(54) Title: OLIGONUCLEOTIDE PRIMERS FOR NUCLEOTIDE INDEXING OF POLYMORPHIC PCR PRODUCTS AND METHODS FOR THEIR USE

(57) Abstract: The present invention relates to a new oligonucleotide for use as PCR primer, comprising three different portions, i.e. a portion complementary to universal primers for automated sequencers, a read start portion and a portion complementary to a region of the target sequence to be amplified, to methods for detection of polymorphisms or polymorphism patterns of nucleotide sequences amplified by PCR with said oligonucleotides, methods for detection and analysis of said polymorphisms or polymorphism patterns and to kits comprising products useful for carrying out said methods.

## OLIGONUCLEOTIDE PRIMERS FOR NUCLEOTIDE INDEXING OF POLYMORPHIC PCR PRODUCTS AND METHODS FOR THEIR USE

### DESCRIPTION

The present invention relates to a new oligonucleotide for use as PCR primer, comprising three different portions, i.e. a portion of a universal primer for automated sequencers, a read start portion and a portion complementary to a region of the target sequence to be amplified, methods for the detection or the identification of polymorphisms or polymorphism patterns of nucleotide sequences amplified by PCR with the said oligonucleotides, methods for detection and analysis of the said polymorphisms or polymorphism patterns and to kits comprising products useful for carrying out the said methods.

### STATE OF THE ART

The identification the detection and the analysis of nucleotide sequences polymorphisms are widely used and of great interest at least for medical, forensic, genetic, alimentary analysis. In fact, polymorphic patterns in determined loci allow the detection of genetic diseases, allow the genotypic characterization of individuals and their recognisability within a species, allow the identification of species in animal and plant families or classes or orders or phyla from tissue or cells samples.

The usefulness of polymorphisms analysis is not limited to the above and widens at a very high rate together with the evolution of biotechnological tools and knowledge. It is of great interest, hence, to easy and speed up the detection and/or the identification of polymorphisms inside a nucleotide sequence and to facilitate as much as possible the analysis of the detected or of the identified polymorphisms.

PCR is widely used for the amplification of polymorphic sequences at given loci and provides sufficient material for carrying out sequence analysis also from very low amounts of starting nucleic acid material. Moreover, automated sequencers further facilitate the work (and improve the health) of the researchers as nucleotide sequences can now be obtained without the use of hazardous denaturing gels and/or radio-labelled nucleotides. Even the genetic banks available online as well as sequence alignments programs freely available online further facilitate the work of the researchers speeding up the detection of known polymorphisms or the identification of new polymorphisms and the analysis of polymorphic patterns in general.

As already said, a very important use of polymorphism analysis is in the medical field, both for the detection of various diseases in individuals and even in analysis before birth, and for the detection and identification of new polymorphisms related to specific diseases, hence nucleotide polymorphisms can be used as genetic

2

markers and/or for the development of new pharmaceutical compounds.

As polymorphic patterns are characteristics of individuals at certain loci, of species at other loci, of families, of orders and of phyla at even other loci, detection, identification and analysis of specific nucleotide polymorphisms or of polymorphic patterns in one or more loci can allow the assessment of paternity or maternity and can provide aid in the identification of criminals.

For the same reasons above, as polymorphisms and/or polymorphic patterns at certain loci can allow the identification of different species, polymorphisms analysis can also be carried out in order to track food origins and to identify the composition of a certain food product and to verify whether certain products are or are not composed of the claimed material.

Up to date, there are various polymorphisms that are considered by the researchers, the most "obsolete" are the Restriction Fragment Length Polymorphisms (RFLPs), which are based on the presence or absence of a given restriction site and on the difference in the length of the digested fragments obtained by digestion with a specific restriction enzyme, triplets polymorphisms, as the polymorphisms present in the human X fragile site, Short Tandem Repeats polymorphisms (STRs) as well as Single Nucleotide Polymorphisms (SNPs).

The use of automated sequencers often causes some problem to the researchers when the polymorphism of interest is located very close to the 5' of the amplicon obtained by PCR, as some of the nucleotides at the 5' are "lost" by the sequencer and no information is hence provided on the polymorphism of interest. Usually, a 5' lengthening of the primer, that shifts the position of the nucleotides of interest in the amplicon, allows a sequencing result that provides information on the nucleotides of interest by the automated sequencer.

When the sequence is obtained by the sequencer, the information is used in one or more sequence alignments against the target sequence and, if needed, against the various amplicons obtained, in order to identify whether polymorphisms, and which polymorphisms are present. The analysis of the sequences and of the possible polymorphisms present therein and of polymorphic patterns when a pool of sequences is under study, is at present a step (or several steps) time consuming and toilsome that cannot be fully automated and requires each time the work of the researcher for determining whether polymorphisms are present and for the screening thereof and the identification of specific polymorphic patterns.

Up to date, oligonucleotides for PCR (especially multiplex PCRs) exist, that renders the sequencing task simpler by the addition, at the 5' of the said primers, of a sequence complementary to universal primers for automated sequencers which

3

allows the use of a single sequencing primer for sequencing the amplified product of a multiplex PCR, as disclosed in EP0832290.

An example of methods and primers used for the detection of polymorphisms that allow the detection of 5 species belonging to the *Scombridae* family is described in Bottero et al 2007 in which the region of the *cytochrome b* gene amplified in this work is the same as fragment A of the present invention (see section VIII). Fragment A or A plus sequence information is needed in some cases in order to be able to distinguish some variants in different species (e.g. *Thunnus thynnus* and *Thunnus maccoyii*).

In another example, a kit called "BIOFISH SEQ Tuna" marketed by Biotools (http://www.biotools.net), allows the distinction of 9 different species within the *Scombridae* family by the analysis of a fragment of the *cytochrome b* mitochondrial gene and DNA sequencing from an amplified fragment of 379bp.

Some technical details of this assay are not known (e.g. primers sequence, identification level of the assay with regard to the intraspecies variability, etc.).

The strategy of the primers design and the method used in the present invention enable the user to detect and/or identify and also to analyse nucleotide sequence polymorphisms in the forensic, alimentary, medical and genetic field. The method includes the use of the Polymerase Chain Reaction (PCR) to amplify polymorphic regions of the genome from total cellular DNA and subsequent sequencing of the PCR products and identification of SNPs (Single Nucleotide Polymorphism) or other DNA polymorphisms in samples in which DNA may be also very degraded (e.g. highly processed meat products). The primers herein described are also used in a molecular method for the amplification of small regions of the mitochondrial *cytochrome b* gene for the identification of 17 fish species in the *Scombridae* family.

## DESCRIPTION

The present invention, discloses new single stranded oligonucleotides comprising, ordered from 5' to 3', a portion of a universal primer for automated sequencers located at the 5' of said oligonucleotide, a read start portion and a portion complementary to a region of the target sequence to be amplified; a method for the detection of polymorphisms or polymorphisms patterns of nucleotide sequences amplified by PCR comprising the steps of:

    a. sequencing nucleotide sequences amplified with one or more pair, forward and reverse, of oligonucleotides according to the description with an automated sequencer using a universal primer comprising said 5' portion;

    b. detecting in said sequences said read start portion;

4

c. indexing nucleotides of said amplified sequences, from said detected sequence said read start portion;

d. detecting in said indexed sequences polymorphisms or polymorphisms patterns;

a method for the analysis of polymorphisms or polymorphisms patterns of nucleotide sequences amplified by PCR comprising the steps of:

a. sequencing nucleotide sequences amplified with one or more pair, forward and reverse, of oligonucleotides according to the description with an automated sequencer using a universal primer comprising said 5' portion;

b. detecting in said sequences said read start portion;

c. indexing nucleotides of said amplified sequences, from said read start portion;

d. detecting in said indexed sequences polymorphisms or polymorphisms patterns;

e. analysing the data obtained at step d;

a method for the detection and identification of new polymorphisms or polymorphisms patterns of nucleotide sequences amplified by PCR comprising the steps of:

a. sequencing nucleotide sequences from given samples amplified with one pair, forward and reverse, of oligonucleotides according to the description with an automated sequencer using a universal primer comprising said 5' portion;

b. detecting in said sequences said read start portion;

c. indexing nucleotides of said amplified sequences, from said read start portion;

d. detecting in said indexed sequences the presence polymorphisms or polymorphisms patterns in comparison to a reference sequence;

e. sequencing nucleotide sequences from the same samples of step a. amplified each with a different pair, forward and reverse, of oligonucleotides according to the description, each pair of oligonucleotides having a different read start, with an automated sequencer using a universal primer comprising said 5' portion;

f. detecting in said sequences said read start portions;

g. indexing nucleotides of said amplified sequences, from said read start portions, each sample being recognisable by a specific read start portion;

h. identifying in said indexed sequences the polymorphisms or polymorphisms patterns in comparison to a reference sequence and,

5

optionally,

i. analysing the data obtained at step h;

a method for the detection and identification of new polymorphisms or polymorphisms patterns of nucleotide sequences amplified by PCR comprising the steps of:

a. sequencing nucleotide sequences from different samples each sample being amplified with a different pair, forward and reverse, of oligonucleotides according to the description and each pair of oligonucleotides having a different read start, with an automated sequencer using a universal primer comprising said 5' portion;

b. detecting in said sequences said read start portions;

c. indexing nucleotides of said amplified sequences, from said read start portions, each sample being recognisable by a specific read start portion;

d. identifying in said indexed sequences the polymorphisms or polymorphisms patterns in comparison to a reference sequence and, optionally,

e. analysing the data obtained at step d;

a software capable of carrying out the methods described above;

a computer readable storage support wherein the software above is stored;

a kit comprising one or more aliquot of at least two oligonucleotides, forward and respective reverse, oligonucleotides of the invention, and optionally one or more aliquot of a universal primer for automated sequencer wherein said universal primer comprises a region that is complementary to said 5' portion.

DETAILED DESCRIPTION OF THE FIGURES

Figure 1a represents an example of a PCR primers Cocktail with the primers of the description, said primers comprising three essential portions: a white portion (in the Fw, Forward, and Rw, Reverse, primer) herein named "PUP" (Partial Universal Primer) which is a portion of a universal primer for automated sequencers located at the 5', a dotted portion called RS (Read Start) signal and a gray portion of the gene (or sequence) of interest to be amplified (GOI) and a suitable pair of sequencing primers three portions as well, i.e. a white portion (in the Fw and Rw primer) called UP (Universal Primers), a dotted portion called RS (Read Start) signal and a small gray portion of the gene of interest (GOI).

Figure 1.b shows the same sequence primers of figure 1.a.

Figure 1.c shows another suitable pair of primers comprising only the: white UP portion (in the Fw and Rw primer). When the sequencing primers of figure 1.c are used, the PCR primer shall comprise a white UP portion of a length suitable for

6

annealing with the UP sequencing primer.

Figure 1.d shows a further sequencing primer pair comprising only the: white UP portion (in the Fw and Rw primer) as well as the dotted (Read Start) signal.

Figure 2 represent a scheme of the method of the specification in which PCR primers as represented in Figure 1a and sequencing primers as represented in Figure 1.b are used for the detection of 17 species of the family *Scombridae*, the UP and PUP portions of figure 1 are represented by a U succession, the "read start" sequence is represented with a waved line (∽∽∽∽), triangles, crosses, stars, etc represent polymorphic sites.

Figure 3 shows an agarose gel with the results of the PCR assay B (see table 3 for oligonucleotides that amplify fragment B) run on DNA from different species in the *Scombridae* family. Lane 1 to 15 commercially available species of the *Scombridae* family: lane 1 *Thunnus thynnus*, lane 2 *Thunnus alalunga*, lane 3 *Katsuwonus pelamis*, lane 4 *Thunnus albacares*, lane 5 *Thunnus obesus*, lane 6 *Scomber colias*, lane 7 *Scomber japonicus*, lane 8 *Scomber australasicus*, lane 9 *Thunnus albacares*, lane 10 *Thunnus obesus*, lane 11 *Scomber scombrus*, lane 12 *Thunnus thynnus*, lane 13 *Sarda sarda*, lane 14 *Thunnus albacares*, lane 15 *Thunnus albacares,* lane 16 negative PCR control, lane 17 molecular weight marker (GeneRuler™ 100bp DNA Ladder plus, Fermentas life sciences). The PCR amplicon B from lane 1 to 8 is obtained using DNA extracted from highly processed canned meat in oil, the PCR amplicon in lane 9 is obtained using DNA from highly processed canned natural meat, the PCR amplicon from lane 10 to 13 is obtained using DNA extracted from fresh fish, the PCR amplicon in lane 14 is obtained using DNA extracted from sauce with 6% of Thunnus sp., in lane 15 from sauce with 20% of *Thunnus sp*. A PCR fragment of 162 bp is visible from lane 1 to 15.

Figure 4 shows an agarose gel with the results of PCR assay A+B (see table 5 for oligonucleotides that amplify fragment A+B) run on DNA from different species in the *Scombridae* family. Lane 3 to 6 commercially available species of the *Scombridae* family. Lane 1 negative PCR control, lane 2 molecular weight marker (GeneRuler™ 100bp DNA Ladder plus, Fermentas life sciences), lane 3 *Thunnus albacares*, lane 4 *Thunnus thynnus,* lane 5 *Scomber scombrus,* lane 6 *Thunnus obesus*. The PCR amplicon A+B from lane 3 to 6 is obtained using DNA extracted from fresh fish. A PCR fragment of 291 bp is visible from lane 3 to 6.

Figure 5 shows an agarose gel with the results of the PCR assay A (see table 7 for oligonucleotides that amplify fragment A) run on DNA from different species in the *Scombridae* family. Lane 1 to 13 and line 15 are commercially available specie of the *Scombridae* family: lane 1 *Thunnus thynnus*, Lane 2 *Thunnus alalunga*, lane 3

*Katsuwonus pelamis*, lane 4 *Thunnus albacares*, lane 5 *Thunnus obesus*, lane 6 *Scomber colias*, lane 7 *Scomber japonicus*, lane 8 *Scomber australasicus*, lane 9 *Thunnus obesus*, lane 10 *Scomber scombrus*, lane 11 *Thunnus thynnus*, lane 12 *Sarda sarda,* lane 13 *Auxis rochei,* lane 14 *sardine,* lane 15   *Thunnus albacares,* 16 *sardine,* lane 17 negative PCR control, lane 18 molecular weight marker (GeneRuler™ 100bp DNA Ladder plus, Fermentas life sciences). The PCR amplicon A from lane 1 to 8 is obtained using DNA extracted from highly processed canned meat in oil, the PCR amplicon from lane 9 to 13 is obtained using DNA extracted from fresh fish, the PCR amplicon in lane 14 and 16 is obtained using DNA extracted from canned sardine in oil, the PCR fragment in lane 15 is obtained using DNA extracted from fresh fish. A PCR fragment of 156 bp is visible from lane 1 to 13 and in lane 15.

Figure 6 fragment of sequence of *cytochrome b* gene of two different species of the *Scombridae* family. The circles highlight different SNPs that discriminate two different species in the *Scombridae* family.

Figure 7 shows the fragments of the *cytochrome b* gene amplifiable with the primers of the invention.

Figure 8 shows the genetics details on the amplified fragments.

DETAILDE DESCRIPTION OF THE SEQUENCES

SEQ ID NOs 1-9, 12-15, 17-24; and 26-30 of the sequence listing are primers for amplification or sequencing of *cytochrome b* gene of the *Scombridae* family according to the present description, the sequence being disclosed also in tables 3, 5, 7 and 9; SEQ IDs 10, 11, 16, 25 and 31 are sequencing primers according to the present description, the sequences being disclosed also in tables 4-10.

DETAILED DESCRIPTION OF THE INVENTION

Glossary in the meaning of the specification:

**Read start portion (RSP)**. Herein also referred as Read start sequence, or merely Read start. In the meaning of the present specification, "read start portion" "Read start sequence" or "Read start" is a portion of an oligonucleotide, that is a PCR primer, that is located before the region of the primer that is complementary to a portion of a target sequence to be amplified, that allows an easy localisation of the sequence of interest, and an indexing of each nucleotide of the same, in all the amplicons obtained by amplifying with a PCR primer comprising an RSP. The RSP allows the determination of a position immediately downstream said RSP, or a given number of nucleotides downstream said RSP, from which an indexing of the nucleotides of the amplified sequence can be carried out, either by a researcher either by a computer in which a suitable software is installed. Different primer pairs sharing

8

the same read start portion will generate amplified sequences all sharing the same read start.

Primer pairs differing in the read start sequence will allow not only a common indexing of amplicons of the same region of interest from different samples, but also the identification, based on the RSP in each amplicon, of the samples from which it derives.

The complementary sequence of the read start portion, or complementary read start portion (cRSP), or complementary read start sequence, will be present in one of the strands of the amplicon or in the amplified sequence, when oligonucleotides herein provided are used.

The read start portion is a sequence of about 4 to 12 bp of length, characterised in that it does not anneal with the target sequence or with the 3' universal primer in case of sequence primer 1.c, (see figure 1.b) at the PCR conditions to be used. The length and the nucleotide sequence of a RSP or read start sequence can change as explained below, also depending on the length of the GOI (gene (or sequence) of interest to be amplified) and PUP (Partial Universal Primer) portions sequences (see figure 1.a). A suitable read start sequence can easily and readily be identified by the skilled person by the use of simple computer programs that enable the user to verify that a sequence does not anneal with the sequences indicated above and, of course with the reverse primer. Programs of said kind are also freely available online (e.g. programs available online such as BioEdit: http://www.mbio.ncsu.edu/BioEdit/BioEdit.html or ClustalX: http://www-igbmc.u-strasbg.fr/BioInfo/ClustalX/Top.html). Moreover, on several species, the full sequencing of the entire genome allows the determination of "rare sequences" that could be used as starting points for when a read start sequence is to be designed.

The skilled person can design suitable read start portions without use of inventive skill as several public or commercially available databases provide for data on the sequences to be amplified, and easy to use and easily available programs allows a fast identification of a small sequence that does not anneal with the target sequence of interest and with the universal primer to be used.

**Universal primer for automated sequencer.** Universal primers for automated sequencer are known in the art and (e.g. Universal Forward 20mer 5' GTTGTAAAACGACGGCCAGT 3' and Universal Reverse 20mer 5' CACAGGAAACAGCTATGACC 3' available on http://www.genome.ou.edu/protocol_book/protocol_adxCDE.html).

These primers are called universal because they can be used to amplify or sequence any insert that is put in the multiple cloning site. Universal primers are

9

really not 'universal' in the sense that they will bind to anything. In fact, universal primers are PCR/sequencing primers that bind to a sequence found in many plasmid cloning vectors, most of which are derived from pUC vectors (which in turn come from pBR322). These sequences were defined as good PCR and sequencing sites as they flank the multiple cloning site where an inserted DNA sequence would be put. According to this description, a universal primer for automated sequencer is an oligonucleotide that does not hybridize, at the PCR conditions to be used, with the sequence to be amplified and is hence unrelated to the target sequence to be amplified.

**Forward and respective reverse oligonucleotide**. By forward and respective reverse oligonucleotide the present specification indicates an oligonucleotide pair (PCR primers) specifically designed to amplify a target sequence, hence, given a target sequence, by "forward and respective reverse oligonucleotide", a primers pair (forward and reverse), each specifically designed for use together with the other for the amplification of said target sequence, is intended.

**Indexing the nucleotides of an amplified sequence.** Indexing nucleotides of an amplified sequence is, herein, the process of establishing an access key to each nucleotide of said sequence directly identifying its position. Hence, each nucleotide of said sequence is identified by an index, said index being a number, a symbol, and, of course, not being the mere letter indicating the nature of the nucleotide itself (e.g. A T G C U N etc). Symbols (normally letters) customarily used in the art for identifying the chemical nature of a nucleotide are excluded as indexes. The indexing of an amplified sequence could be, by way of example, giving to each amplified nucleotide complementary to the target sequence, the same (position) numbering that the said nucleotide has in the original target sequence, i.e. amplification is carried out on nucleotides 345 to 451 of a target sequence, the indexing can be set so to allow the identification, in the sequences resulting from the amplification, of each amplified nucleotide corresponding to the nucleotide in the target sequence, with the same numbering of the corresponding nucleotide of the target sequence. This means, in the example given, that if a read start portion is used, the first nucleotide of the sequences resulting from the amplification after the complementary read start portion will be indexed as 345, the second as 346 and so on.

**Amplicon**. In the meaning of the present specification, the term amplicon is used to define a nucleotidic fragment generated by a PCR amplification.

**Complementary**. In the meaning of the specification a nucleotide sequence complementary to another sequence is a sequence of polynucloetides related by the sequence to which it is defined as "complementary" via a perfect match by the base-

10

pairing rules. For example, the sequence complementary to the DNA sequence T-C-G-A is A-G-C-T. As known in genetics, a given sequence defines its complementary sequence.

Hence, in the present description "a portion complementary to a region of the target sequence to be amplified" means that the said portion of the primer perfectly matches a defined number of consecutive nucleotides of the sequence of interest to be amplified or its complementary strand.

**Polymorphisms pattern or polymorphic pattern**. By polymorphisms pattern or polymorphic patters it is intended that, with reference to a given nucleotide sequence, more than one single nucleotidic variation can be present at one or mere site (ex. more than one SNP), each possible combination of the said polymorphisms is intended as a polymorphic pattern. By way of example, different species can often be recognised by the fact that they exhibit, in a certain sequence, a certain number of SNPs, each specie presenting characteristics SNPs, i.e. a specie specific polymorphisms pattern or polymorphic pattern.

**Detection of, or, detecting polymorphisms or polymorphisms patterns**. In the specification by detecting polymorphisms or polymorphisms patterns is intended as detecting known polymorphisms or polymorphisms patterns as well as detecting (i.e. detecting the existence of) new polymorphisms or polymorphisms patterns. A correct assessment with reference to a given sequence, of new polymorphisms or polymorphisms patterns (i.e. where and how the nucleotides change) is herein defined as identifying new polymorphisms or polymorphisms patterns.

**Analysing data obtained** (by the detection of, or, detecting polymorphisms or polymorphisms patterns). The analysis of the data obtained by the detection or by detecting polymorphisms or polymorphisms patterns is intended as an elaboration of said data such as establishing a genotype, using said data for confronting them against known polymorphisms or polymorphisms pattern (i.e. identifying whether the amplified sequences display polymorphisms or polymorphisms patterns corresponding to the polymorphisms or polymorphisms patterns of a given individual, or of a given species, family, order, class or phylum, or corresponding to polymorphisms or polymorphisms patterns known to be linked to a specific disease, or to identify new genetic markers for a given pathology, etc.). Also, confrontation of the data with a reference sequence can allow detection of the presence of new polymorphisms or polymorphisms patterns, or a punctual identification of each of said new polymorphisms or polymorphisms patterns.

**Computer readable storage support**. By computer readable storage support it is intended any support suitable for the storage of software that allow the

installation of said software on a computer, e.g. CD, DVD, TAPES, USBPen, EPROM, disks, hard disks, etc. and/or the software can be downloadable from a network.

**Complex food matrix** herein defines a foodstuff product containing a certain natural product (i.e. meat, fish, vegetable etc.) wherein said product has been processed (i.e. smoked, cooked, fractioned etc.). By way of example a complex food matrix related to tuna could contain tuna, little tunny, false albacore, mackerels, pelamyd, i.e. derivatives form all species of the *Scombridae* family in the form of canned fish, in water or in oil, in the form of pastes, smoked slices, salami, pâtés, sauces etc.

**Primer cocktail**. By primer cocktail it is intended a mix of different primer pairs to be used in the same PCR amplification, an example is given by primer pairs capable of selectively amplifying a given allele of a polymorphic sequence, a primer cocktail in this case could be represented by primer pairs, used in the same amplification process, each pair amplifying, when present in the target sequence, a specific polymorphism.

**Nucleotides letter code convention** in the present specification, according to the standard nucleotide translation conventions, when a nucleotide is indicated with the letter Y, a pyrimidine, i.e. C or T are indicated, when the letter R is used, it indicates a purine, i.e. A or G.

| Nucleotide translation conventions | | | | |
|---|---|---|---|---|
| **Nucleotide(s)** | **IUB/GCG**[1] | **Staden**[2] | **Sanger**[3] | **Stanford**[4] |
| A | A | A | A | A |
| C | C | C | C | C |
| G | G | G | G | G |
| T | T/U | T | T | T/U |
| (AC) | M | 5 | 5 | j |
| (AG) | R | R | R | R |
| (AT) | W | 7 | 7 | L |
| (CG) | S | 8 | 8 | M |
| (CT) | Y | Y | Y | Y |
| (GT) | K | 6 | 6 | K |
| (AGC) | V | - | - | - |
| (ACT) | H | - | - | - |
| (AGT) | H | - | - | - |
| (CGT) | D | - | - | - |
| (AGTC) | X/N | - | X | - |

12

[1]Internation Union of Biochemistry convention translates sequences for GENBANK, University of Wisconsin Genetic Computing Group, EMBL and the National Biomedical Research Foundation.

[2]The Staden convention is used by the Cambridge University database.

[3]The Sanger convention is not currently used.

[4]The Stanford convention is used by the Stanford University database.

The oligonucleotides of the description comprise, ordered from 5' to 3' a portion at the 5', of a universal primer for automated sequencers, a read start portion and a portion complementary to a region of the target sequence to be amplified.

The three portions indicated above can also be directly linked one to another, optionally, restriction fragment sites or other sites of interest can be introduced into the oligonucleotides, preferably in upstream the read start portion.

The region of the universal primer for automated sequencers has at least two functions; the first one is to elongate the oligonucleotide at the 5', thus elongating the amplified fragment. This elongation can have also two effects, the first effect is to elongate amplified fragments rendering their size a bit larger, the effect being quite useful when short fragments are amplified as short fragments are, in general, more difficult to sequence. Moreover, when an automated sequencer is used, there is a loss in the sequencing of some nucleotides after the sequence primer. In theory, the first base after the sequence primer is legible, but in real experimentation, technical limitations of the automated sequencer do not normally allow to read the sequence up to at least the first 16 bp. So the second effect is that elongation can enable the user to increase the distance from the last nucleotide at the 3' of the sequencing primer and the start of the region of interest where a SNPs or other polymorphisms might be positioned. In fact, the longer the portion of the universal primer at 5' in the PCR oligonucleotides of the description, the shorter the sequence primers at the 3' can be.

Hence, the sequencing primer will be long enough also if it does not overlap completely (i.e. is shorter) at the 3' with the universal primer region of the oligonucleotide of the invention. About 15 bp of complementary sequence between the 5' PCR primer and 3' sequence primer is a suitable embodiment of the invention.

Hence, where needed, the presence of the portion of the universal primer for automated sequencing can be also used to further elongate the amplified fragment and to obtain a more reliable sequencing.

In fact, the sequencing step is usually carried out by further amplifying the amplified fragments using suitable primers, an unlabelled deoxyribonucleotide (dNTPs) mix and labelled dideoxyribonucleotides in order to obtain stopped products each time a dideoxyribonucleotide is inserted in the amplified fragment, thus

13

obtaining, at the end of the PCR amplification a pool of fragments, each labelled at the last nucleotide (each nucleotide having a different marker) that can be purified and automatically read by the sequencer.

5      Elongation of the amplified fragment can often provide a more reliable sequencing as the smallest fragments are often lost in the above mentioned purification.

In case the user of the primers of figure 1 desires to further elongate the amplified fragment obtained by the PCR with the primers of the invention, the region at the 5' of the oligonucleotide of the invention, will be conveniently designed in
10     order to consist of the nucleotides located near the 3' (or including the 3') of the universal primer in order to obtain a longer product in the PCR reaction for sequencing.

Furthermore, the presence of a sequence in the amplicon, suitable for annealing with a universal primer for automated sequencers renders the sequencing
15     step very simple especially when a cocktail of primers is used in the PCR amplification preceding the sequencing PCR (which is quite common when polymorphism analysis is performed). The 5' portion of the oligonucleotides herein described allows the sequencing of several different amplicons originated from different oligonucleotide pairs (the oligonucleotides being oligonucleotides of the
20     invention) with one single kind of primer in a single sequencing process.

The universal primer will not hybridize at the PCR annealing conditions with the target sequence, the length of the portion of universal primer in the oligonucleotide herein described will vary, form about 5 to about 10-15-20 nucleotides depending whether the sequencing primer is designed according to figure
25     1.b, 1.c, or 1.d. When the sequencing primer also comprises the read start portion or even some nucleotide complementary to the sequence to be amplified (that could be in a multiplex in which no polymorphisms are expected in the first nucleotides of the amplified sequence), 5 -10 nucleotides of the universal primer will be sufficient at the 5' of the oligonucleotide herein described, if the sequencing primer comprises
30     only the universal primer sequence and the read start, than the number of nucleotides of the universal primer sequence in the oligonucleotides herein described will be increased to a number of about 10-20 in order to provide a better annealing at the sequencing step, the same applies when the sequencing primer consists only of a universal primer sequence.

35      The read start portion and the region complementary to a part of the sequence to be amplified are preferably directly bound to each other. The direct contact within the complementary read start portion and the amplified sequence complementary to

14

the sequence of interest in the amplicon renders the indexing of the amplified target sequence easier.

In particular, the presence of the read start sequence in the oligonucleotides herein described, allows the indexing of the nucleotides complementary to the sequence of interest in all amplicons. The presence of a common read start sequence in all different amplicons (of the same portion of a gene or a region of interest) undoubtedly fastens the procedures for reading and interpreting the amplified sequences obtained. The presence of an easy to detect common sequence in all amplicons even allows the automation of the sequence indexing, of the collection and interpretation of the data obtained from all amplicons.

At the same time, when different samples are tested in order to detect and identify new polymorphisms or new polymorphic pattern, using primer pairs that differs only in the read start sequence will allow at the same time the amplification of the same region of interest, an identical indexing of the nucleotides of each amplicon (in fact, the position of the RSP with respect to the portion complementary to the sequence of interest will not vary) and, the identification of the origin of each amplicon (i.e. from which of each sample the amplicon derives).

When amplification on different samples with the same oligonucleotide pair or when amplification with a cocktail of primer is made, the oligonucleotides comprising the read start portion herein described will introduce in all the amplicon a read start sequence whose effect is practically exerted on each amplified sequence of interest.

The presence in all amplicons of a known read start sequence will allow also automation of the indexing of the sequence as a start point for indexing can be established in all amplicons following the read start sequence. Hence, the first nucleotide of the amplified sequence that will be indexed will be located downstream the read start sequence. This indexing can be easily performed with a computer using a software capable of recognising a predetermined read start sequence in the sequenced amplicon (the software can also be capable of sequencing and, during or after sequencing recognising the read start sequence in the amplicon) and setting the program for assigning a given indexing group of symbols. If, by way of example a certain sequence of interest has been amplified from nucleotide 87 to nucleotide 212, the program can be set so to index the nucleotides located immediately downstream the read start sequence in the amplicon automatically assessing the numbering from 87 to 212 to the amplicon. This automated indexing will allow also a fast computer analysis of specific nucleotides of interest in case one or more polymorphisms are expected in the amplicon, or, if polymorphisms are to be identified in a given

15

amplicon, the indexing will allow direct comparison within different amplicons without the need of aligning each amplicon on a sequence database as each nucleotide of interest will be indexed and hence identifiable. The oligonucleotides of this description can be used even as primers for the construction of a cDNA library, a

5    read start region will render the analysis of the clones obtained much more easy to make.

Is it also possible to amplify, in the same PCR reaction, two or more different polymorphic regions with PCR primers comprising different portions complementary to a part of the sequence of interest and a different read start sequence. For example

10    the PCR primers that amplify a portion A of a polymorphic region will comprise the read start signal A in the forward primer and A' in the reverse primer, the PCR primers that amplify a portion B of a region will comprise the read start signal B in the forward primer and B' in the reverse primer and so on. It is possible to index different amplicons using different read start.

15    When more different polymorphic regions are amplified by PCR with PCR primers comprising different read start sequences it is possible to selectively sequence a portion A or a portion B and so on using respectively the sequence primers comprising the read start A or A' or B or B' and so on. Moreover, using different read start sequences it is possible to amplify the same polymorphic region

20    in different sources in the same PCR reaction.

The read start portion of the oligonucleotide herein provided is an nucleotide sequence, which sequence of bases is invented following some criteria: like the universal primer, it does not hybridize with the sequence to be amplified at the PCR conditions established by the researcher and is hence unrelated to the target sequence

25    to be amplified, it does not hybridize with the universal primer either at the PCR conditions established by the researcher and is hence unrelated to the universal primer too. The CG-content should be considered in order to obtain nearly similar melting temperatures for the PCR primers forward and reverse. The length of the read start sequence can vary from about 4 to 12 nucleotides depending on whether

30    the sequencing primer is designed according to figure 1.b or 1.d and on the length of the portion of the PCR primers that is complementary to the region or gene of interest. The read start sequence in the forward PCR primers will be different from that in the reverse PCR primers. The read start portion in the forward sequence primer will be the same as the read start portion in the 5' of amplicon that must be

35    amplified and viceversa.

In an embodiment of the invention, the single stranded oligonucleotide described, may comprise one or more polymorphism in order to anneal only to the

16

complementary sequence comprising the same.

Where a detection of a certain polymorphism has to be carried out on a sample of nucleotide sequences, the presence of a sequence complementary to a polymorphism of a certain locus, will allow the amplification of the sequence of interest only if said polymorphism is present. In this kind of analysis, it often happens that loci with high variability are selected. It is often useful to carry out a single PCR reaction with more than one primer pair, i.e. when a certain number of polymorphisms are to be tested in a single individual (i.e. paternity or maternity assessment, or individuation of a potential criminal) or when a certain number of polymorphisms that enables to detect the presence or the absence of a certain number of species or specie variants in a sample (i.e. identification of species comprised in a food sample) for verifying food composition.

In a particular embodiment the polymorphisms comprised in the oligonucleotide can be one or more SNP, although other polymorphisms can be used and the oligonucleotide region complementary to a region of the target sequence or flanking the target sequence can be designed in order to comprise the said other polymorphisms.

Often, when there are SNPs in the region complementary to the portion of the target sequence, the sequence of the oligonucleotide will be written using also Y and R letters to indicate that, in fact, for each of the said letters 2 possible oligonucleotides can be synthesised and used, when each possible nucleotide at an SNP site provides the researcher with useful indication (i.e. a certain SNP is present in *Thunnus albacares* and not in *Katsuwonus pelamis*) oligonucleotides representing each possible combination will be used in order to detect all polymorphisms present in the amplified sample.

Hence for one Y or R nucleotide in the sequence listing two sequences are provided, supposing a y at position 13 from the 5' of the oligonucleotide sequence, that will indicate a sequence in which the nucleotide 13 from the 5' is a c, and another sequence in which the nucleotide 13 from the 5' is a t and so on.

The three portions of the oligonucleotide herein described are not particularly limited in their length, of course a skilled person will know the average length of PCR oligonucleotides acceptable as primers (averagely between 15 and 50 nucleotides), the length of each portion will vary, as already said above, as indicated in figure 1, for what concerns the 5' region, depending on the sequencing primer to be used. The length of the PCR and sequence primers contributes to the cost of a PCR: the cost for the synthesis of an oligonucleotide that is ≤35 bases in length is usually cheaper than the cost for the synthesis of an oligonucleotide that is ≥35 bases

17

in length. In fact, for all oligonucleotides > 35 bp it is recommended to use an additional method of purification, which increase the cost of the primer. Therefore, if one has to design a test using a cocktail of primers, the length of the primers should be considered for cheaper PCR reaction. In any case, the length of the universal primer region will be from about 5 to about 20 nucleotides identical to the universal primer and will comprise the 3' of the universal primer used.

For PCR amplification it is preferred when the region complementary to the target sequence or a flanking sequence thereof is of at least 15.

On the other hand, the oligonucleotides of the invention can be used also as PCR sequencing primers, as shown in Tables 4, 6, 8 and 10. In this case the length of the universal primer region will be closer to about 20 nucleotides, the read start sequence length remains unchanged and the length of the region complementary to the target sequence to be amplified or to a flanking region thereof will be more close to the about 5 nucleotides values as exemplified in the tables and drawn in Figure 1.a and 1.b.

The read start sequence will preferably be of a length comprised between 4 and 12 nucleotides, whereas the region complementary to the target sequence or to a flanking region thereof will be of about 15 to 30 nucleotides.

As oligonucleotides for PCR are usually of a size between 15 and 50 nucleotides, the relative sizes of each portion (universal primer, read start, complementary to target) of the oligonucleotides herein described will preferably give, when added up, a final product (the oligonucleotide of the invention) not shorter than about 15 nucleotides and not longer than about 50 nucleotides, by way of example, the oligonucleotide could comprise a region of 13 nucleotides of universal primer, a read start of 5 nucleotides and a region complementary to a part of the target sequence of 15 nucleotides.

All single number in the ranges cited for each portion of the oligonucleotide is intended as disclosed in the description (i.e. read start can be of 4, 5, 6, 7, 8, 9, 10, 11, or 12 nucleotides) and the same applies to each other component of the oligonucleotide of the invention), all possible combination providing a final length of the oligonucleotide between 15 and 50 nucleotides are intended as part of the description, the skilled person will not need all single detailed number specified as the acceptable length of each part of the oligonucleotide is indicated as well as the fact that a length of the complete oligonucleotide is preferred in any possible length in the range of 15-50 nucleotides which is the average range of PCR oligonucleotides. Functional exceptions to this preferred range are acceptable provided that the primer can be purified and proves to provide the functional

18

characteristics of the oligonucleotides herein described.

In an embodiment of the invention, the oligonucleotides for PCR amplification can be selected in the group consisting of SEQ ID NO: 1-9, 12-15, 17-24, 26-30. These oligonucleotides, enable, in particular, the detection of several species or specie variants of the *Scombridae* family in foods supposed to comprise tuna.

The oligonucleotides of table 3, from SEQ ID NO:1 to SEQ ID NO: 9 amplify a 162 bp fragment (herein called "fragment B"), from bp 382 to 521 of the mitochondrial *cytochrome b* gene of the *Scombridae* family. The PCR conditions are described in example 6 table 12  The reference gene that used is DQ080287 from *Thunnus albacares*, said fragment is not indicated in the art as a region to be amplified for detecting the presence of a given *Scombridae* species or intraspecies variant in a sample. The sequences listed above amplify a fragment "B" (as defined above) of the gene. The amplified fragments were sequenced with the oligonucleotides listed in table 4, SEQ ID NO:10 and SEQ ID NO:11. The variable region in fragment B used to discriminate the species in the *Scombridae* family is from nucleotide 404 to 498 nucleotide. The comparison of the amplified sequences used to set up the assays (see table 11 example 5 for the number of sequences used in each species) showed that 36 positions were variable ones. In tables 15 and 16 the diagnostic positions are indicated. In order to identify the species in a sample, the nucleotide sequence of the variable region in fragment B of the sample in exam is compared to the same region in the different 17 species of *Scombridae* family (see table 15 and 16). The species and the variant intraspecies are identified when an exact match is found. In a small percentage of cases (two cases), there can be, when only fragment B is amplified, some possible misidentifications that do not allow an exact identification of the species or of the intraspecies variants. By way of example, in the first case a variant of *Thunnus maccoyii* (TMAC2) has the same haplotype as some variant of *Thunnus thynnus*  (e.g. TTHY1, TTHY2, TTHY3, TTHY4 and TTHY5) There are only few information about the sequence of *cytochrome b* in *Thunnus maccoy*. In this case, in order to distinguish the two species it is necessary to amplify fragment "A plus" or "A+" of the above mentioned gene (fragments A, A plus, B and A+B are defined in the specification). If in position 375 there is a G the species is *Thunnus thynnus*, however, if in position 375 there is an A the species is *Thunnus maccoyii*. There is also a rare haplotype of *Thunnus albacares* (TALB4), present in only 2% of the sequences of *Thunnus albacares* deposited in public database up to now that can be confused with *Thunnus obesus* (TOBE1) that was never found by the present inventors. However, to exclude any chance of doubt it is

19

necessary to analyse the SNP in position 330 in fragment A: if there is a G the species is *Thunnus albacares*, if instead there is an A the species is *Thunnus obesus*.

Hence, when only fragment B is amplified with the primers of the invention indicated in table 3, the results obtained by the method of the invention allow the identification of polymorphisms identifying from 15 to 17 species depending on the variants that are present for the species above. In most cases, given the frequencies of the superposed variants indicated above and identifiable in table 15 or 16, the primers of Table 3 will be sufficient for the identification of 17 species, in rare cases a further indication is given in fragment A+B (SNPs table 15) or in the fragment A+ or A plus (Table 15 or 17) as identified below.

More precisely, fragment A+B can be amplified, or a fragment called A+ is amplified, the reverse primer for the amplification of A+ is shifted of a few bases downstream with respect to the reverse primer for the amplification of fragment A, that allows the detection of a rare SNP that is located in a region complementary to the reverse primer for A and not to the reverse primer for A+ that is downstream said SNP and allows amplification and discrimination of it.

The oligonucleotides of table 5, i.e. SEQ ID NO:2, 7, 8, 9, 12-15 amplify a 291 bp fragment (herein called "fragment A+B"), from bp 253 to 521, of the mitochondrial *cytochrome b* gene (as above) of the *Scombridae* family, Said fragment comprises fragment B that is not indicated in the art as a region to be amplified for detecting the presence of a given *Scombridae* specie or intraspecies variant in a sample. The sequences listed above amplify a fragment A+B, this fragment was obtained from nearly 59% of the samples amplified from canned tuna, but an amplicon of 291 bp was obtained from all fresh fish or frozen fresh fish samples tested. This fragment is longer than fragment B and can provide further information. The amplified fragments were sequenced with the oligonucleotides of table 6, SEQ ID NO:16 and SEQ ID NO:11. The variable region in fragment A+B used to discriminate the species in the *Scombridae* family is from nucleotide 273 to nucleotide 498. The analysis of the alignment of the sequences used to set up the assays showed that 84 positions were variable ones. In table 15 the diagnostic positions are indicated. In order to identify the species in a sample, the nucleotide sequence of the variable region in fragment A+B is compared to the same region in the different 17 species of the *Scombridae* family (see table 15). The species and the variant intraspecies (or haplotype) is identified when an exact match is found.

In this case all species are immediately identified without an additional test.

The oligonucleotides of table 7, i.e. SEQ ID NO:12-15 and 17-24 amplify a 156 bp fragment (herein called "fragment A"), from bp 253 to 386, of the

20

mitochondrial cytochrome b gene of the *Scombridae* family (reference sequence indicated above), said fragment being already indicated in the art as a region to be amplified for detecting the presence of a given *Scombridae* specie or intraspecies variant in a sample. The sequences listed above amplify a fragment A (as defined

5      above) of the gene. The amplified fragments were sequenced with the oligonucleotides of table 8, SEQ ID NO: 25 and SEQ ID NO:16. The variable region in fragment A used to discriminate the species in the *Scombridae* family is from nucleotide 273 to nucleotide 363, the SNPs of interest are indicated in Table 15 and 17. The variable region in fragment A allows, as described above, to discriminate the

10     rare haplotype TALB4 from TOBE1 The region amplified in fragment A is the same region amplified by Bottero *et al.* 2007, but the primers are modified in order to identify all species and variant intraspecies.

The oligonucleotides in Table 9 i.e. SEQ ID NO 12-15 and 26-30, amplify a 151 bp fragment (herein called "fragment A+"), from bp 253 to 403, of the

15     mitochondrial cytochrome b gene of the *Scombridae* family (reference sequence indicated above)., Said fragment being larger than the fragment A indicated above. The sequences listed above amplify a fragment A+ (as defined above) of the gene , the SNPs of interest are indicated in Table 15 and 17. The amplified fragments were sequenced with the oligonucleotides of table 10, SEQ ID NO: 16 and SEQ ID

20     NO:31. The variable region in fragment A+ is used to discriminate the species in the *Scombridae* family is from nucleotide 273 to nucleotide 381. The variable region in fragment A+ allows, as described above, to discriminate some variants of *Thunnus thynnus* from TMAC2 and the rare haplotype TALB4 from TOBE1.

When processed samples are used (e.g. canned Tuna) and the most complete

25     information is required, fragments A+ and B can be amplified separately (long fragments are less likely to be amplifiable in processed food samples as the DNA is more degraded) and the information obtained will be the one obtainable form fragment A+B but the use of the oligonucleotides of table 3 and of the oligonucleotides in table 9 instead of the oligonucleotides of table 5, allows the

30     amplification of smaller informative fragments that are more likely to be amplifiable in processed food samples.

The present description also discloses a method for the detection of polymorphisms or polymorphisms patterns of nucleotide sequences amplified by PCR comprising the steps of:

35          a. sequencing nucleotide sequences amplified with one or more pair, forward
            and reverse, of oligonucleotides of the invention, with an automated

21

sequencer using a primer comprising, at the 3', said universal primer portion or collecting nucleotide sequences thus obtained;

b. detecting in said sequences said read start sequence;

c. indexing nucleotides of said sequences from said detected start sequence;

d. detecting in said indexed sequences polymorphisms or polymorphisms patterns.

In step a. sequences amplified with primer pairs of the invention are submitted to an automated sequencing i.e. to the above described PCR with a sequencer using a primer comprising, at the 3', the afore mentioned universal primer portion and to the read out by the automated sequencer of the labelled fragment obtained by the sequencing PCR, or with any of the primers as exemplified in figure 1.

The sequencing can be part of the detection method but the method can be carried out also starting from sequences obtained by any means for sequencing or by an automated sequencer having a software independent from the software described below. By way of example, when a software implementing the method is used, the software can be capable of carrying out the automated sequencing, storing the sequences and so on, or, the sequences of the amplified fragments as described in step a. can be obtained separately, and processed by the software for carrying out the method above.

In step b. the read start sequence present into the primers of the invention is detected either manually or by a computer using a software capable of detecting said sequence in the amplicons.

In step c. the nucleotides of the amplified sequence are indexed starting from a given position downstream the read start portion (or read start sequence). If the sequence complementary to the sequence of interest is adherent, i.e. immediately downstream without spacing nucleotides, to the read start portion, than the indexing can begin at the nucleotide immediately downstream the read start sequence. If another known sequence is inserted, in the primer, within the read start portion and the sequence complementary to the target sequence to be amplified, than the indexing could also be shifted immediately downstream said inserted sequence, unless, the user wants, for some reason to start the indexing from the inserted sequence. The presence of the read start, also allows the user to shift the indexing start further downstream (i.e. x nucleotides after the read start sequence) and all these possibility can be easily performed in an automated way by the above mentioned software wherein said software, capable of detecting and localising the read start sequence is also capable of indexing the following nucleotides at a predetermined

22

indexing start that can be from nucleotide 1 after the read start sequence to nucleotide x after the read start sequence, x, being obviously an integer.

The method of the invention, enables the user to easily individuate a nucleotide in the indexed amplified sequence, independently from the number of primer pairs used and is particularly handy in case of amplifications with a cocktail of primers. Although the easy identification of a "read start" is useful in any amplified sequence, the possibility of rapidly and even automatically scanning a pool of amplicons and of identifying in each of said amplicons a common read start sequence is, as evident to any skilled person, extremely useful. The presence of a read start sequence common to all the amplicons allows an automation of the indexing of the sequence that usually needs the intervention of a skilled person. A software can be easily designed to recognise one or more read start sequences and to detect one of more of said read start sequences in a group of amplicons according with the settings selected by the user.

In one embodiment, each group of amplicons to be sequenced and indexed in the same reaction will share the same read start sequence.

In the indexed sequence thus obtained it will be very easy to detect the presence or absence of a given polymorphism in a given position or even to easily localise and identify the existence of new polymorphisms in a target sequence by mere comparison of each nucleotide having a given index number where new polymorphisms are searched, or by immediate checking out of which nucleotide is present at a given index, in order to verify whether a given polymorphism is present at a given site and what polymorphism is present at said given site.

As explained in more detail in the following description, the oligonucleotides of the invention also enable an easy and fast identification (herein intended also as characterisation) of newly detected polymorphisms or polymorphism patterns.

The use of a cocktail of primers comprising a read start sequence, in order to detect the presence or absence and to identify known polymorphisms in a sample, allows a direct indexing of the amplified sequences based on the read start sequence, and allows hence, to compare directly given indexes without the need of performing a sequence alignment. Even the use of a single oligonucleotide pair for amplification, comprising a read start sequence, in order to verify the presence of a polymorphism or a mutation (e.g. patient vs. healthy individuals) on different samples of the same species, allows again, a direct indexing of each amplified sequence and direct comparison index by index. When the comparison is performed by a software, as explained in detail in the method below, it will be extremely easy to detect a. whether there are different nucleotides at a given index (i.e. whether there is a polymorphism

23

or a mutation), b. what is the difference at said index (i.e. which nucleotides are found at that index), c. even what is the frequency of each nucleotide at said index (the frequency being easily calculated by the software on the basis of the number of times a given polymorphism is detected in the pool of amplicons).

Hence, the oligonucleotides of this description can be also used in a method for the detection and the identification of new polymorphisms or polymorphisms patterns of nucleotide sequences amplified by PCR comprising the steps of:

a. sequencing nucleotide sequences from given samples amplified with one pair, forward and reverse, of oligonucleotides according to the description with an automated sequencer using a universal primer comprising said 5' portion;

b. detecting in said sequences said read start portion;

c. indexing nucleotides of said amplified sequences, from said read start portion;

d. detecting in said indexed sequences the presence polymorphisms or polymorphisms patterns in comparison to a reference sequence;

e. sequencing nucleotide sequences from the same samples of step a. amplified each with a different pair, forward and reverse, of oligonucleotides according to the description, each pair of oligonucleotides having a different read start, with an automated sequencer using a universal primer comprising said 5' portion;

f. detecting in said sequences said read start portions;

g. indexing nucleotides of said amplified sequences, from said read start portions, each sample being recognisable by a specific read start portion;

h. identifying in said indexed sequences the polymorphisms or polymorphisms patterns in comparison to a reference sequence.

When the detection/identification method is carried out as described above, sequences will be obtained on the products of an amplification by PCR with the oligonucleotides of this description will be carried out on all the samples to be tested (this can be useful, by way of example, when the aim of the research is to identify a disease gene, or to identify within a gene known to be related with a disease, possible polymorphisms or polymorphic patterns related to the disease). The firs PCR reaction will have been carried out with the same oligonucleotide pair of the description on all samples (samples could be DNAs from healthy individuals and from patients having a given disease or from patients having a given disease only, or from groups of individuals with a different phenotype etc.). The nucleotides of the sequences will be indexed after detection of the read start sequence as explained

24

above, hence, also by means of suitable software as explainer above. The result of the indexing and of the comparison of each indexed nucleotide against a reference sequence (e.g. a wild type sequence present in healthy individuals) having the same indexes will show the presence of different nucleotides at a particular index, if

5    present. If differences are present, a further step of sequencing nucleotide sequences from the same samples of step a. amplified each with a different pair, forward and reverse, of oligonucleotides according to the description, each pair of oligonucleotides having a different read start, will allow the same indexing as above, with the difference that, due to the different read start for each sample, the results can

10   be punctually related to a given samples. This allows a characterisation of each sample and hence will indicate whether a given new polymorphism detected or a given polymorphic pattern is significantly related to a certain disease or to a certain phenotype etc.

     Alternatively, the method for the detection/identification of new

15   polymorphisms or polymorphisms patterns of nucleotide sequences amplified by PCR comprising the steps of:

          a. sequencing nucleotide sequences from different samples each sample being
          amplified with a different pair, forward and reverse, of oligonucleotides
          according to the description and each pair of oligonucleotides having a

20        different read start, with an automated sequencer using a universal primer
          comprising said 5' portion;

          b. detecting in said sequences said read start portions;

          c. indexing nucleotides of said amplified sequences, from said read start
          portions, each sample being recognisable by a specific read start portion;

25        d. identifying in said indexed sequences the polymorphisms or
          polymorphisms patterns in comparison to a reference sequence;

     Hence, the preliminary steps for the simple detection of the presence or absence of polymorphisms can be avoided.

     The methods of the invention hence allows, at a first stage the detection of a

30   polymorphism or of a polymorphic pattern by a first comparison of each nucleotides having a given index on all amplicons obtained with a single PCR reaction (or more PCR reactions with the same primers and on the same kind of samples), but it also allows the interpretation of the data obtained by the index comparison for the identification of known or new polymorphisms as indicated above and, in case of

35   known polymorphisms, as in the case of species detection, even the detection of the species is directly obtained by the method by the relation between each given polymorphism and the species or variants to detect (the same could apply to the

25

detection of a disease or of a maternity or paternity probability as evident to the skilled person).

The software cited above will easily compare each nucleotide having a given index of the indexed sequences against a given target sequence or even with each other, immediately providing a tool (i.e. the index) to identify the position or the relative position of a polymorphism, or to identify the presence or absence of a given polymorphism at a given site.

Hence, the description also discloses a method for the analysis of polymorphisms or polymorphisms patterns of nucleotide sequences amplified by PCR comprising the steps of:

a. sequencing nucleotide sequences amplified with one or more pair, forward and reverse, of oligonucleotides according to the description with an automated sequencer using a universal primer comprising, at the 3', said universal primer portion;

b. detecting in said sequences said read start sequence;

c. indexing nucleotides of said sequences from said detected start sequence;

d. detecting in said indexed sequences polymorphisms or polymorphisms patterns;

e. analysing, as above indicated, the data obtained at step d.

In a specific embodiment, the method of the invention allows the detection of up to 17 *Scombridae* species in complex food matrices.

For tuna, several molecular works have been published which allow to distinguish some species in the *Scombridae* family (table 1).

Table 1: shows a summary of the molecular works published in the scientific journals aimed to identify different species in the *Scombridae* family. The first column lists the molecular methods used by the authors; the second column shows the length of the amplified fragments. In the third column is the name of the analysed gene. The fourth column indicates the number of species identified. The fifth column contains some "critical note" of the method, and the last column lists the author's name and the year of the publication.

Table 1. Molecular works for species discrimination in the *Scombridae* family

| METHOD | LENGTH OF THE AMPLICONS (bp) | GENE/S | SPECIE IDENTIFIED | NOTES | AUTHOR/ YEAR |
|---|---|---|---|---|---|
| PCR with specific primers | 262, 246, 113 | MT-CYB | TALB, TOBE, KPEL | Only 3 species, Long amplicons | Michelini 07 |
| PCR with specific primers | 146, 128, 95 | MTCYB MTATP6 rRNA 12S | ATHA, AROC | Acrylamide gel | Infante 04 |
| PCR with specific primers | 225, 204 | MT-CYB | TTHY, SSAR | Agarosium gel | Lockely 99 |
| PCR-RFLP | 278 | MT-CYB | TTHY, TALB, TOBE, TALA, KPEL | nested, laborious 5 restriction enzymes | Pardo 04 |
| PCR-RFLP | 464 | MT-CYB | 36 different species | Long amplicons, 7 restriction enzymes | Hold 01 |
| PCR-RFLP | 126, (350 of Barlet 91) | MT-CYB | TTHY, TALB, TOBE, TALA, KPEL, SSAR | Acrylamide gel and 3 restriction enzymes | Quintero 98 |
| SEQUENCING | 528 (and 213, 242, 221) | MT-CYB | TALB, TOBE, KPEL | Only 3 species | Terol 02 |
| SEQUENCING | 464, 59 | MT-CYB | TOBE, TMAC, TALA, EAFF, EALL, ATHA, KPEL, SSAR, SCHI | It doesn't differentiate TTHY TALB, Cloning is inevitable | Unseld 95 |
| PCR-SSCP | | | | | Rehbein 98 only abstract |
| Real-Time PCR | (<100pb) | MT-CYB | TALB, TALA | Taqman, | Lopez 05 |
| Real-Time PCR | 208 | ATPase6 | TTHY, TALB, TOBE, TALA | No primers | Dalmasso 07 |
| PER (primers extention assay, SNAP-SHOT) | 132 | MT-CYB | TTHY, TALB, TOBE, TALA, KPEL | It doesn't differentiate all 17 specie and intraspecie variants | Bottero 07 |

These methods are usually based on the amplification of mitochondrial genes (especially the *cytochrome b* gene , MT-CYTB) and rely on Polymerase Chain Reaction (PCR) with specific primers (Michelini *et al.* 2007), Polymerase Chain Reaction followed by fragments restriction analysis (PCR-RFLP, Pardo *et al.*, Hold *et al.* 2001, Quinterio *et al.* 1998), sequencing of mitochondrial DNA amplified by PCR (Terol *et al.* 2002), PCR and analysis of conformation changes caused by single strand DNA polymorphisms (PCR-SSCP, Rehbein *et al.* 98), real-time PCR (Lopez *et al.* 2005, Dalmasso *et al.* 2007, Quintero *et al.* 98) or PCR and multiplex primer extension assay (PER o SNAP SHOT Bottero *et al.* 2007).

Prior to the present invention, there was no molecular method enabling the identification of 17 species of the *Scombridae* family, allowing for intraspecies variability, applicable to complex food matrices and allowing its automation.

So far, not every molecular method developed for the identification of fish species works on highly processed products. In general, one of the most efficient methods of conservation of animal meat, particularly fish meat, is based on the

27

prolonged cooking and/or processing in autoclave. These processes increase the DNA degradation and the difficulties of extraction and analysis.

Molecular methods that amplify and analyse long DNA fragments (> 200-300 pairs of bases, bp) do not work with complex food matrices which have undergone such thermal/pressure processes.

Some methods are extremely laborious (e.g. PCR-RFLP, PCR-SSCP) and their automation is impossible. Other ones (e.g. Real-time PCR) can distinguish just a few species. The method recently developed by Dalmasso *et al.* 2007, can discriminate 4 species belonging to the *Scombridae* family. The primers and the probes sequences have not been published. However, this method does not allow to discriminate the *K. pelamis* species which may be used as fake Tuna by fraud. The method developed by Bottero *et al.* 2007, can distinguish only 5 species among which is *K. pelamis*, works on complex food matrices, but does not entirely allow for intraspecies variability, therefore it is not able to identify and discriminate some variants.

As to the methods based on sequencing, that of Terol *et al.* 2002 can distinguish only three species of the *Scombridae* family and moreover, the size of the amplified fragment (528 bp) makes it inapplicable to complex food matrices (it works only on fresh and frozen meat). So, in order to amplify canned tuna, three PCR reactions, which amplify three overlapping fragments that constitute the original 528 bp sequence, are described. The method of Hunseld *et al.* 1995, instead, takes into account the state of degradation of DNA in complex matrices, can discriminate 9 species though does not allow for intraspecies variability, thus some variants of the species will not be detected; furthermore, the amplified fragment does not discriminate two important commercial species such as *Thunnus thynnus* and *Thunnus albacares*. However, the amplified fragment is too short (59 bp, primers excluded) and cloning is a necessary step to obtain a sequence of the amplicon.

The methods based on PCR-RFLP are laborious and require the use of various restriction enzymes and a result analysis system, which are somewhat complicated. Usually, a region of the *cytochrome b* gene is amplified in different species (max. 6 species of the *Scombridae* family, according to the published works), then the amplified fragment is digested by various restriction enzymes and the band pattern obtained, characteristic of a certain species, is displayed on gel. In some cases, the amplified fragments digested by the enzymes are displayed on acrylamide gel as this type of gel provides a higher resolution than agarose gel. However, the preparation of the acrylamide gel takes longer then the preparation of the agarose gel and is more dangerous for the operator. As a matter of fact, acrylamide is a strong

28

neurotoxin which is easily absorbed by the skin when it is not yet polymerized.

The method herein described, is designed in such a way to be applied to transformed products in which DNA is very degraded.

The amplified fragments with the method can be directly sequenced with no loss of information for the bases at the edge of the amplicon of the amplified gene fragment: the primers were designed to create a region between the primers, used for the sequence reaction, and the region of the *cytochrome b* of interest, for the discrimination of the species.

The primers contain the "read start" region which is a determined sequence of bases which, being present in the primers, will be present in all the amplicons. Such region can be used by a dedicated software, as already indicated, which allows the automated indexing of the amplicons sequences thus making the sequence results analysis easier and faster.

This method, so far, allows to discriminate 17 species of *Scombridae* family but can be implemented, by designing further suitable oligonucleotides according to the invention, to detect all the 24 commercialized species.

Table 2: lists the species in the *Scombridae* family identified with the method described in this invention. In the first column is the scientific name of each species, the second column lists the common name from the website fish base (http://www.fishbase.org), in the last column is the Italian specie name from ichthyic commercial list.

Table 2: this table lists the fish species in the *Scombridae* family identified by the method herein disclosed.

| Scientific Name | Common name (fishbase UK FAO) | Italian specie name (from ichthyic commercial list) |
|---|---|---|
| Thunnus thynnus | Atlantic bluefin tuna | (Tonno rosso) |
| Thunnus albacares | Yellowfin tuna | (Tonno a pinne gialle) |
| Thunnus atlanticus | Blackfin tuna | (Tonno atlantico) |
| Thunnus maccoyii | Southern bluefin tuna | (Tonno australe) |
| Thunnus orientalis | Pacific bluefin tuna | (Tonno orientale) |
| Thunnus obesus | Bigeye tuna | (Tonno obeso) |
| Thunnus alalunga | Albacore | (Alalunga) |
| Thunnus tonggol | Longtail tuna | (Tonno indopacifico) |
| Katsuwonus pelamis | Skipjack tuna | (Tonnetto striato) |
| Euthynnus alletteratus | Little tunny | (Tonnetto o Alletterato) |
| Auxis thazard thazard | Frigate tuna | (Biso tropicale) |
| Auxis rochei rochei | Bullet tuna | (Biso o Tombarello) |
| Sarda sarda | Atlantic bonito | (Palamita) |
| Scomber scombrus | Atlantic mackerel | (Sgombro) |
| Scomber japonicus | Chub mackerel | (Lanzardo o sgombro occhione) |
| Scomber colias | Atlantic chub mackerel | (Lanzardo o sgombro occhione) |
| Scomber australasicus | Blue mackerel | (Sgombro australe) |

29

The *Scombridae* family comprises 54 salt-water fish species, belonging to the *Perciformes* order, which are well known and widely caught for human feeding. Among these, 24 (tuna, little tunny, mackerel, bullet tuna, Atlantic bonito, etc.) are traded in Europe. The test can be extended to identify all the 24 commercialized species.

A "primer's cocktail" is used in the first amplification reaction (PCR-1) performed on a small region of the *cytochrome b* gene (MT-CYB), which is in all species. The PCR-1 primers were designed according to the description and where tuned for two purposes: 1) Introduce a sequence "read start" signal which would allow direct indexing (even automated indexing) of the nucleotides of the amplified sequences thus making the results analysis easier and faster, 2) Increase the length of the amplified fragment so that the amplicon could be easily sequenced.

Then, a single pair of universal primers, comprising the universal primer region at the 5' of the oligonucleotides for amplification (herein also defined as PCR-1) is used in the sequencing amplification reaction (herein also defined PCR-2).

The method of the invention can be carried out using the primer groups listed in the following tables. As indicated in the tables, the oligonucleotides according to the description can be used to amplify fragment B of the mitochondrial *cytochrome b* gene in the *Scombridae* family, fragment A+B, fragment A or A plus.

When the species identification in the *Scombridae* family is from canned tuna or from other samples where DNA is degraded the best choice to identify the species is to amplify fragment B. In rare cases of doubt (see above ex. TMAC2 and some variants of *Thunnus tynnus*) it is necessary to amplify also fragment A plus. The PCR primers and the PCR reaction conditions described in this invention are able to amplify all species and variant intraspecies both for fragment B and for fragment A plus.

Fragment B allows a better discrimination of species and variant intraspecies than fragment A.

In case of fresh fish or frozen fresh fish the best choice is to amplify fragment A+B which obviously contains all information of fragments B and A+B.

Depending on the oligonucleotides used a different fragment of the mitochondrial *cytochrome b* gene is amplified.

30

Table 3, that follows, discloses the oligonucleotides that amplify fragment B (i.e. bp382-521) of the *Scombridae* mitochondrial *cytochrome b* gene.

| Primer name | sequence 5' to 3' | Seq Id No. |
|---|---|---|
| USS1-UniT-Fw | *GCCAGT*GATTACTTCGTYGGCTACGTYCTTCCCT | 1 |
| USS1-UniT-Rw | *TGACCC*GATCGTRGCATTGTCTACTGARAAGCC | 2 |
| USS1-UniK-Fw | *GCCAGT*GATTACTTCGTCGGYTACGTACTTCCCT | 3 |
| USS1-UniR-Fw | *GCCAGT*GATTACTTCGTCGGCTACGTACTYCCAT | 4 |
| USS1-UniE-Fw | *GCCAGT*GATTACTTCGTCGGTTACGTACTCCCTT | 5 |
| USS1-UniSA-Fw | *GCCAGT*GATTACTTCGTCGGYTACGTYCTTCCCT | 6 |
| USS1-UniK-Rw | *TGACCC*GATCGTGGCGTTGTCTACTGAAAAGCC | 7 |
| USS1-UniR-Rw | *TGACCC*GATCGTGGCGTTGTCTACTGAGAAGCC | 8 |
| USS1-UniSA-Rw | *TGACCC*GATCGTRGCATTGTCTACAGAGAAGCC | 9 |

In bold and italic nucleotides of a Universal primer for automated sequencer, in normal text the read start sequence and in bold the sequence complementary to the sequence of interest.

Table 4, that follows, discloses primers suitable for sequencing the amplicons obtained by the amplification of complex food matrix comprising material deriving from *Scombridae* family, with the oligonucleotides of table 3.Primers consisting of the sole universal primer can be used, in this case the oligonucleotides of Table 2 can be elongated at the 5' with further nucleotides of the universal primer.

| Primer name | sequence 5' to 3' | Seq Id No. |
|---|---|---|
| USS2-UniT-Fw | *GTTGTAAAACGACGGCCAGT*GATTACTTCGT | 10 |
| USS2-UniT-Rw | *CACAGGAAACAGCTATGACC*CGATCGTRGC | 11 |

In bold and italic nucleotides of the Universal primer for automated sequencer, in normal text the read start sequence and in bold the sequence complementary to the sequence of interest.

Table 5, that follows, discloses the oligonucleotides that amplify fragment A+B (i.e. bp253-521) of the *Scombridae* mitochondrial *cytochrome b* gene.

| Primer name | sequence | Seq Id No. |
|---|---|---|
| USBotA-CytbsF | *GCCAGT*GATTACAACGGGGCCTCTTTCTTCTT | 12 |
| USBotB-CytbsF | *GCCAGT*GATTACAACGGCGCYTCCTTCTTCTT | 13 |
| USBotCbis-CytbsF | *GCCAGT*GATTACAACGGTGCTTCYTTCTTCTT | 14 |
| USBotDbis-CytbsF | *GCCAGT*GATTACAACGGGGCCTCCTTCTTCTT | 15 |
| USS1-UniT-Rw | *TGACCC*GATCGTRGCATTGTCTACTGARAAGCC | 2 |
| USS1-UniK-Rw | *TGACCC*GATCGTGGCGTTGTCTACTGAAAAGCC | 7 |
| USS1-UniR-Rw | *TGACCC*GATCGTGGCGTTGTCTACTGAGAAGCC | 8 |
| USS1-UniSA-Rw | *TGACCC*GATCGTRGCATTGTCTACAGAGAAGCC | 9 |

In bold and italic nucleotides of a Universal primer for automated sequencer, in normal text the read start sequence and in bold the sequence complementary to the sequence of interest.

Table 6, that follows, discloses primers suitable for sequencing the amplicons

31

obtained by the amplification of complex food matrix comprising material deriving from *Scombridae* family with the oligonucleotides of table 5, primers consisting of the sole universal primer can be used, in this case the oligonucleotides of Table 5 can be elongated at the 5' with further nucleotides of the universal primer.

| Primer name | sequence | Seq Id No. |
|---|---|---|
| USbot2-Fw | *GTTGTAAAACGACGGCCAGT*GATTAC**AACGG** | 16 |
| USS2-UniT-Rw | *CACAGGAAACAGCTATGACC*CGATC**GTRGC** | 11 |

In bold and italic nucleotides of a Universal primer for automated sequencer, in normal text the read start sequence and in bold the sequence complementary to the sequence of interest.

Table 7, that follows, discloses the oligonucleotides that amplify fragment A (i.e. bp253-386) of the *Scombridae* mitochondrial *cytochrome b* gene.

| Primer name | sequence | Seq Id No. |
|---|---|---|
| USBotA-CytbsF | *GCCAGT*GATTAC**AACGGGGCCTCTTTCTTCTT** | 12 |
| USBotB-CytbsF | *GCCAGT*GATTAC**AACGGCGCYTCCTTCTTCTT** | 13 |
| USBotCbis-CytbsF | *GCCAGT*GATTAC**AACGGTGCTTCYTTCTTCTT** | 14 |
| USBotDbis-CytbsF | *GCCAGT*GATTAC**AACGGGGCCTCCTTCTTCTT** | 15 |
| UU1BotAbis-CytbsR | *ATGACC*GACCA**ACGAAGGCGGTYATCATAACTAG** | 17 |
| UU1BotBbis-CytbsR | *ATGACC*GACCA**ACGAAAGCAGTCATCATTACTAG** | 18 |
| UU1BotE-CytbsR | *ATGACC*GACCA**ACGAATGCAGTYATCATRACCAG** | 19 |
| UU1BotD-CytbsR | *ATGACC*GACCA**ACGAAAGCGGTCATCATTACTAA** | 20 |
| UU1BotF-CytbsR | *ATGACC*GACCA**ACGAATGCAGTCATCATGACTAG** | 21 |
| UU1BotG-CytbsR | *ATGACC*GACCA**ACGAAGGCGGTTATCATTACTAG** | 22 |
| UU1BotH-CytbsR | *ATGACC*GACCA**ACGAAGGCAGTCATCATTACGAG** | 23 |
| USBotI-CytbsR | *ATGACC*GACCA**ACGAATGCGGTTATTATTACGAG** | 24 |

In bold and italic nucleotides of a Universal primer for automated sequencer, in normal text the read start sequence and in bold the sequence complementary to the sequence of interest.

Table 8, that follows, discloses primers suitable for sequencing the amplicons obtained by the amplification of complex food matrix comprising material deriving from *Scombridae* family with the oligonucleotides of table 7, primers consisting of the sole universal primer can be used, in this case the oligonucleotides of Table 7 can be elongated at the 5' with further nucleotides of the universal primer.

| Primer name | sequence | Seq Id No. |
|---|---|---|
| USbot2-rw | *CACAGGAAACAGCTATGACC*GACCA**ACGAA** | 25 |
| USbot2-Fw | *GTTGTAAAACGACGGCCAGT*GATTAC**AACGG** | 16 |

In bold and italic nucleotides of a Universal primer for automated sequencer, in normal text the read start sequence and in bold the sequence complementary to the sequence of interest.

32

Table 9, that follows, discloses the oligonucleotides that amplify fragment A plus (i.e. bp253-403) of the *Scombridae* mitochondrial cytochrome B gene.

| Primer name | Sequence | Seq Id No. |
|---|---|---|
| **USBotA-CytbsF** | *GCCAGT*GATTACAACGGGGCCTCTTTCTTCTT | 12 |
| **USBotB-CytbsF** | *GCCAGT*GATTACAACGGCGCYTCCTTCTTCTT | 13 |
| **USBotCbis-CytbsF** | *GCCAGT*GATTACAACGGTGCTTCYTTCTTCTT | 14 |
| **USBotDbis-CytbsF** | *GCCAGT*GATTACAACGGGGCCTCCTTCTTCTT | 15 |
| USAP-UniT-Rw | *ATGACC*GACCAAGGGAAGRACGTAGCCRACGAA | 26 |
| USAP-UniK-Rw | *ATGACC*GACCAAGGGAAGTACGTARCCGACGAA | 27 |
| USAP-UniR-Rw | *ATGACC*GACCAATGGRAGTACGTAGCCGACGAA | 28 |
| USAP-UniE-Rw | *ATGACC*GACCAAAGGGAGTACGTAACCGACGAA | 29 |
| USAP-UniSA-Rw | *ATGACC*GACCAAGGGAAGRACGTARCCGACGAA | 30 |

In bold and italic nucleotides of a Universal primer for automated sequencer, in normal text the read start sequence and in bold the sequence complementary to the sequence of interest.

Table 10, that follows, discloses primers suitable for sequencing the amplicons obtained by the amplification of complex food matrix comprising material deriving from *Scombridae* family with the oligonucleotides of table 9, primers consisting of the sole universal primer can be used, in this case the oligonucleotides of Table 9 can be elongated at the 5' with further nucleotides of the universal primer.

| Primer name | Sequence | Seq Id No. |
|---|---|---|
| **USbot2-Fw** | *GTTGTAAAACGACGGCCAGT*GATTACAACGG | 16 |
| **USAP2-rw** | *CACAGGAAACAGCTATGACC*GACCAANGG | 31 |

In bold and italic nucleotides of a Universal primer for automated sequencer, in normal text the read start sequence and in bold the sequence complementary to the sequence of interest.

The description also discloses software capable of carrying out the methods described above. The software will comprise procedures and steps for

      a. receiving as an input one or more nucleotide sequence, or performing an automated sequencing with a primer comprising, at the 3', said universal primer portion and, storing and/or processing the received or obtained sequences;

      b. detecting the read start portion in each of said sequences;

      c. indexing the nucleotides of each the amplified sequences as indicated above in the description of the method for detecting polymorphisms or polymorphic patterns, starting from the read start sequence detected;

      d. detecting in said indexed sequences polymorphisms or polymorphisms patterns, by comparing indexed nucleotides of different sequences, at specified indexes as explained above in the description of the method;

      e. classifying said detect polymorphisms or polymorphisms patterns as not known and/or known. The software comprises means for inputting and/or storing

33

known polymorphisms, so to obtain a polymorphism database to be used for a subsequent recognition thereof;

f. recognising the known polymorphism. within such database, providing a report with the results (i.e. in the case of the method used for the detection of up to 17 species of the *Scombridae* family) either the polymorphisms or polymorphic patterns identified, or even, if an input for the association of a given polymorphisms or polymorphic pattern with a given result (i.e. polymorphism X equals species, disease, or other, Y) has been provided to the software, more elaborated results can be provided by the software (e.g. direct identification of a given species in a sample, etc.);

g. if the polymorphism is not yet known, the software can further provide a set of features (specific nucleotides, frequency...) of the unknown detected polymorphism, and means for storing such set of features in the said database.

Basically, the read start sequence of the invention can overcome the need of aligning the amplified sequences. It is clear that the user might chose to use the said read start also as a tool for a better and faster alignment, as the read start sequence, although providing a new tool for sequence analysis, does not forbid the use of the indexed sequence in more standard ways. Hence the indexed sequences can be used both in the method of the invention and in a more classic method, without need of carrying out two different amplifications.

A computer readable storage support wherein the software above is stored can be any kind of computer-readable storage support suitable for storing a software, such as CD, DVD, TAPES, USBPen, EPROM, disks, hard disks, etc.. It is to be understood that the method of the invention can be provided as a service through a web-service via an online connection and/or the software can be downloadable from a network.

Also disclosed herein is a kit comprising one or more aliquot of at least two, forward and respective reverse oligonucleotides for PCR of the description, and optionally one or more aliquot of a sequencing primer comprising, at the 3', the said universal primer portion, the kit can comprise the oligonucleotides of one or more of the following groups: SEQ ID NO: 1-9; SEQ ID NO: 12-15, 2, 7, 8, 9; SEQ ID NO: 12-15 and 17-24 and/or SEQ ID NO: 12-15, 26-30.

In an embodiment of the description, when one or more of the groups of sequences listed above are used, sequencing primer selected, respectively, in the groups of: SEQ ID NO 10 and 11, SEQ ID NO 16 and 11, SEQ ID NO 25 and 16 and/or SEQ ID 16 and 31 may be comprised in the kit.

Moreover, the kit may comprise the computer readable storage support

34

wherein the software described above is stored, the said support being capable, when used, of running the said software on a computer.

Each of the single embodiments described herein might be disclaimed without altering the spirit and the scope of the invention.

Table 15, below, shows all the SNPs identifiable by the oligonucleotides of the invention in an indexed matrix, in the fragment "A+B", and, individually in fragment A, A plus and B.

The first column indicates the species in letters and the arbitrary name of the polymorphism identified (letters plus numbers), the first line indicates the indexing of each specific nucleotide of the amplicons, the SNPs indicated by the bold characters are identified in the A fragment (up to nucleotide 363 included), the SNPs indicated by the normal characters (up to 403 included), summed to the SNPs in bold are identified in fragment A plus, the SNPs indicated by the underlined characters (from nucleotide 402 to 498 included) are identified in fragment B.

All the SNPs listed in the table are identified in fragment "A+B".

Note that, nucleotide 402 is shown only in plain text to indicate that it is comprised in fragment A plus, the underlining has not been used for this position but 402 is comprised as well in fragment B.

35

| | 273 | 276 | 279 | 280 | 282 | 285 | 286 | 287 | 288 | 291 | 294 | 297 | 303 | 306 | 312 | 315 | 318 | 321 | 324 | 326 | 327 | 328 | 329 | 330 | 333 | 342 | 343 | 345 | 348 | 349 | 351 | 352 | 353 | 354 | 357 | 360 | 363 | 364 | 366 | 369 | 372 | 375 | 378 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| REF | T | C | C | A | C | C | T | T | C | C | C | U | A | T | C | C | T | C | A | A | C | A | A | G | A | C | A | C | A | G | A | G | T | A | C | A | C | C | A | T | G | G | C |
| TTHY1 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| TTHY2 | . | . | T | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . |
| TTHY3 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| TTHY4 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . |
| TTHY5 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| TTHY6 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| TTHY7 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| TALB1 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| TALB2 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| TALB3 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| TALB4 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| TALB5 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . |
| TALB6 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| TALB7 | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . |
| TATL1 | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| TATL2 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | T | G | . | . | . | . | . | . | . |
| TMAC1 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . |
| TMAC2 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . |
| TMAC3 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . |
| TMAC4 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . |
| TQR1 | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | A | . | . | . | . | . | . | . | . | . | G | . | . | T | . | . | . | . | . | . |
| TQR2 | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . |
| TQR3 | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | A | . | . | . | . | . | . | . | . | . | G | . | . | T | . | . | . | . | . | . |
| TALA1 | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | C | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . |
| TALA2 | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . |
| TALA3 | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . |
| TALA4 | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . |
| TALA5 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | A | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . |
| TALA6 | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | T | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . |
| TALA7 | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | C | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . |
| TALA8 | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | A | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . |
| TTON1 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| TTON2 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | T | . | T | . | . | . | . | . | . | . |
| TTON3 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . |
| TTON4 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . |
| TTON5 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . |
| TOBE1 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . |
| TOBE2 | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | A | A | T | . | . | . | . | . | . |

36

| | 273 | 276 | 279 | 280 | 282 | 285 | 286 | 287 | 288 | 291 | 294 | 297 | 303 | 306 | 312 | 315 | 318 | 321 | 324 | 326 | 327 | 328 | 329 | 330 | 333 | 342 | 343 | 345 | 348 | 349 | 351 | 352 | 353 | 354 | 357 | 360 | 363 | 364 | 366 | 369 | 372 | 375 | 378 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TOBE3 | | | | | | | | | | | | | | | | | | | | | | | | A | | | | | | | | | | | A | | T | | | | | | |
| TOBE4 | | | T | | | | | | | | | | | | | | | | | | | | | A | | | | | | | | | | | A | | | | | | | | |
| TOBE5 | | | T | | | | | | | | | | | | | | | | | | | | | A | | | | | | | | | | | A | | | | | | A | | |
| TOBE6 | | | T | | | | | | | | | | | | | | | | | | | | | A | | | | | | | | | | | A | | | | | | | | |
| TOBE7 | | | T | | | | | | | | | | | | | | | | | | | | | A | | | | | | | | | | | A | | | | | | | | |
| TOBE8 | C | | T | | | | | | | | | | | | | | | | | | | | | A | | | | | | | | | | | | | | | | | | | |
| KPEL1 | C | | | | T | | | C | | T | T | | T | C | | T | A | | C | | | | | A | | | | | T | | | | | T | A | A | | | | A | | | T |
| KPEL2 | C | | | | T | | | C | T | T | T | | T | C | | T | C | | C | | | | | A | | | | | T | | | | | C | A | T | | | | A | | | T |
| KPEL3 | C | | | | T | | | C | | T | T | | T | C | | T | C | | C | | | | | A | | | | | T | | | | | C | A | T | | | | A | | | T |
| KPEL4 | C | | | | T | | | C | | T | T | | T | C | | T | C | | C | | | | | A | | | | | T | | | | | C | A | T | | | | A | A | | T |
| KPEL5 | C | | | | T | | | C | | T | T | | T | C | | T | C | | C | | | | | A | | | | | T | | | | | C | A | T | | | | A | | | T |
| KPEL6 | C | | | | T | | | C | | T | T | | T | C | | T | C | | C | | | | | A | | | | | T | | | | | C | A | T | | | | A | | | T |
| KPEL7 | C | | | | T | | | C | | T | T | | T | C | | T | C | | C | | | | | A | | | | | T | | | | | C | A | T | | | | A | | | T |
| KPEL8 | C | | | | T | | | C | | T | T | | T | C | | T | C | | C | | | | | A | | | | | T | | | | | C | A | T | | | | A | | | T |
| KPEL9 | C | | | | T | | | C | | T | T | | T | C | | T | C | | C | | | | | A | | | | | | | | | | C | A | T | | | | A | A | | T |
| KPEL10 | C | | | | T | | | C | | T | T | | T | C | | T | C | | C | | | | | A | | | | | T | | | | | C | A | T | | | | A | | | T |
| KPEL11 | C | | | | T | | | C | | T | T | | T | C | | T | C | | C | | | | | A | | | | | T | | | | | C | A | T | | | | A | | | T |
| KPEL12 | C | | | | T | | | C | | T | T | | T | C | | T | C | | C | | | | | A | | | | | T | | | | | C | A | T | | | | A | | | T |
| KPEL13 | C | | | | T | | | C | | T | T | | T | | | T | C | | C | | | | | A | | T | | | | | | | | C | A | T | | | | A | | | T |
| KPEL14 | C | | | | | | | | | | | | | C | | T | C | | C | | | | | A | | | | | T | | | | | C | A | T | | | | A | | | T |
| KPEL15 | | | | | | | | | | | | | | C | | T | C | | C | | | | | A | | | | | T | | | | | C | A | T | | | | A | | | T |
| KPEL16 | C | | | | T | | | C | | T | T | | T | C | | T | C | | C | | | | | A | | | | | T | | | | | C | A | T | | | | A | | | T |
| KPEL17 | C | | | | T | | | C | | T | T | | T | C | | T | C | | C | | | | | A | | | | | T | | | | | C | A | T | | | | A | | | T |
| KPEL18 | C | | | | T | | | C | | T | | | T | C | | T | C | | C | | | | | A | | | | T | T | | | | | C | A | T | | | | A | | | T |
| KPEL19 | C | | | | | | | | | | | | | | | T | C | | C | | | | | A | | | T | T | T | | | | | C | A | T | | | | A | | | T |
| KPEL20 | C | | | | | | | | | | | | | | | | C | | C | | T | | | A | | | | T | T | | | | | C | A | T | | | | A | | | T |
| KPEL21 | C | | | | T | | | C | | T | T | | T | | | | C | | T | | | | | A | | | | | T | | | | | C | C | | | | | C | | | A | |
| EALL1 | C | T | | | T | | | C | A | T | T | | T | | | | C | | T | | | | | A | | | | T | T | | | | | T | T | | | G | | C | | | T |
| EALL2 | | T | | | | | | C | A | | | | C | | | | C | | C | | | | | A | | | | T | T | | | | | T | T | | | G | | C | | | T |
| EALL3 | C | | | | T | | | C | A | | | | C | | | | C | | T | | | | | A | | | | T | C | | | | | T | T | | | G | | C | | | T |
| ATHA1 | C | | | | | | | C | A | | A | | C | T | | | C | | C | | | | | A | | | | | C | | | | | T | T | | | G | | C | | | T |
| ATHA2 | C | | | | | | | C | A | | A | | C | T | | | C | | C | | T | | | A | | | | | C | | | | | T | T | | | G | | C | | | T |
| ATHA3 | C | | | | | | | C | A | | A | | C | T | | | C | | C | | | | | A | | | T | | C | | | | | T | T | | | G | | C | | | T |
| ATHA4 | C | | | | | | | C | A | | A | | C | T | | | C | | C | | | | | A | | | | | C | | | | | T | T | | | A | | C | | | T |
| ATHA5 | C | | | | | | | C | A | | A | | C | T | | | C | | C | | | | | A | | | | | C | | | | | T | T | | | G | | C | | | T |
| ATHA6 | C | | | | | | | C | A | | A | | C | T | | | C | | C | | | | | A | | | | | C | | | | | T | | | | G | | C | | | T |
| ATHA7 | C | | | | | | | C | A | | A | | C | C | | | C | | C | | | | | A | | | | | C | | | | | T | T | | | G | | C | | | T |
| ATHA9 | C | | | | | | | C | A | | A | | C | C | | | C | | C | | | | | A | | | | | C | | | A | | T | T | | | G | | C | | | T |
| ATHA9 | | | | | | | | | | | | A | | C | | | C | | C | | | | | A | | | | | C | | | | | T | T | | | G | | C | | | T |

37

| | 273 | 276 | 279 | 280 | 282 | 285 | 286 | 287 | 288 | 291 | 294 | 297 | 303 | 306 | 312 | 315 | 318 | 321 | 324 | 326 | 327 | 328 | 329 | 330 | 333 | 342 | 343 | 345 | 348 | 349 | 351 | 352 | 353 | 354 | 357 | 360 | 363 | 364 | 366 | 369 | 372 | 375 | 378 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ATHA10 | | | | | | | | | | | | | | | | | C | | C | | | | | A | | | | | C | | | | | T | T | | G | | | C | | | T |
| ATHA11 | | | | | | | | | | | | | | | | | C | | C | | | | | A | | | | T | C | | | | | T | T | | G | | G | C | | | T |
| ATHA12 | | | | | | | | | | | | | | | | | C | | C | | | | | A | | | | | C | | | | | T | T | | G | | G | C | | | T |
| ATHA13 | C | | | | | | | | | | | | | | | | C | | C | | | | | A | | | | | C | | | | | T | T | | G | | G | C | | | T |
| ATHA14 | C | | | | | | | | A | | | A | C | | | | C | | C | | | | | A | | | | T | C | | | | | | | | G | | G | C | | | T |
| AROC1 | C | | | | | | | C | A | | | A | C | | | | | | C | | | | | A | | | | | T | | | | | C | T | | A | | G | C | | A | T |
| AROC2 | C | | | | | | | C | A | | | A | C | | | | | | C | | | | | A | | | | | T | | | | | C | | | A | | G | C | | | T |
| AROC3 | C | | | | | | | C | A | | | A | C | | | | | | C | | | | | A | | | | | T | | | | | C | | T | A | | G | C | | | T |
| AROC4 | C | | | | | | | C | A | | | A | C | | | | | | C | | | | | A | | | | | T | | | | | C | | | A | | G | C | | | T |
| AROC5 | C | | | | | | | C | A | | | A | C | | | | | | C | | | | | A | | | | | T | | | | | C | T | | A | | G | C | | | T |
| AROC6 | C | | | | | | | C | A | | | A | C | | | | | | C | | | | | A | | | | | T | | | | | C | T | | A | | G | | | | T |
| AROC7 | C | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | A | C | T | T | A | | G | C | | A | T |
| AROC8 | C | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | A | C | T | T | A | | G | C | | A | T |
| AROC9 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | | | | | | C | | | A | | G | C | | A | T |
| AROC10 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | T | | A | | G | C | | | T |
| AROC11 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | | | A | | G | C | | | T |
| AROC12 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | T | T | A | | G | C | | | T |
| AROC13 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | T | | A | | G | C | | | T |
| AROC14 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | T | | A | | G | C | | | T |
| AROC15 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | | | A | | G | C | | | T |
| AROC16 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | T | | A | | G | C | | | T |
| AROC17 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | C | | | | | C | | | A | | G | C | | | T |
| AROC18 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | T | | A | | G | C | | | T |
| AROC19 | | | | | | | | | | | | | | | | | | | C | | T | | | A | | | | | T | | | | | C | | | A | | G | C | | A | T |
| AROC20 | | | | | | | | | | | | | | | | | | | C | | T | | | A | | | | | T | | | | | C | | T | A | | G | C | | A | T |
| AROC21 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | | | A | | G | C | | A | T |
| AROC22 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | | | A | | G | C | | A | T |
| AROC23 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | | T | A | | G | C | | A | T |
| AROC24 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | C | | | | | C | T | | A | | G | C | | A | T |
| AROC25 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | | T | A | | G | C | | A | T |
| AROC26 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | | T | A | | G | C | | A | T |
| AROC27 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | | T | A | | G | C | | A | T |
| AROC28 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | | | A | | G | C | | A | T |
| AROC29 | | | | | | | | | | | | | | | | | | | C | | | | | A | G | | | | T | | | | | C | | T | A | | G | C | | A | T |
| AROC30 | | | | | | | | | | | | | | | | | | | C | | | | | A | G | | | | T | | | | | C | | T | A | | G | C | | A | T |
| AROC31 | | | | | | | | | | | | | | | | | | | C | | | | | A | | | | | T | | | | | C | | T | A | | G | C | | A | T |
| AROC32 | | | | | | | | | A | | | A | C | | | | | | C | | | | | A | | | | | T | | | | | C | | | A | | G | C | | A | T |
| AROC33 | C | | | | | | | C | A | | | A | C | | | | | | | | | | | | | | | | T | | | | | C | T | | A | | G | C | | A | T |
| AROC34 | C | | | | | | | C | A | | | A | C | | | | | | | | | | | | | | | | T | | | | | C | T | | A | | G | C | | | T |

38

| | 273 | 276 | 279 | 280 | 282 | 285 | 286 | 287 | 288 | 291 | 294 | 297 | 303 | 306 | 312 | 315 | 318 | 321 | 324 | 326 | 327 | 328 | 329 | 330 | 333 | 342 | 343 | 345 | 348 | 349 | 351 | 352 | 353 | 354 | 357 | 360 | 363 | 364 | 366 | 369 | 372 | 375 | 378 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AROC35 | C | . | . | . | . | . | . | C | A | . | . | A | C | . | . | . | C | . | C | . | . | . | . | A | . | . | . | . | T | . | . | . | . | C | T | . | A | . | G | C | . | . | T |
| AROC36 | C | . | . | . | . | . | . | C | A | . | . | A | C | . | . | . | . | . | C | . | . | . | . | A | . | . | . | . | T | . | . | . | . | C | T | . | A | . | G | C | . | . | T |
| SSAR1 | C | . | T | . | . | . | . | . | . | T | T | T | C | . | . | . | C | . | C | . | T | . | . | A | . | . | . | T | C | . | . | . | . | C | . | . | T | T | A | A | . | . | . |
| SSAR2 | C | . | T | . | . | . | . | . | . | T | T | T | C | . | . | . | C | . | C | . | . | . | . | A | . | . | . | T | C | . | . | . | . | C | T | . | T | T | A | A | . | . | . |
| SSAR3 | C | . | . | . | . | . | . | . | . | T | T | T | C | . | . | . | C | . | C | . | . | . | . | A | . | . | . | T | C | . | . | . | . | C | T | . | T | T | A | A | . | . | . |
| SSAR4 | C | . | . | . | . | . | . | . | . | . | T | A | C | . | . | . | C | . | C | . | . | . | . | A | . | . | . | T | T | . | . | . | . | C | . | . | T | T | A | A | . | . | . |
| SSAR5 | C | . | . | . | . | . | . | . | . | . | T | A | C | . | . | . | C | . | C | . | . | . | . | A | . | . | . | T | T | . | . | . | . | C | . | . | T | T | A | A | . | . | . |
| SSCO1 | C | T | . | C | . | . | C | . | . | . | . | A | C | G | . | . | A | . | T | T | T | G | T | A | G | . | G | . | T | . | G | . | . | C | T | C | . | . | C | A | A | A | T |
| SSCO2 | C | T | . | C | . | . | C | . | . | . | . | A | C | G | . | . | A | . | T | T | T | G | T | A | G | . | G | T | T | . | G | . | . | C | T | C | . | . | C | A | A | A | T |
| SSCO3 | C | T | . | C | . | . | C | . | . | . | . | A | C | G | . | . | A | . | T | T | T | G | T | A | G | . | G | T | T | . | G | . | . | C | T | C | . | . | C | A | A | A | T |
| SSCO4 | C | T | . | C | . | . | C | . | . | . | . | A | C | G | . | . | A | . | T | T | T | G | T | A | G | . | G | T | T | . | G | . | . | C | T | C | . | . | C | A | A | A | T |
| SSCO5 | C | T | . | C | . | . | C | . | . | . | . | A | C | . | . | . | A | . | T | T | T | . | T | . | G | . | G | . | T | . | . | . | . | C | T | C | . | . | C | A | A | A | T |
| SSCO6 | C | T | . | C | . | . | C | . | . | . | . | A | C | G | . | . | C | . | T | T | T | G | T | . | . | . | . | T | T | . | G | . | . | C | T | C | . | . | C | A | A | A | T |
| SJAP1 | C | T | . | . | T | . | . | . | . | . | . | G | . | . | . | . | C | . | T | T | . | . | T | . | . | . | . | T | T | . | G | . | . | T | . | C | . | . | C | A | A | A | . |
| SJAP2 | C | T | . | . | T | . | . | . | . | . | . | G | . | . | . | . | C | . | T | T | . | . | T | . | . | . | . | T | T | . | G | . | . | T | T | C | . | . | C | A | A | A | . |
| SJAP3 | . | T | . | . | T | . | C | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | T | . | . | . | . | T | T | . | G | . | . | T | T | C | . | . | C | A | A | A | . |
| SJAP4 | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | T | . | T | . | T | . | . | . | . | T | T | . | G | . | . | T | T | C | . | . | C | A | A | A | . |
| SJAP5 | C | T | . | C | T | . | C | . | . | . | . | G | T | . | . | T | C | . | T | T | . | . | T | . | G | . | . | T | T | . | . | . | . | T | T | C | . | . | C | A | A | A | . |
| SCOL1 | C | T | . | C | T | . | C | . | . | . | . | A | T | . | . | T | C | . | T | T | T | . | T | . | . | . | . | T | T | . | . | . | . | T | T | C | . | . | C | A | A | A | . |
| SCOL2 | C | T | . | C | T | . | C | . | . | . | . | G | T | . | . | T | C | . | T | T | T | . | T | . | . | . | . | T | T | . | . | . | . | T | T | C | . | . | C | A | A | A | . |
| SCOL3 | C | T | . | C | T | . | C | . | . | . | . | G | T | . | . | T | C | . | T | T | T | . | T | . | . | . | . | T | T | . | . | . | . | T | T | C | . | . | C | A | A | A | . |
| SCOL4 | C | T | . | C | T | . | C | . | . | . | . | G | T | . | . | T | C | . | T | T | T | . | T | . | . | . | . | T | T | . | . | . | . | T | T | C | . | . | C | A | A | A | . |
| SCOL5 | C | T | . | C | T | . | C | . | . | . | . | A | T | . | . | T | C | . | T | T | T | . | T | . | . | . | . | T | T | . | . | . | . | T | T | C | . | . | C | A | A | A | . |
| SCOL6 | C | T | . | C | T | . | C | . | . | . | . | G | T | . | . | T | C | . | T | T | T | . | T | . | . | . | . | T | T | . | . | . | . | T | T | C | . | . | C | A | A | A | . |
| SCOL7 | C | T | . | C | T | . | C | . | . | . | . | G | C | . | . | T | C | . | T | T | T | . | T | . | . | . | . | T | T | . | . | . | . | T | T | C | . | . | C | A | A | A | . |
| SCOL8 | C | T | . | C | T | . | C | . | . | . | . | G | C | . | . | T | C | . | T | T | T | . | T | . | . | . | . | T | T | . | . | . | . | T | T | C | . | . | C | A | A | A | . |
| SCOL9 | C | T | . | C | T | . | C | . | . | . | . | G | C | . | . | T | C | . | T | T | T | . | T | . | . | . | . | T | T | . | . | . | . | T | T | C | . | . | C | A | A | A | . |
| SCOL10 | C | T | . | C | T | . | C | . | . | . | . | G | C | . | . | . | C | . | T | T | . | . | T | . | . | . | . | . | T | . | . | . | . | T | T | C | . | . | C | A | A | A | . |
| SAU31 | C | T | . | C | T | . | C | . | . | . | . | G | C | . | . | . | C | . | T | T | . | . | T | A | . | . | . | T | T | . | C | . | . | T | T | C | . | . | C | A | A | A | . |
| SAU32 | C | T | . | . | T | . | C | . | . | . | . | G | C | . | . | . | C | . | T | T | . | . | T | . | . | . | . | T | T | . | C | . | . | T | T | C | T | . | C | A | A | A | . |
| SAU33 | C | T | . | . | T | . | C | . | . | . | . | G | C | . | . | . | C | . | T | T | . | . | T | . | . | . | . | T | T | . | . | . | . | T | T | C | T | . | C | A | A | A | . |
| SAU34 | C | T | . | . | T | . | C | . | . | . | . | G | C | . | . | . | . | . | T | T | . | . | T | . | . | . | . | . | T | . | . | A | C | T | A | C | . | . | C | A | . | . | . |
| SAU35 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| SAU36 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| SAU37 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | T | . | . | T | . | . | . | . | T | T | . | C | . | . | T | T | C | . | . | C | A | A | A | T |
| SAU38 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | T | . | . | . | . | . | T | . | . | . | . | T | T | C | . | . | C | A | A | A | T |

39

| REF | 381 | 387 | 390 | 396 | 399 | 402 | 405 | 408 | 409 | 411 | 414 | 416 | 417 | 419 | 420 | 423 | 426 | 429 | 431 | 432 | 438 | 441 | 444 | 447 | 449 | 450 | 453 | 456 | 459 | 462 | 465 | 468 | 471 | 474 | 477 | 480 | 483 | 486 | 489 | 492 | 495 | 498 |
|------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | C | T | C | C | T | C | A | A | C | A | G | C | T | T | C | A | A | T | C | C | T | T | C | T | C | A | C | A | A | C | T | T | A | T | T | C | T | A | A | C | A | A |
| TTHY1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TTHY2 | | | | | | | | | | | | | | | | | | | | | | | | C | | | | | | | | C | | | | | | | | | | |
| TTHY3 | | | | C | | | | | | | | | | | | | | | | | | | | C | | | | | | | | C | | | | | | | | | | |
| TTHY4 | | | | | | | | | | | | | | | | | | | | | | | | C | | | | | | | | C | | | | | | | | | | |
| TTHY5 | | | | | | | | | | | | | | | | | | | | | | | | C | | | | | | | | C | | | | | | | | | | |
| TTHY6 | | | | | | | | | T | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TTHY7 | | | T | | | | | | | | | | | | | | | | | | | | | | | | G | | | | | | | | | | | | | | | |
| TALB1 | | | | | | | | | | | | | | | | | | C | | | | | | | | | | | | | | | | | | | | | | | | |
| TALB2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TALB3 | | C | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | C | | | | | | | | | |
| TALB4 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TALB5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TALB6 | | | T | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TALB7 | | | T | | | | | G | | | | | | | | | | C | | | | | | | | | | | | | | | | | | | | | | | | |
| TATL1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TATL2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TMAC1 | | | | | | | | | | | | | | | | | | | | | | | | C | | | | | | | | C | | | | | | | | | | |
| TMAC2 | | | | | | | | | | | | | | | | | | | | | | | | C | | | | | | | | C | | | | | | | | | | |
| TMAC3 | | | T | | | | | | | | | | | | | | | | | | | | | C | | | | | | | | | | | | | | | | | | |
| TMAC4 | | | | | | | | | | | | | | | | | | | | | | | | | | | G | | | | | | | | | | | | | | | |
| TORI1 | | | | | | | | | | | | | C | | | | | | | | | | T | C | | | | | | | | C | | | C | | | | | | | |
| TORI2 | | | | | | | | | | | | | C | | | | | | | | | | T | C | | | | | | | | C | | | C | | | | | | | |
| TORI3 | | | | | | | | | | | | | C | C | | | | | A | | | | T | C | | | | | | | | C | | | C | | | | | | | |
| TALA1 | | | | | | | | | | | | | C | | | | | | | | | | | C | | | | | | | | C | | | C | | | | | | | |
| TALA2 | | | | | | | | | | | | | C | | | | | | | | | | | C | | | | G | | | | C | | | C | | | | | | | |
| TALA3 | | | | | | | | A | A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TALA4 | | | | | | | G | | | | | | C | | | | | | | | | | | C | | | | | | | | C | | | C | | | | | | G |
| TALA5 | | | | | | | G | | | | A | | C | | | | | | | | | | | C | | | | | | | | C | | | C | | | | | | G |
| TALA6 | | | | | | | G | | | | | | C | | | | | | | | | | | C | | | | | | | | C | | | C | | | | | | G |
| TALA7 | T | | | | | | | | | | | | C | | | | | | | | | | | C | | | | | | | | C | | | C | | | | | | G |
| TALA8 | | | | | | | | | | | | | C | | | | | | | | | | | C | | | | | | | | C | | | C | | | | | | G |
| TTON1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TTON2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TTON3 | | | | | | | | A | | | | | | | | | | | | | | | | | | | T | | | | | | | | | | Q | | | | | |
| TTON4 | | | | | | | | | | G | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | G |
| TTON5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TOBE1 | | C | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TOBE2 | | | | | | | | | | | | | | | | | | | | | | | | C | | | | | | | | C | | | C | | | | | | G |

40

| | 381 | 387 | 390 | 393 | 396 | 399 | 402 | 405 | 408 | 411 | 414 | 417 | 420 | 423 | 426 | 429 | 432 | 435 | 438 | 441 | 444 | 447 | 450 | 453 | 456 | 459 | 462 | 465 | 468 | 471 | 474 | 477 | 480 | 483 | 486 | 489 | 492 | 495 | 498 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TOBE3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | G |
| TOBE4 | C | | | | | | | | | | | | | | | | | | | | | | | | | | | C | | | | C | | | | | | | |
| TOBE5 | C | | | | | | | | | | | | | | | | | | Q | | | | | | | | | C | | | | C | | | | | | | |
| TOBE6 | C | | | | | | | | | | | | | | | | | | | | | | | | | | | C | | | | C | | | | | | | |
| TOBE7 | C | | | | | | | | | | | | | | | | | | | | | | | | | | | C | | | | C | | | | | | | |
| TOBE8 | C | | | | | | | | | | | | | | | | | | | | | | | | | | | C | | | | C | | | | | | | |
| KPEL1 | T | C | | A | | | G | | | | A | | | | | | | | C | | | | | A | | A | | A | A | A | C | C | | C | | G | | G | G |
| KPEL2 | T | C | | A | | | G | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | G |
| KPEL3 | T | C | | A | | | G | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | G | G |
| KPEL4 | T | C | | A | | | G | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | G |
| KPEL5 | T | C | | A | | | G | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | G | G |
| KPEL6 | T | C | | A | | | G | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | |
| KPEL7 | T | C | | A | | | G | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | |
| KPEL8 | T | C | | A | | | G | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | |
| KPEL9 | T | C | | A | | | G | | | | A | | | | | | | | C | | Q | | | | | G | | A | A | A | C | C | | C | | G | | | |
| KPEL10 | T | C | | A | | | | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | G |
| KPEL11 | T | C | T | A | | | G | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | G |
| KPEL12 | T | C | | A | | | G | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | G |
| KPEL13 | T | C | T | A | | | G | | | | A | | | | T | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | G |
| KPEL14 | T | C | | A | | | | | | | A | | | | | | | | C | | | | | A | | A | | A | A | A | C | C | | C | | G | | | G |
| KPEL15 | T | C | | A | | | G | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | G |
| KPEL16 | T | C | | A | | | G | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | G |
| KPEL17 | T | C | | A | | | G | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | G |
| KPEL18 | T | C | T | A | | | G | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | G |
| KPEL19 | T | C | T | A | | | G | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | G |
| KPEL20 | T | C | | A | | | | | | | A | | | G | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | G |
| KPEL21 | T | C | | A | | | G | | | | A | | | | | | | | C | | | | | | | A | | A | A | A | C | C | | C | | G | | | G |
| EALL1 | C | T | T | | | C | | | | | | A | T | G | T | C | C | | A | | C | T | T | I | T | A | C | A | A | T | C | C | T | C | | G | T | G | G |
| EALL2 | C | T | T | | | C | | | | | | A | T | | T | C | C | | A | | C | T | T | I | G | A | C | A | A | T | C | C | T | C | | A | T | G | G |
| EALL3 | C | | | | | C | | | | | | C | | | | | | | | | | | | | | A | | | | | | | | | | | | | |
| ATHA1 | A | C | A | | | | | | | | | A | T | T | T | C | | | C | | C | | T | T | A | G | G | G | G | C | C | G | T | A | | G | | G | G |
| ATHA2 | A | C | A | | | | | | | | | A | T | T | | C | | | C | | C | | T | T | A | A | C | C | G | C | C | G | T | A | | G | | G | G |
| ATHA3 | A | C | A | | | | | | | | | A | T | T | T | C | | | C | | C | | T | T | A | G | G | C | G | C | C | G | T | A | | G | | G | G |
| ATHA4 | A | C | A | | | | | | | | | A | T | T | T | C | | | C | | C | | T | T | A | A | C | C | G | C | C | G | T | A | | G | | G | G |
| ATHA5 | A | C | A | | | | | | | | | A | T | T | | C | | | C | | C | | | T | A | A | C | C | G | C | C | G | T | A | | G | | G | G |
| ATHA6 | A | C | A | | | | | | | | | A | T | T | T | C | | | C | | C | | T | T | A | A | C | C | G | C | C | G | T | A | | G | | G | G |
| ATHA7 | A | C | A | | | | | | | | | A | T | T | T | C | | | C | | C | | T | T | A | A | C | C | G | C | C | G | T | A | | G | | G | G |
| ATHA8 | A | C | A | | | | | | | | | A | T | T | | C | | | C | | C | | T | T | A | A | C | C | G | C | C | G | T | A | | G | | G | G |
| ATHA9 | A | C | A | | | | | | | | | A | T | T | T | C | | | C | | C | | | T | A | A | C | C | G | C | C | G | T | A | | G | | G | G |

| | 381 | 387 | 390 | 396 | 399 | 402 | 405 | 408 | 411 | 414 | 416 | 417 | 419 | 420 | 423 | 426 | 429 | 431 | 432 | 438 | 441 | 444 | 447 | 448 | 450 | 453 | 456 | 459 | 462 | 465 | 468 | 471 | 474 | 477 | 480 | 483 | 486 | 489 | 492 | 493 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ATHA10 | A | C | . | A | . | A | . | . | . | . | . | A | . | T | . | T | C | . | . | . | C | . | C | . | . | T | . | A | C | C | G | C | C | C | . | A | . | . | . | G |
| ATHA11 | A | C | . | A | . | A | . | . | . | . | . | A | . | T | . | T | C | . | . | . | C | . | C | . | . | T | . | A | C | C | G | C | C | C | . | A | . | . | . | G |
| ATHA12 | A | C | . | A | . | A | . | G | . | . | . | A | . | T | . | T | C | . | . | . | C | . | C | . | . | T | . | A | C | C | G | C | C | C | . | A | . | . | . | G |
| ATHA13 | A | C | A | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| ATHA14 | A | C | . | A | . | A | . | . | . | . | . | A | . | T | . | T | C | . | . | . | C | . | C | . | . | T | . | A | C | C | G | C | C | C | . | A | . | . | . | G |
| AROC1 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | A | C | A | T | C | C | . | A | . | . | . | G |
| AROC2 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC3 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC4 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | C | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC5 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | G |
| AROC6 | A | C | . | A | C | A | G | . | . | . | . | A | . | . | . | C | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | G |
| AROC7 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC8 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC9 | A | C | . | A | C | A | . | . | G | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC10 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC11 | A | C | . | A | C | A | G | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | T | C | A | T | C | C | . | A | . | . | . | . |
| AROC12 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC13 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | T | C | A | T | C | C | . | A | . | . | . | . |
| AROC14 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | G | C | . | A | . | . | . | . |
| AROC15 | A | C | . | A | C | A | . | . | . | A | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | G | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC16 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC17 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | C | C | . | . | . | G | T | . | . | . | . | . | A | T | C | A | T | C | C | . | A | . | . | . | . |
| AROC18 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC19 | A | C | . | A | C | A | . | . | . | A | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC20 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | G |
| AROC21 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | T | . | C | T | . | . | . | . | . | G | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC22 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | G |
| AROC23 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | G |
| AROC24 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | G | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC25 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | H | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC26 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC27 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | G | C | C | A | T | C | C | . | A | . | . | . | G |
| AROC28 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC29 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | H | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC30 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | G | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | G |
| AROC31 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | T | C | A | T | C | C | . | A | . | . | . | G |
| AROC32 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC33 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | C | C | A | T | C | C | . | A | . | . | . | . |
| AROC34 | A | C | . | A | C | A | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | A | T | C | A | T | C | C | . | A | . | . | . | G |

| | 381 | 387 | 390 | 396 | 399 | 402 | 405 | 408 | 409 | 411 | 414 | 416 | 417 | 419 | 420 | 423 | 426 | 429 | 431 | 432 | 438 | 441 | 444 | 447 | 448 | 450 | 453 | 456 | 459 | 462 | 465 | 468 | 471 | 474 | 477 | 480 | 483 | 486 | 489 | 492 | 495 | 498 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AROC35 | A | C | A | A | C | A | . | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | G | C | C | A | T | C | C | C | A | . | . | . | . | C |
| AROC36 | A | C | A | A | C | A | . | . | . | . | . | . | A | . | . | . | T | C | . | . | . | C | T | . | . | . | . | . | G | C | C | A | T | C | C | A | C | . | . | . | G | C |
| SSAR1 | T | C | T | T | . | . | . | . | . | G | . | . | C | . | . | . | G | . | . | . | . | C | . | C | C | C | . | . | A | C | C | A | . | . | C | C | C | . | . | . | G | C |
| SSAR2 | T | C | T | T | . | . | . | . | . | G | . | . | C | . | . | . | . | . | . | . | . | C | . | C | C | C | . | A | A | C | C | A | . | . | C | C | C | . | . | . | G | C |
| SSAR3 | T | C | T | T | . | . | . | . | . | G | . | . | C | . | . | . | . | . | . | . | . | C | . | C | C | C | . | A | A | C | C | A | . | . | C | C | C | . | . | . | G | C |
| SSAR4 | T | C | T | T | . | . | . | . | . | G | . | . | C | . | . | . | . | . | . | . | . | C | . | C | C | C | . | A | A | C | C | A | . | . | C | C | . | . | . | . | G | C |
| SSAR5 | T | C | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C |
| SSCO1 | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | G | C | . | T | C | C | . | A | C | C | A | A | G | C | C | A | C | A | C | . | G | . | . | T | . | G |
| SSCO2 | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | G | C | . | T | C | C | . | A | C | C | A | A | T | . | C | A | C | A | A | . | G | . | . | T | . | G |
| SSCO3 | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | G | C | . | T | C | C | . | A | C | C | A | A | . | . | C | A | C | A | A | . | G | . | G | T | . | G |
| SSCO4 | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | G | C | . | T | C | C | . | A | C | C | A | A | T | . | C | A | C | A | A | . | G | . | . | T | . | G |
| SSCO5 | . | . | . | . | . | . | . | C | . | G | . | . | C | . | . | . | G | C | . | T | C | C | . | A | C | C | A | G | T | . | C | C | C | . | C | . | G | . | . | . | . | G |
| SSCO6 | . | . | . | . | . | . | . | C | . | G | . | . | C | . | . | . | G | C | . | T | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C |
| SJAP1 | A | C | . | . | . | . | C | C | . | . | . | . | C | . | . | . | . | . | . | T | . | A | T | A | T | T | G | G | T | . | C | C | T | . | C | . | G | . | . | . | . | C |
| SJAP2 | A | C | . | . | . | . | C | C | . | . | . | . | C | . | . | . | G | C | . | T | . | A | T | A | T | T | G | G | T | . | C | C | T | . | A | . | G | . | . | . | . | C |
| SJAP3 | A | C | . | . | . | . | C | T | . | . | . | . | C | . | . | . | G | C | . | T | . | A | . | A | T | T | G | G | T | . | C | C | T | . | C | . | G | . | . | . | . | C |
| SJAP4 | A | C | . | . | . | . | C | T | . | . | . | . | C | . | . | . | G | C | . | T | . | A | T | A | T | T | A | A | T | . | C | C | T | . | A | . | G | . | . | . | . | C |
| SJAP5 | A | C | . | . | . | . | C | T | . | . | . | . | C | . | . | . | G | C | . | T | . | A | . | A | T | T | G | G | T | . | C | . | T | . | C | . | G | . | . | . | . | C |
| SCOL1 | A | C | . | . | . | . | . | O | . | . | . | . | C | . | . | . | G | C | . | T | . | G | . | A | . | C | G | G | . | . | C | C | T | . | C | . | G | . | . | . | . | C |
| SCOL2 | A | C | . | . | . | . | . | T | . | . | . | . | C | . | . | . | G | C | . | T | . | G | . | A | C | C | G | G | . | . | C | C | T | A | A | . | G | . | . | . | . | C |
| SCOL3 | A | C | . | . | . | . | . | T | . | . | . | . | C | . | . | . | G | C | . | T | . | G | . | A | C | C | G | G | . | . | C | C | T | A | A | . | G | . | . | . | . | C |
| SCOL4 | A | C | . | . | . | . | . | T | . | . | . | . | C | . | . | . | G | C | . | T | . | A | . | A | C | C | A | A | . | . | C | C | T | A | A | . | G | . | . | . | . | C |
| SCOL5 | A | C | . | . | . | . | . | T | . | . | . | . | C | . | . | . | G | C | . | T | . | G | . | A | C | C | G | G | . | . | C | C | T | A | A | . | G | . | . | . | . | C |
| SCOL6 | A | C | . | . | . | . | . | T | . | . | . | . | C | . | . | . | G | C | . | T | . | A | . | A | C | C | G | G | . | . | C | C | T | A | A | . | G | . | . | . | . | C |
| SCOL7 | A | C | . | . | . | . | . | O | . | . | . | . | C | . | . | . | G | C | . | T | . | G | . | A | C | C | G | G | . | . | C | C | T | A | A | . | G | . | . | . | . | C |
| SCOL8 | A | C | . | . | . | . | . | T | . | . | . | . | C | . | . | . | . | . | . | T | . | A | T | A | C | C | A | G | . | . | C | . | T | A | A | . | . | . | . | . | . | C |
| SCOL9 | A | C | . | . | . | . | . | C | . | . | . | . | C | . | . | . | G | C | . | . | . | A | . | A | C | C | G | G | . | . | C | C | T | A | A | . | G | . | . | . | G | C |
| SCOL10 | A | G | . | . | . | . | . | C | . | . | . | . | C | . | . | . | G | C | . | T | . | A | . | A | C | C | A | G | . | . | C | . | T | A | A | . | G | . | . | . | G | C |
| SAUS1 | A | C | . | . | . | . | . | O | . | . | . | . | C | . | . | . | G | C | . | . | . | A | . | A | A | T | G | G | . | . | C | C | T | A | A | . | G | . | . | . | G | C |
| SAUS2 | A | C | . | . | . | . | . | T | . | . | . | . | C | . | . | . | . | . | . | T | . | A | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | C |
| SAUS3 | A | C | . | . | . | . | . | T | . | . | . | . | C | . | . | . | G | C | . | . | . | A | . | A | C | C | G | G | . | . | C | C | T | A | A | . | G | . | . | . | . | C |
| SAUS4 | A | C | . | . | . | . | . | T | . | . | . | . | C | . | . | . | G | C | . | . | . | A | . | A | T | T | A | A | . | . | C | C | T | A | A | . | G | . | . | . | G | C |
| SAUS5 | A | C | . | . | . | . | . | T | . | . | . | . | C | . | . | . | G | C | . | T | . | A | . | A | C | C | G | G | . | . | C | C | T | A | A | . | G | . | . | . | . | C |
| SAUS6 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| SAUS7 | A | C | . | . | . | . | . | O | . | . | . | . | C | . | . | . | G | C | . | T | . | A | . | A | T | T | G | G | . | . | C | C | T | A | A | . | G | . | . | . | G | C |
| SAUS8 | A | C | . | . | . | . | . | T | . | . | . | . | C | . | . | . | G | C | . | T | . | A | . | A | T | T | G | G | . | . | C | C | T | A | A | . | G | . | . | . | G | C |

43

Table 16 shows the SNPs identified by fragment B. The first column indicates the species in letters and the arbitrary name of the polymorphism identified (letters plus numbers), the first line indicates the indexing of each specific nucleotide of the amplicons. Position 402 is shared by fragment A plus as shown also in Table 15 and 17

| REF. | 402 | 405 | 408 | 409 | 411 | 414 | 416 | 417 | 419 | 420 | 423 | 428 | 429 | 431 | 432 | 433 | 441 | 444 | 447 | 448 | 450 | 453 | 456 | 459 | 462 | 465 | 468 | 471 | 474 | 477 | 480 | 483 | 486 | 489 | 492 | 495 | 498 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | C | A | A | C | A | G | C | T | T | C | A | A | T | C | C | T | T | C | C | C | A | C | A | C | A | T | T | A | T | T | C | T | A | A | C | A | A |
| TTHY1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TTHY2 | | | | | | | | | | | | | | | | | | | C | | | | | | | | C | | | | | | | | | | |
| TTHY3 | | | | | | | | | | | | | | | | | | | | | | | | | | | C | | | | | | | | | | |
| TTHY4 | | | | | | | | | | | | | | | | | | | C | | | | | | | | C | | | | | | | | | | |
| TTHY5 | | | | | | | | | | | | | | | | | | | C | | | | | | | | C | | | C | | | | | | | |
| TTHY6 | | | | | | | T | | | | | | | | | | | | C | | | | | | | | C | | | | | | | | | | |
| TTHY7 | | | | | | | | | | | | | | | | | | | G | | | | | | | | | | | | | | | | | | |
| TALB1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TALB2 | | | | | | | | | | | | | C | | | | | | | | | | | | | | | | | | | | | | | | |
| TALB3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TALB4 | | | | | | | | | | | | | C | | | | | | | | | | | | | | | | | C | | | | | | | |
| TALB5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TALB6 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TALB7 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TATL1 | | G | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TATL2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TMAC1 | | | | | | | | | | | | | | | | | | | C | | | | | | | | C | | | C | | | | | | | |
| TMAC2 | | | | | | | | | | | | | | | | | | | C | | | | | | | | C | | | | | | | | | | |
| TMAC3 | | | | | | | | | | | | | | A | | | | | | | | | | | | | | | | | | | | | | | |
| TMAC4 | | | | | | | | | | | | G | | | | | | | C | | | | G | | | | C | | | | | | | | | | |
| TORI1 | | | | | | | | C | | | | | | | | | | T | C | | | | G | | | | C | C | | C | | | | | | G | G |
| TORI2 | | | | | | | | C | C | | | | | | | | | T | C | | | | | | | | C | | | C | | | | | | G | G |
| TORI3 | | | | | | | | C | C | | | | | | | | | T | C | | | | | | | | C | | | C | | | | | | G | G |
| TALA1 | | | | | | | | C | | | | | | A | | | | | C | | | | | | | | C | | | C | | | | | | G | G |
| TALA2 | | | | | | | | C | | | | | | | | | | | C | | | | | | | | C | | | C | | | | | | G | G |
| TALA3 | | | | | | | | C | | | | G | | | | | | | C | | | | | | | | | | | C | | | | | | G | G |
| TALA4 | | | | | | | | C | | | | | | | | | | | C | | | | | | | | C | | | C | | | | | | G | G |
| TALA5 | | | | | | | | C | | | | | | | | | | | C | | | | | | | | | | | C | | | | | | G | G |
| TALA6 | | | | | | | | | | | | | | | | | | | C | | | | | | | | C | | | C | | | | | | | |
| TALA7 | | | | | | | | | | | | | | | | | | | C | | | T | | | | | | | | C | | C | | | | | |
| TALA8 | | | | | | | | | | | | | | | | | | | C | | | | | | | | C | | | C | | | | | | | |
| TTON1 | | G | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TTON2 | | G | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TTON3 | | G | | A | G | | | | | | | | | | | | | | | | | T | | | | | | | | | | | | | | | |
| TTON4 | | G | A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | C | | | | | |
| TTON5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TOBE1 | | | | | | | | | | | | | | | | | | | C | | | | | C | | | C | | | C | | | | | | G | G |
| TOBE2 | | | | G | | | | | | | | | | | | | | | C | | | | | | | | C | | | C | | | | | | G | G |
| TOBE3 | | | | | | | | | | | | | | | | | | | | | | | | | | | C | | | | | | | | | | |
| TOBE4 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | 402 | 405 | 408 | 409 | 411 | 414 | 416 | 417 | 419 | 420 | 423 | 426 | 429 | 431 | 432 | 438 | 441 | 444 | 447 | 448 | 450 | 453 | 456 | 459 | 462 | 465 | 468 | 471 | 474 | 477 | 480 | 483 | 486 | 489 | 492 | 495 | 498 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TOBE5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | C | | | | | | | |
| TOBE6 | | | | | | | | | | | | | | | | | C | | | | | | | | | C | C | | | C | | | | | | | |
| TOBE7 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TOBE8 | | | | | | | | | | | | | | | | | | | | | | | | | | C | C | | | C | | | | | | | |
| KPEL1 | | | | | | | | A | | | | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | |
| KPEL2 | | G | | | | | | A | | | | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | G |
| KPEL3 | | | | | | | | A | | | | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | G |
| KPEL4 | | | | | | | | A | | | | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | |
| KPEL5 | | G | | | | | | A | | | | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | G |
| KPEL6 | | G | | | | | | A | | | | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | G |
| KPEL7 | | G | | | | | | A | | | | | | | | | C | | | | | | | G | C | C | A | | | C | | C | | | | | |
| KPEL8 | | G | | | | | | A | | | | | | | | | | | | | | | | A | C | C | A | | | C | | C | | | | | |
| KPEL9 | | G | | | | | | A | | | | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | G |
| KPEL10 | | | | | | | | A | | | | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | |
| KPEL11 | | G | | | | | | A | | | | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | |
| KPEL12 | | | | | | | | A | | | | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | G |
| KPEL13 | | | | | | | | A | | | | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | |
| KPEL14 | | G | | | | | | A | | | | | | | | | C | | | | | A | | A | C | C | A | | | C | | C | | | | | |
| KPEL15 | | G | | | | | | A | | | | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | |
| KPEL16 | | G | | | | | | A | | | | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | |
| KPEL17 | | G | | | | | | A | | | | | | | | | C | | T | | | | | A | C | C | A | | | C | | C | | | | | |
| KPEL18 | | | | | | | | A | | | | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | |
| KPEL19 | | | | | | | | A | | | G | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | G |
| KPEL20 | | G | | | | | | A | | | | | | | | | C | | | | | | | A | C | C | A | | | C | | C | | | | | |
| KPEL21 | | G | | | | | | A | | | | | | C | | | C | | | | | | | A | C | C | A | T | C | C | | C | | | | | |
| EALL1 | T | | | | | | | A | | | | T | C | | | | C | | C | T | | T | | A | C | C | A | T | C | C | T | C | | T | T | | G |
| EALL2 | T | | | | | | | C | | | | T | C | | | | | | C | T | | T | | A | C | C | A | T | C | C | T | C | | | T | | G |
| EALL3 | T | | | | | | | A | | | | T | | | | C | | | C | | T | | | A | C | G | G | | C | C | | A | | | | | G |
| ATHA1 | A | | | | | | | A | T | | | T | C | | | | C | | C | | | T | G | G | G | G | G | G | C | C | | A | | | | | G |
| ATHA2 | A | | | | | | | A | T | | | T | C | | | | C | | C | | | T | G | G | G | G | G | G | C | C | | A | | | | | G |
| ATHA3 | A | | | | | | | A | T | | | T | C | | | | C | | C | | | T | G | A | C | G | G | G | C | C | | A | | | | | G |
| ATHA4 | A | | | | | | | A | T | | | T | C | | | | C | | C | | | T | | A | C | G | G | G | C | C | | A | | | | | G |
| ATHA5 | A | | | | | | | A | T | | | T | C | | | | C | | C | | | T | | A | C | G | G | G | C | C | | A | | | | | G |
| ATHA6 | A | | | | | | | A | T | | | T | C | | | | C | | C | | | T | | A | C | G | G | G | C | C | | A | | | | | G |
| ATHA7 | A | | | | | | | A | T | | G | T | C | | | | C | | C | | | T | | A | C | G | G | G | C | C | | A | | | | | G |
| ATHA8 | A | | G | | | | | A | T | | | T | C | | | | C | | C | | | T | | A | C | G | G | G | C | C | | A | | | | | G |
| ATHA9 | A | | | | | | | A | T | | | T | C | | | | C | | C | | | T | | A | C | G | G | G | C | C | | A | | | | | G |
| ATHA10 | A | | | | | | | A | T | | | | C | | | | C | | C | | | T | | A | C | G | G | G | C | C | | A | | | | | G |
| ATHA11 | A | | | | | | | A | T | | | T | C | | | | C | | C | | | | | A | C | G | G | G | C | C | | A | | | | | G |
| ATHA12 | A | | G | | | | | A | | | | | C | | | | C | | C | | | T | | A | C | G | G | G | C | C | | A | | | | | G |
| ATHA13 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

45

| | 402 | 405 | 408 | 409 | 411 | 414 | 416 | 417 | 419 | 420 | 423 | 426 | 429 | 431 | 432 | 438 | 441 | 444 | 447 | 448 | 450 | 453 | 456 | 459 | 462 | 465 | 468 | 471 | 474 | 477 | 480 | 483 | 486 | 489 | 492 | 495 | 498 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ATHA14 | | A | | | | | | | | T | | T | C | | | | C | | C | | | T | | A | A | C | G | G | C | C | | A | A | | | | |
| AROC1 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | G |
| AROC2 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | A | C | A | T | C | C | | A | A | | | | |
| AROC3 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | T | C | A | T | C | C | | A | A | | | | |
| AROC4 | A | A | | | | | | A | | | | C | | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC5 | A | A | | | | | | A | | | | C | C | | | | C | | | | | | | A | C | C | A | T | C | C | | A | A | | | | G |
| AROC6 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | G |
| AROC7 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC8 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC9 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | T | C | A | T | C | C | | A | A | | | | G |
| AROC10 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | G |
| AROC11 | A | A | | | | | | A | | | | T | C | | | | C | | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC12 | A | G | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC13 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | G |
| AROC14 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC15 | A | A | | | | A | | A | | | | T | C | | | | C | T | | | | | | G | C | C | A | T | C | C | | A | A | | | | |
| AROC16 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC17 | A | A | | | | | | A | | | | T | C | | | | G | T | | | | | | A | T | C | A | T | | | | A | A | | | G | |
| AROC18 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC19 | A | A | | | | | | A | | | G | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | G |
| AROC20 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | G |
| AROC21 | A | A | | | G | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC22 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC23 | A | A | | | | A | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | G |
| AROC24 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | G |
| AROC25 | A | A | | | | | | A | | | G | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC26 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | G |
| AROC27 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC28 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC29 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | G |
| AROC30 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC31 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC32 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | T | C | A | T | | C | | A | A | | | | G |
| AROC33 | A | A | | | | | | A | | | | T | C | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | |
| AROC34 | A | A | | | | | | A | | | | T | | | | | C | T | | | | | | A | C | C | A | T | C | C | | A | A | | | | G |
| AROC35 | A | A | | | | | | C | | | | | C | | | | C | T | | | | | | A | C | C | A | | C | C | | A | C | | | | |
| AROC36 | A | | | | | | | C | | | | G | | | | | C | C | | | | | | A | C | C | A | A | C | C | | A | C | | | | |
| SSAR1 | | | | | G | | | C | | | | | | | | | C | | C | C | | | | A | | | A | | | C | | | C | | | | C |
| SSAR2 | | | | | G | | | C | | | | G | | | | | C | | C | C | | | | A | C | | A | | | C | | | C | | | | C |
| SSAR3 | | | | | G | | | C | | | | | | | | | C | | C | C | | | | A | | | A | | | C | | | C | | | | C |
| SSAR4 | | | | | G | | | C | | | | G | | | | | C | | C | C | | | | A | | | A | | | C | | | C | | | | C |

46

| | 402 | 405 | 408 | 409 | 411 | 414 | 416 | 417 | 419 | 420 | 423 | 426 | 429 | 431 | 432 | 438 | 441 | 444 | 447 | 448 | 450 | 453 | 456 | 459 | 462 | 465 | 468 | 471 | 474 | 477 | 480 | 483 | 486 | 489 | 492 | 495 | 498 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SSAR5 | . | . | . | . | . | . | . | . | . | . | . | G | C | . | . | . | C | . | A | . | C | A | . | A | A | C | A | A | . | C | C | C | . | . | . | . | G |
| SSCO1 | . | . | . | . | G | . | . | C | . | . | . | G | C | T | . | C | . | . | A | . | C | A | . | T | A | C | A | C | A | A | . | . | . | . | T | . | G |
| SSCO2 | . | . | . | . | . | . | . | C | . | . | . | G | C | T | . | C | . | . | A | . | C | A | . | . | A | C | A | C | A | A | . | . | . | . | . | . | G |
| SSCO3 | . | . | . | . | . | . | . | C | . | . | . | G | C | T | . | C | . | . | A | . | C | A | . | T | A | C | A | C | A | A | . | . | . | . | T | . | G |
| SSCO4 | . | . | . | . | . | . | . | C | . | . | . | G | C | T | . | C | . | . | A | . | C | A | . | T | A | C | A | C | A | A | . | . | . | G | T | . | G |
| SSCO5 | . | . | . | . | . | . | . | C | . | . | . | G | C | T | . | C | . | . | A | . | C | A | . | T | A | C | A | C | A | A | . | . | . | . | T | . | G |
| SSCO6 | . | . | . | . | . | . | . | C | . | . | . | G | C | T | . | C | . | . | A | . | C | A | . | . | A | C | A | C | A | A | . | . | . | . | . | . | G |
| SJAP1 | . | . | C | . | . | . | . | C | . | . | . | G | C | . | . | . | A | T | A | . | T | G | . | . | . | C | C | T | . | C | . | . | G | . | . | . | C |
| SJAP2 | . | . | C | . | . | . | . | C | . | . | . | G | C | . | . | . | A | . | A | . | T | G | . | . | . | C | C | T | . | C | . | . | G | . | . | . | C |
| SJAP3 | . | . | C | . | . | . | . | C | . | . | . | G | C | . | . | . | A | T | A | . | T | G | . | . | . | C | C | T | . | C | . | . | G | . | . | . | C |
| SJAP4 | . | . | C | . | . | . | . | C | . | . | . | G | C | . | . | . | A | . | A | . | T | A | . | . | . | C | C | T | . | A | . | . | G | . | . | . | C |
| SJAP5 | . | . | T | . | . | . | . | C | . | . | . | G | C | . | . | . | G | . | A | . | . | . | . | . | . | C | C | T | . | A | . | . | G | . | . | . | C |
| SCOL1 | . | . | T | . | . | . | . | C | . | . | . | . | C | T | . | . | G | . | A | . | C | G | . | . | . | C | C | T | . | A | . | . | G | . | . | . | C |
| SCOL2 | . | . | T | . | . | . | . | C | . | . | . | . | C | T | . | . | G | . | A | . | C | G | . | . | . | C | C | T | . | A | . | . | G | . | . | . | C |
| SCOL3 | . | . | T | . | . | . | . | C | . | . | . | . | C | T | . | . | G | . | A | . | C | G | . | . | . | C | . | T | . | A | . | . | G | . | . | . | C |
| SCOL4 | . | . | T | . | . | . | . | C | . | . | . | . | C | T | . | . | A | . | A | . | C | A | . | . | . | C | . | T | . | A | . | . | G | . | . | . | C |
| SCOL5 | . | . | C | . | . | . | . | C | . | . | . | . | C | T | . | . | G | . | A | . | C | G | . | . | . | C | C | T | . | A | . | . | G | . | . | . | C |
| SCOL6 | . | . | T | . | . | . | . | C | . | . | . | . | C | T | . | . | A | . | A | . | C | G | . | . | . | C | C | T | . | A | . | . | G | . | . | . | C |
| SCOL7 | . | . | T | . | . | . | . | C | . | . | . | . | C | T | . | . | G | . | A | . | C | G | . | . | . | C | . | T | . | A | . | . | G | . | . | . | C |
| SCOL8 | . | . | T | . | . | . | . | C | . | . | . | G | C | T | . | . | G | . | A | . | C | G | . | . | . | C | C | T | . | A | . | . | G | . | . | . | C |
| SCOL9 | . | . | T | . | . | . | . | C | . | . | . | G | C | T | . | . | G | . | A | . | C | G | . | . | . | C | C | T | . | A | . | . | G | . | . | . | C |
| SCOL10 | . | . | T | . | . | . | . | C | . | . | . | . | C | . | . | . | A | . | A | . | . | . | . | . | . | . | . | T | . | . | . | . | G | . | . | . | . |
| SAUS1 | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| SAUS2 | . | . | C | . | . | . | . | C | . | . | . | G | C | T | . | . | A | . | A | . | T | G | . | . | . | C | C | T | . | A | . | . | G | . | . | G | C |
| SAUS3 | . | . | C | . | . | . | . | C | . | . | . | G | C | T | . | . | A | . | A | . | C | G | . | . | . | C | C | T | . | A | . | . | G | . | . | . | C |
| SAUS4 | . | . | T | . | . | . | . | C | . | . | . | G | C | T | . | . | A | . | A | . | T | G | . | . | . | C | C | T | . | A | . | . | G | . | . | G | C |
| SAUS5 | . | . | T | . | . | . | . | C | . | . | . | . | C | T | . | . | A | . | A | . | T | G | . | . | . | C | C | T | . | A | . | . | G | . | . | . | C |
| SAUS6 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| SAUS7 | . | . | C | . | . | . | . | C | . | . | . | G | C | T | . | . | A | . | A | . | T | G | . | . | . | C | C | T | . | A | . | . | G | . | . | G | C |
| SAUS8 | . | . | T | . | . | . | . | C | . | . | . | G | C | T | . | . | A | . | A | . | T | G | . | . | . | C | C | T | . | A | . | . | G | . | . | . | C |

Table 17, below, indicates the SNPs identified by fragments A and A plus. The first column indicates the species in letters and the arbitrary name of the polymorphism identified (letters plus numbers), the first line indicates the indexing of each specific nucleotide of the amplicons, the SNPs indicated by the bold characters are identified in the A fragment (up to nucleotide 363 included), the SNPs indicated by the normal characters (up to 402 included), summed to the SNPs in bold are identified in fragment A plus

47

48

49

50

The following examples are given for a better understanding of the description and not to limit the claims.

EXAMPLES

**Example 1.**

**Design of the primers for the detection of several *Scombridae* species:**

Each primer in the cocktail used in PCR-1 is made up of 3 parts:

i) one part complementary to the *cytochrome b* gene (MT-CYB, represented in Figure 2 by a continuous line with symbols, e.g. triangles, crosses, stars, etc. to distinguish a primer from the others)

ii) one part is the "read start" sequence (represented in Figure 2 by a waved ⋀⋁⋀⋁ line,)

iii) one part is a universal primer sequence (represented in Figure 2 by a U succession, UUUU)

The "read start" signal is useful as it inserts a region in the amplicon sequence which is the same in the amplicon of each of the 17 species and can be used to make the results analysis easier. A dedicated software application (not yet available but easy to develop) may be used to read a specific set of SNPs at a determined point of the amplicon. For instance, from a fasta output of the amplicon's sequence, the software will be able to extrapolate the SNPs positions at different points and output a combination of SNPs associated to the name of the species and the intraspecies variant.

To discriminate some variants, e.g. *Thunnus Thynnus* and *Thunnus maccoyii*, the information of two regions of the gene are combined by performing two assays.

Two different methods adopting the same strategy but using primers with different length were tested. Both methods allow to amplify all the same fragments of the *cytochrome b* gene. For economic reasons and based on the results obtained so far, the method using shorter primers was chosen to be described in detail. With this method it is possible to provide the whole sequence of the amplified fragment, without loss of information about the sequence of bases in the discriminating region of the amplicon.

**Example 2**

**Assays**

Four assays that amplify two adjacent regions of the *cytochrome b* gene were made:

one assay amplifies a tiny fragment called B, a second assay amplifies a fragment called A and a third assay amplifies a fragment which includes fragments A and B called A+B and a fourth assay amplifies a fragment called A plus (or A+).

51

Each fragment is amplified using the above described method (see figure 2).

Fragment A: amplifies 91 bp (without primers) of the *cytochrome b* gene, the amplification product obtained by PCR-1 is 156 bp long including the primers length, the amplification product obtained by PCR-2 is 184 bp long including the primers length.

Fragment B: amplifies 95 bp (without primers) of the *cytochrome b* gene, the amplification product obtained by PCR-1 is 162 bp long including the primers length, the amplification product obtained by PCR-2 is 192 bp long including the primers length.

Fragment A+B: amplifies 226 bp (without primers) of the *cytochrome b* gene, the amplification product obtained by PCR-1 is 291 bp long including the primers length, the amplification product obtained by PCR-2 is 320 bp long including the primers length.

Fragment A+: amplifies 151 bp (without primers) of the *cytochrome b* gene, the amplification product obtained by PCR-1 is 174 bp long including the primers length, the amplification product obtained by PCR-2 is 203 bp long including the primers length.

**Example 3**

**a. Analysis of fresh or slightly processed (DNA not much degraded) samples**

Fragment A+B is the longest (and obviously the most informative) and allows to better discriminate the intraspecies variability. From highly processed samples (e.g. tuna sauces or some tinned tuna in oil) it is not possible to get an amplification product of fragment A+B. In facts, according to literature data, it is possible to obtain an amplification product with the size of fragment A+B (Pardo *et al.* 2004) by means of "nested" PCR.

**b. Analysis of highly processed food matrices (DNA much degraded)**

For this kind of samples, fragment B (more informative than fragment A) is amplified. To discriminate some variants (e.g. *Thunnus thynnus* and *Thunnus maccoyii*) in some cases it is necessary to perform a further assay amplifying also fragment A+.

**Example 4**

**Construction of family Scombridae Genetic Database**

Fresh fishes specie morphologically identified by sanitary experts: 1 *Thunnus thynnus,* 1 *Auxis rochei,* 1 *Euthynnus alletteratus?,* 1 *Sarda sarda,* 1 *Thunnus albacares,* 1 *Scomber scombrus.*

Commercial canned tuna samples with the name of the specie declared on

52

label by the large-scale retail trade: 3 *Katsuwonus pelamis*, 2 *Thunnus thynnus*, 5 *Thunnus albacares*, 1 *Thunnus alalunga*.

Commercial canned tuna samples with the name of the specie NOT declared on label by the large-scale retail trade: 9 *Thunnus sp.*, 7 *Scomber sp.*, 80 *Thunnus sp.*

Tuna sauce samples:1 sample of sauce with 20% of *Thunnus sp.*, 9 samples of sauce with 6% of Thunnus sp.

**Example 5**

**DNA extraction**

The clean DNA used as PCR template can be obtained with two different commercial kit:

1)      GREES DNA KIT FOOD (General Rapid Easy Extraction System DNA Kit, Incura s.r.l., code: IC-02-0095), is a rapid kit for DNA genomics extraction from food matrices. Ten samples can be processed in 2 hours. Following the manufacturer's instructions you can obtain 1-1,5 µg total from 350 mg of initial sample. Good quality DNA with an OD **260/280** ratio of 1.9 to 2 is measured with **NanoDrop**® ND-1000 Spectrophotometer.

2)      Wizard ® Magnetic DNA Purification System for food (Promega corporation, code: FF3750) is a less fast kit for DNA genomics extraction from food matrices. Ten samples can be processed in 3,5 hours. Following the manufacturer's instructions you can obtain 4,5-7,5 µg total from 350 mg of initial sample. The quality of DNA is worse than quality of DNA obtained with GREES DNA KIT FOOD: the OD **260/280** ratio are 1.5 to 1,7 measured with **NanoDrop**® ND-1000 Spectrophotometer.

Table 11: number of sequences used to set up the assays for each specie

*Specie abbreviation/ Number of sequence in public database used to set up the assays*

53

| Scientific Name | Specie Abbreviation | Number of sequences in Public Database |
|---|---|---|
| Thunnus thynnus | TTHY | 28 |
| Thunnus albacares | TALB | 37 |
| Thunnus atlanticus | TATL | 2 |
| Thunnus maccoyii | TMAC | 4 |
| Thunnus orientalis | TORI | 4 |
| Thunnus obesus | TOBE | 26 |
| Thunnus alalunga | TALA | 34 |
| Thunnus tonggol | TTON | 12 |
| Katsuwonus pelamis | KPEL | 36 |
| Euthynnus alletteratus | EALL | 7 |
| Auxis thazard thazard | ATHA | 21 |
| Auxis rochei rochei | AROC | 72 |
| Sarda sarda | SSAR | 20 |
| Scomber scombrus | SSCO | 20 |
| Scomber japonicus | SJAP | 11 |
| Scomber colias | SCOL | 19 |
| Scomber australasicus | SAUS | 19 |

The sequences were aligned with clustalW multiple alignment. The primers were designed in order to recognize all the species and the intraspecies variant.

The reaction conditions to amplify the four developed assay (fragment B, Fragment A and fragment A+B and fragment A+) are described below in succession.

**Example 6**

**Amplicon B: reaction conditions, PCR component and primers**

Table 12: reaction conditions and component of the PCR which amplify the fragment B.

| | Amount per reaction (20μl) | final |
|---|---|---|
| DNA (20ng/ μl) | 1,5 μl | |
| 10X PCR reaction buffer | 2 μl | 1X |
| Mix Primer F (10 μM each primer) | 0,75 μl | 0,375 μM (each primer) |
| Mix Primer R (10 μM each primer | 0,75 μl | 0,375 μM (each primer) |
| 10 mM dNTPs | 0.4 μl | 0,2 mM |
| HotStartTaq DNA (5U/ μl) | 0,1-0,2 μl | 0,5-1 U |
| DNAse-free water | | |

The amplifications condition were as follow: 95°C for 15 min, then 35 cycles of 95°C for 45 s, 60°C for 1 min and 72°C for 45 s, followed by 72°C for 10 min with a final hold at 4°C.

Table 3 discloses the sequences of the primers used to amplify the fragment B.

Species identification was achieved by sequencing and indexing the nucleotides of the amplicons as described above. The primers are given in Table 4 and sequencing standard protocol has been used to sequence fragment B.

**Example 7**

**Amplicon A+B: reaction conditions, PCR component and primers**

54

Table 13: reaction conditions and component of the PCR which amplify the fragment A+B.

|  | Amount per reaction (20μl) | final |
|---|---|---|
| DNA  (20ng/ μl) | 1,5 μl |  |
| 10X PCR reaction buffer | 2 μl | 1X |
| Mix Primer F (5 μM each primer) | 2 μl | 0,5 μM (each primer) |
| Mix Primer R (5 μM each primer | 2 μl | 0,5 μM (each primer) |
| 10 mM dNTPs | 0.4 μl | 0,2 mM |
| HotStartTaq DNA (5U/ μl) | 0,2 μl | 1 U |
| DNAse-free water |  |  |

The amplifications condition were as follow: 95°C for 15 min, then 20 cycles of 94°C for 1 min, 56,7°C for 50 sec and 72°C for 45 s, then other 18 cycles of 94°C for  1 min, 55°C for 50 sec and 72°C for 45 s followed by 72°C for 10 min with a final hold at 4°C.

Table 5 discloses the sequences of the primers used to amplify the fragment A+B.

For amplicon A+B positive samples renders a 291 bp band

Species identification was achieved by sequencing and indexing the nucleotides of the amplicons as described above.. The primers are given in Table 6 and sequencing standard protocol was used to sequence amplicon A+B.

## Example 8

### Amplicon A or A+: reaction conditions, PCR component and primers

Table 14: reaction conditions and component of the PCR which amplify the fragment A.

|  | Amount per reaction (20μl) | final |
|---|---|---|
| DNA  (20ng/ μl) | 1,5 μl |  |
| 10X PCR reaction buffer | 2 μl | 1X |
| Mix Primer F (5 μM each primer) | 2 μl | 0,5 μM (each primer) |
| Mix Primer R (5 μM each primer) | 2 μl | 0,5 μM (each primer) |
| 10 mM dNTPs | 0.4 μl | 0,2 mM |
| HotStartTaq DNA (5U/ μl) | 0,1-0,2 μl | 0,5-1 U |
| DNAse-free water |  |  |

The amplifications condition were as follow: 95°C for 15 min, then 20 cycles of 94°C for  1 min, 56,7°C for 50 sec and 72°C for 45 s, then other 18 cycles of 94°C for  1 min, 55°C for 50 sec and 72°C for 45 s followed by 72°C for 10 min with a final hold at 4°C.

Table 7 discloses the sequences of the primers used to amplify the fragment A.

For amplicon A positive samples renders a 156 bp band

Species identification was achieved by sequencing and indexing the nucleotides of the amplicons as described above. The primers are given in Table 8

55

and sequencing standard protocol was used to sequence amplicon A.

Table 9 discloses the sequences of the primers used to amplify the fragment A plus.

For amplicon A plus positive samples renders a 174 bp band

Species identification was achieved by sequencing and indexing the nucleotides of the amplicons as described above. The primers are given in Table 10 and sequencing standard protocol was used to sequence amplicon A plus.

56

CLAIMS

1.      Single stranded oligonucleotide comprising, linked from 5' to 3', a 5' portion of at least 5 consecutive nucleotides of a universal primer for automated sequencers, a read start portion and a portion complementary to a region of the target sequence to be amplified or flanking the sequence thereof.

2.      Single stranded oligonucleotide according to claim 1 wherein said portion complementary to a region of the target sequence to be amplified comprises one or more polymorphism, and wherein said region to be amplified comprises one or more polymorphism.

3.      Single stranded oligonucleotide according to claims 1 or 2 wherein said polymorphisms comprise one or more SNP.

4.      Single stranded oligonucleotide according to anyone of claims 1-3, wherein said oligonucleotide is of a length of about 15 to 50 nucleotides and wherein said portion of universal primer is of a length of about 5-20 nucleotides and the read start portion is of a length of about -12 nucleotides and said complementary portion is of a length of about 10 to 25 nucleotides.

5.      Single stranded oligonucleotide according to claim 3 or 4 selected from the group consisting of SEQ ID NO: 1-9, 12-15, 17-24, 26-31.

6.      A method for the detection of polymorphisms or polymorphisms patterns of nucleotide sequences amplified by PCR comprising the steps of:

        a. sequencing nucleotide sequences amplified with one or more pair, forward and reverse, of oligonucleotides according to anyone of claims 1-4 with an automated sequencer using a primer comprising, at the 3', said universal primer portion or collecting nucleotide sequences thus obtained;

        b. detecting in said sequences said read start sequence;

        c. indexing nucleotides of said sequences from said detected start sequence;

        d. detecting in said indexed sequences polymorphisms or polymorphisms patterns.

7.      A method for the analysis of polymorphisms or polymorphisms patterns of nucleotide sequences amplified by PCR comprising the steps of:

        a. sequencing nucleotide sequences amplified with one or more pair, forward and reverse, of oligonucleotides according to anyone of claims 1-4 with an automated sequencer using a universal primer comprising, at the 3', said universal primer portion or collecting nucleotide sequences thus obtained;

        b. detecting in said sequences said read start sequence;

        c. indexing nucleotides of said sequences from said detected start sequence;

        d. detecting in said indexed sequences polymorphisms or polymorphisms

57

patterns;

e. analysing the data obtained at step d.

8.      The method according to claim 7 wherein said sequences are amplified from simple or complex food matrices with the primers having SEQ ID NOs 1-9, optionally also primers of the groups having SEQ IDs 12-15, 2, 7, 8 and 9; SEQ IDs12-15 and17-24; and/or SEQ IDs 12-15 and 26-30;

the analysis of step e. allowing the detection of up to 17 *Scombridae* species.

9.      A method for the detection and identification of new polymorphisms or polymorphisms patterns of nucleotide sequences amplified by PCR comprising the steps of:

a. sequencing nucleotide sequences from given samples amplified with one pair, forward and reverse, of oligonucleotides according to the description with an automated sequencer using a universal primer comprising said 5' portion;

b. detecting in said sequences said read start portion;

c. indexing nucleotides of said amplified sequences, from said read start portion;

d. detecting in said indexed sequences the presence polymorphisms or polymorphisms patterns in comparison to a reference sequence;

e. sequencing nucleotide sequences from the same samples of step a. amplified each with a different pair, forward and reverse, of oligonucleotides according to the description, each pair of oligonucleotides having a different read start, with an automated sequencer using a universal primer comprising said 5' portion;

f. detecting in said sequences said read start portions;

g. indexing nucleotides of said amplified sequences, from said read start portions, each sample being recognisable by a specific read start portion;

h. identifying in said indexed sequences the polymorphisms or polymorphisms patterns in comparison to a reference sequence;

10.     A method for the detection and identification of new polymorphisms or polymorphisms patterns of nucleotide sequences amplified by PCR comprising the steps of:

a. sequencing nucleotide sequences from different samples each sample being amplified with a different pair, forward and reverse, of oligonucleotides according to the description and each pair of oligonucleotides having a different read start, with an automated sequencer using a universal primer comprising said 5' portion;

58

b. detecting in said sequences said read start portions;

c. indexing nucleotides of said amplified sequences, from said read start portions, each sample being recognisable by a specific read start portion;

d. identifying in said indexed sequences the polymorphisms or polymorphisms patterns in comparison to a reference sequence;

11. A software capable of carrying out the method of anyone of claims 6-10.

12. A computer readable storage support wherein the software of claim 11 is stored.

13. A kit comprising one or more aliquot of at least two, forward and respective reverse oligonucleotides of anyone of claims 1-4, and optionally one or more aliquot of a universal primer for automated sequencer wherein said universal primer comprises said 5' portion.

14. The kit according to claim 13, wherein said oligonucleotides consists one or more of the following oligonucleotides groups SEQ ID NO: 1-9; SEQ ID NO: 12-15, 2, 7, 8 and 9; SEQ IDs12-15 and17-24; and/or SEQ IDs 12-15 and 26-30

15. The kit of claim 14 wherein said universal primers for automated sequencer are of SEQ ID NO 10 and 11 when said oligonucleotides are SEQ ID NO: 1-9; of SEQ ID NO 16 and 11 when said oligonucleotides are SEQ ID NO: 12-15, 2, 7, 8 and 9; of SEQ ID NO 25 and 16 when said oligonucleotides are SEQ ID 12-15 and 17-24 and of SEQ ID NO 16 and 31 when said oligonucleotides are SEQ IDs 12-15, 26-30.

16. The kit of anyone of claims 13 to 15 further comprising the support of claim 12.
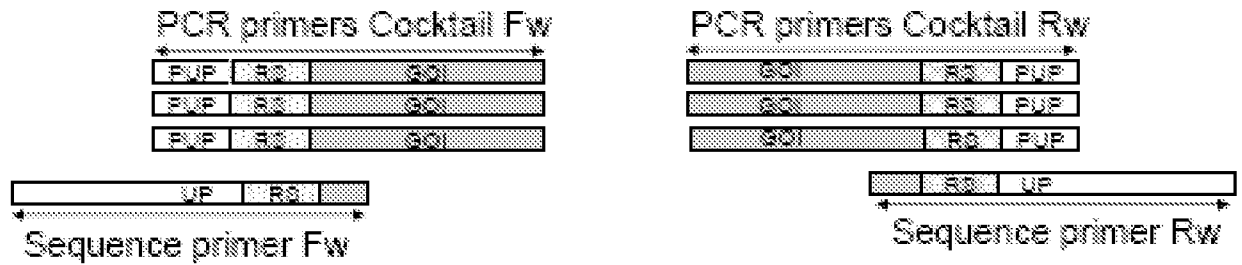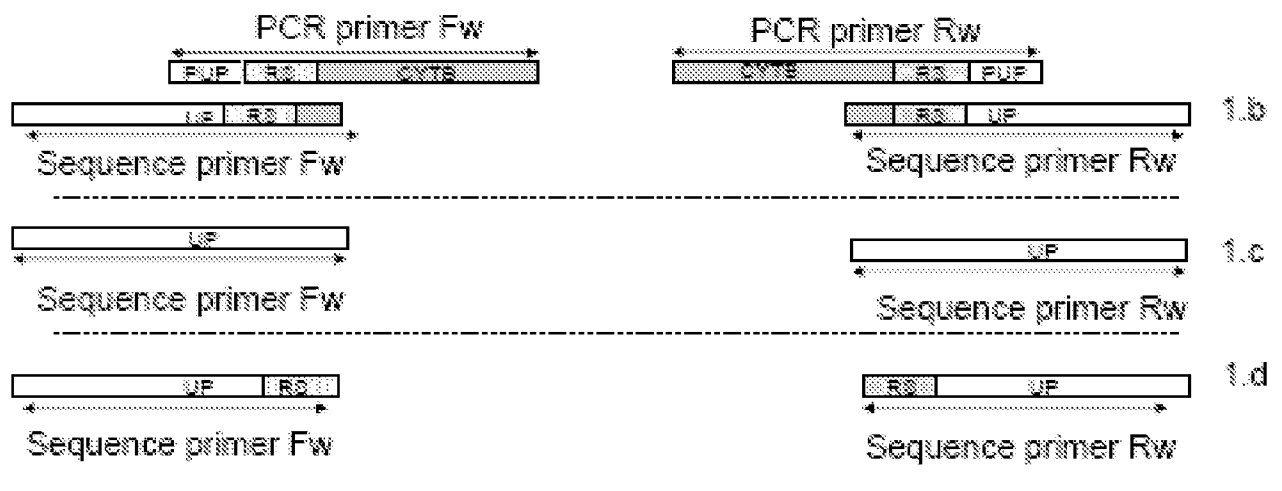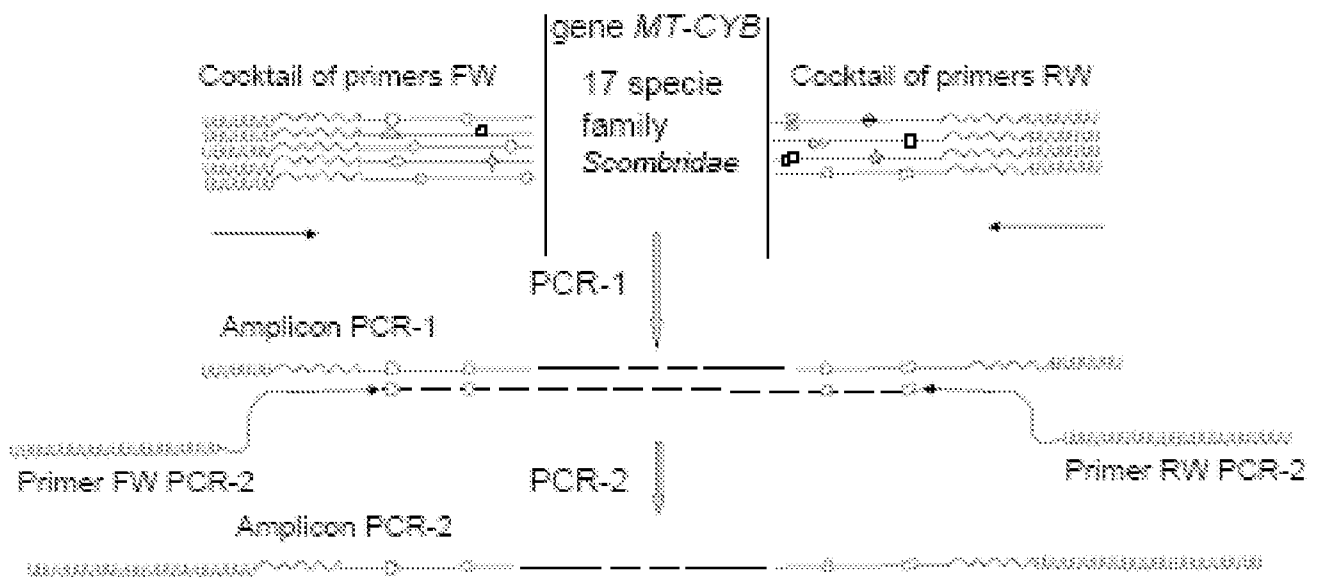
Figure 1.a
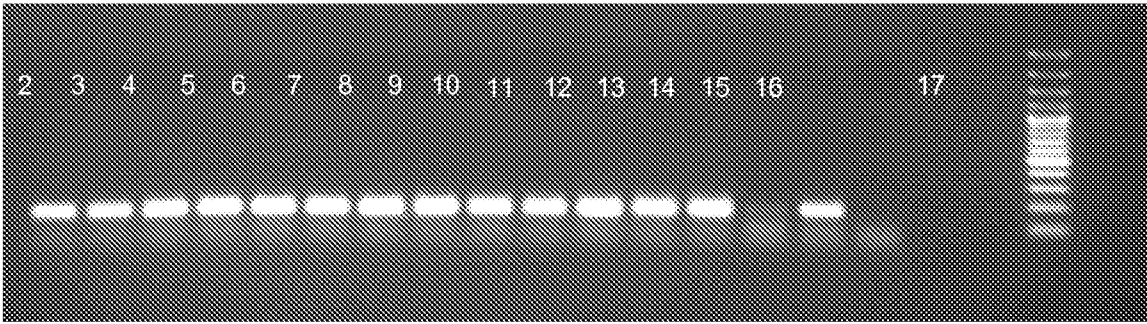


Figure 1
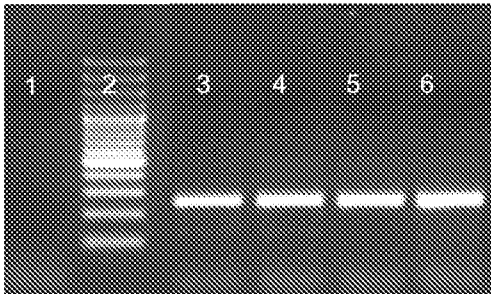
Figure 2
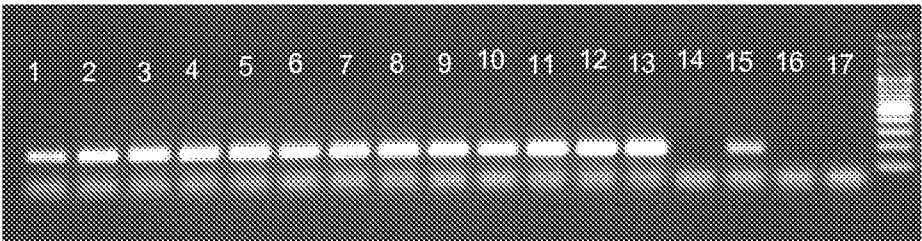
Figure 3



Figure 4



Figure 5
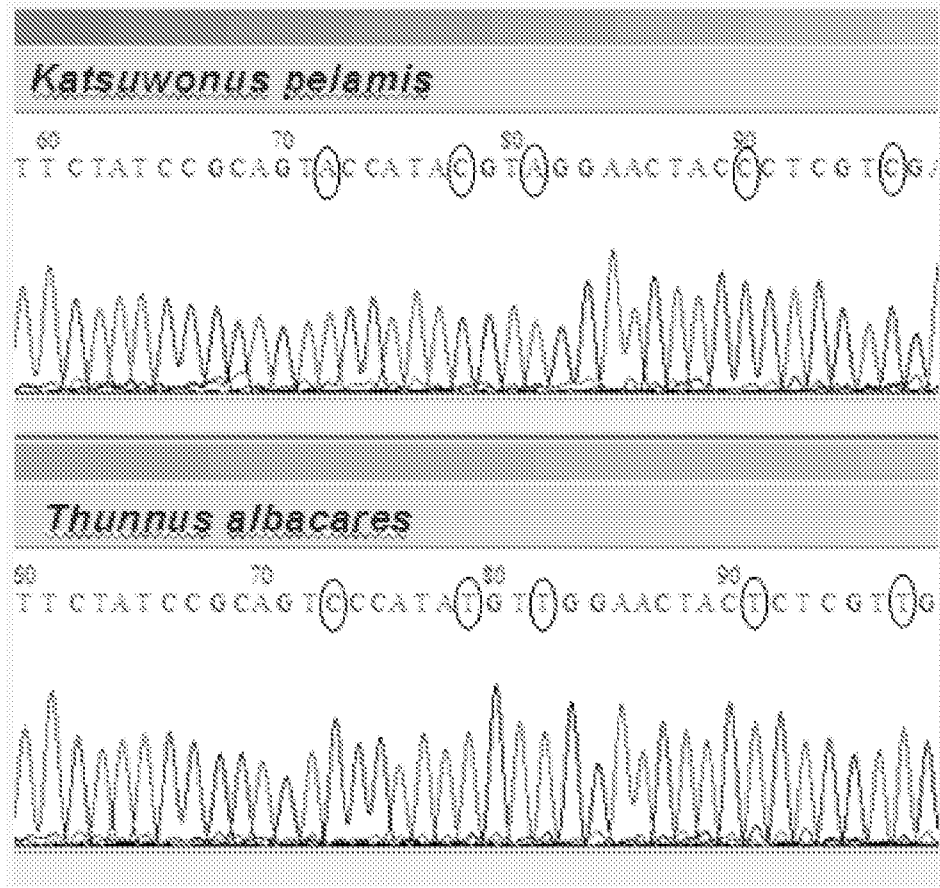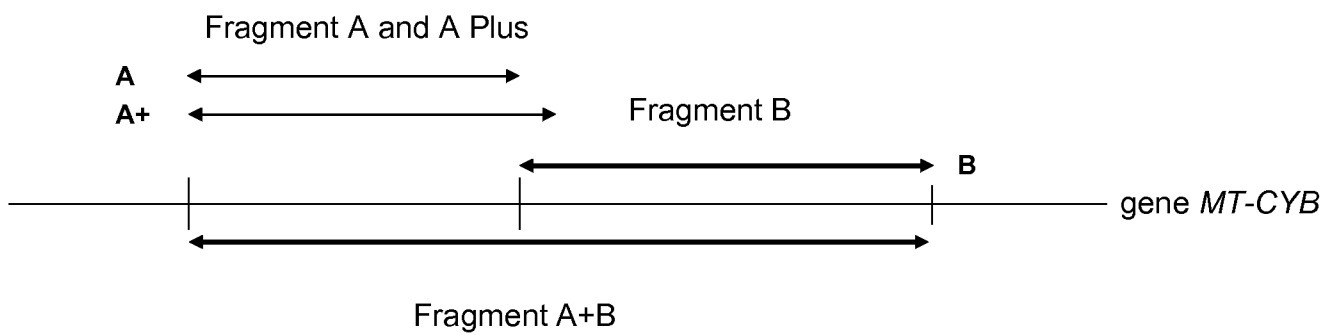
Figure 6



Figure 7

5/5

**Primers for fragment A**

Fw A= 12 + 253-272

Rw A= 364-386 +11

Sequence

from 253 to 386= 133+ 23 (pb add to primers 12pb+11pb)= 156 bp

Length of fragment A= 156 pb

Length of sequence fragment A= 156+14 (fw seq primer)+14 (rw seq primer)=184 bp


**Primers for fragment B**

FW B= 12 + 382-403

RW B= 499-521+10

from 382 to 521=140+22 (pb add to primers 12pb+10pb)= 162 bp

Length of fragment B= 162 pb

Length of sequence fragment B= 162+14(fw seq primer)+15 (rw seq primer)= 192 bp


**Primers for fragment A+B**

FW A+B= 12 .+ 253-272

RW A+B= 499-521+10

from 253 to 521= 269+ 22 (pb add to primers 12pb+10pb)= 291 bp

Length of fragment A+B= 291 bp

Length of sequence fragment A+B= 291+14(fw seq primer)+15 (rew seq primer)= 320 bp


**Primers for fragment A plus**

Fw A = 12 + 253-272

RW A PLUS=382-403+11

from 253 to 403=151+23 (pb add to primers 12pb+12pb)= 174 bp

Length of fragment A PLUS= 174 pb

Length of sequence fragment A PLUS= 175+14(fw seq primer)+14 (rw seq primer)= 203 bp


Figure 8

# INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2008/065804

## A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, EMBASE, WPI Data

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | WO 01/06012 A (PACKARD BIOSCIENCE COMPANY [US]; ENGLERT DAVID F [US]) 25 January 2001 (2001-01-25) the whole document | 1-16 |
| X | WO 2007/037678 A (KEYGENE NV [NL]; VAN EIJK MICHAEL JOSEPHUS THER [NL]; VAN TUNEN ADRIAN) 5 April 2007 (2007-04-05) the whole document | 1-16 |
| | -/-- | |

| X | Further documents are listed in the continuation of Box C. | X | See patent family annex. |

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 23 March 2009 | 31/03/2009 |

| Name and mailing address of the ISA/ | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Fax: (+31–70) 340–3016 | Bradbrook, Derek |

Form PCT/ISA/210 (second sheet) (April 2005)

| C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|---|---|---|
| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| A | BOTTERO ET AL: "Differentiation of five tuna species by a multiplex primer-extension assay" JOURNAL OF BIOTECHNOLOGY, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 129, no. 3, 13 April 2007 (2007-04-13), pages 575-580, XP022027147 ISSN: 0168-1656 cited in the application abstract; table 1 | 1-16 |
| A | WO 96/41012 A (GENZYME CORP [US]; SHUBER ANTHONY P [US]) 19 December 1996 (1996-12-19) the whole document | 1-16 |

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

| Patent document cited in search report | | Publication date | Patent family member(s) | | Publication date |
|---|---|---|---|---|---|
| WO 0106012 | A | 25-01-2001 | AT | 409236 T | 15-10-2008 |
| | | | AU | 6345700 A | 05-02-2001 |
| | | | CA | 2379132 A1 | 25-01-2001 |
| | | | EP | 1194592 A1 | 10-04-2002 |
| WO 2007037678 | A | 05-04-2007 | AU | 2006295556 A1 | 05-04-2007 |
| | | | CA | 2623539 A1 | 05-04-2007 |
| | | | CN | 101313078 A | 26-11-2008 |
| | | | EP | 1929039 A2 | 11-06-2008 |
| WO 9641012 | A | 19-12-1996 | AT | 232242 T | 15-02-2003 |
| | | | AU | 703417 B2 | 25-03-1999 |
| | | | AU | 6163696 A | 30-12-1996 |
| | | | CA | 2223729 A1 | 19-12-1996 |
| | | | DE | 69626111 D1 | 13-03-2003 |
| | | | DE | 69626111 T2 | 24-12-2003 |
| | | | EP | 0832290 A1 | 01-04-1998 |
| | | | JP | 11507226 T | 29-06-1999 |
| | | | US | 5882856 A | 16-03-1999 |
| | | | US | 6207372 B1 | 27-03-2001 |