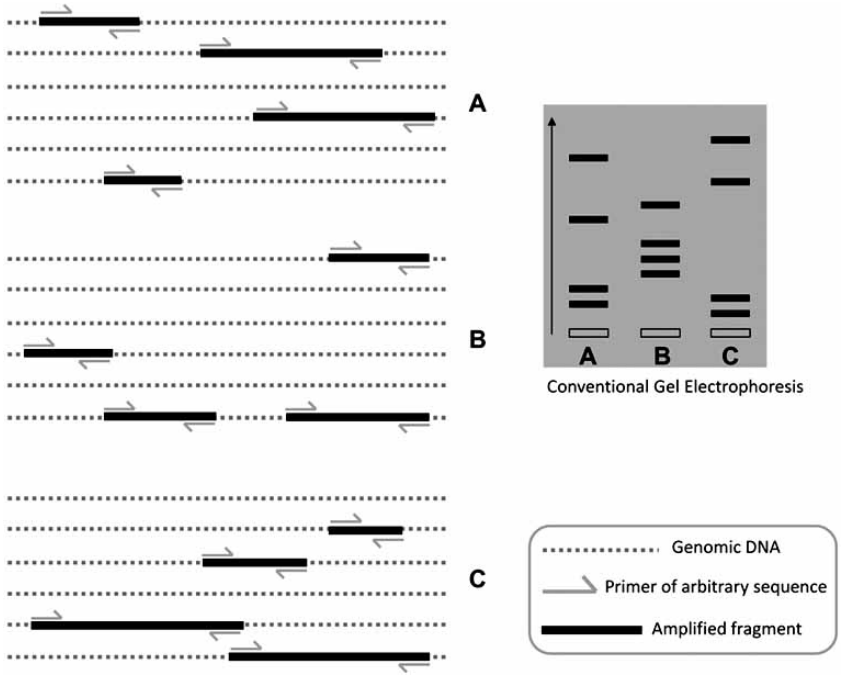


Fig. (4). Schematic representation of the Random Amplified Polymorphic DNA (RAPD) method. Species (A, B and C) are differentiated by the annealing of a single primer of arbitrary nucleotide sequence to different genomic regions. The amplified segments of DNA are separated and visualized in a conventional electrophoretic gel.

been described as a tool for species identification (e.g. [75-79]). Usually, a conventional PCR-based method consists in the design of PCR primers that will only originate an amplification product in the presence of DNA from the target species. The process of designing species-specific primers is now straightforward due to the vast number of genomic sequences available and software programs that assists in primer designing.

A drawback of this technique is that it does not provide information about the presence of biological material from species that are not the target of the primers. A positive result may give an idea about the presence of a particular

species, but a negative result gives no information about the origin of the sample (except that it does not belong to the species for which the assay has been designed for). To avoid this problem, prior sequence knowledge is necessary to derive specific primers for all species suspected to be present in the sample. An additional disadvantage is the need of performing an electrophoresis after the PCR to verify the amplification success of expected target sequences.

The specificity and sensitivity of this kind of approach can be enhanced by performing a nested PCR, in which the target region is first amplified with an outer primer pair followed by a second amplification using an internal primer pair. Applying two rounds of PCR markedly enhances the specificity of PCR analysis because the inner primers only anneal if the proper template has been amplified with the outer primers. The chance of amplifying unspecific genomic regions is reduced with nested PCR as compared to conventional PCR since undesired sequences amplified in the first round of PCR are not likely to contain a sequence to which the primers for the second amplification reaction will bind.

The possibility of carryover contamination represents a significant problem with the classical nested PCR approach since the tubes from the first amplification reaction must be opened to remove an aliquot for the second round of PCR. A number of single-tube nested PCR methods have been developed in order to overcome this problem [80-82]. An additional drawback is the cost involved in doing two rounds of PCRs.

Real-time PCR. The basic goal of real-time PCR is the detection of a specific DNA sequence in a sample by measuring the accumulation of amplified products during the PCR using fluorescent technology. An important benefit of this method is the capability to quantify the starting amount of a specific DNA sequence in the sample (this approach is also known as quantitative PCR).

The ability to monitor the progress of DNA amplification in real time depends on the chemistries and instrumentation used. Generally, chemistries consist of special fluorescent probes that must associate a fluorescent signal to the amplification of DNA. Several types of probes exist, including DNA-binding dyes like ethidium bromide, hydrolysis probes (5´-nuclease probes), hybridization probes, molecular beacons, PNA light-up probes, etc. A complete description of the increasing number of chemistries used to detect PCR products during real-time PCR can be found in several reviews [36,83,84].

A hydrolysis probe consists of a dual-labeled oligonucleotide with a reporter and quencher dye attached. As long as the probe is intact, no fluorescence is released by the reporter molecule when exposed to the appropriate wavelength of light due to the interaction with the quencher (the quencher deactivates the reporter by fluorescence resonance energy transfer). If the target of interest is present in the sample (for instance, DNA from a particular species), the probe anneals specifically between the forward and reverse primer sites during PCR. During amplification, the annealed probe is degraded by the action of DNA polymerase (5´-3´ exonuclease activity) and the reporter and quencher separate, allowing the reporter´s energy and fluorescent signal to be released. Either species-specific probes or primers can be used for identification of species [36,85-87].

Molecular beacons have a similar mode of action with the reporter and the quencher hold in close proximity by a stem-loop structure [88]. When the probe hybridizes to a perfectly matched sequence, it undergoes a spontaneous conformational change (the secondary structure opens) increasing the distance between the quencher and the reporter, restoring the fluorescence.

The real-time PCR has the advantage over conventional PCR-based identification systems of working without post-PCR handling, with a minimised risk of carryover contamination in the laboratory. It also offers an increased sensitivity by permitting the discrimination of spurious PCR amplifications from non-target DNAs and is a relatively fast genotyping method, with some platforms affording high-throughput automation. Most severe disadvantages of real-time PCR methods are the incompatibility of certain platforms with some fluorescent dyes, the restricted multiplex capability and the high cost of most reagents and instrumentation.

Sequencing of PCR products. The DNA sequencing analysis is currently the most used method for molecular species identification. The advent of rapid and cost-effective PCR-linked DNA sequence analysis has circumvented the need for screening of genomic libraries and cloning of DNA fragments. Various novel technologies are being optimized in order to reduce even more the costs of DNA sequencing by several orders of magnitude (for instance, microelectrophoretic, hybridization or cyclic-array sequencing methods; see [5] for an overview).

The identification is achieved by comparing the sequence of a genomic region found in the target sample with a comprehensive reference database. Ideally, the structure of the DNA region to be analysed must consist of a variable sequence (informative enough to discriminate species) flanked by highly conserved regions (ideal to design universal PCR primers that amplify in a large number of species).

In order to attain a correct identification it is crucial to consult a reliable database, namely one that guarantees that (a) the reference specimen was correctly identified by a taxonomic expert or by other molecular methods, (b) the same sequences were obtained in independent studies, preferentially from the full distribution range of the species and that (c) most related species have distinct DNA profiles. A common way to assign a particular sequence to its species of origin is to perform a BLAST search on the vast GenBank sequence database [see Box **5**] for online links and other databases and web resources). However, care must be taken when assigning the questioned sequence to the species with the highest similarity, because several gaps and false sequences are known to be present in these databases [89]. Moreover, this approach does not provide any information and can lead to false identifications if the target sample belongs to a previously uncharacterized species.

Another important aspect for delineating a protocol for species discrimination in a particular taxonomic group is the mutation rate of the selected genomic region. A region with a considerably high mutation rate (such as most regions of animal mtDNA) should be selected if the aim of the study is to discriminate closely related species. The fast accumulation of diversity will be highly informative and appropriate for identifying lineages with a low coalescence time. Conversely, a more conserved region should be selected to discriminate species at deepest branches of the phylogenetic tree. The probability of having back and parallel mutations that might confound the identification is considerably lower in conserved regions. In all cases, to allow a precise allocation of an individual to a described taxon, accurate identifications depend on a low intraspecies variation when compared with the one found between species. Although different regions have been targeted for species identification procedures, most studies rely on sequence information from nuclear ribosomal RNA genes [90,91] [Box **3**] and mtDNA regions [Box **1**].

An advantage of sequencing ribosomal RNA (rRNA) genes is the presence of conserved region (for instance, 18S rRNA in eukaryotes and the 16S rRNA in prokaryotes) adjacent with highly variable segments (such as the internal transcribed spacers) allowing the resolution of relationships among both distantly and closely phylogenetic related species, respectively. This fact is very important if no information about the taxonomical group under study is available *a priori*.

The list of studies using mtDNA cytochrome b gene for species identification is extensive [92,93]. This gene shows a high level of congruence with species limits and can be amplified in several vertebrate species under standard conditions by using a single pair of universal primers [94-96].

Recently, a DNA-based barcoding system for all animal species has been proposed based on 650 to 750 bp of the mtDNA cytochrome c oxidase (COI) gene [97]. The "DNA barcoding" concept is not an entirely new idea but, for the first time, it has been proposed to work at large-scale under well-defined standardized protocols [see Box **5**] for online link). It has been projected both to assign unknown individuals to species and to facilitate the species-discovery process [98-100]. The approach is controversial, with critics questioning both the method and its applications [101,102]. Most important concerns are related to the use of a single gene in delineating and identifying species and the extent of separation between intra- and interspecies variations [101]. Moreover, the COI system is obviously limited to eukaryotic species with mtDNA.

A general drawback of DNA-sequencing approaches is that, in order to provide enough information for a secure discrimination, most of them rely on the sequencing of large DNA regions, usually over 300 bp (e.g. [92,93]). The PCR amplification of such large regions is difficult to obtain from samples with low quality and/or low amounts of DNA. The total amount of DNA available for analysis can be increased by performing a whole genome amplification prior to the sequencing. However, only small amplicons will be generated by this procedure if the sample has low quality DNA [see Box **4** for details].

The targeting of only a single DNA region could be problematic since a failure in the amplification of that region due to, for instance, the occurrence of a polymorphism in a primer binding region, may originate a false or null result. This problem can be overcome by using degenerate primers that pinpoint at polymorphic areas in the primer binding sites, in cases where these variants have been previously identified. Moreover, DNA sequencing methods do not allow the discrimination and identification of biological material

Box (5). Online links to databases and resources

http://www.ncbi.nlm.nih.gov/blast - Basic Local Alignment Search Tool of the National Center for Biotechnology Information. A web tool that compares nucleotide sequences to GenBank/EMBL/DDBJ database and calculates the statistical significance of matches.

http://www.tolweb.org – The Tree of Life Web Project. A web based resource with information about biodiversity and phylogeny.

http://www.gbif.org – The Global Biodiversity Information Facility. A web site that provides information on species biodiversity as well as scientific and common names.

http://species.wikimedia.org - Wikispecies. An open directory with diverse information about species from all taxonomic groups.

http://www.barcodinglife.com - The Barcode of Life Data Systems for management, analysis, and use of DNA barcodes.

http://www.mycobank.org - MycoBank. An on-line database for fungal taxonomy.

http://www.pathooligodb.com - A database and web based resource with information about oligonucleotides for detection of several pathogens.

http://www.ipni.org - The International Plant Names Index. A database of names and associated bibliographical details of all seed plants, ferns and fern allies.

Online site for patent searching:

http://www.espacenet.com/

http://www.wipo.int/pctdb/en/search-simp.jsp

http://www.google.com/patents

http://www.freepatentsonline.com/

from different species mixed in a same sample, unless fragments are cloned before sequencing to separate each molecule of DNA.

A solution to overcome some limitations of single-gene approaches is the use of multilocus sequence analysis (MLSA), a method based on the sequencing of multiple protein-coding genes already used in the characterization of prokaryotes [103]. MLSA overcomes the distorting effect of stochastic sequence variation of single-gene approaches and the influence of recombination and horizontal gene transfer. Additionally, it is easier to detect misleading results from genomic regions affected by natural selection. Genes selected to be used in MLSA should be ubiquitous and with no genomic duplications. A phylogenetic approach could be used to identify the species of origin of an individual by constructing a phylogenetic tree with concatenated sequences of multiple genes. The phylogenetic inference can be established by the clustering pattern of the species of interest with related species. In this case, MLSA provides a higher discriminatory power than other methods used for generating phylogenetic trees. An obvious limitation to the use of this approach is the necessity of sequencing several genomic regions, a fact that can be cumbersome.

DNA microarrays or DNA chips. It consists of small glass microscope slides, silicon chips or nylon membranes containing a large number of immobilized DNA fragments arranged in a regular pattern. A DNA microarray provides a medium for matching a reporter probe of known sequence against the DNA extracted from the target sample of unknown origin. Probes can include synthetic oligonucleotides, amplicons or larger DNA/RNA fragments selectively spotted or addressed to individual test sites in the microarray. The microarray is scanned or imaged to obtain a complete hybridization pattern generated by the release of a fluorescent, chemiluminescent, colorimetric or radioactive signal associated with the binding of the probe to the target DNA sequence [104].

A DNA microarray built with species-specific DNA sequences can be used for identifications purposes [105,106]. For instance, the DNA extracted from the target sample could be labelled with a specific fluorescent molecule and hybridized to the microarray DNA. A positive hybridization is detected with appropriate fluorescence scanning/imaging equipment (fluorescent spots are visualized). The DNA microarray hybridization methodology can also be directed for the screening of samples for species-specific single nucleotide polymorphisms (SNPs).

Advances in printing technology have enabled the production of microarrays containing hundreds of thousands of probes (high-density microarrays may have up to 10^6 test sites in a 1-2-cm^2 area), revealing the potential to achieve sensitive and high-throughput species identifications [104]. PNA probes and molecular beacons can also be applied to the microarray technology for a rapid and large-volume systematic analysis of genetic information. Nevertheless, DNA microarrays require specialized robotics and imaging equipment that generally are not available in most laboratories. Advanced bioinformatic tools are also necessary to reduce the complex data into useful information.

CURRENT & FUTURE DEVELOPMENTS

Current developments in nucleic acid detection technologies are guided by two general trends: miniaturization of genotyping instruments and high-throughput sample analysis. A broad spectrum of highly innovative automated assays have been devised based on conventional genotyping techniques (DNA hybridization or sequencing, real-time PCR, etc) to provide reliable, rapid and low-cost DNA screenings. Methods such as multiplex minisequencing [107,108], microarray primer extension [109], coded microspheres [110] or sequence-coded oligonucleotide ligation assay [111] have a high multiplexing capacity and are rapidly approaching feasible commercial prices [6].

A number of recent biotechnological achievements offer great potentials for the development of a portable DNA diagnostic device for the rapid identification of species from low amount of biological material. The requirement of a PCR amplification step prior to the DNA scrutiny is one of the principal factors that limit both the size and the throughput capacity of current genotyping instruments. A number of isothermal strategies to amplify nucleic acids have been developed [112] without requiring temperature cycling instruments [Box **2**]. The simplicity and isothermal nature of these methods is suitable for the development of a hand-held DNA diagnostic device, although great improvements have to be accomplished to achieve the sensitivity of conventional PCR-based methods [see Box **2** for details].

The most promising area of emerging technologies that will certainly contribute to the resolution of several technical challenges related with species identification is the recently coined "Nanobiotechnology" [113,114]. This interdisciplinary field of research brings together life scientists and engineers to apply the tools and processes of nanofabrication to the build of multifunctional devices and systems for studying biomolecules. The development of "nano-objects" such as carbon nanotubes and quantum dots are becoming of great importance in improving the current capacity for simultaneously screening multiple nucleic acid sequences in a rapid and accurate fashion from submicroliter volumes [113]. The next generation of methods for DNA analysis should be capable of working without the requirement of the nucleic acid amplification step. For instance, an approach to directly read multiple polymorphic sites on single DNA molecules has been recently proposed, using atomic force microscopy with a high-resolution single-walled carbon nanotube probe [115,116]. Further refinement of nanotechnological methodologies might permit a high-throughput analysis from single DNA molecules.

Another important breakthrough for the construction of a portable species identification gadget is the development of a new generation of nanomaterial-based electrochemical biosensors. Electrical DNA hybridization biosensors are now capable of converting DNA-DNA recognition events into an electronic signal-transduction process [117-119]. Further work is needed to realize the full potential of this new class of biosensors for the analysis of large DNA sequences and its applicability in species identifications.

A promising technological platform for genome analyses could also emerge from the construction of optical maps

from simply extracted, unfractionated genomic DNA molecules. The system creates ordered restriction maps of an entire genome using randomly selected individual DNA molecules mounted on specially prepared surfaces and it might be useful in the future for developing new ways to study interspecies genomic variations [120-125].

Although it is very difficult to anticipate all future developments in DNA-based technologies, they probably will enable a fast processing, an increased genotyping sensibility at reduced cost and a radical miniaturization of devices. However, there is no doubt that the utility of any future method will only be possible under a coherent scientific understanding of population genetics, evolution, systematics, ecology and molecular biology. It is important to bear in mind that an effortless access to genetic data does not necessarily imply an increase in scientific knowledge.

ACKNOWLEDGEMENTS

We apologize to the many researchers and inventors whose work we have not been able to cite in this review owing to space constraints. We thank the referees for their helpful comments and suggestions. This work was partially supported by a research grant to FP (SFRH/BD/19585/2004) from Fundação para a Ciência e a Tecnologia. IPATIMUP is partially supported by "Programa Operacional Ciência e Inovação 2010" (POCI 2010), VI Programa Quadro (2002-2006). Authors´ web site: http://www.portugene.com/

REFERENCES

[1] Hey J. The mind of the species problem. Trends Ecol Evol 2001; 16: 326-329.
[2] Kimura M. Evolutionary rate at the molecular level. Nature 1968; 217: 624-626.
[3] Sunnucks P. Efficient genetic markers for population biology. Trends Ecol Evol 2000; 15: 199-203.
[4] Ding C, Cantor CR. Quantitative analysis of nucleic acids-the last few years of progress. J Biochem Mol Biol 2004; 37: 1-10.
[5] Shendure J, Mitra RD, Varma C, Church GM. Advanced sequencing technologies: methods and goals. Nat Rev Genet 2004; 5: 335-344.
[6] Syvanen AC. Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet 2001; 2: 930-942.
[7] Ecker, D., Griffey, R., Sampath, R., Hofstadler, S., McNeil, J.: US20067108974 (**2006**).
[8] Fofanov, Y., Pettitt, B., Li, T., Tchoumakov, S.: US2005255459 (**2005**).
[9] Hershfield, B.: US5674687 (**1997**).
[10] McClelland , M., Welsh, J., Sorge, J.: US5487985 (**1998**).
[11] Alonso A, Albarran C, Martin P, *et al.* Usefulness of microchip electrophoresis for the analysis of mitochondrial DNA in forensic and ancient DNA studies. Electrophoresis 2006; 27: 5101-5109.
[12] Bataille M, Crainic K, Leterreux M, Durigon M, de Mazancourt P. Multiplex amplification of mitochondrial DNA for human and species identification in forensic evaluation. Forensic Sci Int 1999; 99: 165-170.
[13] Bellis C, Ashton KJ, Freney L, Blair B, Griffiths LR. A molecular genetic approach for forensic animal species identification. Forensic Sci Int 2003; 134: 99-108.
[14] Olive D, Bean P. Principles and applications of methods for DNA-based typing of microbial organisms. J Clin Microbiol 1999; 37: 1661-1669.
[15] Darling JA, Blum MJ. DNA-based methods for monitoring invasive species: a review and prospectus. Biol Invasions 2007; 9: 751-765.
[16] Wells JD, Stevens JR. Application of DNA-based methods in forensic entomology. Annu Rev Entomol 2008; 53: 103-120.
[17] Gasser RB. Molecular tools--advances, opportunities and prospects. Vet Parasitol 2006; 136: 69-89.
[18] McManus DP, Bowles J. Molecular genetic approaches to parasite identification: their value in diagnostic parasitology and systematics. Int J Parasitol 1996; 26: 687-704.
[19] Pfenninger M, Schwenk K. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. BMC Evol Biol 2007; 7: 121.
[20] Bolin I, Lonroth H, Svennerholm AM. Identification of *Helicobacter pylori* by immunological dot blot method based on reaction of a species-specific monoclonal antibody with a surface-exposed protein. J Clin Microbiol 1995; 33: 381-384.
[21] Razin S, Rottem S. Identification of *Mycoplasma* and other microorganisms by polyacrylamide-gel electrophoresis of cell proteins. J Bacteriol 1967; 94: 1807-1810.
[22] Skarpeid HJ, Kvaal K, Hildrum KI. Identification of animal species in ground meat mixtures by multivariate analysis of isoelectric focusing protein profiles. Electrophoresis 1998; 19: 3103-3109.
[23] Ahmed FE. Detection of genetically modified organisms in foods. Trends Biotechnol 2002; 20: 215-223.
[24] Nsubuga AM, Robbins MM, Roeder AD, *et al.* Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. Mol Ecol 2004; 13: 2089-2094.
[25] Smith LM, Burgoyne LA. Collecting, archiving and processing DNA from wildlife samples using FTA(R) databasing paper. BMC Ecol 2004; 4: 4.
[26] Rogers NL, Cole SA, Lan HC, Crossa A, Demerath EW. New saliva DNA collection method compared to buccal cell collection techniques for epidemiological studies. Am J Hum Biol 2007; 19: 319-326.
[27] Seifarth, W., Leib-Moesch, C., Baust, C.: US20067115367 (**2006**).
[28] Falkow, S., Moseley, S.: US4358535 (**1982**).
[29] Scholl, D., Jollick, J.: US4652517 (**1987**).
[30] Torok, T., King, A., Rockhold, D., Royer, C.: US5401630 (**1995**).
[31] Janssen F, Hägele G, Buntjer J, Lenstra J. Species identification in meat by using PCR-generated satellite probes. J Ind Microbiol Biotec 1998; 21: 115-120.
[32] Kim, C., Thompson, D., Sarto, G.: WO31626 (**1996**).
[33] Moter A, Gobel UB. Fluorescence in situ hybridization (FISH) for direct visualization of microorganisms. J Microbiol Meth 2000; 41: 85-112.
[34] Zarling, D., Calhoun, C., Sena, E.: US5719023 (**1998**).
[35] Nielsen, P., Buchardt, O., Egholm, M., Berg, R.: US5773571 (**1998**).
[36] Monis P, Giglio S, Keegan A, Thompson R. Emerging technologies for the detection and genetic characterization of protozoan parasites. Trends Parasitol 2005; 21: 340-346.
[37] Mothershed EA, Whitney AM. Nucleic acid-based methods for the detection of bacterial pathogens: Present and future considerations for the clinical laboratory. Clin Chim Acta 2006; 363: 206-220.
[38] Stender H, Fiandaca M, Hyldig-Nielsen JJ, Coull J. PNA for rapid microbiology. J Microbiol Meth 2002; 48: 1-17.
[39] Petersen M, Wengel J. LNA: a versatile tool for therapeutics and genomics. Trends Biotechnol 2003; 21: 74-81.
[40] Braasch DA, Corey DR. Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. Chem Biol 2001; 8: 1-7.
[41] Koshkin AA, Singh SK, Nielsen P, *et al.* LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. Tetrahedron 1998; 54: 3607-3630.
[42] Obika S, Nanbu D, Hari Y, *et al.* Stability and structural features of the duplexes containing nucleoside analogues with a fixed N-type conformation, 2´-O,4´-C-methyleneribonucleosides. Tetrahedron Lett 1998; 39: 5401-5404.
[43] Kubota K, Ohashi A, Imachi H, Harada H. Improved *in situ* hybridization efficiency with locked-nucleic-acid-incorporated DNA probes. Appl Environ Microbiol 2006; 72: 5311-5317.
[44] Dunbar SA. Applications of Luminex(R) xMAP(TM) technology for rapid, high-throughput multiplexed nucleic acid detection. Clin Chim Acta 2006; 363: 71-82.
[45] Diaz MR, Fell JW. Use of a suspension array for rapid identification of the varieties and genotypes of the *Cryptococcus neoformans* species complex. J Clin Microbiol 2005; 43: 3662-3672.
[46] Diaz MR, Fell JW. High-throughput detection of pathogenic yeasts of the genus *Trichosporon*. J Clin Microbiol 2004; 42: 3696-3706.

[47] Dunbar SA, Vander Zee CA, Oliver KG, Karem KL, Jacobson JW. Quantitative, multiplexed detection of bacterial pathogens: DNA and protein applications of the Luminex LabMAP(TM) system. J Microbiol Meth 2003; 53: 245-252.
[48] Innogenetics: EP0905258 (**1999**).
[49] Arnold, L., Nelson, N.: US6004745 (**1999**).
[50] Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 1980; 32: 314-331.
[51] Schwartz, DC, Cantor CR. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell 1984; 37: 67-75.
[52] Mullis, K.: US4683202 (**1987**).
[53] Grolla, A., Czub, M.: CA2522629 (**2004**).
[54] Silverstein, S., Lungu, O., Wright, T.: US5543294 (**1996**).
[55] Wolf C, Rentsch J, Hubner P. PCR-RFLP analysis of mitochondrial DNA: a reliable method for species identification. J Agric Food Chem 1999; 47: 1350-1355.
[56] Zehner R, Zimmermann S, Mebs D. RFLP and sequence analysis of the cytochrome b gene of selected animals and man: methodology and forensic application. Int J Legal Med 1998; 111: 323-327.
[57] Leggett, C., Whitehouse, E., Reeves, R.: US20026395475 (**2002**).
[58] Stull TL, LiPuma JJ, Edlind TD. A broad-spectrum probe for molecular epidemiology of bacteria: ribosomal RNA. J Infect Dis 1988; 157: 280-286.
[59] Liao D. Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. J Mol Evol 2000; 51: 305-317.
[60] Zabeau, M., Vos, P.: EP0534858 (**1993**).
[61] Blears M, De Grandis S, Lee H, Trevors J. Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. J Ind Microbiol Biot 1998; 21: 99-114.
[62] Mueller UG, Wolfenbarger L. AFLP genotyping and fingerprinting. Trends Ecol Evol 1999; 14: 389-394.
[63] Livak, K., Rafalski, J., Tingey, S., Williams, J.: US5126239 (**1992**).
[64] Williams J, Kubelik AR, Livak K, Rafalski J, Tingey S. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl Acids Res 1990; 18: 6531-6535.
[65] Grosz, R., Jensen, M.: US5660981 (**1997**).
[66] Lee J, Chang JG. Random amplified polymorphic DNA polymerase chain reaction (RAPD PCR) fingerprints in forensic species identification. Forensic Sci Int 1994; 67: 103-107.
[67] Lehmann PF, Lin D, Lasker BA. Genotypic identification and characterization of species and strains within the genus *Candida* by using random amplified polymorphic DNA. J Clin Microbiol 1992; 30: 3249-3254.
[68] Partis L, Wells R. Identification of fish species using random amplified polymorphic DNA (RAPD). Molecular and Cellular Probes 1996; 10: 435-441.
[69] Salem H, Ali B, Huang T-H, *et al*. Use of random amplified polymorphic DNA analysis for economically important food crops. J Integr Plant Biol 2007; 49: 1670-1680.
[70] Buntjer J, Lenstra J. Mammalian species identification by interspersed repeat PCR fingerprinting. J Ind Microbiol Biot 1998; 21: 121-127.
[71] Jurka J, Zietkiewicz E, Labuda D. Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the mesozoic era. Nucleic Acids Res 1995; 23: 170-175.
[72] Melancon, S., Labuda, D., Tang, J., Vanasse, M.: WO34979 (**1996**).
[73] Versalovic J, Koeuth T, Lupski R. Distribution of repetitive DNA sequences in eubacteria and application to finerpriting of bacterial genomes. Nucl Acids Res 1991; 19: 6823-6831.
[74] Hulton CS, Higgins CF, Sharp PM. ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. Mol Microbiol 1991; 5: 825-834.
[75] Beck, J.: WO42609 (**1999**).
[76] DeSalle, R., Birstein, V.: WO43618 (**1997**).
[77] Kusama, T., Kadowaki, K., Nomura, T.: CA2461899 (**2005**).
[78] Song, W.-Y., Kim, H.-M., Schaad, N.: US20026423499 (**2002**).
[79] Zenser, N., Song, K., Zhang, F.: US2007264639 (**2007**).
[80] Grosz, R., Jensen, M.: US5340728 (**1994**).
[81] Yourno, J.: US5556773 (**1996**).
[82] Lastrucci, R.: US20077273730 (**2007**).
[83] Lockley AK, Bardsley RG. DNA-based methods for food authentication. Trends Food Sci Tech 2000; 11: 67-77.
[84] Mackay I. Real-time PCR in the microbiology laboratory. Clin Microbiol Infect 2004; 10: 190-212.
[85] Buelte, M., Schoenenbruecher, H., Abdulmawjood, A.: WO05061726 (**2005**).
[86] Clarebout, G.: WO07137650 (**2007**).
[87] Rothman, R., Yang, S., Lin, S., Kelen, G.: WO02070728 (**2002**).
[88] Tyagi S, Kramer F. Molecular Beacons: probes that fluoresce upon hybridization. Nat Biotech 1996; 14: 303-308.
[89] Forster P. To Err is Human. Ann Hum Genet 2003; 67: 2-4.
[90] Clarridge J. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev 2004; 17: 840-862.
[91] Dodge, D., Smith, D.: US20006054278 (**2000**).
[92] Hsieh HM, Chiang HL, Tsai LC, *et al.* Cytochrome b gene for species identification of the conservation animals. Forensic Sci Int 2001; 122: 7-18.
[93] Parson W, Pegoraro K, Niederstatter H, Foger M, Steinlechner M. Species identification by means of the cytochrome b gene. Int J Legal Med 2000; 114: 23-28.
[94] Avise JC, Walker D. Species realities and numbers in sexual vertebrates: perspectives from an asexually transmitted genome. Proc Natl Acad Sci USA 1999; 96: 992-995.
[95] Johns GC, Avise JC. A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. Mol Biol Evol 1998; 15: 1481-1490.
[96] Kocher, TD, Thomas WK, Meyer A, *et al*. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci USA 1989; 86: 6196-6200.
[97] Hebert PD, Cywinska A, Ball SL, Dewaard JR. Biological identifications through DNA barcodes. Proc Biol Sci 2003; 270: 313-321.
[98] Hajibabaei M, Singer GA, Hebert PD, Hickey DA. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet 2007; 23: 167-172.
[99] Hebert PD, Ratnasingham S, Dewaard JR. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc Biol Sci 2003; 270 Suppl 1: S96-S99.
[100] Hebert PD, Gregory TR. The promise of DNA barcoding for taxonomy. Syst Biol 2005; 54: 852-859.
[101] Meyer CP, Paulay G. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol 2005; 3: e422.
[102] Moritz C, Cicero C. DNA barcoding: promise and pitfalls. PLoS Biol 2004; 2: e354.
[103] Gevers D, Cohan FM, Lawrence JG, *et al.* Opinion: re-evaluating prokaryotic species. Nat Rev Microbiol 2005; 3: 733-739.
[104] Heller MJ. DNA microarray technology: devices, systems, and applications. Annu Rev Biomed Eng 2002; 4: 129-153.
[105] Tiedje, J., Cho, J.-C.: WO02101094 (**2002**).
[106] Wiesinger-Mayr, H., Pichler, R., Bodrossy, L., Noehammer, C.: WO08003114 (**2008**).
[107] Pastinen T, Partanen J, Syvanen AC. Multiplex, fluorescent, solid-phase minisequencing for efficient screening of DNA sequence variation. Clin Chem 1996; 42: 1391-1397.
[108] Soderlund, H., Syvanen, A.-C.: US20006013431 (**2000**).
[109] Pastinen T, Raitio M, Lindroos K, *et al.* A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. Genome Res 2000; 10: 1031-1042.
[110] Cai H, White PS, Torney D, *et al.* Flow cytometry-based minisequencing: a new platform for high-throughput single-nucleotide polymorphism scoring. Genomics 2000; 66: 135-143.
[111] Grossman PD, Bloch W, Brinson E, *et al.* High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation. Nucleic Acids Res 1994; 22: 4527-4534.
[112] Gill P, Ghaemi A. Nucleic acid isothermal amplification technologies - a review. Nucleoside Nucleotide Nucleic Acids 2008; 27: 224-243.
[113] Fortina P, Kricka LJ, Surrey S, Grodzinski P. Nanobiotechnology: the promise and reality of new approaches to molecular recognition. Trends Biotechnol 2005; 23: 168-173.
[114] Lowe CR. Nanobiotechnology: the fabrication and applications of chemical and biological nanostructures. Curr Opin Struct Biol 2000; 10: 428-434.
[115] Lieber, C., Hafner, J., Cheung, C., Kim, P.: WO0226624 (**2002**).

[116] Woolley AT, Guillemette C, Li CC, Housman DE, Lieber CM. Direct haplotyping of kilobase-size DNA using carbon nanotube probes. Nat Biotechnol 2000; 18: 760-763.
[117] Kerman K, Kobayashi M, Tamiya E. Recent trends in electrochemical DNA biosensor technology. Meas Sci Technol 2004; 15: R1-R11.
[118] Wang J. Nanomaterial-based electrochemical biosensors. Analyst 2005; 130: 421-426.
[119] Wang J. Electrochemical biosensors: towards point-of-care cancer diagnostics. Biosens Bioelectron 2006; 21: 1887-1892.
[120] Lim A, Dimalanta ET, Potamousis KD, *et al.* Shotgun optical maps of the whole Escherichia coli O157:H7 genome. Genome Res 2001; 11: 1584-1593.
[121] Schwartz, D., Wu, T.: US20036607888 (**2003**).
[122] Hughes S, Arneson N, Done S, Squire J. The use of whole genome amplification in the study of human disease. Prog Biophys Mol Bio 2005; 88: 173-189.
[123] Hawkins TL, Detter J, Richardson P. Whole genome amplification - applications and advances. Curr Opin Biotech 2002; 13: 65-67.
[124] Lasken R, Egholm M. Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. Trends Biotechnol 2003; 21: 531-535.
[125] Barber AL, Foran DR. The utility of whole genome amplification for typing compromised forensic samples. J Forensic Sci 2006; 51: 1344-1349.